

# Journal of Multimedia

ISSN 1796-2048

Volume 9, Number 5, May 2014

## Contents

---

### REGULAR PAPERS

3-Parameter Hough Ellipse Detection Algorithm for Accurate Location of Human Eyes <i>Qiufen Yang, Huosheng Hu, Weihua Gui, Shuren Zhou, and Can Zhu</i>	619
Unsupervised Learning and Linguistic Rule Based Algorithm for Uyghur Word Segmentation <i>Turdi Tohti, Winira Musajan, and Askar Hamdulla</i>	627
Short Text Classification: A Survey <i>Ge Song, Yunming Ye, Xiaolin Du, Xiaohui Huang, and Shifu Bie</i>	635
Multi-agent Remote Sensing Image Segmentation Algorithm <i>Chen Jing and Wang Haifeng</i>	644
Concept Tree Based Information Retrieval Model <i>Yuan Chunyan</i>	652
An Abnormal Speech Detection Algorithm Based on GMM-UBM <i>Jun He, Ji-chen Yang, Jianbin Xiong, Guoxi Sun, and Ming Xiao</i>	660
On Charactering of Word-of-Mouth Propagation in Heterogeneous Online Social Networks <i>Li Niu, Xiaoting Han, and Yongjun Xu</i>	668
Image Retrieval via Relevance Vector Machine with Multiple Features <i>Zemin Liu and Wei Zong</i>	676
Low SNR Speech Recognition using SMKL <i>Qin Yuan</i>	682
Electronic Commerce Data Mining using Rough Set and Logistic Regression <i>Li Xiuli, Zhao Rui, and Xiao Yan</i>	688
Eye State Recognition Algorithm GAHMM of Web-based Learning Fatigue <i>Qiufen Yang, Zhenjun Li, Canjun Li, Xianlin Yang, and Can Zhu</i>	694
Restoration Technique of Video Motion Image Estimation Based on Wavelet <i>Ruibin Chen</i>	701
New Video Target Tracking Algorithm Based on KNN <i>Ding Ma and Zhezhou Yu</i>	709
Face Detection and Location System Based on Software and Hardware Co-design <i>Hua Cai, Yong Yang, Fuheng Qu, and Jianfei Wu</i>	715
Video Image Object Tracking Algorithm based on Improved Principal Component Analysis <i>Wang Liping</i>	722
Data Interpretation Technology for Continuous Measurement Production Profile Logging <i>Junfeng Liu, Heng Li, and Yingming Liu</i>	729
Rival Penalized Image Segmentation <i>Shaojun Zhu, Jieyu Zhao, and Lijun Guo</i>	736

---



# 3-Parameter Hough Ellipse Detection Algorithm for Accurate Location of Human Eyes

Qiufen Yang

School of Information Science and Engineering, Central South University, Changsha, China  
Email: yqf5569@sohu.com

Huosheng Hu

School of Computer Science & Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK  
Email: hhu@essex.ac.uk

Weihua Gui

School of Information Science and Engineering, Central South University, Changsha, China  
Email: whgui@csu.edu.cn

Shuren Zhou

Computer & Communication Engineering School, Changsha University of Science & Technology, Changsha, China  
Email: zsr@csust.edu.cn

Can Zhu

School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha, China  
Email: zhc@csust.edu.cn

**Abstract**—Accurately positioning the Human Eyes plays an important role in the detection of the fatigue driving. In order to improve the performance of positioning of human face and eyes, an accurately positioning method of the human eyes is proposed based on the 3-parameter Hough ellipse detection. Firstly, the human face area is divided by using the skin color clustering and segmentation algorithm. Then, the segmented image is filtered by using its geometric structure and the approximate positions of the human face area and eyes are calculated. Finally, on the basis of the spinning cone-shape eye model, the position of human face and eyes is accurately determined by using the 3-parameter Hough transformation ellipse detection algorithm. The different images of human face are used to test the performance of the proposed method. The experimental results show that the extreme value of upper and lower eyelids and the actual position is 0.104 and the proposed algorithm has higher positioning accuracy.

**Index Terms**—Skin Color Clustering; Positioning of Human Eyes; Hough Transformation; Ellipse Detection

## I. INTRODUCTION

Along with rapid economic development, the number of automobiles in the whole globe has been constantly increasing; therefore, the problem of road safety becomes a hot issue. Besides, fatigue driving is one of the major reasons for traffic accidents. In order to solve this problem, we thus conduct researches about fatigue driving and place our emphasis on human eye tracking in detecting devices for fatigue driving as human eye

tracking serves as an important step in drowsy driving detection system.

Domestic researches on fatigue driving have become more and more since 2003, focusing on the state information of the head or face of the drivers. In 2001, from the perspective of the image recognition technology, Zheng Pei et al. applied the PERCLOS fatigue parameters to develop the measurement system of motor driver fatigue (Li Zengyong and Wang Chengzhu, 2001). Wang Xiaojuan started the research on the combination of the eye state information and mouth state information(Cootes T F and Taylor C J, 1995), capturing the head position information through the mobile camera mounted on the platform, then finding the eyes and the mouth position in the image, and extracting and combining all the data information to determine fatigue. Therefore, it has important research value to use the camera to continuously observe the image of driver's facial features such as eyes, ears, nose and mouth and then determine whether the driver is fatigue driving in accordance with existing facial recognition technology. Federal Highway Administration uses the PERCLOS method, which determines the fatigue degree of eyes in accordance with the duration of eyes being closed during a certain period, and its precondition is the accurate positioning of human eyes. Therefore, it is particularly important to conduct accurate positioning of eyes in accordance with the color image sequences captured by the camera.

The positioning method of human eyes generally consists of two steps: the first step is coarse positioning,

which is to find the possible spot of eyes on the image or initially determine the approximate position of human eyes; the second step is accurate positioning, which uses certain rules or verification method to determine the exact position of two eyes. At present, the positioning algorithms for human eyes include: the method proposed by Reinders et al. which uses the neural network and the micro characteristic of eyes to position the feature of human face [1], its shortage is that it requires a high calculated quantity for multi-scale detection of human eyes; Zhu et al. used the integral image to find the candidate points of pupil and then used SVM for detection [2], which seldom used the shape information of eyes; Liu et al. used the geometric characteristics of iris to detect human eyes, then conducted pairing in accordance with certain rules, and then used the neural network to conduct verification [3]; Bala et al. proposed a eye positioning method based on genetic algorithm and decision tree [4], which cannot effectively address the situation of face rotation, and it does not have adequate recognition ability of objects similar to human eyes; Feng and Yuen proposed a multi-thread method to position eyes [5]; Huang and Weshler used the SVM method to obtain the posture of human face and the position of human eyes [6]; Wang Shoujue et al. used the geometry complexity to rapidly position human eyes [7]; based on edge extraction, Zhang et al. used the Hough transformation [8] and circle detection algorithm to position the iris of eye, but the generated image is not necessarily consistent with the actual shape of eyes.

Therefore, all the accurate positioning algorithms for human eyes proposed up until now require a high calculated quantity, some are difficult to realize, and some are sensitive to the rotation, translation or size change of image, which are highly limited. In this paper, an innovative accurate positioning method for human eyes based on the color space of skin color clustering and the Hough transformation ellipse detection proposed. First of all, through the skin color clustering and segmentation algorithm, this method divides the human face area to obtain the binary image of human face; then, the skin color clustering and segmentation algorithm is used to conduct connectivity analysis of the binary image that includes the human face area, then, geometric filtration is conducted, and the centroid point of hole in the candidate human face area is calculated to find possible human eye pair; based on approximate detection of the human face area and eyes, the spinning cone-shape eye model is proposed to analyze the parameters of elliptical calculation, which overcomes the shortage of traditional 5-parameter space ellipse method, and the 3-parameter space method is used; at last, the Hough transformation ellipse detection algorithm is used to conduct accurately positioning of human eyes. With the relative error scale of 0.104, this algorithm can reach an accuracy of 100%.

## II. COLOR MODEL OF SKIN COLOR CLUSTERING

### A. Color Space of Skin Color Clustering

In order to segment the human face area from the

whole image, a reliable skin color model is built, which applies to various factors that affect the skin detection result, such as different skin colors, lighting conditions and covering.

The RGB space consists of the red, green and blue components, and because in this space, the color information and brightness information are mixed, while the difference of skin color is mainly caused by brightness, therefore, this color model is not suitable. Through statistics, we find that the  $C_b$  and  $C_r$  components present very stable clustering characteristics, as shown in Figure 1. By using the clustering between the chromaticity and brightness of skin [1] [2], the chromaticity and brightness components are selected to divide the  $YCbCr$  color space and conduct skin color modeling.

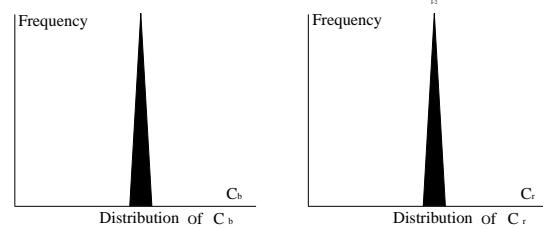


Figure 1. Distribution of  $C_b$  and  $C_r$  components in skin color

In Fig. 2, the distribution of gray level differences over images of eyes is shown. The gray level differences, calculated over different scales of skin, are shown in the Figure It can be seen that the distribution of gray level differences for the class of eye images is well approximated by the  $YCbCr$  color space. Further, it can be seen that the width  $k$  of the distribution increases with the scale.

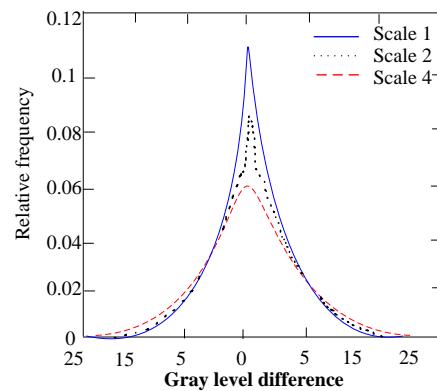


Figure 2. The distribution of gray level differences

### B. Gaussian Model of Skin Color Clustering

After projecting the obtained facial image into the  $YCbCr$  space, the number of different  $C_b-C_r$  pairs is calculated, the final count value is used as the vertical coordinate, and with  $C_b$  as the horizontal ordinate and  $C_r$  as the longitudinal coordinate, a three dimensional stereogram is drawn(Figure 3) [2].

For any pixel, its brightness component  $Y$  and the two chromaticity components  $Cb$  and  $Cr$  are statistically independent, which present Gaussian distribution. Because the brightness component  $Y$  is susceptible to the change of light, therefore, only the two stable chromaticity components  $Cb$  and  $Cr$  are used to build the Gaussian model  $G(m, V^2)$ :

$$m = (\overline{Cr}, \overline{Cb}) \quad (1)$$

$$\overline{Cr} = \frac{1}{N} \sum_{i=1}^N Cr_i \quad (2)$$

$$\overline{Cb} = \frac{1}{N} \sum_{i=1}^N Cb_i \quad (3)$$

$$V = \begin{bmatrix} \sigma_{Cr,Cr} & \sigma_{Cr,Cb} \\ \sigma_{Cb,Cr} & \sigma_{Cb,Cb} \end{bmatrix} \quad (4)$$

In which,  $\overline{Cr}$  and  $\overline{Cb}$  are the corresponding mean values of  $Cr$  and  $Cb$ , and  $V$  is the covariance matrix.

Through the built skin color model, we transform a color image into a grayscale image, the grayscale value corresponds to the possibility of this point belonging to the skin area, and then, the grayscale image can be further transformed into a binary image by selecting appropriate threshold value, in which, 0 and 1 refer to the non-skin area and skin area respectively.

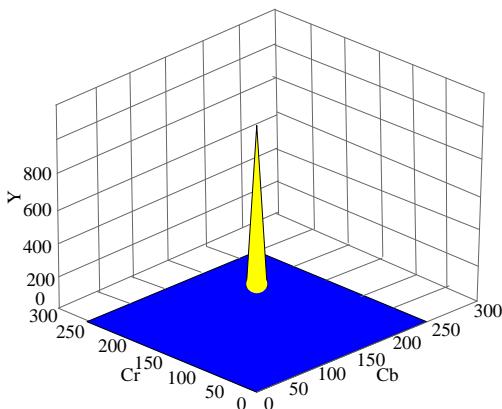


Figure 3. Distribution of skin color in the  $YCbCr$  space

### III. AREA SEGMENTATION DURING COARSE POSITIONING OF HUMAN EYES

Because the human face is a connected area, we also conduct connectivity analysis of the binary image that includes the human face area obtained earlier in the color space of skin color clustering. For the noise existing in the inside and background of image, the  $3 \times 3$  mid-value is adopted to conduct filtering and de-noising, and the expansion operator of morphology is used for filling to make the non-filled area as small as possible.

Based on the color segmentation algorithm proposed in the literature [12], the skin color clustering and

segmentation algorithm is proposed, and the purpose of this algorithm is to ensure various clustering areas have consistent color.

The skin color clustering and segmentation algorithm is as the following:

(1) Divide the skin color image into  $m \times n$  blocks, and calculate the mean value of skin color pixel  $y_i Cb_i Cr_i$  of each block.

(2) Set the two-dimensional array  $clus$  to mark the clustering area of each block, and the initial value is set as 0.

(3) Scan the image from top to bottom and left to right, find the first block with a  $clus$  value of 0, use it as a new clustering area, then set  $clus$  as 1, and initialize  $Cb$ ,  $Cr$  as well as the mean values of  $\overline{Cr}$  and  $\overline{Cb}$  of the area.

(4) Successively examine all skin color blocks with a  $clus$  value of 0 in the surrounding area of Block 8, if its variance with the  $Y$  component of all blocks in current

$$\text{area is smaller than } c_1, \frac{Cb_i \overline{Cb} + Cr_i \overline{Cr}}{\sqrt{Cb_i^2 + Cr_i^2} + \sqrt{(\overline{Cb})^2 + (\overline{Cr})^2}} \geq c_2$$

it should be included into the area, the  $clus$  value should be set as 1, and  $\overline{Cr}$ ,  $\overline{Cb}$  as well as the mean values of  $\overline{Cr}$  and  $\overline{Cb}$  should be recalculated; repeat (4) until all connected blocks are scanned.

(5) Record the searched area, return to Step (3), conduct a new round of search, and add 1 to the  $clus$  value of each block until the  $clus$  values of all skin color blocks are not 0.

#### A. Geometric Filtration

In addition to the facial skin, the skin color area extracted with the skin color clustering and segmentation algorithm also includes the skin of arm and shoulder, which even includes moving external surface that does not belong to human body. However, the human face has unique features, and the non-human face area can be filtered through the following principle:

(1) The human face occupies a certain proportion in the image, which is the common “three parts and five organs”, calculate the size of connected area in the detected candidate area of human face, set the threshold value as  $T$ , when the connected area is smaller than  $T$ , it is regarded as non-human face area, and this area will be abandoned and blackened.

(2) The front human face area is close to an ellipse, and under non-extreme situation, the length-width ratios of the bounding rectangles of the profile, looking up face and looking down face are all within a certain scope. Through multiple experiments, we obtained the scope of length-width ratio is  $[0.5, 2.5]$ , and therefore, the candidate area outside of this scope can be filtered and blackened.

(3) Because the color and skin color of eyebrows and eyes (sometimes the nostrils and mouth area) have significant difference on both chromaticity and brightness, therefore, these areas form one or more than one hole in the facial skin area. However, this kind of situation is rare in other skin area, so in accordance with the Euler

number criterion, the non-human face area can be filtered. Set E as the Euler number, C as the number of connected area and H as the number of hole, then for the human face area, there is:  $E = C - H \leq 1 - 1 = 0$ .

The area with  $E > 0$  should be abandoned and blackened.

(4) The lips area is not considered for now.

In accordance with the region segmentation algorithm and the binary image after geometric filtration, see Figure 4.



Figure 4. Binary image after region segmentation

### B. Coarse Positioning of Human Eyes

The candidate areas of human face obtained above all contain a certain number of holes. Through pairing of the holes in each area, all possible two eye pairs are formed. Define a data structure set  $\{\text{eyes} \mid \text{eyes} = \{L(x,y), R(x,y), d, \theta\}\}$ , and describe these possible two eyes. In it,  $d$  refers to the distance between two holes,  $\theta$  refers to the included angle between the line linking the two holes and the horizontal axis, in the meantime, this angle is also the inclination angle of human face, and  $L$  and  $R$  refer to the centroid coordinates of the two hole areas respectively. The centroid coordinate is calculated in accordance with

$$\bar{x} = \frac{\sum_{i=0}^{n=1} \sum_{j=0}^{m=1} jB[i, j]}{A}, \quad \bar{y} = \frac{\sum_{i=0}^{n=1} \sum_{j=0}^{m=1} iB[i, j]}{A}, \quad \text{the following}$$

formula, in which,  $A = \sum_{i=0}^{n=1} \sum_{j=0}^{m=1} B[i, j]$  refers to the area of hole.

Search each candidate area of human face respectively, find all possible two eye pairs with  $|\theta| < 45^\circ$  and  $L(x) < R(x)$ , and add them to the above set. Here, for the defined size, it is considered that the inclination angle of driver's face captured by the camera won't be too big, and the front image is preferred to reduce the computation overhead.

## IV. ACCURATE POSITIONING ALGORITHM OF HUMAN EYES

Through the above process, the real human face area has been found, and the approximate position of two eyes is found during positioning of the five sense organs. In the following, a new Hough transformation ellipse detection algorithm will be used to accurately and dynamically generate the whole curve shape of eyes.

### A. Human Eye Model

At present, related positioning methods for human eyes select the eyeball center or the iris center as the positioning point, and see Figure 4 for the eyeball structure. Eyeball is a sphere with a radius of  $R$ : the iris is at the front of eyeball with a radius of  $r$ ; the distance between the eyeball center and the iris center is  $d$ , and there is  $R^2 = r^2 + d^2$ . In accordance with the description in literature [13], the radiiuses of the eyeball and iris are constant values, and  $R/r$  is also a constant. Therefore, we only need to measure the radius of iris— $r$ , and we can obtain  $R$  and  $d$ . The front structure of eye is defined as the spinning cone shape, the middle part is iris, the endpoints at two sides are the interior angle point and exterior angle point respectively, and the upper and lower arcs refer to the upper eyelid and lower eyelid respectively, as shown in Figure 5.

In our algorithm, the interior angle point, exterior angle point, upper eyelid, lower eyelid and the centroid point obtained through the above coarse positioning of human eye area are used to draw the longitudinal section of ellipse, so it does not need to conduct accurate positioning of eyeball or iris. First of all, the 5 parameters of the ellipse are analyzed, and the Hough transformation ellipse detection algorithm is used to generate the eye curve and accurately position the area of human eyes.

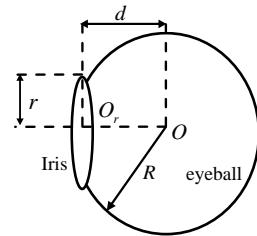


Figure 5. Eyeball structure

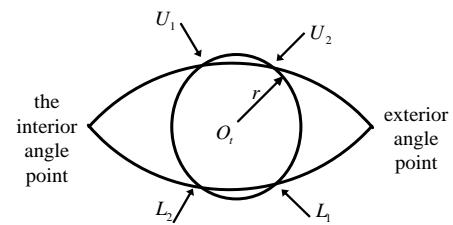


Figure 6. Spinning cone-shape eye model

### B. Ellipse Detection Algorithm

The ellipse detection algorithm has always been a key issue during image processing, and just due to this reason, there are various ellipse detection methods now. Because it requires 5 parameters to completely define an ellipse, therefore, it requires a five-dimensional parameter space to detect an ellipse, which is a very time consuming work. Literature [14] used a new ellipse detection algorithm, which uses the long axis of ellipse to rapidly and effectively find the parameters of ellipse, and it only needs a one-dimensional accumulative array to accumulate the lengths of the short axes of ellipse. In this way, the required computation storage space is much

smaller than the previous algorithms. We used the method of Literature [14] into our system.

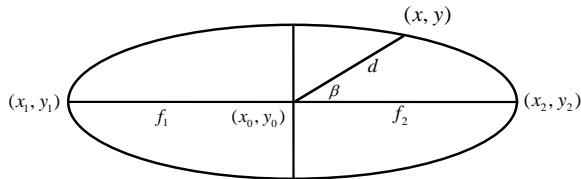


Figure 7. Geometrical characteristic of ellipse

In the ellipse in Figure 6, there are 5 ellipse parameters: the centroid point  $(x_0, y_0)$ , the ellipse direction angle  $\alpha$  and the lengths of long axis and short axis  $(a, b)$ . We add addition information to each boundary point or use special boundary point, and in this way, we only need a small amount of boundary points to determine the position of an ellipse. The interior angle point  $(x_1, y_1)$  and exterior angle point  $(x_2, y_2)$  represent the two endpoints of long axis (which can only be the two endpoints of long axis). The 4 parameters of ellipse can be calculated in accordance with the following method:

$$x_0 = \frac{x_1 + x_2}{2} \quad (5)$$

$$y_0 = \frac{y_1 + y_2}{2} \quad (6)$$

$$a = \sqrt{\frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2}} \quad (7)$$

$$\alpha = \arctan \frac{y_2 - y_1}{x_2 - x_1} \quad (8)$$

Assume  $f_1$  and  $f_2$  are the two foci of ellipse,  $(x, y)$  is a random third point on the ellipse, and we can use  $(x, y)$  to calculate the fifth parameter of this ellipse. Apparently, the distance between  $(x, y)$  and  $(x_0, y_0)$  is smaller than the distance between  $(x_1, y_1)$  and  $(x_0, y_0)$  or the distance between  $(x_2, y_2)$  and  $(x_0, y_0)$ .

In this way, the computational formula for short axis is:

$$b^2 = \frac{a^2 d^2 \sin^2 \beta}{a^2 - d^2 \cos^2 \beta} \quad (9)$$

In which,  $\cos \beta = \frac{a^2 + d^2 - f^2}{2ad}$ ,  $d$  refers to the distance between  $(x, y)$  and  $(x_0, y_0)$ , and  $\beta$  represents the included angle between  $(x, y)$  and  $(x_0, y_0)$ .

#### C. Hough Transformation Ellipse Detection

Hough transformation is a very effective shape analysis method, which is insensitive to stochastic noise, and it has been widely used in detection of straight line, circle and ellipse. Its basic idea is to transform the spatial

domain of image to the parameter space and use a certain parameter form that satisfies most boundary points for description.

Curve on the image (area boundary). The Hough transformation detection technique calculates the parameters of boundary curve in accordance with local measurement, so it has great fault tolerance and robustness to the interruption of area boundary caused by noise interference or being covered by other target [15].

The general form to express the parameters of Hough transformation analytic curve is:

$$f = (\vec{X}, \vec{a}) = 0$$

At this moment, point  $\vec{X} = [x, y]^T$  is a point on the ellipse, and point  $\vec{a} = [d, \beta, b]^T$  corresponds to the parameter of ellipse. The ellipse of image space corresponds to one point in the parameter space  $(d, \beta, b)$ . A set point  $(x, y)$  restrains a group of parameters  $(d, \beta, b)$  of ellipse that pass through this point, which is equivalent to restraining the track of point  $(d, \beta)$  that generates a series of ellipses. When point  $(x, y)$  moves along this series of ellipses in the image space, for each point on the ellipses boundary, the parameter variation in corresponding parameter space forms a spinning cone track. This is consistent with the human eye model described in Section IV. Conduct appropriate quantification to the parameter space of Hough transformation, a three-dimensional accumulator array is obtained, and in the array, each small cubic lattice corresponds to the discrete value of parameter  $(d, \beta, b)$ .

Through the above ellipse detection algorithm, accurate positioning of human eyes can be rapidly and stably conducted based on detection of human face. The specific realization measures are as the following:

(1) For the color image, first of all, transform the image into grey-scale image in accordance with the color space of skin color clustering.

(2) Then, the grayscale image can be further transformed into a binary image by selecting appropriate threshold value.

(3) In accordance with the connectivity of human face and the clustering feature of skin color, the skin color clustering and segmentation algorithm in this paper is used to conduct region segmentation of human face, then, geometric filtration is conducted to the segmentation image, and the centroid point of hole in the candidate human face area is calculated to find possible human eye pair.

(4) In this paper, the Hough transformation ellipse detection algorithm is used to accurately position the human face. Conduct shape detection to the ellipse boundary existing in the image space, firstly, calculate the gradient information of the intensity of each point on the image, then, obtain the edge in accordance with appropriate threshold value, and then add 1 to the accumulator of edge  $(d, \beta, b)$  and small cubic lattice.

## V. EXPERIMENT RESULT

For the Hough transformation ellipse detection algorithm to accurately position the curve of human eyes based on the skin color clustering space, we selected 50 images from the JAFFE (Japanese Female Facial Expression) database [10] of Japanese Kyushu University (200 images) and the Internet (300 images) respectively to test the human eye detection of different images of human face. For the standard human face database, due to the simple background and standard posture, it can rapidly find the position of human eyes with this algorithm; for the latter with multiple postures and backgrounds, some even have certain noise disturbance, this algorithm can also accurately position human eyes, but there exists certain pixel error, and the test method is as the following:

First of all, find the centroid point of human eye pair in the binary image in Figure 3, as shown in Figure 8.

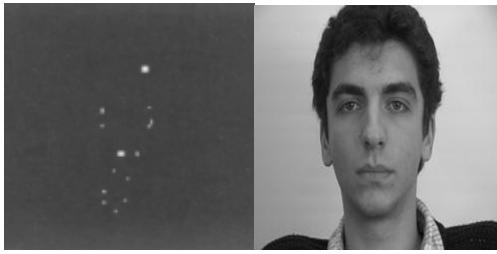


Figure 8. Centroid point of characteristic part of human face

First of all, manually find the coordinates  $L(x, y)$  and  $R(x, y)$  of the centroid points of the left and right human eyes, then, find the interior angle point and exterior angle point of human eye area in the binary image in Figure 3, and conduct accurate positioning of the profile curve of human eyes with the algorithm in this paper. Then, manually find the actual positional coordinates of the extreme points of upper and lower eyelids (take the left eye for example)  $A_{up}(x_{up}, y_{up})$  and  $B_{down}(x_{down}, y_{down})$ , then, on the accurate positioning image of human eyes in Figure 8 obtained with this algorithm, find the computation positions of the extreme points of upper and lower eyelids  $A_{up}(x_{up}, y_{up})$  and  $B_{down}(x_{down}, y_{down})$ , calculate the Euclidean distances between them— $d$  and  $d'$  respectively, and then calculate the relative errors:

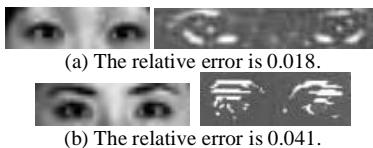


Figure 9. Sample diagram of the test result from the JAFFE human face database

By making use of the data obtained from the test, the relative error curve can be drawn, which is shown in the Figure. The ordinate represents the relative error  $d$  (eye) values, and the abscissa refers to the percentage (%) of eye image number tested in corresponding relative error  $d$  (eye) among the total number of images (500).

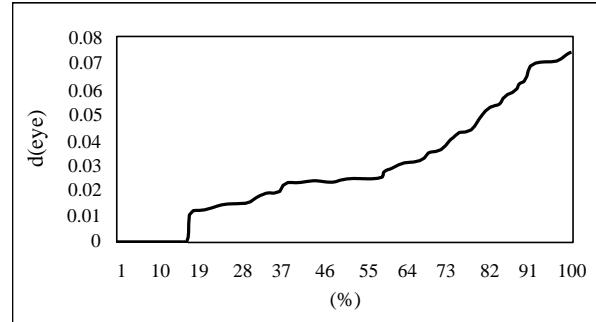


Figure 10. The relative error curve

Several experimental result figures are shown as in Fig. 9, some of which belong to pictures of daily life. (a) refers to the original image, (b) shows the skin color similarity image, (c) is the binary image after morphological filter processing, and (d) represents the image of human face and eye positioning results (the human face is marked with the green frame, and the human eyes are marked with the red frame).

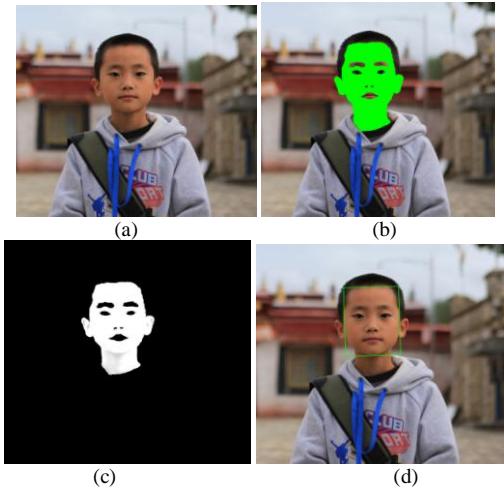


Figure 11. Execution process of the test result from a complicated background

Table 1 has listed the execution time comparison of each step of the precise positioning fig.11 (a)~(d) of this algorithm and the relevant algorithm (operating in a computer with the CPU frequency of 1.8GHz).

TABLE I. EXECUTION TIME COMPARISON OF EACH STEP OF THE PRECISE POSITIONING FIGURE 11 (A) OF THIS ALGORITHM AND THE RELEVANT ALGORITHM

	Positioning of the true human face region (ms)	Precise eye positioning (ms)	Total execution time (ms)
Literature [24] method	92	289	381
Literature [13] Hough transformation method	125	388	513
Method of this paper	68	208	276

In accordance with the data of test results, we can see that the above algorithm can be used to rapidly and

accurately position human eyes during human face detection. During the test to accurately position human eyes with the Hough ellipse detection algorithm, the maximum relative error  $d(\text{eye})$  between the obtained extreme values of upper and lower eyelids and the actual position is 0.104. In other words, among the 50 images used for test, within the relative error scope of 0.104, it is regarded that human eyes have been accurately positioned, and the accuracy of this algorithm can reach 0.104. In the meantime, during the test to accurately position human eyes with the algorithm proposed in this paper in a complicated background, we find this algorithm can also conduct rapid and accurate positioning. We can see the algorithm proposed in this paper has a low time complexity.

As can be seen from Table 1, the algorithm provided by this paper requires less time for the positioning of true human face region and precise eye positioning compared with the literature [24] method and the literature [13] Hough transformation method, the increased rate of total execution time is by 38.04% and 98.9% respectively. Wherein, the average computation time refers to the mean obtained by 10 times of algorithm. Through contrast, it can be seen that the computing result of Hough transformation of this paper is basically in consistency with the classical Hough transformation calculation, but the speed of the former has been doubly increased, which is significant. This is mainly because 5 parameters are required to be tested by the classical Hough transformation ellipse test method. The calculated amount of it is  $(n*n)^5$ , in which, n refers to the size of image. However, the amount of the algorithm provided by this paper is only  $(n*n)^3$ . The larger the image is, the more ellipses shall be detected, and the higher the efficiency of this algorithm will be.

## VI. CONCLUSION

Fatigue driving has always been a main cause of car accidents, so it has important realistic significance to detect the driver's fatigue state and reduce the accidents caused by fatigue driving. Positioning of human eyes is the precondition to build the detection system of fatigue driving. In this paper, an innovative method for accurate positioning of human eyes based on the color space of skin color clustering and Hough transformation ellipse detection is proposed, in other words, the region segmentation method based on the clustering color space is used to conduct coarse positioning of eyes first, and then, the accurate positioning of eyes based on the Hough transformation of ellipse detection is conducted. This can significantly improve the computation speed of eye positioning, which can also increase the accuracy and robustness of positioning. How to conduct fatigue test to the already accurately positioned human eyes is our next research task.

## ACKNOWLEDGMENT

This research is partially supported by the Science and Technology Planning Project of Hunan Province (No.

2012GK3096, No. 2012FJ4334); the Natural Science Foundation of Hunan Province of China (No. 12JJ6057); and Scientific Research Fund of Hunan Provincial Education Department (No. 12C1158, No. 13B132, No. 10C0374, NO. 13C613 and NO. 13C614).

## REFERENCES

- [1] M. Reinders, R. Koch, and J. Gerbrands, "Locating facial features in image sequences using neural networks," *FG*, pp. 230-235, 1996.
- [2] Z. W. Zhu, K. K. Fujimura, and Q. Ji, "Real-time eye detection and tracking under various light conditions," in *Proceedings of ACM SIGCHI Symposium on Eye Tracking Research and Applications*, New Orleans: Academic, 2002, pp. 139-144.
- [3] X. Liu, F. Xu, and K. Fujimura, "Real-time eye detection and tracking for driver observation under various light conditions," in *IEEE Intelligent Vehicle Symposium*, Versailles: Academic, 2001, pp. 18-20.
- [4] J. Bala, K. DeJong, and J. Huang, Visual routine for eye detection using hybrid genetic architectures," in *Proceedings of the 13<sup>th</sup> International Conference on Pattern Recognition*, E. Backer and E. Gelsema, Eds. Los Alamitos: IEEE CS Press, 1996, pp. 606-610.
- [5] G. C. Feng, P. C. Yuen, "Multi cues eye detection on gray intensity image," *Pattern recognition*, vol. 34, pp. 1003-1046, May 2001.
- [6] J. Huang, D. Li, and X. Shao, "Pose discrimination and eye detection using support vector machines(svms)," in *Proceeding of NATO-ASI on Face Recognition: From Theory to Application*, 1998, pp. 528-536.
- [7] C. Wenming and F. Hao, "Study of an algorithm for face pose adjustment based on eye location," in *Proceedings of the 5<sup>th</sup> World Congress on Intelligent Vontrol and Automation*, 2004, pp. 4190-4194.
- [8] J. Zhang, X. F. Yang, and R. L. Zhao, "Eye Detection Based on Hough Transform," *Journal of Computer Engineering and Application*, vol. 27, pp. 43-71, 2005.
- [9] X. H. Sun, G. Y. Chen, C. X. Zhao, and J. Y. Yang, "Gaze Estimation of Human Eye Based on Hough Transform and Gradient Information," *Journal of Chinese Computer Systems*, vol. 6, pp. 1123-1128, 2007.
- [10] L. Guo, Z. S. Tang, "Specification and verification of the triple-modular redundancy fault-tolerant system," *Journal of Software*, vol. 14, pp. 28-35, Jan 2003.
- [11] S. K. Singh and D. S. Chauhan, "A robust skin color based face detection algorithm," *Tamkang Journal of Scienc and Engineering*, vol. 6, pp. 227-234, April 2003.
- [12] T. Kanungo, D. M. Mount, and N. S. Netanyahu, "An efficient K-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881-892, July 2002.
- [13] Z. H. Zhou and X. Geng, "Projection functions for eye detection," *Pattern Recogntion*, vol. 37, pp. 1049-1056, May 2004.
- [14] J. G. Wang, S. Eric, and V. Ronda, "Eye gaze estimation from a single image of one eye," in *Proceedings of the 9th IEEE International Conference on Computer Vision Nice*, France: Academic, 2003, pp. 136-143.
- [15] T. Q. Wang, G. F. Xing, and B. Jiang, "Algorithm for Detection and Locating Human Face Based on Region Segmenation in Complex Background," *Computer Engineering and Degisn*, vol. 11, pp. 2090-2092, 2004.
- [16] W. Zhong, Z. M. Liu, and J. L. Zhou, "Study on Precise Eyes Location in Face Detection," *Journal of Computer Engineering and Application*, vol. 36, pp. 73-76, 2004.

- [17] G. S. Xiang,X. Y. Wang, and D. T. Liang, "Real-time detection and tracking algorithm based on the color and character of human face," *Opto-Electronic Engineering*, vol. 4, pp. 44-48, 2007.
- [18] V. Hemige, "Object-Oriented design of the groupware layer for the ecosystem information system," University of Montana, 1995.
- [19] A. Rose, M. Perez, and P. Clements, *Modechart toolset user's guide, Technical Report*, NML/MRL/5540-94-7427, Austin: University of Texas at Austin, 1994.
- [20] Bing Feng, Xiao Qing Ding, Off-line handwritten Chinese character recognition with hidden Markov models, *in: 5th International Conference on Signal Processing Proceedings, WCCC-ICSP 2000*, vol. 3, pp. 1542–1545.
- [21] A.V. Nefian, Embedded Bayesian networks for face recognition, *Proc. of the IEEE International Conference on Multimedia and Expo*, Vol. 2, 26-29 August 2002, Lusanne, Switzerland, pp. 133-136
- [22] Petar S. Aleksic, Member, IEEE, and Aggelos K. Katsaggelos, Fellow, IEEE. Automatic Facial Expression Recognition Using Facial Animation Parameters and Multistream HMMs. *IEEE Transactions on Information Forensics and Security*, Vol. 1, No. 1, pp. 3-11, March. 2006.
- [23] Michael J L, Julien Budynek, Shigeru A kamatsu. Automatic Classification of Single Facial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21 (12) pp. 1357-1362.
- [24] Zhong Wei, Liu Zhiming, Zhou Jiliu. Study on Precise Eyes Location in Face Detection. *Computer Engineering and Applications*, 2004, 36, pp. 72-76.

**Qifeng Yang** received her B. Eng. and M. Eng. degrees degree from Nanjing University, Jiangsu, China and National Defense Science and Technology University, ChangSha, Hunan, China in 1996 and 2004, respectively. She is currently pursuing her Ph. D. degree in the School of software, Central South University,

Changsha, Hunan, China. Her current research interests are in the areas of data fusion and computer vision.

**Huosheng Hu** received his Ph. D. degree from Oxford University, UK. He is a Professor in Computer Science at the University of Essex, UK, leading the Human Centred Robotics Group. His research interests include biologically inspired robotics, service robots, human-robot interaction, evolutionary robotics, data fusion, artificial life, embedded systems, pervasive computing and RoboCup. He is also a Chartered Engineer, a senior member of IEEE and a member of IEE, AAAI, IAS, IASTED and ACM.

**Weihua Gui** received the degree of the B. Eng. (Automatic Control Engineering) and the M. Eng. (Control Science and Engineering) from Central South University, Changsha, China in 1976 and 1981, respectively. From 1986 to 1988 he was a visiting scholar at Universitat-GH-Duisburg, Germany. He has been a academician of Chinese Academy of Engineering, a full professor in the School of Information Science & Engineering, Central South University, Changsha, China, since 1991. His main research interests are in modeling and optimal control of complex industrial process, distributed robust control, and fault diagnoses.

**Shu-Ren Zhou** received his PhD in 2009 from Central South University, China. He received his MS and BS in 2004 and 1999 respectively from the Changsha University of Science& Technology and Central South University, China. His currently a post-doctoral at the National University of Defense Technology. His current research interests include image processing, pattern recognition, human pose estimation and computer vision.

**Can Zhu** received his Ph. D. in Computer Application from Central South University, China, in 2009. He is engaged as a lectorate of Changsha University of Science & Technology, China. His research interests involve modeling and optimization concerned transportation planning and management.

# Unsupervised Learning and Linguistic Rule Based Algorithm for Uyghur Word Segmentation

Turdi Tohti and Winira Musajan

School of Information Science and Engineering, Xinjiang University, Urumqi, China

Email: turdy@xju.edu.cn, winira@xju.edu.cn

\*Askar Hamdulla

School of Software Engineering, Xinjiang University, Urumqi, China

\*Corresponding author, Email: askar@xju.edu.cn

**Abstract**—Inter-word spaces based traditional word segmentation method not very appropriate for multi-word structured semantic words due to the fact that it will split the semantic words into several fragments that inconsistent with its original meaning. So, this will be a bottleneck problem in Uyghur text analysis and text understanding applications. This paper puts forward a new idea and related algorithms for segmentation of Uyghur multiword structured semantic words. In this algorithm, the word based Bi-gram and contextual information are derived from large scale raw text corpus automatically, and according to the association rules between Uyghur words, the liner combinations of mutual information, difference of t-test and dual adjacent entropy are taken as a new measurement(*dmd*) to estimate the agglutinative strength between two adjacent Uyghur words. The experimental result on large-scale open tests shows that the proposed algorithm achieves 88.21% segmentation accuracy.

**Index Terms**—Uyghur Language; Semantic Word; Combined Statistics; Word Association Rule; Semantic Word Segmentation

## I. INTRODUCTION

Word segmentation is the first step and also is a key step in natural language processing (NLP), what method to use and its difficulty is different in different language environment. But the ultimate goal is to get the minimum use of linguistic units that express specific independent semantics.

Uyghur language belongs to Turkish language group of Altaic language family, and also belongs to the agglutinating language on structure grammar, is a kind of alphabetic writing. Looking on the surface, Uyghur text is a word sequences that separated by inter-word spaces and on this feature is similar to English. For this reason, word segmentation always have been ignored in Uyghur natural language processing and uses the inter-word space as a natural separator simply to obtain the words in the text, is the only word segmentation method so far. But from the Uyghur word's semantic independence and integrity perspective, can be divided into two categories that *single-word structured semantic word* (SSSW) and *multi-word structured semantic word* (MSSW).

**Definition 1** (SSSW): Is a Uyghur word also is a no space alphabetic string, complete and independent on its semantics, can be obtains by *inter-word spaces based word segmentation* (ISWS).

**Definition 2** (MSSW): Stable combination of several single-word strings and satisfying the following conditions:

1) Association patterns of two or more words (Commonly is double word or three word structured) and separated by inter-word spaces.

2) Complete and independent on its semantics, and inseparable on its structure.

However, as the Uyghur natural language processing related works constantly goes deeper and wider range of development, ISWS began to expose its potential pitfalls and limitations [1].

In Uyghur Web search, because a MSSW will be split into several fragments that inconsistent with its original meaning, so these word fragments cannot be play to the rule of keywords in text indexing, and also leads to a lot of problems that large scale of word list , big index size, and even low search precision [2].

In the text classification and clustering that the words taken as features, high dimensionality and inter-class cross features is the main factor to restrict the performance of classification and clustering algorithms [3], and the traditional word segmentation method (ISWS) would makes this situation even more serious in Uyghur classification and clustering [4].

Word segmentation also is a bottleneck in machine translation [5], keyword extraction and unknown word processing [6], as well as Uyghur personal name identification (names former surname, separated by a space) and so on [7]. So, the needs of research a kind of automatic word segmentation method and be able to extract the Uyghur words or words associations that stable on its structure grammar, specific and independent on its semantics is increasingly prominent and urgent.

Word segmentation research has a long history in Chinese natural language processing [8], and formed more mature technology and practical segmentation tools [9]. However, the booming popularity of WWW and electronic publications puts forward a series of new

issues on the Chinese automatic segmentation research. Especially the demands of unknown words processing, open environmental adaptability and robustness of word segmentation system have become increasingly prominent. Therefore, more and more scholars have realized that the massive electronic text with extreme ease should be an important resource, and directly obtaining certain applicable knowledge from raw corpus by machine learning method should be an important supplement of automatic word segmentation [10-16].

Our work is closest to that of Maosong S and Sili W, but differs in two important ways. First, we introduced another statistics called dual adjacent entropy (*dae*), and the liner combinations of difference of t-test (*dts*), mutual information (*mi*) and *dae* as a new measurement to estimate the agglutinative strength of adjacent Uyghur words. Second, we introduced the word association rule (WAR) of Uyghur language features and improved the segmentation accuracy further more.

The basic task of our work is to achieve a Uyghur semantic word segmentation algorithm based on unsupervised learning and non-dictionary strategy, in addition to stemming pretreatment, all statistic information required by the algorithm are automatically derived from raw corpus without manual intervention, and the purpose is to test the algorithm effectiveness in the open environment that completely simulate in practical application.

We have a large corpus collected from the network and formal publications, including the text corpus from Internet (20 classes), all content of the "Xinjiang Daily" in March and April 2008, and eight books (about history, culture, law, politics, economy, etc). In our work, the large corpus is divided into three corpuses and each corpus contains certain percentage of the content in the above large corpus.(1) Uyghur Raw Corpus (URC): contains a total of 9,443,290 Uyghur words and punctuations without segmentation.(2) Uyghur Cooked Corpus(UCC<sub>1</sub>)for closed test: contains a total of 135708 Uyghur words and punctuations after word segmentation in manual way.(3) Uyghur Cooked Corpus(UCC<sub>2</sub>)for open test: contains a total of 154411 Uyghur words and punctuations after word segmentation in manual way. All corpuses above are provided by Xinjiang University Key Laboratory of Intelligent Information Processing and also been processed by stem extraction [17].

## II. INTER-WORD POSITION JUDGMENT BASED ON STATISTICAL MESUREMENT

As a semantics specific and independent linguistic unit, internal association degree of a semantic word is relatively close and its relations with the external context are relatively loose [18]. So, it is possible to judge that the words are independent of each other or strongly associated and forms a semantic word according to the association degree between adjacent Uyghur words [19].

The basic approach is to use a statistical measurement *S* to observe the association degree between adjacent words, if  $S > T$  (*T* for threshold), then keep the "connected" between them, otherwise insert a separator

to separate them (because of Uyghur words are separated by spaces, so ":" is used as the separator instead of space), then the adjacent words are "disconnected" and cannot be form a semantic word. For example,  $W_1 \sim W_n$  is a word string of *n* words, and inter-word position judgment ("connected" or "disconnected") based on statistical measurement *S* as shown in Fig. 1.

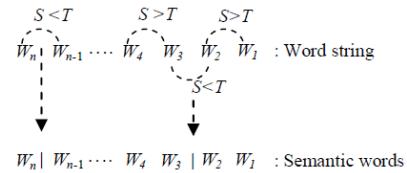


Figure 1. Inter-word position judgment based on statistical measurement

It is easy to calculate the statistical measurement based on the words Bi-gram information be derived from the large scale corpus, such as mutual information, difference of t-test, coupling degree, and so on. On the inter-word position judgment, we can take a kind of basic statistics or some combination of them as the statistical measurement *S*, to estimate the agglutinative strength between adjacent Uyghur words. Of course, uses the combined statistics can be make more accurately judgment than using alone a basic statistics, because it is the comprehensive reflections of association capacity of adjacent words.

In our research, the judgment accuracy of "connected" and "disconnected" between words are taken as an inspection object, constantly adjusting the threshold *T* and other parameters until the segmentation algorithm is at its best, and the segmentation accuracy  $\alpha$  is defined as shown in formula (1):

$$\alpha = \frac{Pos_{con}(x, y) + Pos_{discon}(x, y)}{Pos(x, y)} \quad (1)$$

In formula (1), *x* and *y* are arbitrary adjacent Uyghur words,  $Pos_{con}(x, y)$  is stands for the inter-word position numbers of correctly "connected",  $Pos_{discon}(x, y)$  is stands for the inter-word position numbers of correctly "disconnected", and the  $Pos(x, y)$  is stands for the total number of inter-word position in the text.

### A. Basic Statistical Measurement: Mutual Information

According to the principles, mutual information (*mi*) between adjacent Uyghur words *A* and *B* is defined as shown in formula (2):

$$mi(A, B) = \log_2 \frac{P(A, B)}{P(A)P(B)} \quad (2)$$

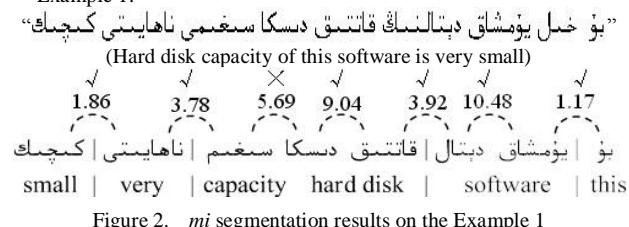
Among them,  $P(A, B)$  is the probability of the adjacent words "A B" appears in the large corpus,  $P(A)$  and  $P(B)$  are the probability of word *A* and *B* appears in the large corpus.

$mi(A, B)$  reflects the closeness between adjacent words *A* and *B*, if  $mi(A, B) \geq 0$ , then "A B" is strongly associated; if  $mi(A, B) \approx 0$ , then "A B" is weakly associated; If  $mi(A, B) < 0$ , then "A B" is mutually exclusive. With the

increase of  $mi(A,B)$ , the degree of association is also increased, and if  $mi(A,B)$  is greater than a given threshold  $T_{mi}$ , then we can believe that the “A B” is a semantic word.

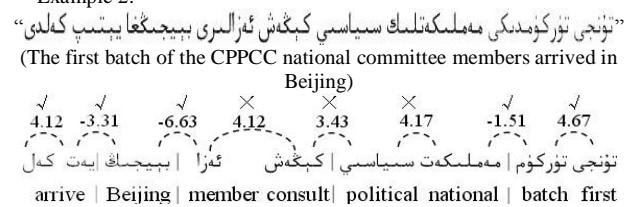
We training the Uyghur words Bi-gram model on the raw corpus URC and examine the  $mi$  judgment distribution about inter-word positions “connected” or “disconnected” on the UCC<sub>1</sub>. Value of  $mi$  changes in the range between -6.75 and 21.01, when the threshold  $T_{mi}$  is 4 (mean value of  $mi$  is 3.63 according to URC statistics) the value of  $\alpha$  is up to a maximum of 75.26%. Such as, the judgment of each position on Example 1 is substantially correct, misjudged only one position (As shown in Fig. 2).

Example 1:

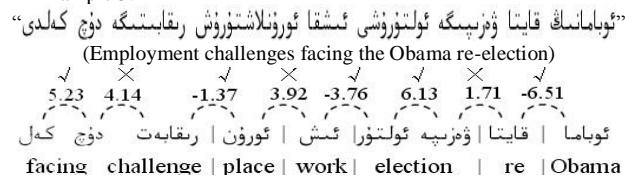
Figure 2.  $mi$  segmentation results on the Example 1

It is can be seen from the formula (2),  $mi$  reflects the static association capacity between the adjacent word A and B, and regardless of the context in which they are. So, there will be wrong judgment if we rely solely on the  $mi$  as absolute measurement. Such as, misjudgment occurred all in example 2 and 3, appears low segmentation accuracy (as shown in Fig. 3 and Fig. 4).

Example 2:

Figure 3.  $mi$  segmentation results on the Example 2

Example 3:

Figure 4.  $mi$  segmentation results on the Example 3

#### B. Basic Statistical Measurement: Difference of T-Test

The concept of t-test was introduced at the earliest and uses to measuring the association degree of an English word A and any other two words X and Y [20]. According to the definition, a Uyghur word string “X A Y” and its t-test value calculation as shown in formula (3):

$$t_{x,y}(A) = \frac{p(Y|A) - p(A|X)}{\sqrt{\sigma^2(p(Y|A)) + \sigma^2(p(A|X))}} \quad (3)$$

In this formula,  $p(Y|A)$  and  $p(A|X)$  are the Bi-gram probability of adjacent words “A Y” and “X A”,  $\sigma^2(p(Y|A))$  and  $\sigma^2(p(A|X))$  are the variance of them respectively. It is could be seen by this formula that, if  $t_{x,y}(A) > 0$ , the association strength of A and with its subsequent Y is greater than with its precursor X, then A should be to break with the X and to be connect with the Y. if  $t_{x,y}(A) < 0$ , the association strength of A and with its precursor X is greater than with its subsequent Y, then A should be to break with the y and to be connect with the X. if  $t_{x,y}(A) = 0$ , then the association strength between A and X is equal to between A and Y, so it is couldn't be able to determine if they are should be break or connect.

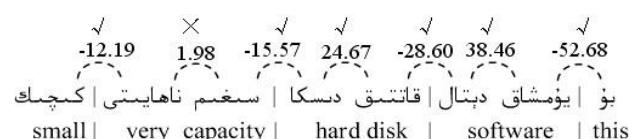
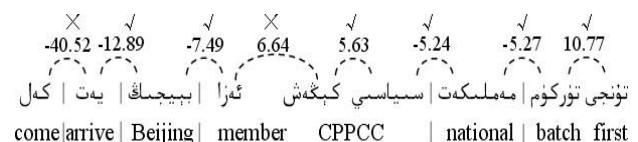
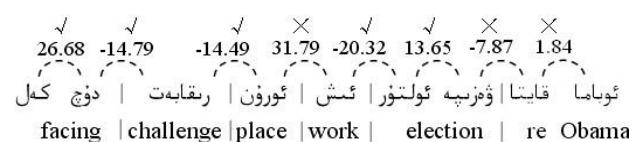
The t-test is a statistical measurement that based on word statistics and rather than based on the inter-word position, so in order to directly used to the Chinese word segmentation and measuring the association probability of adjacent Chinese words, the t-test formula has been improved and proposed the concept of difference of t-test ( $dts$ ), and applied to Chinese word segmentation combined with  $mi$  [10].

According to the definition, the adjacent words “A B” in Uyghur word string “X A B Y” and its  $dts$  value calculation as shown in formula (4):

$$dts(A,B) = t_{x,y}(A) - t_{A,Y}(B) \quad (4)$$

When  $dts(A,B) > T_{dts}$  ( $T_{dts}$  is threshold), the inter-word position of “A B” is more tending to judge for “connected”, and to judge for “disconnected” otherwise. We examined the  $dts$  judgment distribution about inter-word positions “connected” or “disconnected” still on the UCC<sub>1</sub>.

Changes of  $dts$  in the range between -264.14 and 108.41, and the value of  $\alpha$  is up to a maximum of 78.14% when  $T_{dts} = 0$ . Compared with the  $mi$ ,  $dts$  have higher judgment accuracy, but there are some difference with  $mi$  on the positions judgments in example 1, 2, and 3(shown in Fig. 5, Fig. 6, and Fig. 7).

Figure 5.  $dts$  segmentation results on the Example 1Figure 6.  $dts$  segmentation results on the Example 2Figure 7.  $dts$  segmentation results on the Example 3

### C. Basic Statistical Measurement: Dual Adjacent Entropy

As a frequently used language unit, semantic words have a certain degree of circulation in the real text and could be applied in a variety of different context environment, rather than a temporary combination of a special context [21-22]. Therefore, it is possible to estimate the independent linguistic unit possibility of adjacent words according to its contextual environment variability, so as to determine the inter-word position of them is “connected” or “disconnected”.

The contextual variability of adjacent words can be measured by information entropy [23]. To calculate the right and left information entropy of adjacent words “A B”, if they are greater than a given threshold, then can believe that there have varied contextual environment of “A B” and they will be a independent language unit [24], so we can make a judgment that the A and B are “connected”, otherwise “disconnect”.

However, it was found in our segmentation research that, the left and right adjacent entropy were taken as a measurement cannot overall segmentation out the semantic words that three-word structured. For example, making a judgment on the inter-word position of A and B in three-word structured semantic word “A B C”, “A B” may be have varied changes in its left adjacency, but its right adjacency is fixed to the C. According to the definition of information entropy, the right adjacent entropy of “A B” is equal to 0(minimum), and then the inter-word position between A and B (also between B and C) will be wrongly judged to “disconnect”.

For the above situation, we proposed a kind of statistical measurement called dual adjacent entropy (*dae*) and very suitable for the research needs of this paper.

**Definition 3(dae):** For the Uyghur word string “x A B y”, x and y are the contextual right and left adjacency of “A B” each appears in text and constitutes a dual adjacency of  $\langle x, y \rangle$ , then we will have an adjacency set that includes all dual adjacency of “A B” and defines as  $S_{da} = \{ \langle x_i, y_i \rangle \}$ . Supposing the  $m$  is stands for the total number of dual adjacency in  $S_{da}$ ,  $c$  is stands for the species number(the number of no repeated dual adjacency) ,  $n_i$  is stands for the frequency of each kind of dual adjacency ,then the information entropy (dual adjacent entropy)calculation for dual adjacency set of “A B” as shown in formula(5):

$$dae(A, B) = -\sum_{i=1}^c \frac{n_i}{m} \log\left(\frac{n_i}{m}\right) \quad (5)$$

It is informed by the formula that, the minimum theoretical value of *dae* is equal to 0(when  $c=1$ ), and the maximum theoretical value is  $\log(m)$  (when  $c= m$ ), the greater value of  $dae(A, B)$  is indicating that the contextual environment of “A B” is more variable and there is greater possibility of became an independent linguistic unit. If the  $dae(A, B)$  is smaller, shows that the independence of “A B” is not strong, and there is may be an accidental association between A and B. Therefore, when  $dae(A, B) > T_{dae}$  ( $T_{dae}$  is threshold), the position

between A and B is more tending to “connected”, otherwise “disconnected”.

For example, a double-word structured semantic word “A B” appears a total of five times in corpus, and its contextual language environments respectively are “X A B Y”, “Y A B C”, “Z A B X”, “W A B Y”,and“ V A B X”, dual adjacency set of “A B”is  $S_{da} = \{ \langle X, Y \rangle, \langle Y, C \rangle, \langle Z, X \rangle, \langle W, Y \rangle, \langle V, X \rangle \}$ ,  $m=5$ ,  $c=5$ , then dual adjacent entropy of “A B” is:

$$dae(A, B) = -\frac{1}{5} \log \frac{1}{5} - \frac{1}{5} \log \frac{1}{5} - \frac{1}{5} \log \frac{1}{5} - \frac{1}{5} \log \frac{1}{5} - \frac{1}{5} \log \frac{1}{5} \\ - \frac{1}{5} \log \frac{1}{5} = 0.699$$

We examined the *dae* judgment distribution about inter-word positions “connected” or “disconnected” still on the UCC<sub>1</sub>. Changes of *dae* is in the range between 0.06 and 1.37, and the value of  $\alpha$  is up to a maximum of 73.23% when the  $T_{dae} = 0.60$ .

Compared with *mi* and *dts*, we have slightly lower segmentation accuracy by *dae* judgment, and most of the errors occurred on the positions that should be “disconnected”, but it will work well on the unknown words that rarely appeared in the corpus.

For example, Uyghur word A and B are two independent linguistic units that frequently uses in the real text and they also will be adjacently to constitute a specific new word (double-word structured semantic word) “A B”, because of the rarely appearance of “A B” in the large corpus, there will be the maximal of *count* (A) and *count* (B), and the minimal of *count* (A B). In this case, the *mi* and *dts* almost making a wrong judgment on the inter-word position, but the judgment of *dae* is correct, because it is only considered the change diversity of contextual language environment and has little to do with word frequency in *dae* judgment.

A practical example illustrates this situation in our work, the new word “فۇش زۇكام”(bird flu) appears in the corpus a total of 17 times, the word“فۇش”(bird) appears 2378 times ,and the word“زۇكام”(flu) appears 4927 times. Therefore,  $mi = 3.78$ ,and  $dts = 0.48$ ,then the inter-word position of “فۇش زۇكام” is “disconnected” if judged by *mi* or *dts*,but the  $dae = 0.96$ ,then the inter-word position is judged to “connected” by *dae*. In the following judgment on example 4, *mi* and *dts* all are wrong, only the judgment of *dae* is correct (as shown in Fig. 8, Fig. 9, and Fig. 10).

Example 4:  
“ئالىملار قۇش زۇكىمى ۋېرىسىنى تەتقىق قىلىپ ياساب چىقىسى”

(Scientists developed the bird flu virus)

✓	✓	✓	✓	✗	✓	✓
0.63	0.23	1.07	0.15	0.63	0.96	0.15

ئالىم | قۇش زۇكام ۋېرىسۇن | تەتقىق قىلىپ ياساب چىقىسى

create | research | virus bird flu | scientist

Figure 8. *dae* segmentation results on the Example 4

However, it is the *mi*, *dts* or *dae*, because they all are the basic statistical measurement and only considered one

aspects of statistical information, so there are some inevitable limitations if uses only one kind of them to measuring inter-word position.

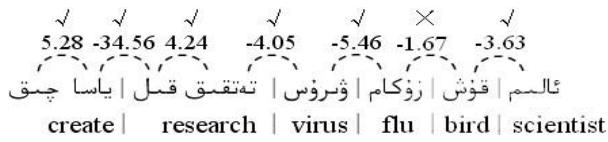


Figure 9. *dts* segmentation results on the Example 4

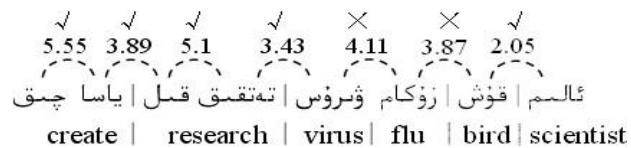


Figure 10. *mi* segmentation results on the Example 4

#### D. Combined Statistical Measurement: Dmd

In our work, we carried on word segmentation on the corpus UCC<sub>1</sub> separately used alone the basic statistical measurements. Among them, *dts* showed the higher segmentation accuracy ( $\alpha = 78.14\%$ ), followed by *mi* ( $\alpha = 75.26\%$ ), and is *dae* ( $\alpha = 73.23\%$ ).

The case has been successful in Chinese word segmentation that to combine the basic statistical measurements and take advantage of each other [10] [11], and it was also been found in our segmentation experiment that there is certain large feasibility of combination complementary. Therefore, we give priority to *dts* and take the linear combination of above three basic statistics as a combined statistical measurement (*dmd*) to estimate the agglutinative strength between two adjacent Uyghur words. Because of the quite difference of value range (*dts*: -264.14~108.41, *mi*: -6.75~21.01, *dae*: 0.06~1.37), so it is also have to be normalizing before make the linear combination of them. As shown in formula (6), (7), and (8):

$$dts^*(A, B) = \frac{dts(A, B) - \mu_{dts}}{\sigma_{dts}} \quad (6)$$

$$dae^*(A, B) = \frac{dae(A, B) - \mu_{dae}}{\sigma_{dae}} \quad (7)$$

$$mi^*(A, B) = \frac{mi(A, B) - \mu_{mi}}{\sigma_{mi}} \quad (8)$$

The  $\mu_{dts}$ ,  $\mu_{mi}$ ,  $\mu_{dae}$  are respectively stands for the mean value of *dts*, *mi* and *dae*, and the experimental values of them are sequentially for -6.51, 3.63, and 0.52. The  $\sigma_{dts}$ ,  $\sigma_{mi}$ ,  $\sigma_{ea}$  are respectively stands for the mean square error of *dts*, *mi*, and *dae*, and the experimental values of them are sequentially for 24.29, 3.54, and 0.31. Then make a linear combination of them as shown in formula (9):

$$dmd(A, B) = dts^*(A, B) + \lambda \times mi^*(A, B) + \gamma \times dae^*(A, B) \quad (9)$$

In the formula (9),  $\lambda$  and  $\gamma$  are determined by the experiment and be found there is higher segmentation effects when  $\lambda = 0.35$ , and  $\gamma = 0.30$ . The variation range of *dmd* on UCC<sub>1</sub> is between -11.5 and 6.9, and the  $\alpha$  is up to a maximum of 84.31% ( $T_{dmd} = 0$ ), respectively increased by 6.17%, 9.05% and 11.08% compared with separately using of *dts*, *mi* and *dae*.

In our segmentation experiment on the example 4, *dts*( $\text{قوش}, \text{زۇكام}$ )=-1.67, *mi*( $\text{قوش}, \text{زۇكام}$ )=3.87, and *dae*( $\text{قوش}, \text{زۇكام}$ )=0.96, then there have wrong judgments (“disconnected”) made by *dts* and *mi* (*dae* judgment is correct). After normalization, *dmd*=0.38> $T_{dmd}$ , and then the inter-word position ( $\text{قوش}, \text{زۇكام}$ ) is correctly judged to “connected” after uses the *dmd* measurement

For the inter-word position “ $\text{زۇكام}, \text{ۋىرۇس}$ ” also on the example 4, *dts* ( $\text{زۇكام}, \text{ۋىرۇس}$ )=-5.46, *mi* ( $\text{زۇكام}, \text{ۋىرۇس}$ )=4.11, and *dae* ( $\text{زۇكام}, \text{ۋىرۇس}$ )=0.63, then there have wrong judgment(“disconnected”) made by *mi* and *dae* (*dts* judgment is correct). After normalization, *dmd*=-0.07< $T_{dmd}$ , and then the inter-word position( $\text{زۇكام}, \text{ۋىرۇس}$ ) is correctly judged to “disconnected” after uses the *dmd* measurement (as shown in Fig. 11).

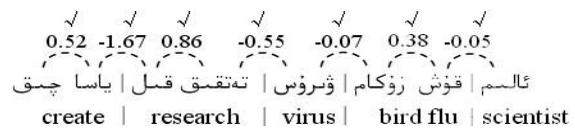


Figure 11. *dmd* segmentation results on the Example 4

It is thus clear that, *dmd* judgment is consistent with *dts*, *mi*, and *dae* when they have agreements on their judgment, and also there is a certain degree of complementarity when their judgment is not consistent.

### III. LINGUISTIC FEATURES AND WORD ASSOCIATION RULE

In our work, using the combined statistical measurement *dmd* to estimate the inter-word positions and the segmentation accuracy reached 84.31%, but there is still some distance with the accuracy of the ideal. Therefore, we look for some information from the Uyghur language itself that could be helpful for inter-word position judgment, and found the following linguistic features different from Chinese.

**Feature 1:** There are small quantities of Uyghur words but frequently appearance in the text that like auxiliary words (بىرلاق, ياكى), conjunctions (ئىدى, ئىكىن), adverbs (دانە, نېپەر), quantifiers (بەك, ئاران), pronouns (مەن, سەن), and interjections (ڭاھ, پاھ) etc, this kind of words has never associate with other words and to constitute a semantic word, be referred to as *independent word* (IW) in this paper.

**Feature 2:** word combinations mainly occur between in the nouns (N), adjectives (ADJ), and verbs (V) in Uyghur language. Among them, Adjectives always as a precursor of a semantic word and it will not appear in the following position when the adjectives associated with

nouns or verbs. Therefore, adjacent words like “N+ADJ” or “V+ADJ” relations never to be constitute a semantic word.

In our work, we have statistics the number of independent words and adjectives and their frequency on the raw corpus URC. There are 592 independent words with total frequency of 1634860, accounted for 17.3% of the corpus. Adjectives are 988 with total frequency of 1170968, accounted for 12.4% of the corpus. We also went to observe the inter-word positions that wrongly judged by *dmd* on the UCC<sub>1</sub>, found that 3.96% of 15.69% errors just occurred on the independent words and adjectives, and all the “*disconnected*” be wrongly judged to “*connected*”.

According to the above situations and also in accordance with the Feature 1 and Feature 2, the rules uses for inter-word position judgment defined as follow.

**Definition 4** (word association rule: WAR): For the adjacent Uyghur words “A B”, IW stands for the independent words, Pos(A,B) stands for the inter-word position of “A B”, ADJ stands for the adjectives, then:

```

If A ∈ {IW} or B ∈ {IW} then
  Pos (A, B) ← “disconnected”
else if B ∈ { ADJ } then
  Pos (A, B) ← “disconnected”
else then
  Pos (A, B) ← “connected” or “disconnected”

```

Therefore, we have established two kind of word list that separately includes independent words and adjectives; the purpose is to judge the inter-word positions whether like feature 1 or feature 2 based on the word list. In this way, on the one hand to reduce the *dmd* computations for considerable part of inter-word positions; on the other hand, to judge inter-word positions use the known rule of linguistic features, this will very helpful to improve the accuracy.

#### IV. SEMANTIC WORD SEGMENTATION ALGORITHM BASED ON DMD AND WAR

After the determination of combined statistical measurement *dmd* and the WAR in accordance with linguistic features, the Uyghur semantic word segmentation algorithm is as follows:

**Step 1:** Do stem extraction for the text will be processed, and all the punctuations replaces with predefined separator “.”.

**Step 2:** For the inter-word position of any adjacent words “A B”, calculate the *dmd* values of all inter-word positions in the text to be processed under the following conditions:

```

if A ∈ {IW} or B ∈ {IW} or B ∈ {ADJ} then
  dmd (A, B) = -1 (less than the  $T_{dmd}$ )
else then

```

$$dmd(A, B) = dts^*(A, B) + \lambda \times mi^*(A, B) + \gamma \times dae^*(A, B)$$

**Step 3:** In turn, to extract the current inter-word position, perform step 4 until the end.

**Step 4:**

```
  If dmd >  $T_{dmd}$  then
```

```
    Pos (A, B) ← “connected”
```

```
  else then
```

```
    Pos (A, B) ← “disconnected”
```

## V. EXPERIMENTS

### A. Experimental Results

Such as mentioned above, we obtained the Uyghur word (stem) Bi-gram module by training on the raw corpus URC, observed and studied the segmentation efficiency of basic statistical measurements and adjusted the threshold, as well as determined the values of  $\lambda$  and  $\gamma$  in the formula (9) on the corpus URC1.

The experiment is designed in order to the following two purposes: the one is to examine the effectiveness of combined statistical measurement *dmd*; and the second is to verify the robustness of the segmentation algorithm in open environment. Therefore, we performed segmentation experiment separately on the closed test set and open test set, and observed the inter-word position correctness under the different measurements judgment.

After stem extraction and punctuations replacement, the number of words and inter-word positions needed to judge in the test set URC<sub>1</sub> and URC<sub>2</sub> are as shown in Table I.

TABLE I. WORDS AND INTER-WORD POSITIONS IN URC<sub>1</sub> AND URC<sub>2</sub>

Test set	words	Inter-word positions
Open test set (URC1)	120948	108998
Closed test set (URC2)	137757	125699

The open test and closed test experimental results under different segmentation strategy are summarized as shown in Table II and Table III.

TABLE II. CLOSED TEST EXPERIMENTAL RESULTS

Segmentation strategy	Inter-word positions correctly judged	Segmentation accuracy $\alpha$ (%)
<i>dts</i>	85174	78.14
<i>mi</i>	82024	75.26
<i>dae</i>	79819	73.23
<i>dmd</i>	91900	84.31
<i>dmd</i> +WAR	96211	88.27

TABLE III. OPEN TEST EXPERIMENTAL RESULTS

Segmentation strategy	Inter-word positions correctly judged	Segmentation accuracy $\alpha$ (%)
<i>dts</i>	97881	77.87
<i>mi</i>	95446	75.93
<i>dae</i>	90936	72.34
<i>dmd</i>	105150	83.65
<i>dmd</i> +WAR	110877	88.21

It is can be seen from the experimental results that, the algorithm performance has not decreased in the open test. In addition, the accuracy of *dmd* segmentation is invariably higher than that *dts*, *mi*, and *dae*, indicating that the proposed measurement *dmd* and its parameters determination are valid. Especially, *dmd* judgment will be more accurate with the assist of WAR.

### B. Problems and Future Work

In the experiment, we also found the following problems affecting the accuracy of Bi-gram statistical

information to a certain extent, and leads wrongly judgment on the inter-word position.

(1) Limitations of the stem extraction tool. Because of the existing stemming algorithm is more dependence on the pre-built stem library, so it is hard to work on the stem extraction for the unknown words.

(2) Spelling errors hard to avoid. There are more similar letters in Uyghur alphabet, and in addition the custom difference of each place people pronunciation (dialect), the correct spelling is not always an easy task in Uyghur language, such as even many Uyghur people cannot exactly spell the word: “ئۇنىۋېرسىتېت” (university). Therefore, spelling error is a main factor influencing the pattern matching and word frequency statistics.

(3) Non-standard abbreviations. Now that, the authorities have set the use of standard specification for the new words from foreign languages, but there are still arbitrariness of non-standard abbreviations on the internet. For example, the standard terminology “كۆمپیوتەر” (computer) is abbreviated to “كۆم”, the “ئېلېكترونلۇق يۈللانما” (E-mail) is abbreviated as “ئېلخەت” or “ئېمایل” etc, This will also influencing the pattern matching and word frequency statistics..

For the limitations of stem extraction, spelling error also is a major factor in addition to the shortcomings of the algorithm itself, and the existing Uyghur NLP methods and tools can not be effective on the fully automatic error detection and correction for batch processing to large text [25], and about problems (3), there have not any related research report up to now.

Obviously, whether it is the imperfections of text pretreatment algorithm or defects of the original text, all this situations will affects the term extraction accuracy and also affects the index quality and search precision. Therefore, as far as possible to eliminate these negative factors before semantic word extraction and to obtain the higher extraction accuracy on the normative text corpus is completely possible, this will be the research emphasis in our future work.

## VI. CONCLUSION

This paper discusses the traditional concept of Uyghur word segmentation, explores the limitations and problems causes by the commonly used method, and puts forward the idea of Uyghur semantic word segmentation and a kind of automatic segmentation algorithm for the first time. In this algorithm, no longer takes the inter-word space as a natural delimiter, but to estimate the agglutinative strength between adjacent Uyghur words and identifies the segmentation positions based on the strategy of combined statistics and rules.

Motivations and reasons of the strategy to use are in the following two aspects: One of the main reasons is that there never have word segmentation idea and the dictionary for word segmentation in Uyghur language. The second reason is that the many successful stories in Chinese word segmentation have proved the effectiveness of this method. Therefore, We have directly obtained words Bi-gram statistics from large-scale raw corpus,

introduced a new statistics that called dual adjacent entropy(*dae*), and take the liner combinations of mutual information, difference of t-test and dual adjacent entropy as a combined measurement (*dmd*) to estimate the inter-word positions which is connected or disconnected. When calculating the *dmd*, we also introduced the language features and improved the segmentation accuracy with the aid of Uyghur word combination rules.

Proposed algorithm shows the higher segmentation accuracy and robustness on the large-scale open test, but there is a lot of works have to continue in the mechanisms and methods, as well as the segmentation system and its verification.

## ACKNOWLEDGMENT

This work has been supported by the National Natural Science Foundation of China (61063022, 61262062, 61163033, and 61262063), Scientific Research Program of the Higher Education Institution of Xinjiang (XJEDU2012I11), and Program for New Century Excellent Talents in University (NCET-10-0969).

## REFERENCES

- [1] T. Turdi, M. Winira, H. Askar. “Efficient Term Extraction and Indexing Approach in Small-Scale Web Search of Uyghur Language”. *Journal of Multimedia*, Vol. 8, pp. 481-488, 2013.
- [2] C. Y. Gang, C. Y. Zhong, J. M. Zhong, L. Chao. “Information Retrieval Oriented Adaptive Chinese Word Segmentation System”. *Journal of Software*, Vol. 17, pp. 356-363, 2006.
- [3] X. Yan, L. J. Tao, W. Bin, S. C. Ming. “A Category Resolve Power-Based Feature Selection Method”. *Journal of Software*, Vol. 19, pp. 82-89, 2008.
- [4] A. Alimjan, I. Turgun, O. Hasan, A. Marhaba. “Machine learning based Uyghur language text categorization”. *Computer Engineering and Applications*, Vol. 48, pp. 110-112, 2012.
- [5] X. Wen, J. Y. Hong. “Semantic MMT model based on hierarchical network of concepts in Chinese-English MT”. *Journal of Networks*, Vol. 8, pp. 237-244, 2013.
- [6] C. J. Guo, C. Hao. “A structured information extraction algorithm for scientific papers based on feature rules learning”. *Journal of Software*, Vol. 8, pp. 55-62, 2013.
- [7] R. Askar, Z. C. Qing, M. Guljamal, M. Rehim, H. Askar. “Approach to recognizing Uyghur names based on conditional random fields”. *Journal of Tsinghua University*, Vol. 53, pp. 873-877, 2013.
- [8] H. C. Ning, Z. Hai. “Chinese Word Segmentation: A Decade Review”. *Journal of Chinese Information Processing*, Vol. 21, pp. 8-19, 2007.
- [9] Z. H. Ping, Y. H. Kui, L. Qun. “HHMM-Based Chinese Lexical Analyzer ICTCLAS”. *In Proc of the 2nd SigHan Workshop*, pp. 184-187, 2003.
- [10] S. M. Song, X. Ming, K. B. TSOU. “Chinese Word Segmentation without Using Dictionary Based on Unsupervised Learning Strategy”. *Chinese Journal of Computers*, Vol. 27, pp. 736-742, 2004.
- [11] W. S. Li, W. Bin. “A Chinese Overlapping Ambiguity Resolution Method Based on Coupling Degree of Double Characters”. *Journal of Chinese Information Processing*, Vol. 21, pp. 14-17, 2007.

- [12] F. H. Xiao, K. S. Lin, Z. X. Juan, X. W. Biao. "Chinese Word Segmentation Research Based on Statistic the Frequency of the Word". *Computer Engineering and Applications*, Vol. 30, pp. 67-69, 2005.
- [13] W. Fang, W. C. Xuan. "Chinese Integrated Word Identification Based on Confidence". *Journal of Chinese Information Processing*, Vol. 23, pp. 17-23, 2009.
- [14] L. Dan, F. W. Guo, Z. Hong. "Bigram Chinese Word Segmentation Model Based on Bayesian Network". *Computer Engineering*, Vol. 36, pp. 12-14, 2010.
- [15] J. J. Hong, Z. S. Zheng, L. Mei. "Analysis and Application of Chinese Word Segmentation Model which Consist of Dictionary and Statistics Method". *Computer Engineering and Design*, Vol. 33, pp. 387-391, 2012.
- [16] Z. M. Shan, D. Z. Long, C. W. Xiang, L. Ting. "Combining Statistical Model and Dictionary for Domain Adaption of Chinese Word Segmentation". *Journal of Chinese Information Processing*, Vol. 26, pp. 8-12, 2012.
- [17] A. Mireguli, A. Mijiti, A. Aisikaer. "A Morphological Analysis Based Algorithm for Uyghur Vowel Weakening Identification". *Journal of Chinese Information Processing*, Vol. 22, pp. 43-47, 2008.
- [18] Z. Hua, Z. Q. Tian. "Micro-blog keyword extraction method based on graph model and semantic space". *Journal of Multimedia*, Vol. 8, pp. 611-617, 2013.
- [19] M. Bo, Y. Y. Ting, Z. Xi, Z. J. Lin. "Semantically enhanced uyghur information retrieval model". *Journal of Software*, Vol. 7, pp. 1315-1320, 2012.
- [20] K. W. Church, W. Gale, P. Hanks, D. Hindle. "Using statistics in lexical analysis". In: Zernik U. ed. *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*. Hillsdale NJ: Lawrence Erlbaum Associates, pp. 115~ 164, 1991.
- [21] F. S. Qin, C. H. Ming. "Context based Approach to Combinational Ambiguity Resolution in Chinese Word Segmentation". *Journal of Chinese Information Processing*, Vol. 21, pp. 13-16, 2007.
- [22] Y. J. De, W. X. Jie, F. X. Zhong. "Comparing of importance of above-context versus below-context for Chinese word segmentation". *Computer Engineering and Applications*, Vol. 47, pp. 117-120, 2011.
- [23] H. S. Ke, W. X. Jie, D. Yuan, Z. T. Zheng, B. Xue. "Apply Normalized Accessor Variety in Chinese Word Segmentation". *Journal of Chinese Information Processing*, Vol. 24, pp. 15-19, 2010.
- [24] H. Min, G. C. Chun, Z. H. Ping, C. X. Qi. "Method of new word identification based on lager-scale corpus". *Computer Engineering and Applications*, Vol. 43, pp. 157-159, 2007.
- [25] Y. Mayire, A. Mijiti, H. Askar. "A Minimum Edit Distance Based Uighur Spelling Check". *Journal of Chinese Information Processing*, Vol. 23, pp. 110-114, 2008.



**Turdi Thohti** born in 1975, received B. E. in 1999 from Nanjing University of China, received M. E. in 2009 from Beijing University of technology of China. He is an associate professor and Graduate Supervisor in the School of Information Science and Engineering, Xinjiang University. He has published more than 30 papers on international journals, national journals, and conferences in recent years. His research interests include natural language processing, information retrieval, web mining and content security. He is a PhD candidate in Xinjiang University, Senior of CCF, member of IEEE CS and CIPSC.



**Winira Musajan** born in 1960, received B. E. in 1983 from Xinjiang University of China. She is a professor and Graduate Supervisor of School of Information Science and Engineering, Xinjiang University. Her main research interests in natural language processing and multilingual information retrieval.



**Askar Hamdulla** born in 1972, received B. E. in 1996, M. E. in 1999, and Ph. D. in 2003, all in Information Science and Engineering, from University of Electronic Science and Technology of China. In 2010, he was a visiting scholar at Center for Signal and Image Processing, Georgia Institute of Technology, GA, USA, tutored by Professor Biing-Hwang (Fred) Juang. Currently, he is a professor in the School of Software Engineering, Xinjiang University. He has published more than 80 technical papers on speech synthesis, natural language processing and image processing. He is an affiliate member of IEEE. member of phonetics branch committee of Association of Chinese Linguistics.

# Short Text Classification: A Survey

Ge Song, Yunming Ye, Xiaolin Du, Xiaohui Huang, and Shifu Bie

Shenzhen Key Laboratory of Internet Information Collaboration, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, 518055, China

Email: carroll0708@qq.com, yeyunming@hit.edu.cn, {duxiaolinhitsz, hxh016}@gmail.com, bieshifu001@sina.com

**Abstract**—With the recent explosive growth of e-commerce and online communication, a new genre of text, short text, has been extensively applied in many areas. So many researches focus on short text mining. It is a challenge to classify the short text owing to its natural characters, such as sparseness, large-scale, immediacy, non-standardization. It is difficult for traditional methods to deal with short text classification mainly because too limited words in short text cannot represent the feature space and the relationship between words and documents. Several researches and reviews on text classification are shown in recent times. However, only a few of researches focus on short text classification. This paper discusses the characters of short text and the difficulty of short text classification. Then we introduce the existing popular works on short text classifiers and models, including short text classification using semantic analysis, semi-supervised short text classification, ensemble short text classification, and real-time classification. The evaluations of short text classification are analyzed in our paper. Finally we summarize the existing classification technology and prospect for development trend of short text classification.

**Index Terms**—Short Text; Text Classification; Feature Selection; Semantic Analysis; Integrated Learning; Semi-Supervised Learning

## I. INTRODUCTION

With the explosion of e-commerce and online communication, short texts become available in many application areas, such as Instant Messages, online Chat Logs, Bulletin Board System Titles, Web Logs Comments, Internet News Comments, SMS, twitter etc. Therefore, successfully processing them becomes increasingly important in many Web and IR applications. However, it is a new challenges that classifying these sorts of text and Web data.

Unlike normal documents, these text and Web segments are usually noisier, less topic-focused, and much shorter, that is, they consist of from a dozen words to a few sentences [1]. Because of the short length, they do not provide enough word co-occurrence or shared context for a good similarity measure [40]. Therefore, normal machine learning methods, which are rely on the word frequency, enough word co-occurrences or shared context to measure the similarity of documents [41], usually fail to achieve desire accuracy due to the data sparseness.

New classifying methods on short text are appeared, such as semantic analysis, semi-supervised short text

classification, ensemble models for short text, and real-time classification. However, compared with a lot of reviews and surveys on text classification, only few of surveys are appeared to discuss the recent researches on short text classification. This paper analyzes the challenges associated with classifying short text and systemic summarizes the existing related methods to short text classification using analytical measures.

After the analysis of the feature and difficulty of short text, we point out the process of short text classification in section II. In section III, short text classification based on semantic analysis is introduced. We describe some algorithms on semi-supervised short text classification in section IV. Section V and Section VI introduce the ensemble model for classifying short text and online short text classification, respectively. We analyze relevant evaluating measures in Section VII. In Section VIII, we summarize the methods for classifying short text.

## II. BACKGROUNDS

### A. Feature of Short Text

Short text has been widely used in many fields, such as mobile short message, instant message, BBS title, news title, online chat record, blog comment, news comment, etc. And its main characteristic of the text length is very short, no longer than 200 characters. As mobile short message which we common used daily is no more than 70 characters, BBS title and news title less than 30. Instant messaging (IM) software supports longer message. For sending message quickly and ensuring it safely, IM software also limits its length, such as Windows Live Messenger of Microsoft allowing the longest message 400 characters. In fact, in daily communication, the instant message is only dozen words.

Generally, the features of short text are as follows [1] [2]:

**Sparcity:** a short text only contains several to a dozen words with a few features, it does not provide enough words co-occurrence or shared context for a good similarity measure. It is difficult to extract its valid language features.

**Immediacy:** short texts are sent immediately and received in real time. In addition, the quantity is very large.

**Non-standardability:** The description of the short text is concise, with many misspellings, non-standard terms and noise.

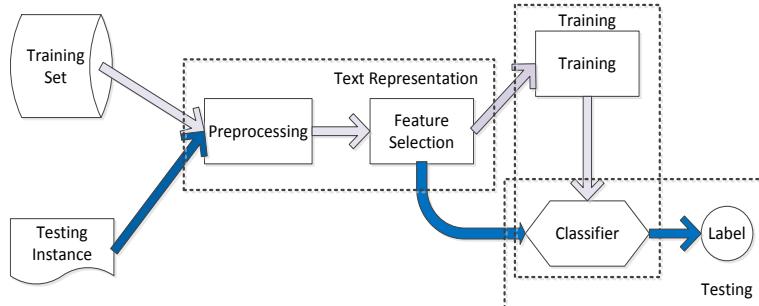


Figure 1. The processing of short text classification

Noises and imbalanced distribution: The application background (such as network security) needs to deal with massive amounts of short textual data. However, we may focus on only a small part (detecting objects) among the large-scale data. Therefore, useful instances are limited, and the distribution of short text is imbalanced.

Large scale data and labeling bottlenecks: It is difficult to manually label all of the large scale instances. Limited-labeled instances may only provide limited information. So how to make full use of these labeled instances and other unlabeled instance has become a key problem of short text classification.

Most of the traditional methods (such as SVM, BAYES, and KNN) are based on the similarity of term frequency, ignoring the feature of short text. These traditional methods may be not deal with the short text classification. Most of them (such as BAYES) may fail to obtain high accuracy if the labeled information is insufficient. In addition, some classification methods based on Vector Space Model (SVM) should use semantic information to improve the performance of the classifiers.

### B. Short Text Classification

With the growing use of digital devices and the fast growth of the number of pages on the World Wide Web, text categorization is a key component in managing information. Automated categorization of text documents plays a crucial role in the ability of many applications to classify and provide the proper documents in a timely and correct manner.

Text classification can be defined simply as follows (Fig. 1): Given a set of documents D and a set of classes (or labels) C, define a function F that will assign a value from the set of C to each document in D. For example in short text classification, D might consist of the set of all classified advertisements in a newspaper, and C would therefore be the set of headings in the classified section of that same newspaper.

Learning to classify text and Web documents has been intensively studied during the past decade. Many learning methods, such as k nearest neighbors (k-NN), Naive Bayes, maximum entropy, and support vector machines (SVMs), have been applied to a lot of classification problems with different benchmark collections and achieved satisfactory results. However, traditional classification method was not good at short text classification, because of the character and difficulty of

short text. Therefore, how to reasonably represent and choose features items, effectively reduce the spatial dimension and noises, and increase classification accuracy become the problem of short text classification.

### III. SHORT TEXT CLASSIFICATION BASED ON SEMANTIC ANALYSIS

At present, the solution of reducing the feature spatial dimension is mainly based on the semantic features and semantic analysis. This is because the processing of text classification is generally in Vector Space Model (VSM), which has the basic assumption that the relationships of words are independent, neglected the correlation between texts. However, short text has weaker capacity of semantic expression, which is needed this correlation. While traditional classification cannot distinguish language fuzziness of natural language, cognates and synonyms, all of which are abundant in short text. Therefore traditional classification methods usually fail to achieve expected accuracy of short text.

Semantic analysis pays more attention to the concept, inner structure semantic level, and the correlation of texts to obtain the logic structure, which is more expressive and objectivity. In the existing researches, classification based on the Latent Semantic Analysis occupies an important position. Using the statistic method, latent semantic analysis extracts potential semantic structure, eliminates the synonymous influence, and reduces feature dimension and noises. Thus, many algorithms based on semantic analysis are proposed to deal with short text classification (More detailed information is shown in Table I) [3] [5-11] [44]. Zelikovitz [3] applies it to short text classification. Qiang Pu etc [5] combine LSA and Independent Component Analysis (ICA) [8] [42] together. Xuan - Hieu Phan etc. [7] establishes large-scale short text classification framework. The framework is mainly based on recent successful latent topic analysis models (such as pLSA and LDA) and powerful machine learning methods like maximum entropy and SVMs. Bing-kun WANG etc. [9] present a new method to tackle the problem by building a strong feature thesaurus (SFT) based on latent Dirichlet allocation (LDA) and information gain (IG) models. Language independent semantic (LIS) kernel [10] is proposed to enhance the language-dependency, when exploiting syntactic or semantic information. It is able to effectively compute the similarity between short-text documents without using grammatical tags and lexical databases. Mengen Chen etc.

[11] propose the method that extracts topics at multiple granularities, which can model the short text more precisely. Transductive LSA [3] [4] is another example of short text classification based on LSA. Transduction makes use of the test examples in choosing the hypothesis of the learner. The recreation of the space with the incorporation of the test examples does choose a representation based upon the test examples. The reduction of dimensionality of the training/test set combined allows the smaller space to more accurately reflect the test set to which it will be applied for classification. The inclusion of the test examples into the original matrix allows LSA to calculate entropy weights of words with the vocabulary and examples and co-occurrences of words in the test set.

In the following subsection, we describe detailed on definition and procession of LSA, pLSA, and LDA. Then we analyze the advantages and disadvantages of these three semantic analysis methods, respectively.

TABLE I. COMPARISON OF SEMANTIC APPROACHES FOR SHORT TEXT CLASSIFICATION

Reference	Model	Datasets	Measure
[3] [4]	Transductive LSA	Physics, NetVet , Business, News, Thesaurus	Accuracy
[5]	ICA, LSA	400 Chinese short-text documents	WS,BS,N, Similarity
[7]	pLSA, LDA	Wikipedia data, MEDLINE	Error, Accuracy,
[9]	SFT, LDA, IG	Chinese corpus, BBS	Precision, Recall, F1
[10]	LIS kernel	English and Korean datasets	Accuracy
[11]	LDA, Multi-Granularity Topics	Search Snippets, Chinese corpus, BBS	Accuracy

#### A. Short Text Classification Using Latent Semantic Analysis (LSA)

Latent Semantic Analysis (LSA) [12] [13] [43] is based upon the assumption that there is an underlying semantic structure in textual data, and that the relationship between terms and documents can be redescribed in this semantic structure form [14]. LSA transforms the vector space into the semantic space. Based on statistical method, LSA extracts and quantifies the semantic structure, eliminates the correlation between terms. LSA can reduce the high-dimensional vector matrix to construct the low-dimensional subspace which can effectively describe the relationship of term-document. Many Dimension reduction methods in LSA are proposed [15] [16], such as Singular Value Decomposition (SVD), Semi-discrete Decomposition (SDD), and Nonnegative Matrix. The most common method, SVD, should be introduced in our paper.

The process of LSA based on SVD is as follows [15] [16]:

Short documents are represented as vectors in a vector space. Therefore, the term-document matrix represents as  $A_{mn} = [a_{ij}]_{m \times n}$ , with each position corresponding to the absence or weighted presence of a term (a row i) in a

document (a column j). This matrix is typically very sparse, as most documents contain only a small percentage of the total number of terms seen in the full collection of documents.

Compute the weight of  $a_{ij}$ . In order to focus on the contribution of each term (or document), we should compute the weight of  $a_{ij}$ . The traditional method is:  $a_{ij} = LW_{ij} \times GW_{ij}$ , where  $LW_{ij}$  is a local weight of the term I in the document j,  $GW_{ij}$  is the global weight of the term I in the entire dataset. We should obtain the local weight of a term by computing the log of the total frequency of the term I in the document j. The global weight of a term equal to the entropy of the term in the dataset. This entropy is based upon the number of occurrences of this term in each document.

The singular value decomposition (SVD) of  $A_{mn} = [a_{ij}]_{m \times n}$  matrix. According to the above analysis,  $A_{mn} = [a_{ij}]_{m \times n}$  matrix is very sparse. Moreover, in this very large space, some documents seem to be closer with each other by sharing common words. But these documents may be not related to each other semantically. Meanwhile, many documents that appear very distant by not sharing any term may actually closer. Because the same concept can be represented by many different words and words can have ambiguous meanings. LSA reduces this large space, and hopefully captures the true relationships between documents. To do this, LSA uses the singular value decomposition (SVD) of the term by  $A_{mn} = [a_{ij}]_{m \times n}$  matrix. SVD of  $A_{mn} = [a_{ij}]_{m \times n}$  is the product of three matrices:

$$A = \sum_{i=1}^r u_i \sigma_i v_i = [u_1 \cdots u_r] \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} \quad (1)$$

where  $u$  and  $v$  are the matrices of the left and right singular vectors and  $\sigma$  is the diagonal matrix of singular values. The diagonal elements of  $\sigma$  are ordered by magnitude, and therefore these matrices can be simplified by setting the smallest k values in  $\sigma$  to zero. The columns of T and D that correspond to the values of  $\sigma$  that were set to zero are deleted.

$$A = U_r \sum_r V_r^T \approx U_k \sum_k V_k^T \quad (2)$$

Classification based on LSA. According to SVD, The form of the reduced space  $U_k \sum_k V_k^T$  is similar to the form of the original vector. Therefore, all of the classification algorithms that are suitable for vector space model can also apply to LSA classification model [17]. Many classification methods that combine LSA and traditional algorithms [18] [19] [20], such as sequence classification algorithm, Naive Bayes, KNN [4], and SVM are proposed to improve the accuracy of classifying short text.

To sum up, the advantage of the latent semantic analysis as follows:

Reduce the feature distribution and noise. According to LSA, K-dimensional semantic space is extract. This semantic space not only keeps the most information of the original vector space, but also reduces the dimension. Meanwhile, LDA eliminate the noise by discarding useless feature. Therefore, LSA can effectively deal with large-scale short test dataset;

Strengthen the semantic relationship. In latent semantic space, vectors no longer simply mean the frequency and distribution. They can describe the semantic relationship between terms and documents. LSA create less dependency on polysemy, synonyms, and common words, which may lead to low accuracy in traditional vector space. Previous researches show that at a relatively low-dimensional space with higher semantic expression, the performance of classification will be improved by similarity analysis

Flexibility. Since terms and documents are mapped to the same K-dimensional space, the LSA model can not only analyze the similarity between term-term (as traditional models do), but also obtain the better effect on analyzing the similarity between document-document and term-document.

Although the LSA model is effective for short text classification, it still has some defects in the as follows:

Lose the structural information as reducing the dimension. Since feature space is the semantic space based on the text information, it may retain the main global information of the original feature matrix, ignoring some features that contribute a lot in the local space, but are insignificance in the global space.

SVD does not have the strict mathematical sense. In addition, it cost more computing time and space complexity in high-dimensional space.

The meaning of document is represented by the linear summation of vectors, ignoring grammar information of words in the phase of information extraction.

LSA can only deal with the visible variable. However, some meanings such as metaphor and analogy cannot be calculated [20] [21].

#### B. Short Text Classification Using Probabilistic Latent Semantic Analysis (pLSA)

Probabilistic Latent Semantic Analysis (pLSA) [20] [21] is proposed by Hofmann at 1999. The principle of pLSA is stronger than LSA. Moreover, It explicitly introduced the concept of "latent topic". The latent topic is defined as the latent variable during a random process. According to fitting the training data, the probability model P is shown below:

$$P(d, w) = P(d)P(w|d) = \sum_{z \in Z} P(w|z)P(z|d)P(d) \quad (3)$$

where z is the latent class variables, d (document) and w (word) are observed variables, which are independent on the conditional probability of latent topic z. The probability model P can be expressed form of SVD as follows [28]:

$$P = U_k \sum_k V_k^T \quad (4)$$

where

$$\begin{aligned} U_k &= (P(d_i | z_k))_i \\ \sum_k &= \text{diag}(P(z_k))_k \\ V_k^T &= (P(w_j | z_k))_{j,k} \end{aligned} \quad (5)$$

For a particular training set, the probability of document  $P(d)$ , the word probability  $P(w)$  is known, and conditional probability  $P(z|d)$ ,  $P(w|z)$  is unknown. According to the principle of maximum likelihood estimation the maximum of the logarithm likelihood function is calculated by expectation maximum algorithm (EM) to fit the following model [20] [21]:

$$L = \sum_{i=1}^N \sum_{j=1}^M f(d_i, w_j) \log P(d_i, w_j) \quad (6)$$

where  $f(d_i, w_j)$  is the frequency of term  $w_j$  in document  $d_i$ .

The feature vector should be considered as the left singular vectors (the relationship between document and latent factors) or right singular vector (the relationship between terms and latent factors) in the procedure of classification.

Compared with the LSA model, pLSA model has the following advantages:

The pLSA model is the approximation model based on probability function with polynomial sampling. pLSA acquires the more stable and better approximation effect. However, the LSA model involves in Gaussian noise hypothesis.

The probability distribution of each variable has been clearly defined in the pLSA model, but LSA fails to define normal probability distribution.

pLSA can determine the optimal k-dimension using the existing statistical method, while LSA is mainly based on heuristics to determine the optimal k-dimension with more computational complexity for model selection.

pLSA model, however, still has some shortcomings:

The pLSA model needs to acquire the prior probability of label, which is computed by training set. However, it does not have a suitable prior probability to describe the unknown test corpus

The parameter space is increased with increasing training instances in pLSA model. This phenomenon may lead to excessive fitting problems. Meanwhile, too many discrete features that are only suitable for training set is existed, but these features cannot properly describe the unknown testing set.

#### C. Short Text Classification using Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) [7] is a probabilistic generative model that builds the linear discriminant function on the input variable. It is looking for a kind of

transformation to acquire the maximum separability between classes and minimum difference within the class, LDA is a generative graphical model as shown in Fig. 2. It can be used to model and discover underlying topic structures of any kind of discrete data in which text is a typical example. LDA was developed based on an assumption of document generation process.

The main process of LDA is shown as: Assume that the text corpus contains several linear latent topics. According to probability inference algorithm, each document should be represented as the probability form on these latent topics. In the training phase, since LDA is the generative model, generative classification algorithms (such as MaxEnt algorithm [23]) are used to build the classification model. That is, the documents in each class independently train the sub-LDA model and share a set of topics, while the topics belong to different class are separated. In the predicting phase, the testing instance is generated by all of the sub-LDA models, and predicts its label by calculating the testing instance generative probability on each class.

It is worth to point out that the key problem of constructing LDA model is how to acquire the distribution information of latent topics within the document. Some algorithms such as Calculus of variations [24], Expectation Maximization (EM) [25], and Gibbs sampling algorithm [26] are used to deal with this question.

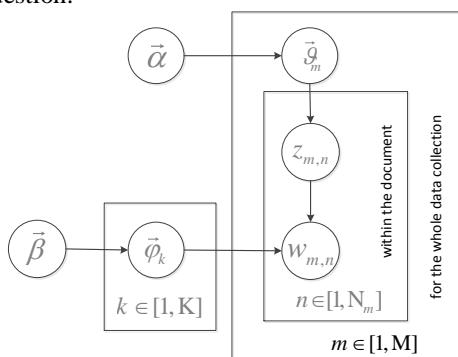


Figure 2. The structure of LDA

Compared with LSA and pLSA, LDA can find the latent structure of “topics” or “concepts” in a test corpus. The advantage of LDA is shown as

The Dirichlet probability distribution is used in LDA. This probability distribution (a continuous distribution) can give the unknown instance a probability of belonging to some topic set;

LDA directly chooses a suitable topic set from the topic distribution compared with pLSA that need a prior probability of unknown instance.

The LDA model has stronger ability of describing the realistic semantic. The LDA model which inherits all the advantages of the pLSA generative model, is more close to realistic semantic environment.

#### IV. SEMI-SUPERVISED SHORT TEXT CLASSIFICATION

Semi-supervised learning refers to the use of both labeled and unlabeled data for training. It contrasts supervised learning (data all labeled) or unsupervised

learning (data all unlabeled). Other names are learning from labeled and unlabeled data or learning from partially labeled/classified data. It is found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. The acquisition of labeled data for a learning problem often requires a skilled human agent to manually classify training examples. The cost associated with the labeling process thus may render a fully labeled training set infeasible, whereas acquisition of unlabeled data is relatively inexpensive. In such situations, semi-supervised learning can be of great practical value [27].

Most of the semi-supervised short text classifications are flexible semi-supervised learning. It can utilize unlabeled data to improve the classifier. However, unlike traditional semi-supervised learning algorithms, the universal data and the training/test data are not required to have the same format. In addition, once estimated, a topic model can be applied to more than one classification problems provided that they are consistent.

Reference [28] proposes a new semi-supervised short text classification algorithm. It uses the unlabeled corpus as “background knowledge” for the learner. A concrete example can be seen in the task of assigning topic labels to technical papers. Any title containing a word on galaxy (such as *galaxy*) should be easily classified correctly as an astrophysics paper, since the feature term in the title is common. However, an article on a less common topic, as for example old white dwarfs, should be able to correctly classify by utilizing a corpus of unlabeled paper abstracts from the same field. Those unlabeled paper abstracts are most similar to both old white dwarfs and to various training titles, each of which is quite dissimilar to it when compared directly.

Background knowledge may provide with a corpus of text that contains information both about importance of words and joint probability of words [29] [6]. We can use this background combined with the training examples to label a new example.

```
SELECT Test.instance, Train.label  
FROM Train AND Test AND Background  
WHERE Train.instance SIM Background.value  
AND Test.instance SIM Background.value
```

If the features of the sample are special fewer, some background information that connected training samples and testing samples may not be found. We can use it as follows:

```
SELECT Test.instance, Train.label  
FROM Train AND Test AND Background as B1  
AND Background as B2  
WHERE Train.instance SIM B1.value  
AND Test.instance SIM B2.value  
AND B1.value SIM B2.value
```

Reference [30] establishes information flow corpus of related topics in HAL space to classify the document title (short text). In that reference, Latent semantic relationship (such as co-occurrence relationship) among topics is represented by information flow. The feature HAL vector is extracted in the HAL model. Feature HAL

vector is very sensitive to semantic information (especially to context semantic information). Then, the degree of the vector is calculated as follows:

$$\deg \text{ree}(c_i \triangleleft c_j) = \frac{\sum_{p_l \in (\text{QP}_\mu(c_i) \cap \text{QP}_\mu(c_j))} w_{c_i p_l}}{\sum_{p_k \in \text{QP}_\mu(c_i)} w_{c_i p_k}} \quad (6)$$

According to dynamic corpus, information inference model is established.

Another algorithm based on contextual information is Distributional term representations (DTRs) [31]. DTRs represent terms by means of contextual information, which are given by document occurrence and term co-occurrence statistics. This algorithm enriches document representations that help to overcome the small-length and high-sparseness short text to some extent. More detailed on semi-supervised approaches are shown in Table II.

TABLE II. COMPARISON OF SEMI-SUPERVISED APPROACHES FOR SHORT TEXT CLASSIFICATION

Reference	Model	Datasets	Measure
[28]	background knowledge	Technical papers, News, Web page titles, Companies	Error rate
[30]	HAL	Reuters	Precision
[31]	DTRs	reduced Reuters R8 news corpus, EasyAbstract, CICLing2002	Error, Accuracy,

## V. ENSEMBLE SHORT TEXT CLASSIFICATION

Single classifier generally based on similarity of items feature to classification is difficult to obtain the very good prediction result of short text classification since it is difficult to compute the similarity if the feature space is sparseness. Ensemble learning methods, on the other hand, through assigning a right weight for each weak classifier, get the weight of each feature; it is suitable for solving short text classification problem.

Reference [1] proposes a new dynamic assembly classification algorithm for short text, to solve the problems of sparse features and unbalanced data of the short text. In this method, to reduce the impact of the sparse features and unbalanced data, a treelike assembly classifier was constructed to support the classification. Then a dynamic strategy is presented to adjust the combinational structure in an adaptive way. Aixin Sun is proposed a short text classification method using very few words [32]. The predicted category label in this method is the majority vote of the search results, which are obtained by searching for a small set of labeled short texts best matching the query words.

A new model [33] is proposed to directly measure the correlation between a short text instance and a domain instead of representing short texts as vectors of weights. Firstly domain knowledge for each user-defined domain is drawn using an external corpus of longer documents. Secondly, the correlation is calculated. Finally, if the

correlation is greater than a threshold, the instance will be classified into the domain.

To address this problem that short texts in Twitter do not provide sufficient word occurrences, another algorithm [34] is proposed with a small set of domain-specific features extracted from the author's profile and text in twitter. The proposed approach effectively classifies the text to a predefined set of generic classes. More detailed information on ensemble models are shown in Table III.

TABLE III. COMPARISON OF ENSEMBLE MODELS FOR SHORT TEXT CLASSIFICATION

Reference	Model	Datasets	Measure
[1]	Dynamic ensemble	Chinese BBS	Precision, Recall, F1
[32]	IR, Voting	Web snippet dataset	Accuracy
[33]	Domain knowledge,	External Corpus, Micro-blog Dataset	Precision, Recall, F1
[34]	Domain specific features	- recent tweets from random users	Accuracy

## VI. REAL-TIME CLASSIFICATION OF LARGE SCALE SHORT TEXT

Immediacy is another feature of short text, which means Short texts are sent immediate and received in real time and usually the quantity is very large. So how to classify large-scale short text data immediately also becomes an important problem. Currently, comparing with several classic classification algorithm, it is often chose Bayes algorithm as online classifier. Naive Bayes algorithm judges categories through calculating probability of text belongs to each category, which is a simple, accuracy and widely used algorithm [39].

TABLE IV. COMPARISON OF REAL-TIME MODELS FOR SHORT TEXT CLASSIFICATION

Reference	Model	Datasets	Measure
[35]	Online and offline	messages	Precision, Recall, Error,
[36]	Extent Bayesian filtering technique	English and Spanish SMS spam collections	ROC curves
[37]	naïve Bayesian	SMS spam collections	Precision, Recall

Reference [35] proposed a spam message filtering system which combined online filtering with offline classifying. The system could filter the messages efficiently according to the features of sending behavior and the content of the messages using Naive Bayes algorithm. Further, the system uses a feedback self-learning mechanism, so the classifiers could improve themselves according to the filtering result. Another spam messages filter systems are based on Native Bayes and SVM [36] [37]. Native Bayes advantage of rapid statistics classification and SVM incremental training feature in this system, and updates the keywords database in time to enhance the self-adaptability. More detailed description on real-time algorithms are shown in Table IV.

## VII. EVALUATION OF SHORT TEXT CLASSIFICATION

How to evaluate the model is another important problem in short text classification. Some measures that are used to compare and evaluate the classification method mainly include [38]:

- Accuracy,
- Precision and recall
- F-measure.
- Macro average and micro average

### A. Accuracy

Accuracy is the most basic classification evaluation measure. The accuracy is computed:

$$\text{accuracy} = \frac{TP + TN}{N} \quad (7)$$

where N is the total number of the testing instances, TP, FP, FN, TN is described in the Table V:

TABLE V. THE CONTINGENCY TABLE FOR CATEGORY C

		True Label	
		Yes	No
Classifier Label	Yes	TP	FP
	No	FN	TN

$$TP = \sum_{i=1}^C TP_i, FN = \sum_{i=1}^C FN_i,$$

$$FP = \sum_{i=1}^C FP_i, TN = \sum_{i=1}^C TN_i,$$

where C is the total number of the classes.

### B. Precision and Recall

Classification effectiveness is usually measured in terms of precision and recall, which are described as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

### C. F-measure

The precision and recall is reciprocal relationship. The purpose of classification is to get precision and recall together. F-measure is considered both of them, whose weather is described by parameters b.

$$F_\beta = \frac{(\beta^2 + 1) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} \times \text{Recall}} \times 100\% \quad (10)$$

If  $\beta = 1$ , it is called breakeven point, which is widely used:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (11)$$

### D. Macro Average and Micro Average

Precision, recall and  $F_1$  are aimed at a class with only local significance. Now commonly comprehensive measures (macro average and micro average) are shown below:

$$\text{Micro Recall} = \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C TP_i + FN_i} \times 100\% \quad (12)$$

$$\text{Micro Precision} = \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C TP_i + FP_i} \times 100\% \quad (13)$$

$$\text{Micro } F_1 = \frac{2 \times \text{Micro Precision} \times \text{Micro Recall}}{\text{Micro Precision} + \text{Micro Recall}} \times 100\% \quad (14)$$

$$\text{Macro Recall} = \frac{1}{C} \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C TP_i + FN_i} \times 100\% \quad (15)$$

$$\text{Macro Precision} = \frac{1}{C} \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C TP_i + FP_i} \times 100\% \quad (16)$$

## VIII. SUMMARY

Nowadays, the rapid development of information dissemination and media, especially the rise and development of instant communication lead to short text are widespread use, such as its application in topic tracking and discovery, catchword analysis, and network security. Short text has its own features, such as sparseness, large-scale, immediacy, and non-standardability. Therefore, normal machine learning methods usually fail to achieve desire accuracy due to the data sparseness.

At present, short text classification algorithm mainly classify in the following aspects:

Reduce feature dimension and extract feature using semantic relationship. Such as an LSA, LDA classification model.

Combined with plenty of unlabeled text, use semi-supervised classification algorithm to solve label bottleneck problems.

Use ensemble classification to improve classification accuracy.

Combine online classification and off-line classification to deal with large scale short texts.

However, short text classification is a challenging field, because many technologies are in the initial stage as well as the difficulties of classification didn't get the excellent solution such as how to design dynamic short text stream

classification model. According to the passage of the application of short text, it produces several other problems such as multi-label short-text classification, comment emotional classification spam filtering, and topic tracking and control.

#### ACKNOWLEDGMENT

This research was supported in part by NSFC under Grant No. 61073195, Shenzhen Science and Technology Program under Grant No. CXY201107010163A, Shenzhen Strategic Emerging Industries Program under Grants No. JCYJ20130329142551746 and No. JCYJ20120613135329670.

#### REFERENCES

- [1] YAN Rui, CAO Xian-bin, LI Kai, "Dynamic Assembly Classification Algorithm for Short Text," *ACTA ELECTRONICA SINICA*, Vol. 37(5), pp. 1019-1024, 2009.
- [2] SU Jin-Shu, ZHANG Bo-Feng, and XU Xin. "Advances in Machine Learning Based Text Categorization," *Journal of Software*, Vol. 17(9), pp. 1848–1859, 2006.
- [3] Zelikovitz, S., Marquez, F., "Transductive Learning for Short-Text Classification Problems using Latent Semantic Indexing," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 19(2), pp. 143-163, 2005.
- [4] Zelikovitz, S., "Transductive LSI for Short Text Classification Problems," *Proceedings of the 17th International Flairs Conference*, pp. 556—561, 2004.
- [5] Qiang Pu and Guo-Wei Yang, "Short-Text Classification Based on ICA and LSA," *Advances in Neural Networks - ISNN 2006*, pp. 265-270, 2006.
- [6] S. Zelikovitz and H. Hirsh, "Using LSI for text classification in the presence of background text," *In Proceedings of 10th International Conference on Information and Knowledge Management*, pp. 113-118, 2001.
- [7] Xuan-Hieu Phan, Le-Minh Nguyen, Susumu Horiguchi, "Learning to Classify Short and Sparse Text & Web with Hidden Topics from Large-scale Data Collections," *WWW 2008 Refereed Track: Data Mining – Learning*, pp. 91-100, 2008.
- [8] BACH F JORDAN M I, "Kernel independent component analysis," *The Journal of Machine Learning Research*, Vol. 3(11), pp. 1-48, 2003.
- [9] Wang, Bing-kun and Huang, Yong-feng and Yang, Wan-xia and Li, Xing, "Short text classification based on strong feature thesaurus," *Journal of Zhejiang University SCIENCE C*, Vol. 13(9), pp. 649-659, 2012.
- [10] Kim, Kwanho and Chung, Beom-suk and Choi, Yerim and Lee, Seungjun and Jung, Jae-Yoon and Park, Jonghun, "Language independent semantic kernels for short-text classification," *Expert Systems with Applications*, Vol. 41(2), pp. 735-743, 2012.
- [11] Chen, Mengen and Jin, Xiaoming and Shen, Dou, "Short text classification improved by learning multi-granularity topics," *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume*, Vol. 3, pp. 1776-1781, 2011.
- [12] Dumais, S. "Combining evidence for effective information filtering," *In AAAI Spring Symposium on Machine Learning and Information Retrieval, Tech Report SS-96-07*, 1996.
- [13] M WBerry, S T Dumais, GW O'Brien, "Using Linear Algebra for Intelligent Information Retrieval," *SIAM Review*, December 1995.
- [14] Ding C. H., Q. A, "Dual Probabilistic Model for latent semantic indexing information retrieval and filtering," *Proceedings of the 22nd Annual Conference on Research and Development in Information Retrieval (ACM SIGIR)*, 1999.
- [15] Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K., and Harshman, R.. "Indexing by Latent Semantics Analysis," *Journal of the American Society for Information Science*, Vol. 41(6), pp. 391-407, 1990,
- [16] Landau T. K., Foltz P. W., Laham D., "An Introduction to Latent Semantic Analysis," *Discourse Processes 25*, pp. 259-284, 1998.
- [17] Andreas Hotho, Steffen Staab, Gerd Stumme., "Explaining Text Clustering Results Using Semantic Structures," *In Principles of Data Mining and Knowledge Discovery. 7th European Conference*, pp. 217-228, 2003.
- [18] Baker L D, McCallum A K., "Distributional clustering of words for text classification," *Proc. ACM -SIGIR-98, Australia: ACM Press*, pp. 96-103, 1998.
- [19] Park H, Howland P, Jeon M. "Cluster Structure Preserving Dimension Reduction Based on the Generalized Singular Value Decomposition," *SLAM Journal on Matrix Analysis and Applications*, Vol. 25 (1), pp. 165-179, 2003.
- [20] Hofmann T., "Probabilistic latent semantic analysis," *In UAI99, Morgan Kaufmann*, pp. 289-296, 1999.
- [21] Hofmann T., "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, Vol. 42, pp. 177-196, 2001.
- [22] Chen Ning, Chen An, Zhou Long xiang, "An Incremental Grid Density Based Clustering Algorithm," *Journal of Software*, Vol. 13 (1), pp. 1-7, 2002.
- [23] Zhao, Wayne Xin and Jiang, Jing and Yan, Hongfei and Li, Xiaoming, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid," *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 56-65, 2010.
- [24] Blei DM, Andrew YN, Michael II, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.
- [25] Minka T, Lafferty J., "Expectation Propagation for the Generative Aspect Model," *Proc of UAI2002*, 2002.
- [26] Griffiths TL, Steyvers M., "Finding scientific topics," *The National Academy of Sciences*, Vol. 101, pp. 5228-5235, 2004.
- [27] K. Nigam, A. K. McCallum, S. Thrun, T. Mitchell, "Text classification from labeled and unlabeled documents using EM," *Machine Learning*, Vol. 39(2-3), pp. 102-134, 2000.
- [28] Zelikovitz S, Hirsh H, "Improving Short-text Classification Using Unlabeled Background Knowledge to Assess Document Similarity," *Proceedings of the Seventeenth International Conference on Machine Learning*, 2000.
- [29] Zelikovitz, Sarah, "Using background knowledge to improve text classification," 2002.
- [30] D Song, P D Bruza, Z Huang et al. "Classifying Document Titles Based on Information Inference," *Proceedings of the 14th International Symposium on Methodologies for Intelligent System*, pp. 297-306, Japan, 2003.
- [31] Cabrera, Juan Manuel and Escalante, Hugo Jair and Montes-y-mez, Manuel, "Distributional term representations for short-text categorization," *Computational Linguistics and Intelligent Text Processing*, pp. 335-346, 2013.
- [32] Sun, Aixin, "Short text classification using very few words," *Proceedings of the 35th international ACM SIGIR*

- conference on Research and development in information retrieval*, pp. 1145-1146, 2012.
- [33] Feng, Xiao and Shen, Yang and Liu, Chengyong and Liang, Wei and Zhang, Shuwu, "Chinese Short Text Classification Based on Domain Knowledge," *International Joint Conference on Natural Language Processing*, pp. 859–863, 2013.
- [34] Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, Murat Demirbas, "Short Text Classification in Twitter to Improve Information Filtering," *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, Jul. 2010
- [35] Huang Wbnlian, Li Shijian, li Jiuxinl, Xu Congfu, "c," *Jorunal of Beijing University of posts and telecommunications*, Vol. 31(3), 2008.
- [36] José Marí Gómez Hidalgo, Guillermo Cajigas Bringas, Enrique Puertas Sáñez, "Content Based SMS Spam Filtering," *DocEng'06*, Amsterdam, The Netherlands, 2006.
- [37] WEI-WEI DENG, HONG PENG, "Research on a naïve Bayesian based short message filtering system," *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, Dalian, pp. 13-16 August 2006.
- [38] Han Jiawei, Kamber M. "DATA MINING-Concepts and Techniques," 2001.
- [39] S. Dubussion, E. Devoine. M. Masson, "A solution for facial expression representation and recognition," *Signal Processing Image Communication*, Vol. 17 (9), pp. 657-673, 2002.
- [40] Xiaojun Quan, Gang Liu et al., "Short text similarity based on probabilistic topics." *Knowl Inf Syst*, pp. 473-491, 2009.
- [41] Rafeeqe P C, Sendhilkumar S. "A survey on Short text analysis in Web," *Advanced Computing (ICoAC), 2011 Third International Conference on IEEE*, pp. 365-371, 2011.
- [42] Li H, "Text Classification Retrieval Based on Complex Network and ICA Algorithm," *Journal of Multimedia*, Vol. 8(4), pp. 372-378, 2013.
- [43] Hu J, Li J, Zeng Z. "SWSCF: a semantic-based Web service composition framework," *Journal of Networks*, Vol. 4(4), pp. 290-297, 2009.
- [44] Jun H Y, Xin J J, You C H. "Chinese Short-Text Classification Based on Topic Model with High-Frequency Feature Expansion." *Journal of Multimedia*, Vol. 8(4), pp. 425-431, 2013.

# Multi-agent Remote Sensing Image Segmentation Algorithm

Chen Jing<sup>1</sup> and Wang Haifeng<sup>1,2</sup>

1. College of Electronics and Information Engineering, Qiongzhou University, Sanya, Hainan 572022, P. R China

2. Information and Computer Engineering College, Northeast Forestry University, Harbin, 150040, P. R. China

Email: chenjingmyname@163.com, wyfxxz@163.com

**Abstract**—Due to fractal network evolution algorithm (FNEA) in the treatment of the high spatial resolution remote sensing image (HSRI) using a parallel global control strategies which limited when the objects in each cycle by traversal of and not good use the continuity of homogenous area on the space and lead to problems such as bad image segmentation, therefore puts forward the remote sensing image segmentation algorithm based on multi-agent. The algorithm in the merger guidelines, combining the image spectral and shape information, and by using region merging process of multi-agent parallel control integral, its global merger control strategy can ensure algorithm has the advantages of parallel computing and fully considering the regional homogeneity, and continuity. Finally simulation experiment was performed with FNEA algorithms, experimental results show that the proposed algorithm is better than FNEA algorithm in dividing the overall effect, has a good stability.

**Index Terms**—Diffusion; Breed; Agent; Parallel Computation

## I. INTRODUCTION

Multi-agent system is a collection of multiple agents; its goal is to transform large and complex system into small, communication and coordination to each other, easy to management system [1]. Its studies involve the agent's knowledge, goals, skills, planning and how to make the agent to take concerted action to solve problems. Researchers mainly study the interaction communication between the agent, coordination and cooperation, conflict resolution, etc., emphasized cooperation group between multiple-agent, rather than the autonomy and play of individual ability, mainly illustrate how to analyze, design, and integrate multiple agent to constitute cooperation system [3-6].

High spatial resolution remote sensing image (HSRI) has rich texture information and shape information, compared with the low resolution image, can get more sophisticated surface feature information [7-9]. However, with the improvement of spatial resolution images, spectrum distribution of features in images is more complex, and of different feature spectrum overlapping, feature information present highly details, the feature classes internal variance in images increases, variance in different features classes decreases, and these characteristics make computer interpretation way must be

transform from the traditional way of pixels oriented spectrum processing to object oriented multi features approach [10-12]. Object-oriented approach is usually obtained multiple pixel object by spatial adjacent, spectral similar through image segmentation. Image segmentation, on the one hand reduces the feature classes internal variance, on the other hand can provide computer interpretation texture and shape information of object, increase the variance between feature classes, to increase the separable of feature category, and to increase the accuracy of classification and identification [13].

For remote sensing image segmentation algorithms, such as regional growth and watershed algorithm is mostly aimed at SAR (Synthetic Aperture Radar) images and middle and low resolution remote sensing images, such as TM and SPOT images, for high resolution remote sensing image segmentation, the research is relatively small [14]. Along with the development of remote sensing technology, the earth resource satellite place provided remote sensing image has higher spatial resolution, breakthrough m level resolution remote sensing image data, have been able to distinguish clearly the details of the ground characteristics, thus in the agricultural, forest, mining, environmental current situation survey, and other social fields has a broad application market [15]. At present, with high spatial resolution remote sensing image automatic analysis and understanding technology, especially the segmentation technology is still very immature, large amount of information in the images can't get the full application [16].

Commonly used remote sensing image segmentation algorithms are mainly the segmentation method based on edges and the segmentation method based on region. Due to the complexity of remote sensing image feature category, the segmentation method based on the edge tends to has not a closed area, is not conducive to extract HSRI object information. Commonly used segmentation algorithm based on region has split-merge algorithm, based on morphological watershed algorithm and fractal network evolution algorithm. Split-merge algorithm is a kind of high efficiency based on quad-tree segmentation method, but it can't make full use of regional homogeneity, and continuity, for feature complex remote sensing image, segmentation effect is poorer; Based on morphological watershed algorithm makes gradient

amplitude figure of image as a topographic map, pixel gradient amplitude value as a pixel altitude, the affect region of each local gradient amplitude minimum form a reception basin, the boundary of reception basin is watershed, but the watershed algorithm is sensitive to remote sensing image weak edge, easy to produce serious over-segmentation phenomenon, through the pre-processing, notation, post-processing methods improve the usability of the algorithm [17]. FNEA algorithm based on local optimum objects merge criterion, combined with the spectral information and shape information of the object to identify, can get better segmentation result. In addition, with the development of computer science, artificial intelligence, and other fields, a large number of new theories and methods, such as fuzzy sets, neural networks, graph theory and level set, morphological theory, has been introduced into the segmentation, the domestic and foreign scholars put forward many threshold fuzzy automatic segmentation, pulse coupled neural network segmentation, graph theory segmentation based on minimum spanning tree, multidimensional level set segmentation, average drift segmentation, improve HSRI segmentation effect. The above segmentation algorithm, FNEA method has been widely used as HSRI core segmentation algorithms in software processing. However, this method uses a parallel global control strategy that restricts each cycle subject by traversal, can not good use of the continuity of homogenous area on the spatial.

In order to improve the effect of HSRI segmentation, this paper proposes a high resolution remote sensing image segmentation algorithm based on multi-agent theory (MARSS). Multi-agent theory (multi-agent) has parallel computing ability and high flexibility in the global control, make its more efficient in processing engineering problems, has been successfully applied to image segmentation, communications, medical image processing and other fields.

This paper mainly in the following aspects as the development and innovative work:

(a) According to the fractal network evolution algorithm is used when dealing with high spatial resolution remote sensing image is to restrict objects in each cycle by traversal of a parallel global control strategies and not good use the continuity on the space of homogenous area, resulting in image segmentation result is bad and poor stability, at the same time because of the multi-agent theory in the global control with parallel computing ability and high flexibility, make its more efficient in processing engineering problems, therefore puts forward the remote sensing image segmentation algorithm based on multi-agent. The algorithm in the merger guidelines combine image spectral and shape information, and by using multi-agent parallel control integral region merging process, in its global merger control strategy can ensure algorithm has the advantages of parallel computing and fully considering the regional homogeneity, and continuity.

(b) in order to further validate the correctness and validity of remote sensing image segmentation algorithm

based on multi-agent is proposed in this paper, a simulation experiment was performed with FNEA algorithms, the contrast evaluation method of high spatial resolution remote sensing image segmentation result adopted the unsupervised evaluation method, the experiment using two groups of data sets of high spatial resolution remote sensing image, a set of Sanya of Hainan province village area, another group is in the Beijing area, the experimental results show that: compared with FNEA, although this algorithm of large difference area of the same material will produces the over-segmentation phenomenon, but algorithms in this paper for small physical differences area, segmentation effect is better. General score (GS) evaluation index taking overall result evaluation to the segmentation results, GS value of this paper is lower than FNEA algorithm, the overall effect is better. Through the intelligent initialization number parameter analysis to this paper's algorithm shows that this algorithm has better stability, and the segmentation result is better than FNEA algorithm as a whole.

## II. PROPOSED ALGORITHM

Agent derived from the concept of people's knowledge of artificial intelligence, has been widely used in physics, biology, computer, social and other fields in complex system simulations. Agent perceive environment, according to the accessed environmental information and their own status, effects on environment, makes there exist very strong interactivity and flexibility in the agent and the environment. At the same time, when many agent working at the same time, there is maintained strong independence between the agents, make the multi-agent has the characteristics of parallel computing in the global control.

Based on the multi-agent system strong interaction with the environment, the advantages of high flexibility, parallel control, using multi-agent diffusion, copy, reproduce, etc. Operator proposed high resolution remote sensing image segmentation method based on multi-agent theory (MARSS). In order to facilitate MARSS methods described, explained for the following variables:

(a) To split HSRI  $e_s = \{r_1, r_2, \dots, r_n\}$  is the working environment of MARSS algorithm,  $r_n$  as the  $n_{th}$  pixel in the image, N is number of image pixels, and when initialization, each pixel  $r_n$  was regarded as a figure spot object;

(b) The figure object collection of high resolution image segmentation is  $o = \{o_1, o_2, \dots, o_t\}$ ,  $o_t$  is the first figure spot  $T$  is the total number of figure spot;

(c) The agent set  $AG = \{Ag_1, Ag_2, \dots, Ag_n\}$ , each agent  $Ag_n$  contains perception  $p_n$ , merger object rules  $r_n$  and  $a_{sn}$  status attribute and behavior collection  $a_{cn}$  4 parts, the specific meaning in the MARSS method as shown in table 1.  $p_n$  and  $a_{sn}$  embodies the strong interactions of agent and image, the setting of  $a_{sn}$  and  $a_{cn}$  makes agent more flexible in the segmentation.

TABLE I. MULTI-AGENT THEORY AND MARSS VARIABLES CORRESPONDING INSTRUCTION

Multi-agent theory	MARSS
Multi-agent set AG	For intelligent AG segmentation
Intelligent individual $Ag_n$	$Ag_n$ Said search the segmentation process executive
Perceptron $p_n$	$p_n$ as Intelligent individual $Ag_n$
Ruler $r_n$	$r_n$ as Intelligent individual $Ag_n$
Status attribute $a_{sn}$	$a_{sn} = (a_r, p_r, p_o, c)$
Action $a_{sn}$	$a_{sn} = \{a_1, a_2, a_3\}, a_1, a_2, a_3$ Representing the diffusion behavior, reproduction, death, for the search, combined with object, state attribute update agent.

For the convenience of description of MARSS theory, take picture 1 for an example to describe the algorithm process. In the whole process of image segmentation, an arbitrary agent  $Ag_n$  in multi-agent system in different environment, through the behavior such as diffusion, reproduction, death search the mutual best merge objects, and with mutual best merger guidelines as object determination criterion for the combining operation, finally achieve the goal of segmentation. Specific as follow:

#### A. Initialization

When initialization (figure 1 (a)), the multi-agent system uses a certain distribution way (such as uniform distribution) makes multi-agent distributed in figure spot object collection  $o = \{o_1, o_2, \dots, o_t\}$ , initialize each pixel is a spot. At this point, the perception of agent  $Ag_n$ ,  $p_n$  pointing to the corresponding objects  $o_T$ , at the same time state attribute  $((a_e, p_r, p_o, c) = (0, \text{false}, \text{true}, \text{true}))$ .

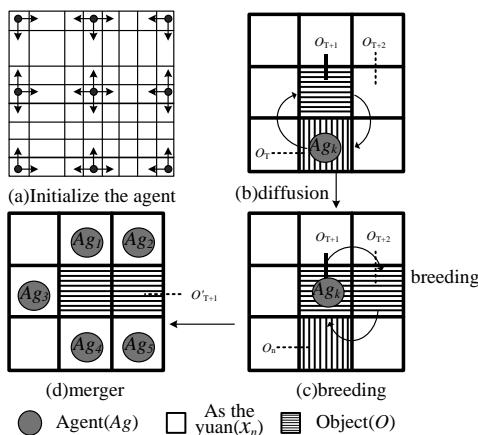


Figure 1. Is a merge strategy controlled by Multi-agent

#### B. The Behavior of the Multi-agent

After the initialization, MARSS through the agent's behavior such as diffusion, reproduction, death for split operation.

##### 1) Agent Spread

According to  $r_n$ , assuming that  $Ag_n$  find candidates for merging of  $o_t$ ,  $o_t + 1$ , but the candidate merger object of  $o_t + 1$  is not  $o_t$  (figure 1 (b)),  $o_t$  and  $o_t + 1$

have large heterogeneity,  $Ag_n$  selective diffusion behavior (figure 1 (c)).  $Ag_n$  spread to  $o_t + 1$ , namely the  $p_n$  point to  $o_t + 1$ ,  $Ag_n$  age increase at the same time, the other properties remain unchanged, the expression as type (1), among them, the  $o_{\min}(o_t)$  represents best combined object identifier

$$\begin{aligned} b_1 : & p(Ag_n) \rightarrow o_t + 1, r_e(Ag_n) \leftarrow r_e(Ag_n) + 1 \\ & \text{if } o_t + 1 = o_{\min}(o_t) \text{ and } o_t \neq o_{\min}(o_t + 1) \end{aligned} \quad (1)$$

##### 2) Agent Spawn

When the candidate merger object that the  $Ag_n$  referred to object  $o_t + 1$  is  $o_t + 2$ , at the same time  $o_t + 2$  is  $o_t + 1$  candidates for merging, which can meet the mutual best merger object condition, then the  $Ag_n$  choose reproduction (figure 1 (c)) reproduce offspring agent, and the child agent join to the original agent collection, at the same time  $Ag_n$  merger  $o_t + 1$  and  $o_t + 2$  (figure 1 (d)), make the  $p_r = \text{true}$ , its expression as shown on the type (2). Among them, the  $\{Ag_n\}_m$  is the offspring intelligence collection that  $Ag_n$  reproduce,  $m$  is the number of objects adjacent with  $Ag_n$ , and  $o_t + 1$  as the new object after merge.

$$\begin{aligned} b_2 : & Ag \leftarrow \{Ag_n\}_m, o_t + 1 = \{o_t + 1, o_t + 2\}, p_r = \text{true} \\ & \text{if } o_t + 2 = o_{\min}(o_t + 1) \text{ and } o_t + 1 = o_{\min}(o_t + 2) \end{aligned} \quad (2)$$

##### 3) Agent Death

When  $Ag_n$  that pointing to  $o_t$  to satisfy any of the following four conditions,  $Ag_n$  choose death,  $Ag_n$ s are cleared from the agent collection, its expression as shown in type (3)

$$a_3 : Ag \leftarrow Ag \setminus \{Ag_n\} \quad (3)$$

(1) Age properties of  $(Ag_n, a)$  exceeds the specified agent age of ( $t_h$ ), namely  $a_e > t_h$ .

(2)  $Ag_n$  has chosen a reproductive behavior, namely,  $p_r = \text{true}$ .

(3)  $Ag_n$  could not find the candidates for merging of  $o_t$ ,  $o_t + 1$ , namely  $c = \text{false}$ .

(4)  $o_t$ , the pointed object of  $Ag_n$  has been incorporated into the other object, namely the  $Po = \text{false}$ .

Agent behavior need to set up two parameters: the agent age  $t_h$  and agent initialization number  $n$ . Among them, choice principles of  $t_h$  for all agents have the ability to traverse the entire image range when diffusion, so  $t_h$  to meet  $4N > N$ ,  $N$  for image size, when  $t_h$  exceeds a certain value, the effects on the segmentation result is very small. Under a certain distribution, parameter  $n$  determines the referents of agent perception  $p_n$ .

### C. The Mutual Best Merge Criteria

Agent need according to the appropriate criteria for judging whether the object is the best combination object, MARSS algorithm in three different behaviors in the segmentation process are adopting mutual best merge norms as the agent objects merge rule  $r_n$ . Mutual best merge criteria include two parts: judgments of candidates for merging, determination of mutual best merge objects.

#### 1) Candidates for Merging

In the mutual best merger guidelines, the determination of candidates for merging sees type (4). According to (4) and (5) ~ (9), agent use the spectral information of pattern spot objects (type (6)) and shape information (type (7) - (9)) to find the candidates for merging.

$$\begin{aligned} o_{\min} &= \arg \min \{c(o_t, o_t + j)\}, o_t + j \in N(o_t) \\ o_{\min}^* &= o_{\min}, \text{if } c(o_t + j, o_{\min}) \leq \text{scale}^2 \end{aligned} \quad (4)$$

In the formula, as candidates for merging;  $o_t$  is the object  $p_n$  pointing to;  $o_t + j$  for  $o_t$  adjacent objects; Scale for the scale of segmentation.  $B(o_t, o_t + j)$  for the  $o_t$  and  $o_t + j$ ,  $j$  merger cost, calculation methods as shown in (5). Merger cost function contains spectral information merge costs CCLR ( $o_t, o_t + j$ ) and shape information merging cost CSHP ( $o_t, o_t + j$ ), the calculation formula respectively for type (6) with type (7), WSHP for the shape information weight

$$B(o_t, o_t + j) = \varpi_{shp} c_{shp}(o_t, o_t + j) + o_t, (1 - \varpi_{shp}) c_{shp}(o_t, o_t) + j, \varpi_{shp} \in [0, 1] \quad (5)$$

$$c_{clr}(o_t, o_t + 1) = \sum_{j=1}^c \varpi_i (l_{mrg} \theta_i^k - j_i \theta_t^k) \quad (6)$$

$$\begin{aligned} c_{shp}(o_t, o_t + j) &= \varpi_{cmp} c_{cop}(o_t, o_t + 1) + (1 - \varpi_{emp}) \\ c_{smk}(o_t, o_t + j), \varpi_{cmp} &\in [0, 1] \end{aligned} \quad (7)$$

Type (6),  $o_t^k, o_{t+j}^k, o_{mrg}^k$  respectively respect  $o_t, o_t + j$  and standard deviation that combined objects on  $\text{mrg}$  in  $k$  band contained pixel;  $n_t, n_t + j, n_{mrg}$  represent the number of pixel in  $o_t, o_t + j, o_{mrg}$ ;  $B$  represents the band number of image;  $w_k$  for the weight of the first  $k$  band. In type (7), containing the compact degree merger price of the object CCMP ( $o_t, o_t + j$ ) and smoothness combined price CSMH ( $o_t$  and  $o_t + j$ ), calculation formula respectively type (9) and (8), WCMP for compact degrees weight coefficient. In type (8)  $l_t, l_t + j, l_{mrg}$  respectively the perimeter of  $o_t, o_t + j, o_{mrg}$ . In type (9),  $b_t$ , respectively the perimeter of the external rectangle  $o_t, o_t + j, o_{mrg}$ .

$$c_{cmp}(o_t, o_t + j) = n_{mrg} \frac{l_{mrg}}{\sqrt{n_{mrg}}} - n_t \frac{l_t}{\sqrt{n_t}} - n_t + j \quad (8)$$

$$c_{smh}(o_t, o_t + j) = n_{mrg} \frac{l_{mrg}}{b_{mrg}} - n_t \frac{l_t}{b_t} - n_t + j \quad (9)$$

Candidates for merging determine formula of agent at the same time, taking into account the image spectral information and shape information, you need to set up four parameters: the first  $k$  band weighting  $w_k$ , shape weight WSHP, compact degree weight WCMP, segmentation scale. In segmentation,  $w_k$  assignment according to the importance of spectrum, a band is important, the weight is heavy. Due to split information is indispensable of the spectra, WSHP value not more than 0.9. WCMP can choose according to feature state, when feature is compact, choose larger value, when features are dispersive, choose smaller value. Scale controls the final figure spot size, larger Scale values; finally get the figure spot correspondingly larger.

#### 2) Determination of Mutual Best Merge Objects

(1) The search for candidate merges objects. For a scale, according to the type (4), if  $o_t$  candidates for merging is  $o_t + 1$ .

(2) Determine the mutual best merge object. According to (4), if the  $(o_t + 1)$  candidate merger object is  $o_t, o_t$  and  $o_t + 1$  are the mutual best combined object.

### D. Segmentation Algorithm

MARSS algorithm process is shown in figure 2.

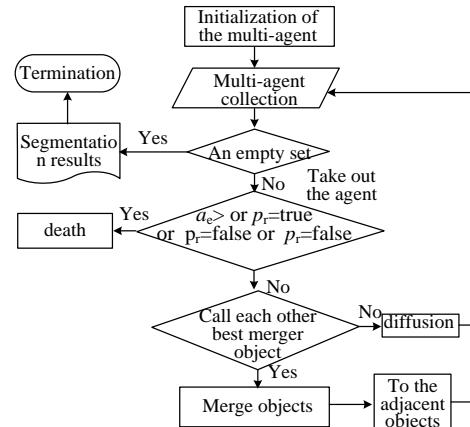


Figure 2. MARSS algorithm flow chart

(a) Initialization. Each pixel  $x N$ , initialized to a figure spot  $o_n$ , get set  $O = \{o_1, o_2, \dots, o_n\}$ ,  $m$  agent consisting of a collection  $Ag = \{A_{gn}\}$   $m$ , attributes  $A_{gn} = (a_e, pr, Po, c) = (0, \text{false}, \text{true}, \text{true})$ , the  $p_n$  point to  $o_n$ .

(b) If the  $Ag = \theta$ , output segmentation results  $o = \{o_1, o_2, \dots, o_t\}$ , or execute Step (3).

(c) From the  $Ag$  take the  $A_{gn}$ . Determine  $A_{gn}$  status, if the  $A_{gn}$  has executed reproduction ( $pr = \text{true}$ ), age more than threshold ( $a_e > t_h$ ),  $A_{gn}$  can not find candidates for merging ( $c = \text{false}$ ), or  $A_{gn}$  pointing to the object  $o_t$  has been merged into other objects ( $Po =$

false), the  $A_{gn}$  death is cleared out agent set Ag, execute step (2), otherwise execute the step (4).

(d) Looking for mutual best merge object,  $A_{gn}$  calculate the candidate merge object of  $o_t$  for  $o_t + 1$ , the candidate object  $o_t + 1$  to find is  $o_t$ ,  $A_{gn}$  to reproduce, or  $A_{gn}$  diffuse. Take Step (2).

### III. EXPERIMENTAL RESULTS

In current remote sensing image segmentation software e Cognition software multi-scale segmentation, namely FNEA segmentation algorithm, shows excellent performance. In order to check the MARSS calculation method have effectiveness, use HSRI two groups of data sets, and were analyzed compared to FNEA test results . HSRI segmentation result contrast evaluation method uses unsupervised evaluation method.

#### A. Segmentation Result Evaluation Method

Evaluation method of image segmentation algorithms can be divided into three categories: visual evaluation, supervision evaluation and non-supervision evaluation. In the supervision and evaluation, access to segmentation reference image with very strong subjectivity, very time consuming, taking high quality reference image of complex remote sensing image is difficulty. In order to quantitatively evaluate the overall effect of the Segmentation results of remote sensing Image, this paper adopted the non-supervision Evaluation method that JOHNSON. B mentioned in Unsupervised Image Segmentation Evaluation and Refinement Using a Multi - scale Approach

$$GS = Var_{norm} + MI_{norm} \quad (10)$$

$$Var_{ar} = \frac{1}{c} \sum_{i=1}^b \frac{1}{h} \sum_{j=1}^h h_i (\sigma_j^i)^2 \quad (11)$$

$$MI = \frac{1}{c} \sum_{j=1}^b \frac{h \sum_{i=1}^k n_i (\theta_i^j)^2}{\sum_k (y_i - y_j)^2 \left( \sum_{i=1}^k \sum_{j=1}^h \sigma_{ij} \right)} \quad (12)$$

In type (10), the  $Var_{norm}$  and  $MI_{norm}$  respectively the values of  $Var_{ar}$  and MI after normalized, the normalized formula as shown in type (13); X as a variable ( $Var_{ar}$  or MI),  $x_{min}$  and  $x_{max}$  respectively the minimum and maximum value of variable,  $x_{norm}$  for normalized variable values

$$x_{norm} = (x - x_{min}) / (x_{max} - x_{min}) \quad (13)$$

Equation (11), for the standard deviation of first figure spot the first k band;  $N_t$  said t figure spot object pixel number; N for spot number; B for the band. In type (12),  $y_{ik}$  for wave degree average of figure spot t; mean to image the first k wave degree average; WTT 'shows

that object t and t' of the adjacent relation, if the object t, t 'adjacent, the WTT' = 1, otherwise the WTT '= 0.

$Var_{ar}$  value is small, which indicates that high homogeneity within the overall segmentation image spot, the segmentation effect is good in figure spot;  $Var_{ar}$  value is large, it shows that in the overall segmentation image spot, the homogeneity is low, the segmentation effect is poor in figure spot, owe segmentation degree is high. MI value is small, shows high heterogeneity between segmentation image blocks, segmentation effect is good between the figure spot; MI value is big, show that the heterogeneity between segmentation image is low, segmentation effect is poor between the figure spot, splitting degree is high. Global score is small, indicating that the overall segmentation effect is good, low degree of owe segmentation and over-segmentation.

#### B. Split Test

Experiment 1 image for Sanya village area of Hainan province in 2012 size of 400 pixels by 400 pixels, QUICKBIRD image of resolution of 2.4 m 4 band, as shown in figure 3 (a)). Trial FNEA algorithm and MARSS algorithm parameter is set to scale = 50, band weight wk = 1 a/b, shape weight WSHP= 0.1 and the compact degree WCMP = 0.5, b for the image band. MARSS algorithm is initialized by uniform distribution, the initial agent number is 2000, and under the distribution for a given initialization number of imaging agent determines the agent referents. FNEA algorithm and MARSS algorithm segmentation results, respectively, as shown in figure 3 (b), (c).

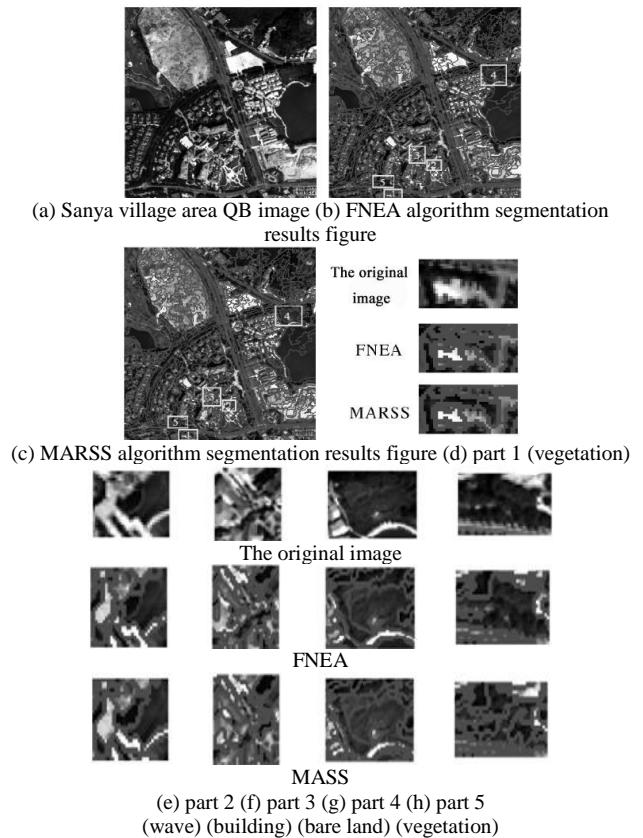


Figure 3. Sanya village region QB image segmentation result

Test 2 data is the QUICKBIRD image in Beijing area 548 pixels by 463 pixels resolution of 2.4 m (as shown in figure 4 (a)). Wave band weighting  $w_k$ , shape weight WSHP and a compact WCMP parameter is respectively set to  $1/b$ , 0.1 and 0.5,  $b$  for the image band. Initial agent for 2000 MARSS, and segmentation results of FNEA in scale of 50, as shown in figure 4 (b) (c).

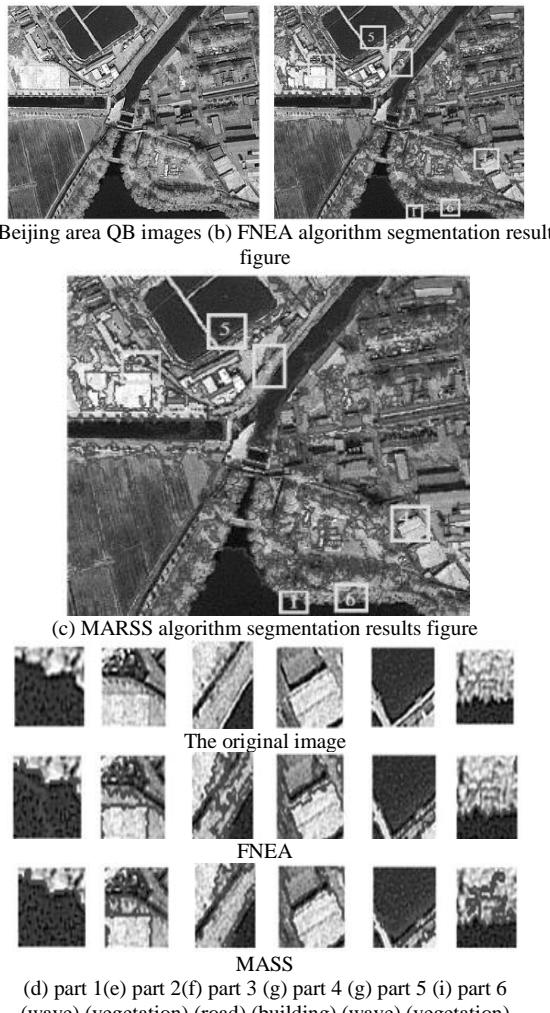


Figure 4. Beijing area QB image segmentation results

In order to clearly contrast the segmentation results of two methods on visual, the test 1 and 2 respectively choose five and six children regions enlarge to compare and analysis, the chosen area contains water, vegetation, roads, construction and other typical feature categories, as shown in figure 3 (d) ~ (h). From figure 3 shows that, compared with MARSS, FNEA in the region that same material differences are large, the segmentation results are good, as shown in figure 3 (h) the vegetation area difference is big, FNEA segmentation result is complete; Compared with FNEA, MARSS algorithm in different material difference smaller region, segmentation effect is better, as shown in figure 3 (d) the vegetation and roads, water and vegetation of figure 3 (e), figure 3 (f) of building with different roof, figure 3 (g) bare land border with the plant light spectrum difference is relatively small, MARSS can also accurately partition it, ensure the

correctness of the follow-up processing (e.g., classification). Also, figure 4 (d) to (g) segmentation result is shown in different material differences is smaller region, MARSS algorithm compared FNEA algorithm segmentation result is better; In figure 4 (h), (I) shown water vegetation in the same material in regions with large differences, MARSS algorithm segmentation results relatively FNEA algorithm is more fragile.

In order to more accurately and objectively evaluate the segmentation results of MARSS algorithm with FNEA algorithm, take section 3.1 GS indicators for quantitative evaluation of the segmentation results of two kinds of algorithm, the result of experiment 1 and experiment 2 as shown in table 2.

TABLE II. STATISTICS OF MARSS ALGORITHM WITH FNEA ALGORITHM SEGMENTATION

	Segmentation method	$MI_{nom}$	$Var_{nom}$	GS
test 1	FNEA	0.234	0.566	0.801
	MARSS	0.216	0.551	0.767
test 2	FNEA	0.233	0.613	0.847
	MARSS	0.153	0.618	0.771

According to the results of Table 2, in experiment 1 and experiment 2, M I norm index of MARSS calculate method were 0.216 and 0.153, respectively, less than FNEA algorithm 0.234 and 0.233, shows that overall MARSS algorithm segmentation effect is good in FNEA algorithm. Experiment 1 MARSS algorithm  $Var_{nom}$  index is 0.551 less than the 0.566 of FNEA algorithm, illustrate MARSS algorithm compared to FNEA algorithm whole owe segmentation effect is better; And test 2 MARSS algorithm  $Var_{nom}$  index 0.613 large than FNEA algorithm, illustrate MARSS algorithm under segmentation effect compared to FNEA algorithm is poorer. But MARSS algorithm in experiment 1 and experiment 2 GS global index were 0.767 and 0.767, lower than FNEA algorithm 0.801 and 0.847, MARSS algorithm in the over segmentation and under segmentation effect is better than FNEA algorithm

From theory and experiment analysis we can see, the article MARSS algorithm of overall segmentation is better than FNEA algorithm because of MARSS algorithm in global objects merge strategy wipe our FNEA all the objects in each cycle can be traversed only one time, can better use the spatial continuity on homogenous area. To illustrate the advantages of MARSS, take 5 as an example to illustrate. As shown in figure 5 (a), in one loop that FNEA algorithm global control, set object o1, o2, o3 for the same feature, o4 is another feature. According to the mutual best merger guidelines, object o1 and o2 as the mutual best merge, the object o1 can be traversed only once, o3 is no longer considered object o1, may exist object o4 was identified as the object of o3 mutual best merge, in the final merged result, object o3 was mistakenly merger with o4 (as shown in figure 5 (b)). As the global control loops, FNEA algorithm in o3 and o4 combined error will be continuously enlarged. For MARSS, as shown in figure 5 (c), (d), the method canceled the constraints of traversal

times, Agent  $A_{gn}$  find the mutual best merger object  $o_1$  and  $o_2$  through diffusion behavior, will perform reproduction,  $o_1$  and  $o_2$  combined object is  $o'_1$ , breeding progeny agent  $Ag_2$  will make perception  $P_2$  pointing to objects,  $o_3$ , and search for the mutual best object  $o'_1$ , and will eventually solvated object  $o_3$ ,  $o'_1$  (as shown in figure 5 (e)). Therefore, the MARSS algorithm more flexible global control can take account of homogenous area on the space continuity, improve the segmentation effect of different substances on a smaller area (e.g., bare land and vegetation area). And MARSS algorithm take account of the space continuity on homogenous area at the same time, easy to create over segmentation effect in the region of large difference of same material (e.g., vegetation), but on the whole segmentation effect, MARSS algorithm can realize the global optimization control and improve the high resolution image segmentation results.

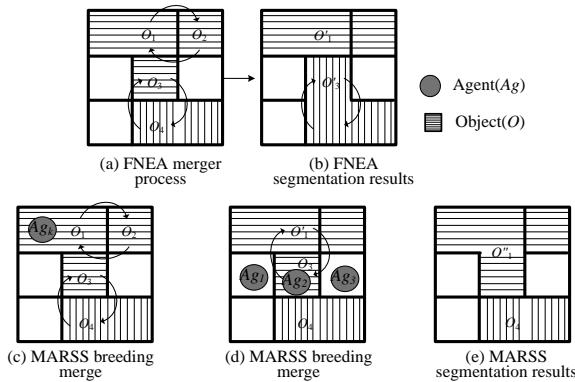


Figure 5. Global merger control comparison diagram of the MARSS and FNEA algorithm

### C. Parameter Analysis

Agent initialization number and segmentation scale two parameters influence MARSS algorithm segmentation results a lot. This section studies the influence of agent initialization number and scale of MARSS algorithm segmentation. When testing, agent initial choose the number of 1000, 2000, 3000, 2000 (respectively write as MARSS1000, MARSS2000, MARSS3000, MARSS4000), scale chose 10 to 100, step length of 10 multiple values. Test results as shown in figure 6.

Test 1 results (figure 6 (a)) shows the initial agent number for 1000, 2000, 4000, of the global index GS that MARSS algorithm on different scale segmentation results were superior to FNEA algorithm. When the initial agent number is 3000, relatively FNEA algorithm, MARSS algorithm except in 90 and 100 two scales of segmentation effect is poor performance, GS in eight other scales value better than FNEA GS of the segmentation results. Comprehended the above test results, to initialize agent number 1000, 2000, 3000, 2000, the segmentation results of scale degrees of 10 to 100, and MARSS algorithm overall segmentation effect is better than FNEA algorithm, and has good stability. Due to the different distribution of features, different HSRI

optimal segmentation scale is different, MARSS algorithm in low dimension and scale is too high, can't get the optimal segmentation result, choosing the right segmentation scale according to the actual image. For experiment 1 and experiment 2 images, MARSS algorithm and FNEA algorithm are similar for images optimal segmentation scale, about 40 or 30.

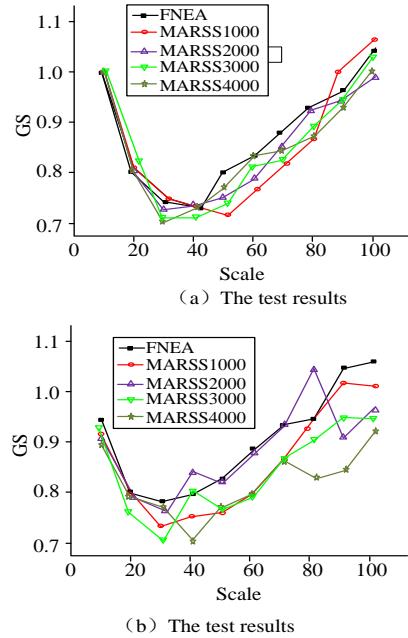


Figure 6. The influence of agent initialization number and segmentation scale parameters to MARSS algorithm segmentation

### IV. CONCLUSION

By using multi-agent and image environment strong interactivity, the advantages of high flexibility, parallel computing, this paper proposes a multi-agent based theory of high spatial resolution remote sensing image segmentation algorithm. The algorithm in the merger guidelines, combining image spectral and shape information, and by using multi-agent parallel control integral region merging process, its global merger control strategy can ensure algorithm has the advantages of parallel computing and fully considering the regional homogeneity, and continuity. Split test results show that, compared with FNEA, although this algorithm for the same material exist large difference area produces the over-segmentation phenomenon, but in this paper, algorithms for small physical differences of different region, segmentation effect is better. using all scored evaluation indexes take overall segmentation result evaluation to the segmentation results, this paper algorithm of GS value lower than FNEA algorithm, which split overall effect is better. Through intelligent initialization number parameter analysis, shows that this algorithm has better stability, and the segmentation result is better than FNEA algorithm as a whole.

### ACKNOWLEDGMENT

This work was supported in part by a grant from the Scientific Research Foundation of Qiongzhou University

(No. QYQN2013-26), the project of science foundation of educational department of HaiNan province (Hjkj2012-46)

## REFERENCES

- [1] Kasman Suhairi, Ford Lumban Gaol, The Measurement of Optimization Performance of Managed Service Division with ITIL Framework using Statistical Process Control. *Journal of Networks*, Vol 8, No 3 (2013), 518-529.
- [2] Guang Yan, Zhu Yue-Fei, Gu Chun-Xiang, Fei Jin-long, He Xin-Zheng, A Framework for Automated Security Proof and its Application to OAEP. *Journal of Networks*, Vol 8, No 3 (2013), 552-558.
- [3] Muhammad J. Mirza, Nadeem Anjum. Association of Moving Objects across Visual Sensor Networks. *Journal of Multimedia*, Vol 7, No 1 (2012), 2-8.
- [4] Pearson S. Taking account of privacy when designing cloud computing services. In *CLOUD '09: Proceedings of the 2009 ICSE workshop on software engineering challenges of cloud computing*, IEEE Computer Society, Washington, DC, USA, 2009. pp. 44-52. <http://dx.doi.org/10.1109/CLOUD.2009.5071532>.
- [5] Mondol, J.-A.M. Cloud security solutions using FPGA. In *Communications, Computers and Signal Processing (PacRim), 2011 IEEE Pacific Rim Conference on*, 2011, pp. 747-752.
- [6] Wang Lina, Gao Hanjn, Liuwei, Peng yang. Detection and management of virtual machine monitor. *Research and development process of Computer*, 2011, pp: 1534-1541.
- [7] Yinghua Xue, Hongpeng Liu, Intelligent Storage and Retrieval Systems Based on RFID and Vision in Automated Warehouse. *Journal of Networks*, Vol. 7, No. 2 (2012), pp: 365-369.
- [8] Huan Zhao, Kai Zhao, He Liu, Fei Yu, Improved MFCC Feature Extraction Combining Symmetric ICA Algorithm for Robust Speech Recognition, *Journal of multimedia*, Vol. 7, No. 1, 2012. pp: 74-81.
- [9] ISO/FDIS 25178-2, Geometrical product specifications (GPS) – Surface texture: Areal – Part 2: Terms, definitions and surface texture parameters, ISO, Geneva, 2010.
- [10] LI La yuan, LI Chun lin, "A multicast routing protocol with multiple QoS constraints," *Journal of Software*, vol. 15, No. 2, 2004, pp. 286-291.
- [11] Q. He, C. Han, "Satellite Constellation Design with Adaptively Continuous Ant System Algorithm," *Chinese Journal of Aeronautics*. vol. 6, No. 4, 2007, pp. 297-303. [http://dx.doi.org/10.1016/S1000-9361\(07\)60047-8](http://dx.doi.org/10.1016/S1000-9361(07)60047-8)
- [12] L. Y. Ren, "Study on Scheduling Optimization in Construction Project of Lagerstroemia Hope City," *Xi'an University of architecture & technology*. vol. 6, No. 2, 2011, pp. 12-17.
- [13] B Yu, Z Z Yang, B Z Yao, "A hybrid algorithm for vehicle routing problem with time windows," *Expert Systems with Applications*, vol. 38, No. 1, 2011, pp. 435-441. <http://dx.doi.org/10.1016/j.eswa.2010.06.082>
- [14] S. Kudekar, T. Richardson, R. Urbanke, "Threshold saturation via spatial coupling: why convolutional LDPC ensembles perform so well over BER," *IEEE Transactions on Information Theory*, 2011, 57(2): 803-834. <http://dx.doi.org/10.1109/TIT.2010.2095072>
- [15] Y. Yona, M. Feder, "Efficient parametric decoder of low-density lattice codes," *IEEE International Symposium on Information Theory: June 28-July 3, 2009, Seoul, Korea. New York, NY, USA: IEEE*, 2009, 8: 744-748.
- [16] B. Kurkoski, J. Dauwels, "Message-passing decoding of lattices using Gaussian mixtures," *IEEE International Symposium on Information Theory: June 6-11, 2008, Toronto, Canada. New York, NY, USA: IEEE*, 2008, 8: 2489-2493.
- [17] Bickson D, Ihler A, Avissar H et al. A low-density lattice decoder via non-parametric belief propagation. *Forty-Seventh Annual Allerton Conference on Communication, Control and Computing: Sep 30-Oct 2, 2009, Illinois, USA. Monticello, IL, USA: IEEE*, 2010, 1: 439-446.

# Concept Tree Based Information Retrieval Model

Yuan Chunyan

Institute of Civil Engineering, Chang'an University, Chang'an University, Shanxi, Xi'an, China 710061  
Email: jancyer@163.com

**Abstract**—This paper proposes a novel concept-based query expansion technique named Markov concept tree model (MCTM), discovering term relationship through the concept tree deduced by term markov network. We address two important issues for query expansion: the selection and the weighting of expansion search terms. In contrast to earlier methods, queries are expanded by adding those terms that are most similar to the concept of the query, rather than selecting terms that are similar to a signal query terms. Utilizing Markov network which is constructed according to the co-occurrence information of the terms in collection, it generate concept tree for each original query term, remove the redundant and irrelevant nodes in concept tree, then adjust the weight of original query and the weight of expansion term based on a pruning algorithm. We use this model for query expansion and evaluate the effectiveness of the model by examining the accuracy and robustness of the expansion methods. Compared with the baseline model, the experiments on standard dataset reveal that this method can achieve a better query quality.

**Index Terms**—Concept Tree; Markov Networks; Query Expansion; Information Retrieval

## I. INTRODUCTION

Since the earliest days of IR, researchers noted the potential defects of keyword retrieval, such as synonymy, polysemy, hyponymy or anaphora. Although in principle these linguistic phenomena should be taken into account in order to obtain high retrieval relevance, the lack of algorithmic models prohibited any systematic study of the effect of this phenomena in retrieval. Instead, researchers resorted to distributional semantic models to try to improve retrieval relevance, and overcome the brittleness of keyword matches.

Most research concentrated on Query Expansion (QE) methods, which typically analyze term co-occurrence statistics in the corpus and in the highest scored documents for the original query in order to select terms for expanding the query terms [1]. Such techniques are used to augment the original query to produce a representation that better reflects the underlying information need. Current, the most commonly methods used to treatment the query are relevance feedback [2], query reconstruction and automatic query expansion [3]. According to document sets, automatic query expansion technique divided into local analysis method and global analysis method. As an early appeared query expansion method, global analysis method has been widely used.

The traditional global analysis method didn't take the indirect relationship between terms into account. And in

traditional approaches of query expansion, a term will be chosen if it associates strongly with a certain query term in a given query. However, the entire query concept is rarely taken into account, in reality, terms which are strongly similar to the query topic are more important and benefit to improve efficiency. Therefore, such kind of terms should be added to expand original query.

Recently, a Markov network model (MNM) for information retrieval was proposed that goes beyond the simplistic bag of words assumption that underlies BM25 [4]. The Markov network retrieval algorithm, in particular, has been shown to be ideal for concept-based information retrieval. The work in Zuo [5] first proposes and implements an information retrieval model based on Markov network. Markov Network is an undirected graphical network that is capable of efficiently representing relevance in knowledge and can be easily gotten from training data. They construct the Markov network to represent the relationships between terms and the relationships between documents learned from training corpus, then do query expansion and re-compute the similarity between documents and queries. Experiments show the result is promising. In Cao [6] they consider the retrieval model as an inference process. Its isotropy take the query as the evidence source, the relevance document is looked upon as active document, improving the performance of information retrieval by adding useful information into retrieval process. Making use of edges without the direction in the network, activating the words closely related with the query as the additional evidential sources into the model. All the models ignore the dependence between query terms, but it is unrealistic.

Associative models consider relationships between terms in addition to the terms themselves. They have been extensively considered and studied for information retrieval, e.g. by Stiles [7], van Rijsbergen [8] and Salton & Buckley [9] among many others. There are many lexical and semantic relations that may be considered for associating a pair of terms. For example: stemming, based on common morphology; synonymy, where aspects of meaning are shared; co occurrence, in which both words tend to appear together; and general association, where a person is likely to give one word as a free-association response to the other.

Each relation may be thought of as an inference step, in which a source word  $v$  has some property  $R(v)$ , and a new word  $w$  can be inferred to have the property value  $R(w)$  with probability  $PR(w|v)$ , based on their shared

relation. For example, if vis the word ‘matrix’ and the property R(.) is ‘relevancy to a query’, then one possible way to calculate PR (w|v) is based on co-occurrence, so that, for example, the term ‘row’ also has some measure of relevance. Note that this is not symmetric: ‘row’ having more senses and being more common, it is less likely to imply relevance of ‘matrix’, unless another term is also present for context, such as ‘column’.

While lexical and semantic relations may be useful individually, it is important to consider how they may be used in combination. One reason for this is a common problem in language processing called sparsity: for co-occurrence relations for example, even with a huge corpus, we only have reliable co-occurrence data for a fraction of all potential term pairs. External semantic resources such as Word Net or stemming dictionaries supply a broad set of terms but are limited in the depth and currency of their vocabulary. By combining multiple relations into chains of inference, we can help bridge the gaps that exist in the data.

A second reason is that the various relations between words represent potentially complimentary sources of evidence that may help to distinguish and disambiguate terms. For example, if ‘bank’ and ‘merger’ are known to be relevant to a query, then the following inference chains would provide evidence that ‘negotiations’ may also be relevant:

1. bank →agreement (C) →negotiate (C)  
→negotiations (M)
2. merger →talks (C) →negotiations (S)

where C, S, and M represent co-occurrence, synonymy, and morphology relations respectively. Note that chains can emphasize different types of evidence at different walk stages. In the above example, co-occurring terms are found first, followed by their synonyms or stems.

This paper proposes an concept tree information retrieval model based on Markov network, our work differs from the related work in one or more of the following aspects: first, we used the Markov network model to express the indirect relationship between terms. Moreover, we adjusted the contribution of the initial query and select the most appropriate expansion terms association with the entire query concept according to a pruning algorithm. This is done using query concept tree extracted from the text collection, selecting expanded term relies on the overall similarity with the query concepts rather than the similarity with a signal query term. We focus our attention on the problem of generating concept tree and pruning algorithm. Our approach is to provide high level suggestions for expanding the original user query with additional context. Especially in the Multi-keyword query, numbers of query terms are semantically related; we can reduce the possibility of ambiguity, and improve the precision.

In this paper, we first give a brief introduction about previous work. Then present a construction method that allows getting a concept tree from Markov network. Section 3 is a core part; we introduce the detailed steps of the pruning algorithm, and describe a probabilistic query expansion and weighting model through the Markov

concept tree. After describing our test setting, some results of experiments carried out with five standard test collections are presented in section IV. Finally, we conclude with the main findings and point out further research and possible applications of the methods presented.

## II. PRELIMINARIES

Markov Network is an undirected graphical network [10] that is capable of efficiently representing relevance in knowledge and can be easily gotten from training data with strong learning and inferring capability. Since non-direction of the edge in Markov network, construct the Markov network is much easier than the discovery of Bayesian networks. Using the non-direction edge to explain the semantic relationship between items in information retrieval is more intuitive and appropriate. In addition, Markov network have powerful learning ability to analyze the instance dataset.

Markov network can be represented by the graph  $G = (V, E)$ ,  $V$  is the set of vertices is the set of edges,  $E = \{(x_i, x_j) | x_i \neq x_j, x_i, x_j \in V\}$ , The nodes in the graph represent the random variables and the edges define the independence semantics of the distribution. The graph satisfies the Markov property, which states that a node is independent of all of its non-neighboring nodes given observed values for its neighbors. That is  $p(x_i | x_j) = p(x_i | x_j, (x_i, x_j) \in E)$ .

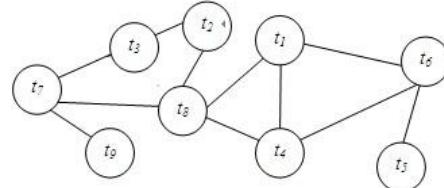


Figure 1. Markov network structure

In the case of query expansion, the target terms are potential expansion terms, the source terms are initial query, and the labels are probabilities of relevance. We then define a random process to propagate the label information through the graph. The stationary distribution of this process gives us a probability distribution over expansion terms. Figure 1 shows a portion of a term network for the query ‘Parkinson’s disease’. Solid connections denote explicit term associations, while the dashed line shows an implied connection inferred between ‘brain’ and ‘dopamine’ based on a short chain through the shared node ‘Parkinson’s disease’. We extend earlier work by Lafferty and Zhai [11] on using Markov chains for query expansion, by using a more flexible family of random walks similar to that described in Toutanova [12].

The Markov chain approach for modeling term associations is related to previous models based on term clustering and spreading activation networks, both of which have a long history that will only be briefly summarized here.

Stiles [13] described heuristics for using sets of term associations in improved indexing, and later Quillian [14] proposed a semantic network of concepts for binary relations between words. Gotlieb and Kumar [15] devised a semantic clustering of index terms using maximal complete subgraphs in a term network, although their method's effectiveness was never evaluated for retrieval. Early work in term clustering for query expansion by Sparck Jones [16] focused on constructing a similarity matrix of single index terms before any user queries were submitted. Wong and Raghavan [17] proposed the use of a matrix of term-term associations in ranking document vectors against query vectors, focusing on the special case of term correlation based on co-occurrence. Salton and Buckley [9], van Rijsbergen [8], and many others explored organizing of associations into networks for expanding the search vocabulary. These networks used various node activation heuristics and decay rules that were intuitively plausible but had limited retrieval success. Crestani [18] gives a summary of earlier work on spreading activation networks.

Stationary distributions have been used previously in information retrieval for ‘influence weighting’ schemes such as PageRank [19] and hub-authority [20] [21], and also for query expansion [11]. Lafferty and Zhai considered a bipartite graph on query terms and documents and calculated an approximate stationary distribution using a random walk. In their scheme, the random walk was defined in terms of words and documents, not words only. Our local co occurrence link is calculated in a similar way, but our random walk framework is more general in that we can use multiple sources of lexical and semantic evidence, not just co-occurrence, with the potential to weight these sources differently at different stages of the walk.

With regard to other query expansion approaches, the idea that we present below of rewarding expansion terms reflecting multiple aspects of the original query was previously noted by Xu and Croft [22] for Local Context Analysis (LCA), which has shown good empirical performance. Their method uses an empirically derived formula to score potential expansion terms that is similar in effect to our probabilistic term scoring. A number of studies have used external resources for query expansion. For example, Voorhees used Wordnet with limited success [23]. Shah and Croft [24] used Wordnet synonyms to perform query expansion for high-precision retrieval, selecting terms with high clarity scores.

In this paper, we pay attention to construct the Markov network of the items; therefore, we will select the appropriate method to measure the relationship between words, so we choose the mutual information after normalized for determining the similarities of all the term pairs ( $t_i, t_j$ ). Fig. 1 illustrates the structure, and the value on the edge stands for the similarity of the two connected nodes. In order to reduce the complexity of the Markov network, we assume that only the similarity value is larger than the threshold  $\beta$  (set according to from 0.04 to 0.32), adding an edge between the two items. The general motivation for using a Markov chain on a network of

terms is that we want to infer a particular property (the label) of a target word given a set (usually small) of labeled source words.

In contrast to many early spreading activation systems, the Markov chain approach to query expansion is relatively simple and offers a well-motivated probabilistic framework that fits well within the language modeling approach to information retrieval. Moreover, its close relationship with semi-supervised learning [25] means that we may make use of insights from that area to help illuminate the nature of query expansion and learn more robust expansion algorithms.

### III. CONCEPT TREE BASED MARKOV INFORMATION RETRIEVAL MODEL

The proposed algorithm is composed of four basic steps. First, build the term Markov network by calculating the similarity between terms; Second, construct concept tree for each original query, each tree is viewed as a concept, each child node is related to their root node with various relationships(such as synonyms, describing the same thing, representing the same concept...), the association is computed through Markov random walk matrix; Third, calculate the similarity between query items according to the association of their concept tree, in order to determine the core query, then remove redundant information node, and adjust the weight of query and the child node of the tree; Fourth, expand the original query with all remaining nodes in their concept trees, compute the similarity between documents and queries.

#### A. Building Concept Tree

Query expansion technique is a combination of computational linguistics, information science and other technique, expanding the initial user query with additional terms related with initial query to avoid synonyms and missing words problem, cutting off the irrelevant expansion terms to address the words mismatch problem in information retrieval, thus improving recall and precision. The core issue of query expansion technology is the expansion of word selection and weight [26, 27] distribution.

This article make semantic extension for each query term through the Markov concept tree, all the child node in the concept tree added to the original query as the expansion words, to compensate for the defects of the poor initial query information.

Markov network is a network of interconnected index which closely related between terms, the directly connected index may be synonyms, may be the words to express the same concept or describe the same thing. Markov network would activate the neighbor nodes of the index automatically, even if not directly connected to each other, there may be exist a strong semantic relationship. When retrieval the document, we find the original query item in the Markov cyberspace, then select its closely related (activated) index together as evidence source to calculate the document probability. So in this article we created a concept tree for each query item, each node in the tree has a certain degree of semantic

association with the root node (the initial query term), we designed the algorithms to create concept tree as follow:

(1) First let  $p = 1$ , search the item  $t_j$  directly connected to the initial query item  $q_m$  in Markov network space by breadth-first algorithm, then create the concept tree with the root node of  $q_m$ , let  $t_j$  be the child of the node  $q_m$ , and add the child node to the list  $L_p$  and list  $L_p$ .

(2) If the list  $L_p$  is not null, let pointer  $q$  points to the first child node  $t_j$ ;

(3) Remove the node  $t_j$  from the  $L_p$ , and find the item  $t_k$  connected with  $t_j$  in Markov network by the breadth-first algorithm; if  $\text{sim}(t_i, t_j) > \sigma$  and  $t_k \notin L_p$ , let  $t_k$  be the child node of  $t_j$ , and join in list  $L_{p+1}$  and  $L_p$  list;

(4) the pointer  $q$  points to the next node of  $L_p$ , if it does not point to end of the list  $L_p$ , go to Step 3; otherwise  $p = p + 1$ , go to Step 2.

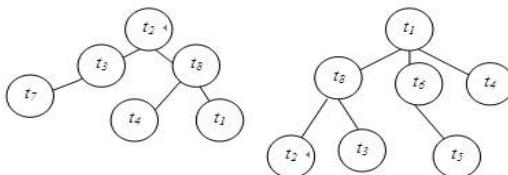


Figure 2. Concept tree

Fig. 2 explain a specific process to create a query concept tree based on the Markov network shown in Fig. 1 Suppose the index item  $t_1$  and  $t_2$  form a query, and then create two concept trees in Fig. 2. In general, the deeper level the child node in, the weaker correlation with the root node, using Shi [28] the transition probability matrix of the Markov network to update the correlation between each child node and the root of tree,  $\text{sim}(q_m, t_j)$  represent the association between the query  $q_m$  with the index term  $t_j$ . The parameters of the transition probability matrix of weight coefficients  $\lambda$  set to 0.5, and optimal random walk steps  $k$  set to 2.

$$\text{sim}(t_i, t_j) = \sum_{t=1}^k \lambda^t P_k(t_i, t_j) \quad (1)$$

$$P_1(t_i, t_j) = \sum_{t \in T} P_0(t_i, d_k) P_0(t_k, t_j) \quad (2)$$

$$P_K = P_{K-1} \times P_0 \quad (3)$$

### B. Pruning Algorithm for Concept Tree

In the multi-keyword query, each query term plays different role in search task, the weight should also be different. Determining the core query terms by the association degree of the query concept trees, pruning the Redundant node and adjusting the semantic weight of each node in concept tree, making query expansion terms closer to the user's query intent, in order to improve the precision. Our method for pruning is composed of three steps: (1) increasing the weight of core query, (2) pruning the redundant node and (3) decreasing the weight of non-

core node. In the following, we discuss each of these 3 steps..

#### 1) Increasing the Weight of Core Query

In the query, if there is some query words semantically related, indicating that the entire query theme will be more inclined to these related query words, in the search task should enhance these words weight. In our approach, using the semantic relevance of the concept tree to calculate the association between the query terms,  $\text{sim}(q_m, q_n)$  represent the relationship between the initial query term  $q_m$  and  $q_n$ .

$$\text{sim}(q_m, q_n) = \text{tree\_sim}(q_m, q_n) \quad (4)$$

$\text{tree\_sim}(q_m, q_n)$  denotes the association between the concept tree of query  $q_m$  with the concept tree of query  $q_n$ , values are in the following range,  $0 \leq \text{tree\_sim}(q_m, q_n) \leq 1$ , the greater value  $\text{tree\_sim}(q_m, q_n)$  is, the more similar between the two concept tree.

$$\text{tree\_sim}(q_m, q_n) = \text{sum\_co}(q_m, q_n) \times \text{sum\_co}(q_n, q_m) \quad (5)$$

$\text{sum\_co}(q_m, q_n)$  represent the degree of overlap information in  $q_m$  concept tree, between two query concept tree of  $q_m$  and  $q_n$ , where  $\text{sum\_co}(q_m, q_n)$  represent the similarity of  $q_m$  and  $q_n$  which can be performed offline according to the formula 1 in precious section.

$$\text{sum\_co}(q_m, q_n) = \frac{\sum_{t_j \in \text{tree}(q_m) \wedge t_j \in \text{tree}(q_n)} \text{sim}(q_m, t_j)}{\sum_{t_i \in \text{tree}(q_m)} \text{sim}(q_m, t_i)} \quad (6)$$

If the correlation between the query terms is great enough, we believe that these query terms are more consistent with the entire query topic, which would play a more important role in search task, it will be defined as the core query, and their weights will be increased. This article will calculate the correlation of each two query pairs, setting the threshold  $\varepsilon$ , if  $\text{sim}(q_m, q_n) > \varepsilon$ , set the initial query  $q_m$  and  $q_n$  as core query, then update the query weight as follows:

$$\text{weight\_adjust}(q_m) = (1 + \text{sim}(q_m, q_n)) \times \text{weight}(q_m) \quad (7)$$

where  $\text{weight\_adjust}(q_m)$  represent the weight of query  $q_m$  after adjustment, and  $\text{weight}(q_m)$  represent the initial weight of query  $q_m$ .

#### 2) Pruning the Redundant Node

If the value of  $\text{sim}(q_m, q_n)$  between two original queries is 0, it indicated that they did not have any semantic relationship. The larger value it is, the more the semantic overlap there will be. We use the whole concept tree instead of single query term as the evidence source to search in document retrieval. So we will adjust the tree to reduce the complexity of the tree, in order to minimize the noise take into query process, we remove the redundant node between the concept trees. The specific steps are as follow, compare semantic overlap between the query terms. For example, if  $\text{sum\_co}(q_m, q_n) > \text{sum\_co}(q_n, q_m)$ , we believe that the semantics of  $q_m$  is more simple, the more able to express the concept of the

entire query, so keep the overlapping node in concept tree of qm, but pruning the concept tree of qn, that is to remove the node which appear in two concept tree at the same time, to update the weight of nodes remain in the qm concept tree moreover.

The weight of the child node should reflect the characteristic of root node, so we take the semantic similarity factor  $\text{sim}(\text{qm}, \text{qn})$  into account.

$$\begin{aligned} \text{weight\_adjust}(t_j) &= (1 + \text{sim}(q_m, q_n)) \times \text{weight}(t_j) \times \text{sim}(q_m, t_j) \\ t_j \in \{ \text{tree}(q_m) \wedge \text{tree}(q_n) \} \wedge \text{sim}(q_m, q_n) &> \varepsilon \end{aligned} \quad (8)$$

### 3) Weaken the Non-Core Terms

If the initial query term is polysemous, its expansion terms added into the concept tree are strongly associated with the initial query, but may be not relevant to the entire query topic, that would bring into lots of noise, and would not make precision improved, even may cause the topic drift; the non-overlapping information nodes in the concept tree of query qm and qn produced in the previous step, which is possible be the non-related concepts, we reduce their weights to avoid semantic drift so as to enhance the precision as follows formula:

$$\begin{aligned} \text{weight\_adjust}(t_i) &= (1 - \text{sim}(q_m, q_n)) \times \text{weight}(t_i) \times \text{sim}(q_m, t_i) \\ t_i \notin \{ \text{tree}(q_m) \wedge \text{tree}(q_n) \} \wedge \text{sim}(q_m, q_n) &> \varepsilon \end{aligned} \quad (9)$$

After above three steps, if  $\text{sim}(\text{qm}, \text{qn})$  value is 0, we believe that there is no semantic relationship between query qm and query qn, treating the node in the query concept tree without any adjustment. Fig. 3 is an example of pruning based on the concept tree shown in Fig. 2.

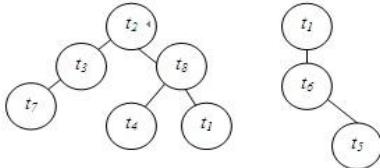


Figure 3. Concept tree after pruning

### C. Probabilistic Query Expansion Model

Giving query q, and document set D, we will calculate the probability of  $p(d_j|q)$  according to the formula 9, where  $d_j$  is contained in document set D, the range of j is,  $1 < j < n$ , and the MT represent the item space of Markov network:

$$p(t_i | d_j) \propto \sum p(t_i | d_j) p(t_i | q) \quad (10)$$

If both the item network and the document network are fixed,  $p(T|MT)$  and  $p(MT)$  is equal to any  $d_j$ ,

$$p(t_i | d_j) \propto \sum p(t_i | d_j) p(t_i | q) \quad (11)$$

The formula 10 is the general form of the search model algorithm, we can derive formula 11:

$$p(t_i | d_j) \propto \sum p(t_i | d_j) p(t_i | q) \quad (12)$$

We compute the weight [29] as follows:

$$p(t_i | d_j) = \frac{idf_i \times dl_j^2}{\alpha \sum_{t_i \in d_j} idf_i \times dl_j^2} \quad (13)$$

$$p(t_i | q) = \frac{\log(tf_i)^2}{\beta \sum_{t_i \in q} \log(tf_i)^2} \quad (14)$$

### IV. EXPERIMENTS RESULTS

In order to better understand the strengths and weaknesses of our technique, we evaluate it on a wide range of datasets, five standard test collections shown in Table 1 were used for our experiments. After extracting all the words from the collections and removing stop words, we used Porter Stemmer algorithm to index both queries and documents included by title and body. Table 1 indicates the number of documents, the number of queries, and the number of terms. It can be seen that the Adi collection is rather small and the Cacm collection is sizable as a test collection. The Med, Cran, Cisi collection is of medium size.

TABLE I. EXPERIMENTAL DATASETS

collection	description	Num of doc	Num of query	Num of terms
Adi	Information science	82	35	893
Med	Medical science	1033	30	8702
Cran	Aviation science	1400	225	4110
Cisi	Library science	1460	76	5494
Cacm	Computer science	3024	64	5041

Term weights in both documents and queries are determined according to the  $\log(\text{tf}).\text{idf}.\text{dl}$  weighting scheme, see also formula (13). The mutual information was used for determine the similarities between all the term pairs in the collections, i.e. build up the Markov network. In addition, the construction method described in section 3 was used to generate the concept tree. Note that this can be achieved very efficiently as the pre-computed the Markov random walk matrix. The entire concept tree are chosen to expand or modify the query according to formulae (3), (4), and (5). Then, for each query, we rank the terms of the collection in decreasing order according to formula (14). The weight of the child node in concept tree cannot be equally treated as root node (initial query term), so introduce the weight adjustment factor  $\eta$  to dynamically adjust the contribution of them:

$$p(d_j | q) \propto \sum_{t_i \in q} (\eta p(t_i | d_j) p(t_i | q) + (1-\eta) \sum_{t_m \in \text{tree}(t_i)} p(t_m | d_j) p(t_m | q))$$

The results were evaluated by applying the average precision of a set of queries at three representative recall points 3-avg (0.2, 0.5, 0.8), and eleven representative recall points 11-avg (0, 0.1, 0.2, ..., 0.9, 1.0). In addition, we now investigate how well our model performs in practice. We provide results for three distinct ranking algorithms: the baseline BM25 [30] model, the basic Markov network model(BMNM) [6], and the Markov concept tree model

(MCTM), to better understand how each model performs across the various data sets.

Table 2 shows the 11-point average precision for the BM25, BMNM, MCTM algorithms, and table3 shows the 3-point average precision for three algorithms, the improvement also provided as the BM25 for baseline. We observe that BMNM, MCTM algorithms yield better precision than the baseline, MCTM ranking yields the largest gains in average precision in all data sets (Especially increases by around 39% and 27% for the Cisi dataset and the Cacm dataset respectively); and we compare the results between BMNM and MCTM, there is a stable improvement as the BMNM for baseline, which is 5.4%, 4.17%, 5.9% and 6.4% for Adi, Med, Cisi, and Cacm respectively. The figures indicate that our concept tree query expansion method makes a considerable improvement in all collections. That is because its disambiguation scheme can be used to successfully describe user information needs, capturing the concepts related to the user search experience. We also found that the number of expansion terms is really much smaller than the BMNM baseline model.

TABLE II. 3- AVG VALUE OF RESULTS ON DIFFERENT DATA SETS

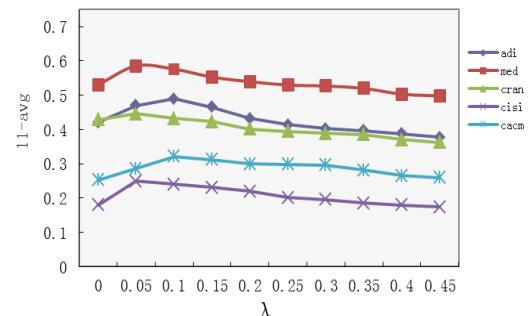
MODEL	BM25	BMNM	MCTM
Adi	43.09	44.81(3.9%)	49.25(14.3%)
Med	54.32	58.67(11.9%)	61.12(12.5%)
Cran	42.15	42.21(0.14%)	42.32(0.41%)
Cisi	18.17	21.04(15.8%)	22.31 (22.8%)
Cacm	22.94	28.31(23.4%)	30.12(31.2%)

It seems that the improvement increases with the size of the collection. In addition, the improvement increases with the number of additional search terms that expand the original query as long as the collection is large enough. Obviously, the large collection contains more domain knowledge than the small ones. As a consequence, the quality of the concept tree created from the large collection is better than the quality of the concept tree belonging to the small collections. Compared to Cisi there is smaller improvements in Med, it's possible that the feature of dataset Med itself caused, there may be less polysemous, synonyms in the medical science, these words express the semantics are clear. The dataset Cran perform special in the experiments that may be caused by its document specialty. This was true across all collections. The reasons for this require further study, but it suggests that how the evidence is weighted by time-step does indeed matter. In this case, terms that co-occur with terms semantically close to the query appear to be more valuable than terms semantically close to many potential co-occurrence terms.

TABLE III. 11-AVG VALUE OF RESULTS ON DIFFERENT DATA SETS

MODEL	BM25	BMNM	MCTM
Adi	42.09	44.02(7.7%)	48.92(16.2%)
Med	53.08	57.02(7.4%)	58.60(10.4%)
Cran	42.88	42.54(-0.7%)	43.26(0.89%)
Cisi	17.96	23.9(33.1%)	24.97(39%)
Cacm	25.29	29.5(23.6%)	32.12(27%)

In our approach, the parameter  $\eta$  reflect the importance of the extended terms. The higher the value of  $\eta$  is, the greater importance the extended term is, and vice versa, the smaller the value of  $\eta$ , the greater importance of the original query terms. In this experiment, the  $\eta$  value under the optimal results is almost small,  $\eta$  values are less than 0.1 in all the document set.  $\eta$  values in this article seem to follow certain rules, changing with the size of the set of documents. In Fig. 4, we study how the  $\eta$  value affects the usefulness of the concept tree expansion method.

Figure 4. Comparison of different  $\eta$  on each dataset

The results in the previous section indicate that optimization is very important, but unfortunately real applications usually lack training data. In this Section we wanted to study whether the parameters can be carried over from one dataset to the other, and if not, whether the extra terms found by QE would make the system more robust to those sub-optimal parameters.

Fig. 4 includes a range of parameter settings, including defaults, and optimized parameters coming from the same and different datasets. The results show that when the parameters are optimized in other datasets, QE provides improvement with statistical significance in all cases.

Overall, the figure shows that our QE method either improves the results significantly or does not affect performance, and that it provides robustness across different parameter settings, even with suboptimal values.

The depth of concept tree  $d$  determines the indirectly relationship between terms, in order to find the most descriptive words to word relevance of different depth, we compared 11-avg value for depth change from 1 to 5 as shown in Fig. 5 To evaluate the efficiencies of different depth of concept tree.

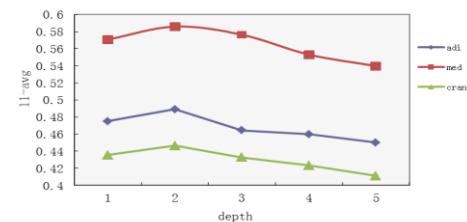


Figure 5. Comparison of different depth of concept tree on Adi, Med, Cran

We can conclude that when depth is 2, all the results reach best. This seems to be the reason information after

indirect related three time will be deviate from the original intent, only the value of depth in Adi make slight improvement of precision, the results of other four dataset decline, obviously the relationship depth between the terms can not exceed more than three, otherwise, it will add noise into meaning of original query. Actually, that is consistent with the optimal random walk steps in [28].

While the number of potentially active term-nodes in the network is large (our vocabulary size was around 6,000), the Markov transition matrices are very sparse: a typical matrix has about 900 non-zero entries. Limiting a walk to a maximum of 3steps typically results in less than 900 (15%) of potential nodes becoming active for title-length queries. Even so, using the network efficiently requires some planning. For example, the first time a word is seen, its score is cached since the aspect probabilities will not change over the lifetime of the query. The time to build the network can be reduced by doing off-line indexing for each article.

## V. CONCLUSION

This paper presents a novel query Expansion method based on a Markov-based system to find related concepts and words. The documents in three datasets were thus expanded with related words, which were fed into a separate index. When combined with the regular index we report improvements over standard test collections using BM25 for those three datasets across several parameter settings, including default values, optimized parameters and parameters optimized in other datasets. In most of the cases the improvements are statistically significant, indicating that the information in the query expansion is useful. Similar to other expansion methods, parameter optimization has a stronger effect than our expansion strategy. The problem with parameter optimization is that in most real cases there is no tuning dataset.

The concept tree retrieval model based on Markov network, by introducing concept tree, and we make the entire query concept into account when select the expansion terms. This model is primarily concerned with two important problems of query expansion, namely with the selection and the weighting of expansion search terms.

The term selection relies on the dependence between the initial query terms and the similarity between terms of the collection. The weigh distribution adjusted according to a concept tree pruning algorithm, which remove the redundant information nodes, increase the weight of core query and core child node, and weaken the weight of non-core child node. The experiments carried out on the five test collections show that consistent improvement in the retrieval effectiveness can be expected.

In our future research we will concentrate on parameter tuning, the method involved in the proposed model to test to the optimal through the experiment, we will try to tune the parameters automatically, such as the EM algorithm. Furthermore, the parameters almost affected by the size of document size, so we want to test the approach in a larger test set, such as TREC.

## ACKNOWLEDGMENT

Project supported by special fund of the Central University number: CHD2012JC087.

## REFERENCES

- [1] Manning, C. D., P. Raghavan, and H. Schütze. 2009. An introduction to information retrieval. *Cambridge University Press, UK*.
- [2] Zhaoman Zhong, Cunhua Li, Yan Guan, Zongtian Liu, "A Method of Query Expansion Based on Event Ontology", *JCIT*, Vol. 7, No. 9, pp. 364 - 371, 2012.
- [3] S. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In Online Proceedings of the Third Text Retrieval Conf, pp. 109–126, 1995.
- [4] MandarMitra, AmitSinghal, ChrisBuckley. Improving automatic query expansion, *SIGIR*, pp. 206 - 214, 1998.
- [5] Zuo jiali, Wang mingwen, Wang xi. Extended information retrieval model based on Markov network. *Journal of Tsinghua University (Science and Technology)*, 45(s1): pp. 1847-1852, 2005.
- [6] Caoying, Wang mingwen, Tao hongliang. Information retrieval model based on Markov Network. *Journal of Shandong University (Natural Science)*, vol. 41, No. 3, pp. 126-130, 2006.
- [7] H. E. Stiles, The association factor in information retrieval, *J. ACM*, 0004-5411 8:2, pp. 271-279, 1961
- [8] K. van Rijsbergen. *Information Retrieval* (2nd edition). Butterworths, London, 1979.
- [9] G. Salton and C. Buckley. On the use of spreading activation networks in automatic information retrieval. *SIGIR*, pp. 147-160, 1998.
- [10] L M de Campos. Independence relationships and learning algorithm for singly connected networks. *J Exp Theory ArtifIntell*, 10(4): pp. 511-549, 1998.
- [11] J. Lafferty and C. Zhai. Document languagemeans, query models, and risk minimization for information retrieval. *SIGIR*, pp. 111-119, 2001.
- [12] K. Toutanova, C. Manning, and A. Y. Ng. Learning random walk models for inducing word dependency distributions. *Proc. of ICML* 2004.
- [13] K. L. Kwok, L. Grunfeld, M. Chan. TREC-8 Ad-Hoc, Query and Filtering Track Experiments using PIRCS, *NIST special publication*, pp. 217-224, 2000.
- [14] M. Quillian. Word concepts: a theory and simulation of some basic semantic capabilities. *Behav. Sci.* 12: pp. 410-430, 1967.
- [15] C. C. Gottlieb, S. Kumar. Semantic clustering of index terms. *JACM* 15(4): pp. 493-513, 1968.
- [16] K. Sparck Jones and E. O. Barber. What makes an automatic keyword classification effective? *JASIS* 22(3) 166-175, 1971.
- [17] S. K. M. Wong and V. Raghavan. Vector space model of information retrieval-a re-evaluation. *SIGIR*, pp. 167-185, 1984.
- [18] F. Crestani, Application of spreading activation techniques in information retrieval. *AI Review* 11 (6), pp. 453-482, 1997.
- [19] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Proc. of 7th Int. WWW Conference*, pp. 107-117, 1998
- [20] J. Kleinberg. Authoritative sources in a hyperlinked environment. *JACM*, 46:pp. 604-632, 1999
- [21] J. Kleinberg and A. Tomkins. Applications of linear algebra in information retrieval and hypertext analysis.

- 18th ACM Symp. on Principles of Database Systems*, pp. 185-193, 1999.
- [22] J. Xu and W. B. Croft. Query expansion using local and global analysis. *SIGIR* 1996, pp. 4-11
- [23] E. M. Voorhees. Query expansion using lexical-semantic relations. *SIGIR* 1994, pp. 61-69.
- [24] C. Shah and W. B. Croft. Evaluating high accuracy retrieval techniques. *SIGIR* 2004, pp. 2-9, 2004.
- [25] A. J. Smola and R. Kondor. Kernels and regularization on graphs. Proc. of COLT 2003, Eds. B. Schölkopf and M. Warmuth, Lecture Notes in Computer Science, Springer.
- [26] Sparck-Jones, K., Notes and references on early classification work. *SIGIR Forum*, 25(1), pp. 10-17, 1991.
- [27] Yonggang Qiu, Hans-peter Frei. Concept based query expansion. *SIGIR*, pp. 160–169, 1993.
- [28] Shisong, Wang mingwen, Tuwei, He shizhu. Extended information retrieval model based on the Markov network cliques. *Journal of Shandong University (Natural Science)*, vol. 46, No. 5, pp. 54-58, 2011.
- [29] Makoto Iwayama, Atsushi Fujii, NorikoKando, Yuzo Marukawa. An empirical study on Retrieval Models for Different Document Genres: Patents and Newspaper Article. *SIGIR*, pp. 251-258, 2003.
- [30] S. E. Robertson and D. A. Hull. The TREC-9 filtering track final report. In *TREC*, pp. 25–40, 2000.

# An Abnormal Speech Detection Algorithm Based on GMM-UBM

Jun He

Guangdong Province Key Lab of Petrochemical Equipment Fault Diagnosis, Guangdong University Petrochemical Technology, Maoming City, Guangdong Province, China  
Email: hejun\_723@126.com

Ji-chen Yang

South China University of Technology, School of Electronic and Information Engineering, Guangzhou, China  
NisonYoung@163.com

Jianbin Xiong, Guoxi Sun, and Ming Xiao

Guangdong Province Key Lab of Petrochemical Equipment Fault Diagnosis, Guangdong University Petrochemical Technology, Maoming City, Guangdong Province, China  
Email: xiongjianbin@21cn.com; sguoxi@126.com; xiaoming1968@163.com

**Abstract**—To overcome the defects of common used algorithms based on model for abnormal speech recognition, which existed insufficient training data and difficult to fit each type of abnormal characters, an abnormal speech detection method based on GMM-UBM was proposed in this paper. For compensating the defects of methods based on model which difficult to deal with the diversification speech. Firstly, many normal utterances and unknowing type abnormal utterances came from different speaker, were used to train the GMM-UBM for normal speech and abnormal speech, respectively; secondly, the GMM-UBM obtained by training normal speech and abnormal speech were used to s core for these testing utterances. From the results show that compared with GMM and GMM-SVM methods under 24 Gaussians and the ratio of training speech and testing is 6:4, the correct classification ratio of this proposed have 6.1% and 4.4% improvement, respectively.

**Index Terms**—Abnormal Speech; Speech Detection; GMM-UBM; Continuous Speech

## I. INTRODUCTION

The issue of speaker recognition has been attracting more and more attention over 40 years in areas of information security and speech signal processing [1-4]. Speaker recognition is a procedure the employs the comparison of an input voice sample with predetermined voices in order to select the one that most closely corresponds to the input voice sample, thus enabling the establishing of an unknown speaker from a group of several speakers. Just like other pattern recognition tasks, speaker verification typically involves three common steps, feature extraction, speaker modeling, and a classification decision. Up today, many speaker recognition systems are developed and can achieve very high accuracy, when the background noise and the speakers' emotion can be controlled or the speaker's

vocal organs worked well [5-7]. But unfortunately, most of speaker recognition systems can not be widely used in the application market, because of their performance declined sharply when they faced the complex background noise and abnormal utterance [8, 9]. Recently, In order to improve the performance of most signal processing algorithms for these kinds abnormal speech, many scholars have do their best to abnormal speech signal processing related research, and make it to be a hot and promising research area [10, 11].

Wright air laboratories of U.S air force and Armstrong medical laboratory were carry out the robust speech recognition research for the fighter utterances [12, 13], from research results show that when under the condition of G-Force, the pilot's head had a great influence on the performance of speech recognition algorithms. Professor Picard put forward the emotional calculation firstly and then many scientific researchers devoted into the emotional research, and got some progress on the speech emotion recognition [14, 15]. In recent years, many illegal molecules, using false voice to crime, and so, this kind of false voice were known as disguised voice [16, 17]. With the deepening of the research on disguise voice [18], disguise voices were divided into the following two types: 1) disguise voice identification with not rely on a given electronic equipment; 2) disguise voice recognition rely on the electronic equipment. Philip Rose etc, committed to disguise voice forensic research [19, 20], also provoked a forensic phonetics research upsurge in the speech signal processing world [21, 22].

In Medical engineering, this kind of abnormal voices, caused by the speaker pronunciation organ lesions, named as pathological voice, which also belongs to abnormal speech. Currently, most researches about abnormal speech were focused on biomedical engineering, aiming to detect the pathological speech from normal speech, in the other words, in order to find lesions in the

vocal organs early through analysis the speech quality [23-25]. Recently, pathological voice detection and tracking have achieved fast development; some pathological voice detection methods were published. These methods for pathological voices detection can be divided into the following two types mainly; 1) pathological voice detection methods based on perturbation [26], such as Absolute Jitter (AJ), Relative Average Perturbation (RAP), Amplitude Perturbation (AP), Shimmer Percent, Absolute Shimmer (AS), etc; 2) detection methods based on noise, such as Harmonics to Noise Ratio (HNR) [27], Normalized Noise Energy (NNE), Glottal to Noise Excitation Ratio (GNER) [28], Sub-Harmonic to Noise Ratio (SHNR), etc.

However, when pronounced the voiceless consonants continuous speech under organs pathological, it made the glottis can not open or close well, and then which will produce the hoarse voice that affecting the accurate for extract disturbance parameters and noise parameters.

In order to overcome the defects of the above two kinds of algorithm to capture the fundamental frequency and hoarse noise under the condition of lesions hardly, many algorithms based model were reported. Such as Gaussian Mixture Model (GMM) and GMM-Support Vector Machine (SVM) [23]. However, this kind of algorithm which existed insufficient data for training model and difficult to fit many types of lesions defect.

To make up for the shortcomings of model-based algorithm, an abnormal speech detection algorithm based on GMM-UBM was proposed in this paper. Aim to overcome the insufficient training data and to descript the commonness of different kind of abnormal voice, kinds of abnormal voices were used to train a GMM-UBM. This can avoid the defect that a single type abnormal voice made. In this paper, text-independent continuous speech of pathological was the mainly research object, such as cold speech. Compared with lesions monophthong, these text-independent continuous speech have more complicated hoarse noise in pronunciation phase, so, it is easy to be user acceptance and has more valubales for do some researching.

The remainder of this paper is organized as follows. In Section II, a brief overview of GMM and GMM-UBM training method is presented. The experiment data and experiment results analysis were introduced in the Section III. Conclusion and future work are presented in Section IV.

## II. GMM-UBM

GMM-UBM is the abbreviation of Gaussian mixture model-universal background model, it proposed by Reynolds originally and used in speaker recognition field [29, 30], aiming to overcome the defects of GMM that mismatch of channel between speaker model and testing speech and background noise [31], especial under the background noise, the matching degree of test utterance and model is very poorer. It was firstly used in speaker recognition system to train and obtain the speaker-independent model. However, in this paper, we attempts use it in abnormal voice detection system. GMM-UBM

was regarded as the abnormal continuous text-independent speech model, which can obtain by training kinds of abnormal utterances. The training and recognition system is shown in Figure 1.

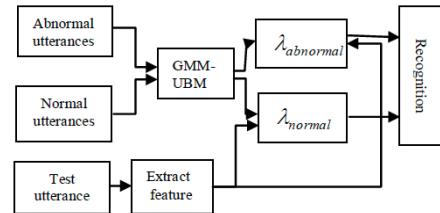


Figure 1. Based on GMM-UBM abnormal utterance detection system figure

### A. GMM Introduction

Any an  $M$  order Gaussian mixture model, its possibility density function can be obtained through summing the  $M$  Gaussian probability density function with weighted, which can be expressed as the following equation:

$$P(X / \lambda) = \sum_{i=1}^M \omega_i b_i(X) \quad (1)$$

where  $X$  is an  $D$  dimension random vector, and  $b_i(X_i)$ ,  $i=1,\dots,M$  is a sub-distribution of random vector  $X$ , also  $\omega_i, i=1,\dots,M$  is the mixed weight of  $b_i(X_i)$ . Each sub-distribution is the joint Gaussian possibility density distribution of  $D$  dimension random vector, which can be expressed as:

$$b_i(X) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i)\right\} \quad (2)$$

where  $\mu_i$  is mean vector,  $\Sigma_i^{-1}$  is Covariance matrix, and the mixed weight value met the following equation :

$$\sum_{i=1}^M \omega_i = 1 \quad (3)$$

And then a complete Gaussian mixture model can be composed of mean vector, covariance matrix and mixed weights, and which can be showed as the following equation:

$$\lambda = \{\omega_i, \mu_i, \Sigma_i\} \quad (4)$$

However, some literatures reports that a GMM can be abbreviated as the following equation:

$$\lambda = \{M, \omega_i, \mu_i, \Sigma_i\} \quad (5)$$

Therefore, for a given time series  $X = \{X_t\}, t = 1, 2, \dots, T$ , whose degree of logarithmic likelihood in the GMM model matching can be defined as:

$$L(X / \lambda) = \frac{1}{T} \sum_{t=1}^T \log P(X_t / \lambda) \quad (6)$$

To obtain a good GMM, a sufficient set of training data used to train the GMM. The entire model training process was a supervision and the optimization of the process. And some criterions used to determine the model parameters. The common used criteria for determine the model parameters was the Maximum Likelihood (Maximum Likelihood, ML) criterion. For a length of T training vector sequence, such as  $X = \{X_1, X_2, \dots, X_T\}$ , the likelihood of GMM can be expressed as the following equation:

$$P(X | \lambda) = \prod_{t=1}^T P(X_t | \lambda) \quad (7)$$

From formula (7) shows that the equation was a nonlinear function of the parameter  $\lambda$ , and it's more difficult to directly find out its maximum. Therefore, the Exception Maximum (Exception Maximum, EM) calculating is often used to estimate model parameter  $\lambda$  of the GMM.

Maximum Likelihood criterion requires the model description must approximate the distribution of training data set extremely, Optimal parameters of M order GMM model with ML criterion satisfies the following formula:

$$\lambda_M^* = \arg \max_{\lambda_M} \left\{ \prod_{t=1}^T p(x_t | \lambda_M) \right\} \quad (8)$$

Therefore, the larger training data set is, the more they can truly describe the speaker's feature distribution, and the closer for the GMM to fit the real distribution of the speaker's features, resulting that the higher speaker recognition ration for the speaker recognition system. The following figure is the fitting curve of Gaussian function under the same Gaussian.

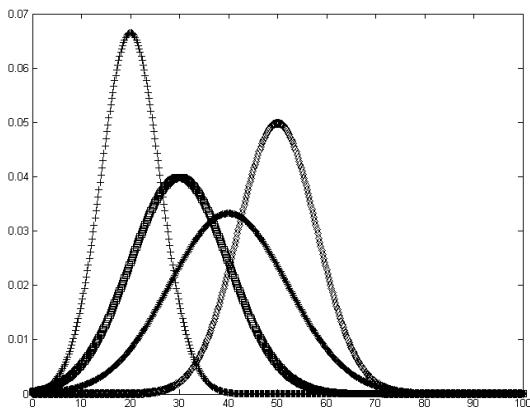


Figure 2. Fitting curve of Gaussian under the same Gaussian

For an observation vector  $x_t$ , the Likelihood of its model  $\lambda$  is denoted as  $p(\lambda | x_t)$ , to estimate the model parameter  $\lambda$ , is to make posterior probability of the Likelihood  $p(x_t | \lambda)$  have a maximum. According to the Bayesian theory which can be expressed as the following equation:

$$p(\lambda | x_t) = \frac{p(x_t | \lambda)p(\lambda)}{p(x_t)} \quad (9)$$

where  $p(\lambda)$  is the possibility of the model  $\lambda$ ,  $p(x_t)$  represents the probability of  $x_t$  that maybe occur. And in general, the probability  $p(\lambda)$  is a constant; the probability  $p(x_t)$  isn't related with the parameter estimation of the model  $\lambda$ . Therefore, maximized probability of  $p(\lambda | x_t)$  is equal to that of  $p(x_t | \lambda)$ . However, the observation vector  $x_t$  is uncertain data; in other words, it's difficult to known that the time series  $x_t$  came from which Gaussian components of a given GMM. The step E of EM algorithm used to obtain the expectation when the number of Gaussian M is at the maximizing. Through continuous E, M iteration, the estimated value of model parameter can be obtained well.

In the model training period, assuming that the mixture degree of GMM was M  $\lambda_0$  is the initial value of the model. And EM algorithm used to iterate for the initial model until convergence the model  $\lambda_0$  is convergence. The value of model  $\lambda_0$  did not affect the GMM training convergence speed and effectiveness, all the iteration of the model training period, according to the following equations update the parameters:

The kth Gaussian weights [32]  $\omega_k$ :

$$\omega_k = \frac{1}{T} \sum_{t=1}^T p(k | x_t, \lambda) \quad (10)$$

The kth Gaussian Variance  $\sigma_k^2$ :

$$\sigma_k^2 = \frac{\sum_{t=1}^T p(k | x_t, \lambda) \cdot x_t^2}{\sum_{t=1}^T p(k | x_t, \lambda)} - \mu_k^2 \quad (12)$$

The kth Gaussian mean  $\mu_k$ :

$$\mu_k = \frac{\sum_{t=1}^T p(k | x_t, \lambda) \cdot x_t}{\sum_{t=1}^T p(k | x_t, \lambda)} \quad (11)$$

where  $p(k | x_t, \lambda)$  is the posterior probability kth Gaussian .

### B. UBM Building

In this proposed, the Bayesian adaptive algorithm was used to train the GMM-UBM. Firstly, GMM-UBM for abnormal speech and normal speech obtained respectively, through training a lot of independent-text abnormal voices and normal voices from different speakers. Bayesian adaptive method is mainly used to solve the performance of speech recognition system dropped sharply, which were caused by the model mismatch in speech recognition processing and the lack

of training speech data. And even in the case of mass training speech data, it can quickly and effectively obtain the speech GMM-UBM.

Assuming, a given UBM and the training speech vector of the model  $X = \{x_1, \dots, x_t\}$ . Firstly, we can confirm the probability distributions of the training speech vector in the UBM. The Calculation method for the probability distribution of train speech vector is given as following equation:

$$p_u(i|x_t) = \frac{w_i p_i(x_t)}{\sum_{j=1}^M w_j p_j(x_t)} \quad (13)$$

And then,  $p_u(i|x_t)$  and  $x_t$  were used to obtain the weights, mean and variance matrix of the UBM, respectively. Finally, update the  $i$ th statistics component of UBM with the calculated results; the expressions for update were given as the following equations:

$$\hat{w}_t = [\frac{a_i^w}{T} + (1 - a_i^w)w_t] \gamma \quad (14)$$

$$\hat{\mu}_t = a_i^m E_t(x) + (1 - a_i^m)\mu_t \quad (15)$$

$$\hat{\delta}_i^2 = a_i^v E_i(x^2) + (1 - a_i^v)(\delta_i^2 + \mu_i^2) - \hat{\mu}_t^2 \quad (16)$$

In order to keep the balance of  $i$ th component for UBM in the update processing, the  $a_i^w$ ,  $a_i^m$  and  $a_i^v$  were used to adjust the weights, mean and variance matrix for new UBM effectively. And then the scale factor  $\gamma$  was used to ensure the sum of all weights was 1.

### III. EXPERIMENTS AND RESULTS ANALYSIS

In this paper, the utterances which came from a cold speaker regarded as the mainly research object, Due to the cold severity of the speaker were different, the voice quality from them were different also. In this proposed, we only regarded all the cold voice as abnormal voice, all speaker voice not in the cold case regarded as normal speech. All about the details of the abnormal speech database design and construction please to refer to the literature [33].

In the following, the experiment setup and results analysis were discussed and analysis.

#### A. Experimental Setup

450 normal utterances from 9 speakers (50 for each speaker) and 910 abnormal utterances from 13 speakers (70 for each speaker) selected from the PANSD [33] as the experimental speech data for this paper. In the experiment, all the experiment data were divided into two types; 1) training speech data, selected from each speaker, average; 2) testing speech data, come from each speaker by average. The whole experiment in this paper, carried out according to the following four schemes, the number of abnormal utterances and normal utterances about training and testing for each experiment scheme was given in the following table I.

All the experimental data was mono channel WAV format, Cooledit Pro2.0 used to adjust the sampling rate to 16 kHz, the quantitative accuracy was 16 bits. The

duration of all experiment utterances data is between 3s and 5s, and with removed the entire silence component by artificial processing. Non-silence segment is firstly divided into frames of 32ms with 50% overlap, and 26 mel-frequency cepstral coefficients (MFCCs) are extracted for each frame. 8, 16 and 24 Gaussian mixture degrees were used to train GMM and GMM-UBM for abnormal utterances and normal utterances, respectively. The kernel function of SVM in GMM-SVM algorithm, which mentioned in this paper, is radial basis kernel function.

In order to analysis and estimate the performance of all the abnormal detection algorithms mentioned in this paper well, regarded the voice from cold speaker as abnormal speech or pathological speech, the voice from well speaker as normal speech. Classification Correct Ratio (CCR) used to estimate the performance all kinds of algorithms mentioned in this paper,  $P_r$  and  $N_r$  denote the recognition ratio for pathological speech and normal speech, respectively, which were defined as the following equation:

$$\begin{aligned} CCR &= \frac{T_n + T_p}{N_n + N_p} \times 100\% \\ P_r &= \frac{T_p}{N_p} \times 100\% \\ Nr &= \frac{T_n}{N_n} \times 100\% \end{aligned} \quad (17)$$

where  $T_n$  denotes the number of utterances that normal utterance was judged as normal utterance,  $T_p$  denotes the number of utterance that abnormal utterance was judged as abnormal utterance. And then  $N_n$ ,  $N_p$  denotes the total testing speech sample for abnormal speech and normal speech, respectively.

#### B. Results Analysis

In the Experimental setup section of this paper, all the experiment data were divided into two parts. Firstly, all the training speech were used to obtain the speech model of abnormal utterance and normal utterance for those algorithms that mentioned in this paper; and then the testing speech used to detect the ability of distinguish between normal voice and abnormal voice for the mentioned algorithms. The CCR,  $P_r$  and  $N_r$  of all the mentioned algorithms were given in the table II as the following:

From the Table II shows that compared with the mentioned method of GMM and GMM-SVM under the same number of Gaussian, the CCR of this proposed algorithm, in any case of the experimental scheme, was improved. Especially, the improvement of CCR of GMM-UBM was 6.1% and 4.4% under the scheme4 with 24 Gaussians. Knowing from the Table II, at the same time, the recognition ratio of normal utterance was higher than that of abnormal utterance. Nevertheless the gap of this proposed between recognition ratio for abnormal utterance and that of normal utterance was min. Because of these methods which based on GMM can not describe the characteristics changes of all the kinds of abnormal utterance effectively. However, in all the experiment

TABLE I. EXPERIMENT SCHEMES AND WITH THE ABNORMAL AND NORMAL UTTERANCES FOR EACH SCHEME

	Scheme1		scheme2		scheme3		scheme4		scheme5	
	Normal	Pathology								
Training	135	273	180	364	225	455	270	546	315	637
Testing	315	637	270	546	225	455	180	364	135	273

TABLE II. PERFORMANCE OF MENTIONED METHODS WITH GIVEN EXPERIMENT SCHEME AND GAUSSIANS MIXTURE DEGREES

Scheme	Gaussian Mixture Degree	GMM			GMM-SVM			GMM-UBM		
		Nr	Pr	CCR	Nr	Pr	CCR	Nr	Pr	CCR
scheme1	8	0.622	0.595	0.604	0.657	0.614	0.628	0.657	0.625	0.636
	16	0.657	0.623	0.635	0.695	0.639	0.658	0.733	0.645	0.674
	24	0.689	0.633	0.651	0.717	0.648	0.671	0.759	0.674	0.702
scheme2	8	0.633	0.632	0.632	0.648	0.643	0.645	0.670	0.650	0.657
	16	0.693	0.656	0.668	0.707	0.665	0.679	0.748	0.716	0.727
	24	0.700	0.659	0.673	0.719	0.676	0.69	0.759	0.720	0.733
scheme3	8	0.650	0.651	0.651	0.672	0.662	0.665	0.694	0.679	0.684
	16	0.7	0.668	0.678	0.728	0.673	0.691	0.794	0.717	0.743
	24	0.711	0.673	0.686	0.75	0.679	0.702	0.806	0.720	0.748
scheme4	8	0.674	0.656	0.662	0.689	0.663	0.672	0.719	0.707	0.711
	16	0.711	0.663	0.679	0.741	0.670	0.694	0.763	0.736	0.745
	24	0.733	0.681	0.699	0.763	0.689	0.713	0.785	0.747	0.760

TABLE III. RECOGNITION RATIO OF THE ENTIRE EXPERIMENT SCHEME FOR SHR ALGORITHM

	recognition ratio	scheme1	scheme2	scheme3	scheme4
SHR	Nr	0.851	0.885	0.906	0.911
	Pr	0.410	0.416	0.448	0.410
	CCR	0.556	0.571	0.599	0.576

schemes the training speech was composed of many unknown kinds of abnormal utterance, such as adenopharyngitis utterance, pharyngitis utterance, acute rhinitis and so. And caused by different factors, the voice characteristics of the change is not the same, which can affect the performance of those methods based on GMM. In order to analysis the performance of this proposed algorithm efficiently, the experiment scheme mentioned above was used to the SHR algorithm, and the results were given as the following Table III.

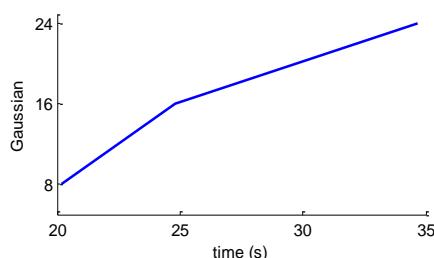


Figure 3. Relation of training time and the Gaussian

From the Table III shows the CCR of SHR algorithm is only 57.6%, compared with that of this proposed, has 18.4% dropped under the experiment scheme4. Because when the speaker pronounced the unknown continuous speech, the glottis can not close well in voiceless consonant period, and the glottis can not open well in resonance period, especially under the condition of pronunciation organs can't work well. However, due to the glottis can not work well, many hoarse noise were made in the speaker utterance which made the SHR

method difficult to extract the harmonic component of abnormal utterance.

Although, the CCR can be improved with its number of Gaussians increasing, however the time price of the system for training the model and testing was very huge, which can not be accepted by the online detection system. The specific time price for training and testing were given as the following figures.

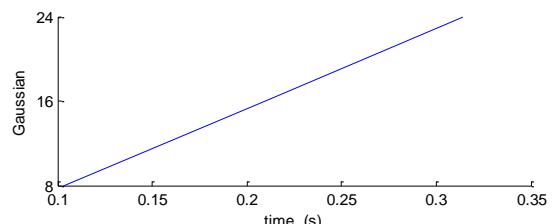


Figure 4. Relation of testing time and the Gaussian

Figure 4 shows the training time of a given training speech that its duration between 20s and 33s. And figure5 shows the testing time of a given testing speech with its duration between 2s and 3 s.

From the Table II and Table III also shows that there were long ways to go in the continuous speech abnormal detection still. Many researchers worked in the speech signal processing and throat disease diagnosis will do them best to work hard in the abnormal speech detection.

As Table II shows that the CCR of the mentioned for abnormal and normal speech is lower than that of their which published in recently literatures. Continuous speech is one of the best factors, resulting in the CCR decline sharp. However, few research report were found in recently published literatures, so only using the experimental results shows in Table II to conclude that this proposed algorithm has the advantages of performance for abnormal speech detection, which are not enough.

In order to further analysis the performance of the mentioned algorithm for abnormal monophthongs speech

detection published in recently literature [34]. 50 abnormal monophthongs, composed of |a| and |o| (the number of |a| is 30) selected from PANSD, and 50 normal monophthongs, consisted of |a| and |o| (the number of |a| is 35) selected from PANSD, were used to the experiment. And the duration of each experiment speech data is between 1s and 2s with removed the unvoiced component by manual.

Under the monophthongs, the performance of mentioned algorithms in this paper for abnormal speech is given as the following table:

TABLE IV. THE CCR OF MENTIONED ALGORITHM IN THIS PAPER FOR ABNORMAL SPEECH

	GMM	GMM-SVM	SHR	GMM-UBM
CCR	0.736	0.758	0.784	0.782

From the Table IV shows that the CCR of this proposed algorithm is lower than that of the SHR algorithm mentioned in this paper, because of when the speaker pronounced monophthongs, the speech signal of vocal cord vibration can show period well, but due to the glottis's opening or closing does not reach the designated position, and resulting produce hoarse noise in speech signal. Although this kind of hoarse noise can influence SHR algorithm to extract the harmonic components in the voice, but the overall impact is not big. However this kind of hoarse noise can make big influence on these algorithms which based on GMM model.

From the Table II and Table III also shows that there were long ways to go in the continuous speech abnormal detection still. Many researchers worked in the speech signal processing and throat disease diagnosis will do them best to work hard in the abnormal speech detection.

Good news obtained in this proposed is that compared with the GMM and GMM-SVM algorithm, the performance of the algorithm proposed in this paper, the improvement of the CCR is 3.6% and 2.4%, respectively. Further shows that only rely on GMM algorithm, it is difficult to fit the distribution curve of speech feature under complex conditions. So, from the Table II and Table III show that this proposed method have a good performance for abnormal utterance and utterance detection, but a long way to go to solve those problems about the continuous abnormal speech detection. Now, although the performance of the algorithm proposed in this paper is not very good, but is a step in terms of continuous abnormal voice detection. Next step, the algorithm improvement and other more advanced study for abnormal speech detection and abnormal speech speaker recognition were our do best direction.

#### IV. CONCLUSION

In order to push the products based on speech recognition to the application market, dealing with the abnormal speech effectively becomes the key and urgent task in speech signal processing area. Nowadays, the abnormal speech signal research is still in its infancy period, how to effectively detect continuous abnormal voice will impel the voice products market is the most fundamental and most crucial step. In this paper, the

method based on the model of abnormal detection for continuous speech is discussed, and from the experimental results shows, the proposed method for detecting abnormal continuous speech has a certainly effect in this paper, but there are still a distance from the abnormal speech detection target. Detection of continuous abnormal speech also needs to continue efforts to explore, to meet the development of speech signal well.

#### ACKNOWLEDGMENT

This work was supported by the national natural science foundation of China (No. 61301300), Guangdong university outstanding young innovative personnel training project, China (LYM 10089), the Science and Technology Planning Project of Guangdong Province, China (2012B010100034) and the Fundamental Research Funds for the Central Universities of SCUT (2012ZB0010).

#### REFERENCES

- [1] S. Furui, "50 years of progress in speech and speaker recognition research," *Eci Transactions on Computer And Information Technology*, vol. 1, 2005.
- [2] A. Fazel and S. Chakrabarty, "An Overview of Statistical Pattern Recognition Techniques for Speaker Verification," *Circuits and Systems Magazine, IEEE*, vol. 11, pp. 62-81, 2011.
- [3] Qing Li, Hon Chung Mak, Jianmin Zhao, Xinzong Zhu. "OXML: an Object XML Database Supporting Rich Media Indexing and Retrieval," *Journal of Multimedia*, 2011, Vol.6, pp. 115-121
- [4] Ji-chen Yang, Lei-an Liu, Qing-wei Qin, Min Zhang. "Audio Event Change Detection and Clustering in Movies," *Journal of Multimedia*, 2013, Vol.8, pp. 113-120
- [5] N. Karamangala and R. Kumaraswamy, "Speaker Recognition in Uncontrolled Environment: A Review," *Journal of Intelligent Systems*, vol. 22, pp. 49-65, 2013.
- [6] R. Togneri and D. Pulella, "An Overview of Speaker Identification: Accuracy and Robustness Issues," *Circuits and Systems Magazine, IEEE*, vol. 11, pp. 23-61, 2011.
- [7] H. J. M. Steeneken and J. H. L. Hansen, "Speech under stress conditions: overview of the effect on speech production and on system performance," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, 1999, pp. 2079-2082 vol.4.
- [8] T. May, et al., "Noise-Robust Speaker Recognition Combining Missing Data Techniques and Universal Background Modeling," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, pp. 108-121, 2012.
- [9] G. Vaziri, et al., "Pathological assessment of patients' speech signals using nonlinear dynamical analysis," *Computers in Biology and Medicine*, vol. 40, pp. 54-63, 2010.
- [10] Y. Jeong, "Robust speaker adaptation based on parallel factor analysis of training models," *Electronics Letters*, vol. 47, pp. 465-467, 2011.
- [11] N. Morales, et al., "Feature Compensation Techniques for ASR on Band-Limited Speech," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, pp. 758-774, 2009.

- [12] M. H, "Voice Control System for Airborne Environments," 1977.
- [13] S. Chakrabarty, et al., "Robust speech feature extraction by growth transformation in reproducing kernel Hilbert space," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, pp. 1842-1849, 2007.
- [14] J.-S. Park, et al., "Feature vector classification based speech emotion recognition for service robots," *Consumer Electronics, IEEE Transactions on*, vol. 55, pp. 1590-1596, 2009.
- [15] K. Jangwon, et al., "An exploratory study of manifolds of emotional speech," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 5142-5145.
- [16] L. DV, "A problem in forensic science," *Biometrika*, vol. Vol. 64, pp. 207-213, 1977.
- [17] C. Zhang and T. Tan, "Voice disguise and automatic speaker recognition," *Forensic Science International*, vol. 175, pp. 118-122, 2008.
- [18] R. D. Rodman, "Speaker recognition of Disguised Voices: A Program for research" *North Carolina State University, North Carolina* 2003.
- [19] P. Perrot, et al., "Voice Disguise and Automatic Detection: Review and Perspectives," *Nonlinear Speech Processing Springer Berlin / Heidelberg*, vol. vol. 4391, pp. pp. 101-117, 2007.
- [20] L. A. Khan, et al., "Speaker verification from partially encrypted compressed speech for forensic investigation," *Digital Investigation*, vol. 7, pp. 74-80, 2010.
- [21] Hollien. (2002). Forensic Voice Identification.
- [22] A. Drygajlo, "Forensic Automatic Speaker Recognition [Exploratory DSP]," *Signal Processing Magazine, IEEE*, vol. 24, pp. 132-135, 2007.
- [23] X. Wang, et al., "Discrimination Between Pathological and Normal Voices Using GMM-SVM Approach," *Journal of Voice*, vol. 25, pp. 38-43, 2011.
- [24] R. Tavares, et al., "Combining entropy measurements and cepstral analysis for pathological voice assessment," in *Biosignals and Biorobotics Conference (BRC), 2011 ISSNIP*, 2011, pp. 1-5.
- [25] L. Arias, et al., "Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients," *Biomedical Engineering, IEEE Transactions on*, vol. 58, pp. 370-379, 2011.
- [26] M. Brockmann, et al., "Reliable Jitter and Shimmer Measurements in Voice Clinics: The Relevance of Vowel, Gender, Vocal Intensity, and Fundamental Frequency Effects in a Typical Clinical Task," *Journal of Voice*, vol. 25, pp. 44-53, 2011.
- [27] Paul.Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonic to noise ratio of a sample sound," *Institute of Phonetic Sciences, University of Amsterdam*, vol. Vol. 17, pp. 97-110, 1993.
- [28] D. Michaelis, et al., "Glottal to noise excitation ratio-A new measure for describe pathological voices," *Acta Acustica United with Acustica*, vol. Vol.83, pp. pp.700-706, 1997.
- [29] R. D. A., et al., "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, pp. 19-41, 2000.
- [30] T. May, et al., "Noise-robust speaker recognition combining missing data techniques and universal background modeling," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. pp. 1-1, 2011.
- [31] Y. Lei and J. H. L. Hansen, "Mismatch modeling and compensation for robust speaker verification," *Speech Communication*, vol. 53, pp. 257-268, 2011.
- [32] J. Yang, "Research on Speaker Information Analysis and Its Application to Multimedia Retrieval," *Ph. D, South China University of Technology Guangzhou*, 2010.
- [33] H. J. L. Y.-x. He and Q.-h. L. Wei, "Speaker Recognition Algorithm for Abnormal Speech Based on Abnormal Feature Weighting," *Journal of South China University of Technology (Natural Science Edition)*, vol. 3, p. 019, 2012.
- [34] P. Henriquez, et al., "Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, pp. 1186-1195, 2009.



**Jun He** was born in Shaoyang, Hunan province China in July, 1978. He received the B.Eng. degree in computer and application from National University of Defense Technology, Changsha, Hunan province, China, in 2002; he received the M. Eng. Degree in computer software and theory from South China Normal University, Guangzhou, Guangdong province, China, in 2008. He received Ph. D. in signal and information processing from South China University of Technology, Guangzhou, Guangdong province, China, in 2012.

Now he is work in Guangdong University of Petrochemical Technology and Guangdong key laboratory of petrochemical equipment fault diagnosis. His currents interest is abnormal speech detection, speaker recognition for abnormal utterance and large vibration machinery fault diagnosis, image processing, pattern recognition, machine learning.

**Ji-chen Yang** was born in Jieshou, Anhui province China in June, 1980. He received the B.Eng. degree in electronic and information engineering from Guangdong University of Petrochemical Technology, Maoming, Guangdong province, China, in 2004, he received the M. Eng. Degree in system engineering from Guangdong University of Technology, Guangzhou, Guangdong, China, in 2007.he received the Ph.D. in Telecommuncation and information system from South China University of Technology, Guangzhou, Guangdong, China, in 2010.

Now he is a post doc. researcher in South China University of Technology. His currents interest is movie audio signal processing.

**Jianbin Xiong** was born in Shaoyang, Hunan province China, in July, 1976. He received B.E. and PH.D in Control Theory and Control Engineering from Guangdong University of Technology, Guangdong, China in 2006 and 2012, respectively. He is now working at the computer and information school, Guangdong University of Petrochemical Technology, China.

His current research interests include signal processing, image processing, information fusion, and computer applications.

**Guoxi Sun** was born in Youyi, Heilongjiang Province, China, in October, 1972. He received B.E degree in Control Technology and Instruments from Tianjin University, Tianjin, Jiangsu province, China, in 1992; he received the M. Eng. Degree in signal and information processing from South China University of Technology, Guangzhou, Guangdong province, China, in 1999; He received Ph.D. in Signals and Systems from South China University of Technology, Guangzhou, Guangdong province, China, in 2006.

Now he is work in Guangdong University of Petrochemical Technology and Guangdong key laboratory of petrochemical equipment fault diagnosis. His currents interest is abnormal speech detection, speaker recognition for abnormal utterance and large vibration machinery fault diagnosis, image processing, pattern recognition, machine learning.

**Ming Xiao** was born in Sichuan, China. He received the B.Sc. in applied mathematics from Dalian University of Technology,

Dalian, China, in 1989, and the M.Sc. in telecommunication and information system and the Ph.D. degrees in signal and information processing from South China University of Technology, Guangzhou, China, in 2005 and 2008, respectively.

His main research interests lie in blind signal processing, image processing, pattern recognition, machine learning, and neuroinformatics.

# On Characterizing of Word-of-Mouth Propagation in Heterogeneous Online Social Networks

Li Niu

Key Laboratory of Ministry of Education for Data Engineering and Knowledge Engineering, Renmin University of China, Beijing, China

School of Information Resource Management, Renmin University of China, Beijing, China  
Email: libraniu@foxmail.com

Xiaoting Han

School of Economics and Management, Beihang University, Beijing, China  
Email: hanxiaoting@buaa.edu.cn

Yongjun Xu

School of Information Resource Management, Remin University of China, Beijing, China  
Email: xyj@ruc.edu.cn

**Abstract**—Recent years have witnessed the explosive growth of research on online social networks (OSNs), which provide a perfect platform for observing the Word-of-Mouth (WOM) propagation. However, recent research regarded OSNs as homogeneous networks, while the real-world OSNs are heterogeneous. This paper considered the heterogeneity of OSNs, proposed a WOM propagation model called HOSN to simulate WOM propagation in different types of nodes in OSNs. Numerical simulation experiments based on agents have been extensively conducted, which show the mechanism of WOM propagation in heterogeneous online social networks. It can be seen from the experimental results that along with the WOM spreading and time passing, more and more users in OSNs review and know the WOM information. Moreover, from the network structure and user behavior perspective, some interesting findings are got. First, the number of initial known nodes has a certain impact on the WOM propagation scope and speed. Then, the number of neighbors has a great impact on the WOM propagation process and results. Finally, percentage of different user type is the decisive factor for WOM propagation either in final propagated number or propagation speed. Findings in this paper attempt to complement the theoretical framework of WOM propagation and complex networks, and thus to further promote the practice of WOM control in OSNs.

**Index Terms**—Information Propagation; Heterogeneous Online Social Networks; Word of Mouth; Network Structure; User Behavior

## I. INTRODUCTION

Along with the development of Web2.0, Online Social Networks (OSNs) such as Facebook, MySpace, LinkedIn and Twitter have become a popular social media platform [1, 2, 3], while they have been developed massively for business and political purposes, such as viral marketing, targeted advertising, political campaigns, and even

terrorist activities [4, 5], especially for the propagation of Word-of-Mouth (WOM).

WOM is the passing of information from person to person by communication. In the modern society, WOM is becoming a major information source which can affect current or potential customer's business choice. WOM is not only conducive to the expansion of brand awareness, more importantly; it can greatly influence the consumers' business behavior and attitudes. Katz and Lazarsfeld found that the influence of the WOM among consumers to switch brands is seven times that of the press, four times that of personal selling and two times that of the radio advertising [6].

WOM in OSNs is regarded as a high-impact marketing method, and this method can effectively spread the product and marketing message. Therefore, research on the propagation of WOM in OSNs in order to study how to motivate and control WOM has enormous essentiality and has become a research hotspot in recent years. However, most of the recent research regarded OSNs as homogeneous networks, thus ignoring the heterogeneity of real-world OSNs applications. In limited studies which considered the heterogeneity of OSNs, the heterogeneous online social networks are converted to single-node-type network or only focus on certain type of heterogenous network. Therefore, from the view of heterogenous complex network, to study the rules of WOM propagation over various types of nodes in OSNs, has a very important theoretical value and practical significance.

## II. RELATED WORK

Online social networks are focused on sharing information or opinions, which has been studied extensively in the context of WOM propagation [7]. The propagation mode of WOM in OSNs has a fundamental difference with that in traditional media. WOM propagation in traditional media is based on content,

while WOM propagation in OSNs is user-centered, in other words, it is based on the friend relationship among users in OSNs. Traditional diseases and information propagation, such as epidemics spread in the population [8], virus spread on computer networks [9], information propagation in the virtual community and blogs [10, 11], rumor diffusion in virtual society [12], can be regarded as propagation behaviors subject to certain laws, which can provide useful references to WOM propagated in OSNs. OSNs has broken the traditional WOM propagation method, which uses relationship among users to change the relationship between user and information, and vice versa, it also uses the relationship between user and information to affect the relationship among users. Therefore, WOM propagation in OSNs has become a research hotspot in recent years.

WOM can be regarded as information, therefore research on WOM propagation in OSNs is based on information propagation theory. Information propagation has been extensive and in-depth studied in the field of epidemiology, sociology and marketing. SIS model and SIR model are proposed to solve the infectious disease diffusion problem [13]. BASS model was proposed to simulate the products' WOM propagation problem [14]. Lopez-Pintado et al. studied the product diffusion in complex social networks based on mean-field theory, and found out that innovation diffusion in complex networks also exists a threshold which closely related to the degree distribution and propagation functions of the network [15]. Westerman et al. studied the effect of system generated reports of connectedness on credibility, and got that curvilinear effects for number of followers exist, such that having too many or too few connections results in lower judgments of expertise and trustworthiness [16]. Agliari et al. studied information spreading in a population of diffusing agents [17]. Karsai et al. studied the effects of different topological and temporal correlations on information spreading in complex communication networks [18]. Jin et al. provided HPC simulations method to study the behaviors of information propagation in complex social networks [19].

It can be seen that there has been a lot of research focus on WOM information propagation in OSNs. Most recent research regarded OSNs as homogeneous networks. However, OSNs in real-world applications are heterogeneous. Heterogeneous networks are composed of multiple types of nodes and links [20]. For example, there are different types of nodes in Facebook as a typical representative of OSNs, such as users, blogs. Nodes of different types are usually associated with user nodes [21], for instance, a user commented on a product. Some scholars have studied the issues of information propagation in heterogeneous networks. Li et al. transform the heterogeneous network to a probabilistic influence graph, in which nodes contains only human types, and then solve the influence maximization problem based on Independent Cascade Model [22]. Sun et al. studied clustering of multi-typed heterogeneous networks with a star network schema and proposed NetClus algorithm to present good ranking and cluster

membership information for each attribute object in each net-cluster nodes to generate high-quality net-clusters [23]. Shi et al. studied the relevance search problem in heterogeneous networks and proposed HeteSim algorithm to evaluate the relatedness of heterogeneous objects in academic networks such as ACM and DBLP [24].

In the above few studies considering the heterogeneity of OSNs, the heterogeneous online social networks are converted to single-node-type network or only focus on certain type of network such as star network. There is no further consideration of WOM propagation in different types of nodes.

### III. MODELING THE WOM PROPAGATION IN HETEROGENEOUS ONLINE SOCIAL NETWORKS

#### A. Model Description

In this paper, a WOM propagation model referred to communicable disease model SIR and SIS [25] named HOSN model is proposed in this section. The basic idea of this model is as follows:

WOM is diffused in heterogeneous OSNs, nodes in HOSN model are divided into two types which are user and product.

Links in HOSN model are divided into three types: links among users, links between products, links among users and products. Thereinto, links among users and products include reviews from users to products and the known relationship between users and products.

Users in HOSN are divided into three types based on user behavior:

Type A users: They don't accept the WOM information in HOSN, and they don't diffuse the WOM information.

Type B users: They accept the WOM information in HOSN, but they don't diffuse the WOM information.

Type C users: They accept the WOM information in HOSN, and they are willing to diffuse the WOM information.

At the initial time ( $t=0$ ), there are few users know the products' WOM information and fewer users have reviewed the products, and most users haven't known nor reviewed the products.

When  $t=t+1$ , the WOM propagation process starts until the total running time arrives, for each review link, the WOM is propagated from both user end  $U$  and product end  $P$ .

For the user end  $U$ , it will visit and affect its user neighbor, the propagation process is:

If the neighbor node is type A, it will reject the WOM information.

If the neighbor node is type B or C and hasn't relationship with  $P$ , it will accept the WOM information and know  $P$  with the probability of  $p_2$ .

If the neighbor node is type C and hasn't relationship with  $P$ , it will accept the WOM information and review  $P$  with the probability of  $p_1$ .

For the product end  $P$ , it will visit and affect its product neighbor  $P'$ , the propagation process is:

If the neighbor node  $P'$  hasn't relationship with  $U$ , user  $U$  review  $P'$  with the probability of  $p_1$ .

If the neighbor node  $P'$  hasn't relationship with  $U$ , user  $U$  know  $P'$  with the probability of  $p_2$ .

In order to describe the WOM propagation model clearly, notations of parameters used in HOSN model is shown in Table I.

TABLE I. NOTATIONS OF PARAMETERS

Notation	Description
$n$	number of user nodes in HOSN
$m$	number of product nodes in HOSN
$\langle k_1 \rangle$	averaged degree of user nodes in HOSN
$\langle k_2 \rangle$	averaged degree of product nodes in HOSN
$n_{ir}$	number of initial reviews
$n_{ik}$	number of initial known links
$p_A$	percentage of type A users
$p_B$	percentage of type B users
$p_C$	percentage of type C users
$n_A$	number of type A users
$n_B$	number of type B users
$n_C$	number of type C users
$p_1$	probability of a user to review a product
$p_2$	probability of a user to know a product
$t$	total running time

### B. Model Algorithm

The process algorithm of HOSN model is shown in Fig. 1.

*Step 1:* When  $t=0$ , Initialize information. Set  $n_{ir}$  links (between user node and product node) to review, and then set  $n_{ik}$  links (between user node and product node) to known. After the initialization, the number of type A, B, C users are:

$$n_A = n \times p_A \quad (1)$$

$$n_B = n \times p_B \quad (2)$$

$$n_C = n \times p_C \quad (3)$$

*Step 2:* When  $t=t+1$ , visit each review link (suppose the user end node is  $U$ , and the product end node is  $P$ ) and do Step 2.1 to Step 2.2. The WOM information propagation process repeats until  $t=T$ .

*Step 2.1:* Visit user end node  $U$ , and then visit user neighbors of  $U$ , and we name the user neighbor  $U'$ , then do Step 2.1.1 to Step 2.1.3.

*Step 2.1.1:* If  $U'$  has relationship with  $P$  or  $U'$  is type A, do nothing; else, go to Step 2.1.2.

*Step 2.1.2:* If  $U'$  is type B, create a known link between  $U'$  and  $P$  with the probability of  $p_2$ ; else, go to Step 2.1.3.

*Step 2.1.3:* Create a review link between  $U'$  and  $P$  with the probability of  $p_1$ ; then estimate if there is relationship with  $U'$  and  $P$ , if not, create a known link between  $U'$  and  $P$  with the probability of  $p_2$ .

*Step 2.2:* Visit product end node  $P$ , and then visit product neighbors of  $P$ , and we name the product neighbor  $P'$ , then do Step 2.2.1 to Step 2.2.2.

*Step 2.2.1:* If there is no relationship between  $U$  and  $P'$ , create a review link between  $U$  and  $P'$  with the probability of  $p_1$ ; go to Step 2.2.2.

*Step 2.2.2:* If there is no relationship between  $U$  and  $P'$ , create a known link between  $U$  and  $P'$  with the probability of  $p_2$ .

*Step 2.3:* Visit next review link until all the review links in HOSN are visited.

*Step 3:* End.

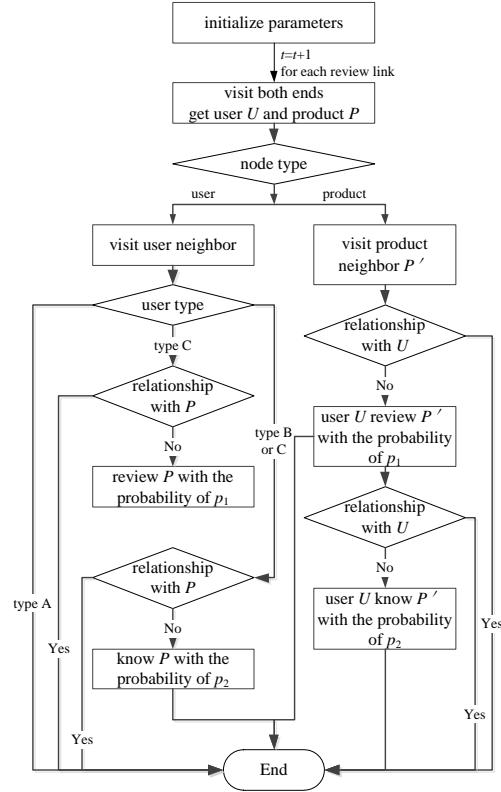


Figure 1. WOM propagation process in HOSN

Suppose the number of different links at time  $t$  is as shown in Table II. According to mean-field theory [26], after an iteration of the above propagation process, when time  $= t+1$ , the number of each type of links is:

TABLE I. NUMBER OF DIFFERENT LINKS AT TIME  $T$  AND  $T+1$

Link type	Number at $t$	Number at $t+1$	Number increased
review	$n_{rt}$	$n_{rt1}$	$n_{ri}$
known	$n_{kt}$	$n_{kt1}$	$n_{ki}$

$$n_{ri} = n_{rt} \times \langle k_1 \rangle \times p_C \times p_1 + n_{rt} \times \langle k_2 \rangle \times p_1 \quad (4)$$

$$n_{ki} = n_{rt} \times \langle k_1 \rangle \times (p_B + p_C) \times p_2 + n_{rt} \times \langle k_2 \rangle \times p_2 \quad (5)$$

$$n_{rt1} = n_{rt} + n_{ri} \quad (6)$$

$$n_{kt1} = n_{kt} + n_{ki} \quad (7)$$

From (4) and (5), we can see that the number of review and known links is increasing while the WOM information is diffusing, and this phenomenon will be simulated in the next section.

From (4)-(7), we can get that the number of review and known links at time  $t+1$  are:

$$n_{rt1} = n_{rt} \times (1 + \langle k_1 \rangle \times p_C \times p_1 + \langle k_2 \rangle \times p_1) \quad (8)$$

$$n_{kt1} = n_{rt} \times (1 + \langle k_1 \rangle \times (p_B + p_C) \times p_2 + \langle k_2 \rangle \times p_2) \quad (9)$$

#### IV. EXPERIMENTS AND RESULTS ANALYSIS IN DEFAULT PARAMETER SETTING

##### A. Methodology

There are several methods to study the WOM propagation in OSNs, such as complex network analysis [27], cellular automata [28] and agent based modeling [29]. In the above three methods, we choose agent based modeling as the method to simulate the WOM propagation process because of its flexibility. Factors which affect the WOM propagation process can be adjusted by agent based modeling method, therefore how the different combination of factors causes different information propagation effect can be compared easily, which can provide strong evidence for spreading WOM. In the WOM propagation simulation process, specific data of each agent can be easily obtained to quantitatively analysis how different user behavior effects the WOM propagation in real-world OSNs.

In order to prove the efficiency of the HOSN model, a network simulating WOM propagating in OSNs is conducted in Netlogo, which is a simulation software based on multi-agents. Nodes in HOSN are modeled by agents, and interaction among agents is used to simulate the WOM propagation mechanism in the proposed HOSN model. In this way, parameters in HOSN model can easily be controlled, which can facilitate the observation of efficiency and effectiveness the model and obtain the data of simulation results for quantitative analysis.

##### B. Experimental Setup

This paper uses a randomly generated dataset for experiments. There are 1000 user nodes and 1000 product nodes in the dataset, other key features of this dataset are summarized in Table III.

TABLE II. PARAMETERS SETTING OF DATA SET

parameter	value
n	1000
m	1000
$\langle k \rangle$	6
$\langle k_2 \rangle$	4
$n_{ir}$	20
$n_{ik}$	40
$p_A$	0.1
$p_B$	0.2
$p_c$	0.7
$p_I$	0.1
$p_2$	0.5
t	1000

The proposed HOSN model is implemented in Netlogo, on a Microsoft Windows 8 Professional platform with 64bit edition. The experimental computer is with an Intel Core i7 2620M CPU, 16 GB DDRII 667 MHz RAM, and a Seagate Barracuda 7200.11 500GB hard disk.

##### C. Experimental Results

The experiment program runs 1000 times, the averaged results are analyzed in this paper. The running effect of the HOSN model is given in Fig. 2 and Fig. 3. In the left part of Fig. 2, the circular nodes represent user nodes, among which the brown ones represent the type A users,

the red ones represent the type B users, the orange ones represent type C users; while the triangular nodes represent the product nodes. It is can be seen in Fig. 2. That there are lines between nodes and nodes, among which the orange ones represent links between users, the blue ones represent links between products, the black ones represent review relationship between users and products, while the green ones represent the known relationship between users and products. In the right part of Fig. 2 shows the final state of the propagation process, in which the violet lines represent the review relationship between users and products while the pink ones represent the known relationship between users and products. It is can be seen that there are a mass of new links emerging while WOM propagating in HOSN.

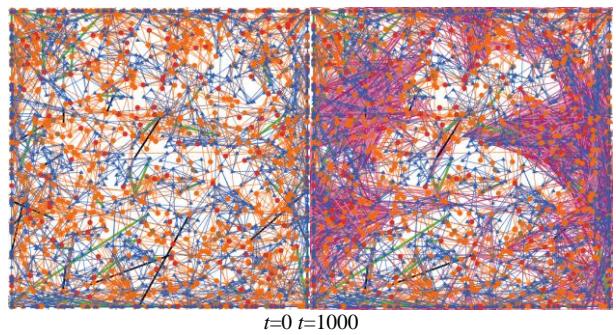


Figure 2. Initial and final status of WOM propagation in HOSN

Fig. 3 shows the percentage change process of the two types of relationship between users and products while the WOM is diffused in HOSN. As shown in the figures, at the initial time ( $t=0$ ), there are only 20 review links and 40 known links between users and products. As time passed, number of known links rapidly increases meanwhile number of review links also increases. At  $t=269$ , the number of review links achieves maximum of 371, and then at  $t=292$ , the number of known links achieves maximum of 1823. After that, number of review links and known links remains unchanged.

#### V. SENSITIVITY ANALYSIS OF PATAMETERS IN HOSN MODEL

##### A. Sensitivity Analysis of Number of Initial Reviews and Initial Known Links in HOSN

In order to examine how  $n_{ir}$  (number of initial reviews in HOSN) and  $n_{ik}$  (number of initial known links in HOSN) affects the WOM propagation in HOSN,  $n_{ir}$  and  $n_{ik}$  are changed to different values as shown in Table IV.

TABLE III. DIFFERENT PARAMETER SETTINGS OF INITIAL LINKS

	$n_{ir}$	$n_{ik}$
Default Situation	20	40
Situation 1	10	40
Situation 2	30	40
Situation 3	20	20
Situation 4	20	60

In situation 1, situation 2, situation 3 and situation 4, the experiment program also runs 1000 times. The experimental results are shown in Fig. 4 - Fig. 6. Fig. 4

shows number of final review and known links in different situations. Fig. 5 shows the changing process of number of reviews in different situations, while Fig. 6 shows the changing process of number of known links in different situations.

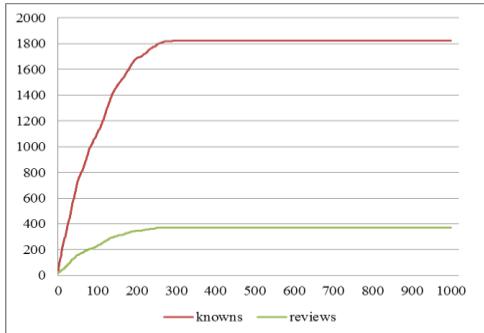


Figure 3. Evolution process of WOM propagation in HOSN

We can see from Fig. 4 that  $n_{ir}$  (number of initial reviews in HOSN) and  $n_{ik}$  (number of initial known links in HOSN) both have influence on the final state, but  $n_{ir}$  affects the number of final links more than  $n_{ik}$ , especially when  $n_{ir}$  decreases, the number of final reviews and known links drops largely. Also we can see that when  $n_{ik}$  increases, the number of final reviews and known links just change little.

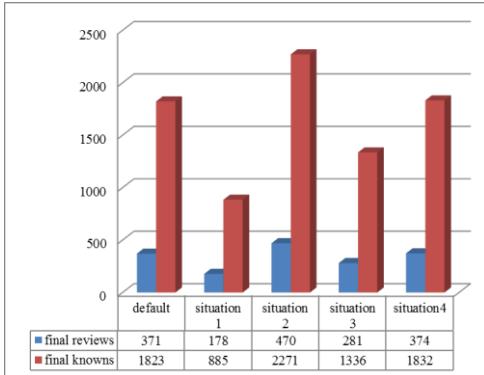


Figure 4. Number of final nodes in different situations

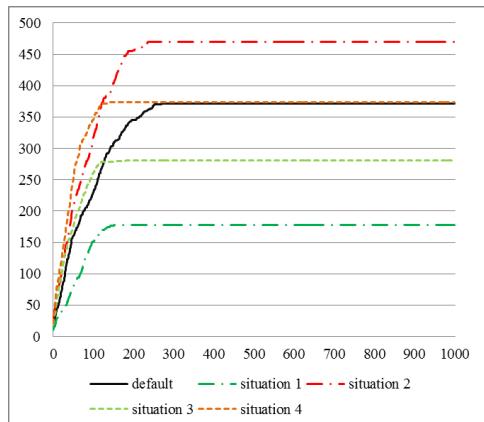


Figure 5. Changing process of number of reviews in different situations

It is can be seen in Fig. 5 that  $n_{ir}$  (number of initial reviews in HOSN) and  $n_{ik}$  (number of initial known links

in HOSN) do have impact on the propagation process. Clearly the propagation process goes faster when  $n_{ir}$  and  $n_{ik}$  increases. It means that when the initial number of reviews and known links increases, the propagation speed will be faster, and vice versa. Compared with the little change of final number of links while  $n_{ik}$  increasing, the growth rate of review and known links goes faster than in the default situation.

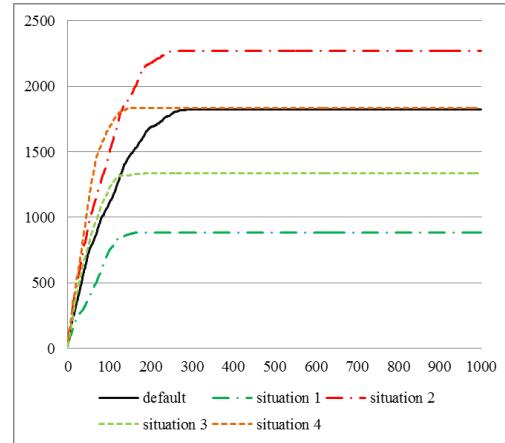


Figure 6. Changing process of number of known links in different situations

By sensitivity analysis of initial number of review and known links in HOSN, we find that  $n_{ir}$  and  $n_{ik}$  has a certain impact on the WOM propagation process and results. When  $n_{ir}$  and  $n_{ik}$  increases, the final number of review and known links will be more and the propagation process will be faster.

#### B. Sensitivity Analysis of Averaged Degree of Nodes in OSNs

In order to examine how the averaged degree of user nodes in HOSN ( $\langle k_1 \rangle$ ) and the averaged degree of product nodes in HOSN ( $\langle k_2 \rangle$ ) affects the information propagation,  $\langle k_1 \rangle$  is changed to 5 and 7 while  $\langle k_2 \rangle$  is changed to 3 and 5 in the HOSN model as shown in Table V.

TABLE IV. DIFFERENT PARAMETER SETTINGS OF  $\langle k_1 \rangle$  AND  $\langle k_2 \rangle$

	$\langle k_1 \rangle$	$\langle k_2 \rangle$
Default Situation	6	4
Situation 1	5	4
Situation 2	7	4
Situation 3	6	3
Situation4	6	5

The experimental results are shown in Fig. 7 - Fig. 9. Fig. 7 shows the final number of review and known links in different situations. We can see from Fig. 7 that the number of final review and known links increases with  $\langle k_1 \rangle$  and  $\langle k_2 \rangle$  increasing. In other words, when nodes in HOSN have more neighbors, the same information will diffuse to more nodes in the end, and vice versa.

Different propagation process based on the above five situations is shown in Fig. 8 and Fig. 9, in which the propagation speed can be observed distinctly. It can be

seen that the propagation process changes a lot when  $\langle k_1 \rangle$  and  $\langle k_2 \rangle$  changes. Clearly the propagation process goes faster when  $\langle k_1 \rangle$  and  $\langle k_2 \rangle$  increases. It means that when the degree of nodes increases, the propagation speed among nodes will be faster.

By sensitivity analysis of averaged degree of nodes in HOSN, we find that  $\langle k_1 \rangle$  and  $\langle k_2 \rangle$  has a certain impact on the information propagation process and results. When  $\langle k_1 \rangle$  and  $\langle k_2 \rangle$  increases, the final number of review and known links will be more and the propagation process will be faster.

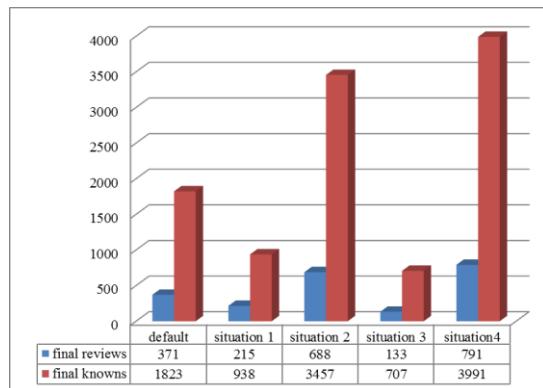


Figure 7. Number of final nodes in different situations

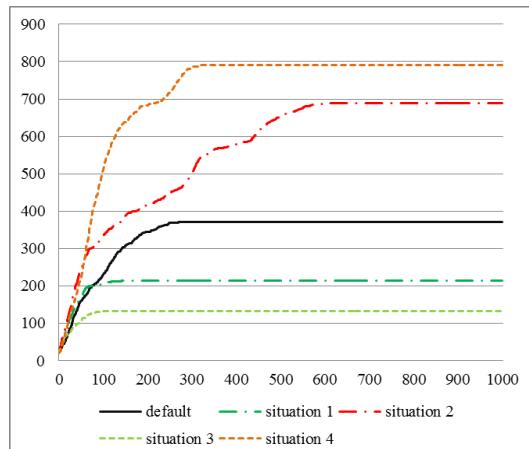


Figure 8. Changing process of unknown nodes

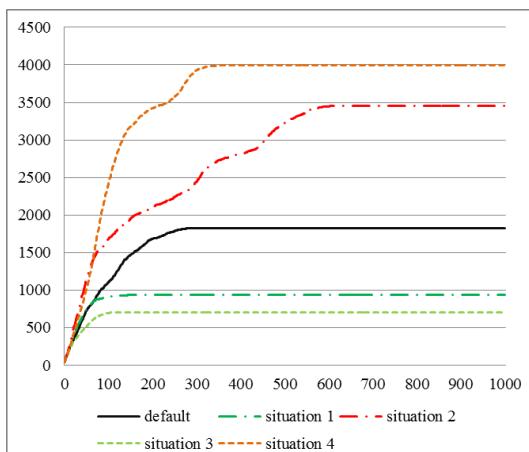


Figure 9. Changing process of known nodes

### C. Analysis of Different User Types

In this section, we change the percentage of different user types in order to analyze how different user behavior affects WOM propagation in HOSN. Two more situations are assumed. In situation 1, we decrease  $p_1$  and  $p_2$  (the percentage of Type A and Type B user), and in contrast, in situation 2,  $p_1$  and  $p_2$  are increased. The parameter setting is shown in Table VI.

TABLE V. DIFFERENT PARAMETER SETTINGS OF USER TYPES

	$p_1$	$p_2$	$p_3$
Default Situation	10%	20%	70%
Situation 1	5%	10%	85%
Situation 2	15%	30%	55%

The experimental results are shown in Fig. 10 - Fig. 12. Fig. 10 shows the final number of review and known links in different situations. We can see from Fig. 10 that the number of final review and known links increases with  $p_3$  increasing and  $p_1$  and  $p_2$  decreasing. In other words, when there are more users who want to accept and diffuse WOM in HOSN, the products' WOM information will be diffused to more users.

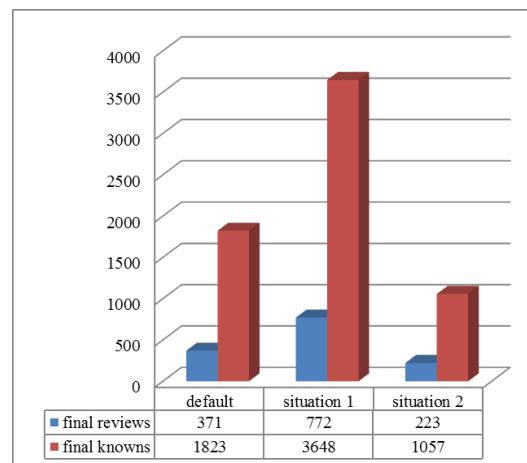


Figure 10. Number of final nodes in different situations

Different propagation process based on the above three situations is shown in Fig. 11 and Fig. 12. We can see in the figures that the propagation process changes in the three situations. Its changing trend goes steeper in situation 1 while gentler in situation 2. It shows that when the number of users who accept and want to diffuse products' WOM information in HOSN increases, the WOM propagation speed will be faster, and in contrast, when the number of users who accept and want to diffuse WOM information in HOSN decreases, the WOM propagation speed will be slower.

By analysis of different user types affecting the WOM information propagation in HOSN, we find out that the percentage of different users has a certain impact on the propagation process and propagation result. It is clearly seen from the experimental results that when there are more users who are willing to accept and diffuse the WOM information in HOSN, users who will finally review and know the products' WOM information will be more, and the propagation speed will be faster.

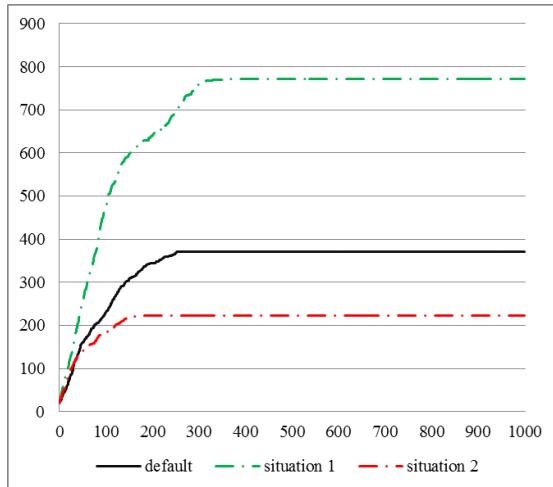


Figure 11. Changing process of unknown nodes

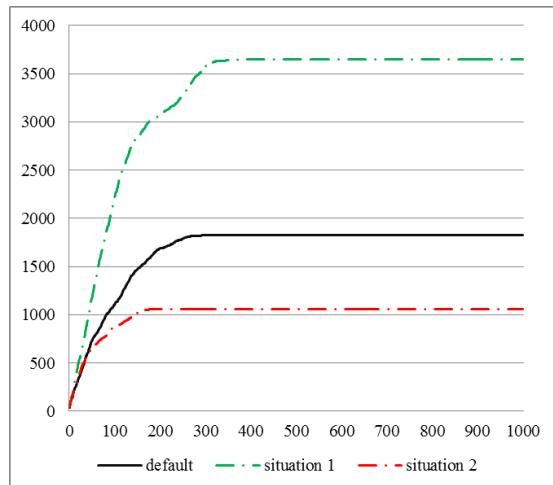


Figure 12. Changing process of known nodes

## VI. SUMMARY

A WOM propagation model named HOSN is proposed in this paper. From the model, we can investigate some WOM propagation rules in heterogeneous online social networks. From the experimental results, it is can be seen that along with the WOM propagation, the number of users who know the products and products' reviews increases and reaches its maximum, then keep an unchanging status. Moreover, numerical simulations are conducted from the network structure and user behavior perspective. Simulation results show some meaningful findings for WOM propagation and control in heterogeneous networks. First, the number of initial review and known links is a decisive factor on the WOM propagation speed and the number of review and known links. Also, the number of users' neighbors and products' neighbors has a great impact on the WOM propagation process and results either in final number of review and known links or propagation speed. Furthermore, we find out that the more users who are willing to accept and diffuse the WOM information, the more users who will finally review or know the WOM information, and the faster the information will be propagated.

## ACKNOWLEDGMENT

This work was supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (14XNJ026).

## REFERENCES

- [1] Y. Y. Ahn, S. Han, H. Kwak, S. Moon, H. Jeong, "Analysis of topological characteristics of huge online social networking services," *Proceedings of the 16th international conference on World Wide Web, WWW '07*: pp. 835–844, 2007.
- [2] Z. Wu and K. Aizawa, "Building Friend Wall for Local Photo Repository by Using Social Attribute Annotation," *Journal of Multimedia*, vol. 9, pp. 4-13, 2014.
- [3] D. Pallegedara and L. Pan, "Investigating Facebook Groups through a Random Graph Model," *Journal of Multimedia*, vol. 9, pp. 25-34, 2014.
- [4] A. Dunne, M. A. Lawlor, J. Rowley, "Young people's use of online social networking sites - a uses and gratifications perspective," *Journal of Research in Interactive Marketing*, vol. 4, pp. 46-58, 2010.
- [5] C. M. K. Cheung, P. Y. Chiu, M. K. O. Lee, "Online social networks: Why do students use facebook?," *Computers in Human Behavior*, vol. 27, pp. 1337–1343, 2011.
- [6] E. Katz and P. F. Lazarsfeld, "Personal influence: The part played by people in the flow of mass communications." New York: Free Press. (1955).
- [7] Xiaoting Han, Li Niu. Word of Mouth Propagation in Online Social Networks. *Journal of Networks*, 2012, 7(10):1670-1676.
- [8] Li Rui-Q, Tang Ming, Hui Pak-Ming, "Acta Physica Sinica", vol. 62(16), pp. 168903, 2013.
- [9] Newman M. E. J., Forrest S., Balthrop J. Email Networks and the Spread of Computer Viruses. *Physical Review E*, 2002, 66(3): 035101
- [10] Xiaoting Han, Li Niu. On Charatering of Information Propagation in Online Social Networks. *Journal of Networks*, 2013, 8(1):124-131.
- [11] Cha M., Pérez J. A. N., Haddadi H. The Spread of Media Content through Blogs. *Social Network Analysis and Mining*, 2011, 2(3): 249-264
- [12] Myers S. A., Zhu C., Leskovec J. Information Diffusion and External Influence in Networks. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2012: 33-41
- [13] R. M. Anderson, R. M. May. "Infectious Diseases of Humans: Dynamics and Control." Oxford University Press, USA, 1992.
- [14] R. Albert, A. L. Barabasi. "Statistical mechanics of complex networks." *Reviews of Modern Physics*, vol. 74(1), pp. 47-97, 2002.
- [15] D. Lopez-Pintado. "Diffusion in Complex Social Networks." *Games and Economics Behavior*, vol. 62(2), pp. 573-590, 2008.
- [16] D. Westermana, P. R. Spenceb, B. V. D. Heide. "A social network as information: The effect of system generated reports of connectedness on credibility on Twitter." *Computers in Human Behavior*, vol. 28(1), pp. 199-206, 2012.
- [17] E. Agliari, R. Burioni, D. Cassi, F. M. Neri. "Efficiency of information spreading in a population of diffusing agents." *Physical Review E*, vol. 73, pp. 046138, 2006.
- [18] M. Karsai, M. Kivel ä, R. K. Pan, K. Kaski, J. Kert ász, A. - L. Barab ási, J. Saram áki. "Small But Slow World: How

- Network Topology and Burstiness Slow Down Spreading.” *Physical Review E*, vol. 83, pp. 025102, 2011.
- [19] J. Jin, S. J. Turner, B. S. Lee, J. Zhong, B. He. “HPC Simulations of Information Propagation Over Social Networks.” *Procedia Computer Science*, vol. 9, pp. 292-301, 2012.
- [20] Sun Y., Han J. Mining Heterogeneous Information Networks: A Structural Analysis Approach. *ACM SIGKDD Explorations Newsletter*, 2013, 14(2): 20-28
- [21] Cai X., Bain M., Krzywicki A., et al. Reciprocal and Heterogeneous Link Prediction in Social Networks. *Advances in Knowledge Discovery and Data Mining*, 2012, 7302: 193-204
- [22] Li C. T., Lin S. D., Shan M. K.. Influence Propagation and Maximization for Heterogeneous Social Networks. *Proceedings of the 21st International Conference Companion on World Wide Web*. New York: ACM Press, 2012: 559-560
- [23] Sun Y., Yu Y., Han J. Ranking-Based Clustering of Heterogeneous Information Networks with Star Network Schema. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York: ACM Press, 2009: 797-806
- [24] Shi C., Kong X., Yu P. S., et al. Relevance Search in Heterogeneous Networks. *Proceedings of the 15th International Conference on Extending Database Technology*. New York: ACM Press, 2012: 180-191
- [25] H. W. Hethcote, “Qualitative Analyses of Communicable Disease Models,” *Mathematical Biosciences*, vol. 28, pp. 335-356, 1976.
- [26] Y. Roudi, J. A. Hertz, “Mean field theory for nonequilibrium network reconstruction,” *Physical Review Letters*, vol. 106, pp. 048702, 2011.
- [27] R.V. Kozinets, K. De Valck, A.C. Wojnicki, S.J.S. Wilner, “Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities,” *Journal of Marketing*, vol. 74, pp. 71-89, 2010.
- [28] J. Goldenberg, B. Libai, “Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth,” *Marketing Letters*, vol. 12, pp. 211-223, 2001.
- [29] T. Smith, J. R. Coyle, E. Lightfoot, A. Scott, “Reconsidering Models of Influence: The Relationship

between Consumer Social Networks and Word-of-Mouth Effectiveness,” *Journal of Advertising Research*, vol. 47, pp. 387-397, 2007.

**Xiaoting Han** was born in Shandong province of China at 7<sup>th</sup> April, 1983. She received her bachelor degree in information management and information system at Beihang University in Beijing of China in 2003, and then received her master degree in management science and engineering at Beihang University in Beijing of China in 2006.

She is now a technician at School of Economics and Management, vice dean of Beijing Demonstration Innovation Practice Base in Higher Education of Economics and Management and dean of Lab of Economics and Management of Beihang University. She is also a Ph.D. candidate currently focused on information diffusion in online social networks, complex networks analysis, web data mining and statistical physics.

**Li Niu** was born in Henan province of China at 23<sup>rd</sup> October, 1982. He received his bachelor degree in information management and information system at Beihang University in Beijing of China in 2005, and then received his doctor degree in system engineering at Beihang University in Beijing of China in 2010.

He is now an assistant professor at School of Information Resource Management of Renmin University of China. His research interests include information resource management, social network analysis, agent-based modeling and decision support system.

**Yongjun Xu** was born in Hunan province of China at 25th March, 1975. He received his bachelor degree in archives science at Xiangtan University in 1997, and then received his master degree and doctor degree both in archives science at Renmin University of China in 2004 and 2007.

He is now an associate professor at School of Information Resource Management, Renmin University of China. His research interests include enterprise archives management, knowledge management.

# Image Retrieval via Relevance Vector Machine with Multiple Features

Zemin Liu

College of Mathematics and Computer Science, Panzhihua University, China

Wei Zong

Massachusetts Avenue, MA 01239, USA

Email: wzong@mit.edu

**Abstract**—With the fast development of computer network technique, there is large amount of image information every day. Researchers have paid more and more attention to the problem of how users quickly retrieving and identifying the images that they may interest. Meanwhile, with the rapid development of artificial intelligence and pattern recognition techniques, it provides people with new thought on the study on complex image retrieval while it's very difficult for traditional machine learning method to get ideal retrieval results. For this reason, we in this paper propose a new approach for image retrieval based on multiple types of image features and relevance vector machine (RVM). The proposed method, termed as MF-RVM, integrates the informative cures of features and the discrimination ability of RVM. The retrieval experiment is conducted on COREL image library which is collected from internet. The experimental results show that the proposed method can significantly improve the performance for image retrieval, so MF-RVM presented in this paper has very high practicability in image retrieval.

**Index Terms**—MF-RVM; Image Feature; Image Retrieval

## I. INTRODUCTION

With the rapid development of computer network technique, there is large amount of image retrieval information every day, that how to effectively organize and manage these images has been a more and more serious research topic. If image retrieval data can't be effectively managed, a large amount of information will be lost [1, 2]. Therefore, the problem of how users quickly retrieving and identifying their required image information are being paid more and more attention by researchers. Recently, retrieval technologies based on data mining about image retrieval emerge as the times requirement, that how to quickly and accurately retrieve these images has become the key point of recent image retrieval technical study [1].

In recent years, with the fast development of artificial intelligence and pattern recognition, it provides people with new thought on the study on complex image retrieval, some classification methods based on goal decomposition and spectral signature, such as fuzzy set, generalized multiple kernel learning (GMKL), Bayesian and neural network classification method [2,3], and other

methods based on ground material property, statistic characteristic for image retrieval data, such as spectral angle mapping method, maximum likelihood method and minimum distance method, are began to be widely applied to image retrieval, but it has a lot of difficulties when being in the face of spectral data with hyper spectral and multi angles [2, 4]. In addition, this method has also some shortcomings for themselves.

For example, the degree of membership for fuzzy set classification method shall be given by experience or experts, and it has high subjectivity, so precision and intelligibility for learning problems of fuzzy system are the first questions to be solved [5, 6]. What's more, it first needs to select the variables according to problems' definition, the data type and feature when separating remote sensing effect with fuzzy technique. For GMKL, it's suitable to classify problems given by data, but its chosen kernels will significantly affect classification results. Moreover, it is difficult to fulfill the assumption of conditional independent assumption for naïve Bayesian classification when handling large-scale classification problems and select the required evaluation function [3, 4]. Also, this method has very complicated learning and training [7].

There are some shortcomings easily appearing for neural network classification method, such as local minimum and the slow rate of convergence. In addition, for the above traditional machine classification learning method [2], they have very high requirements of data regularity and shall be conducted under the assumption of infinite sample size. However, data classifying remote sensing image retrieval cannot usually meet the above requirements, with the characteristics of circumpolar latitude and small sample. That is to say, it is very difficult to obtain the ideal classification results with traditional machine learning method for these data. Furthermore, classification methods based on spectrum cannot provide textural features, just like that presented by identical permutation. Many remote sensing image retrievals are just reflected based on textural features. Those image classification methods based on spectral extraction feature are unable to accurately classify image retrieval [5, 6].

How can we effectively and accurately identify images? If image data can't be effectively managed, a large amount of information will be lost, so they can't be effectively and timely retrieved for use by people when being required. For this reason, researchers have paid more and more attention to problem of how users effectively and quickly identifying and managing images. Based on shortcomings for the above methods, one image retrieval method based on multiple features and RVM [8] is presented in this paper. Since SVM is a convex optimization problem, seen from the properties of convex optimization, local optimum solution solved with convex optimization must be global optimum solution, which is not included in other classification methods. In addition, SVM shows its effectiveness on image retrieval research in the aspects of independent study classification and automatic processing. However, the application of SVM is affected by kernel parameters and its classification performance is highly dependent on kernel parameters [9].

For this reason, a method based on multiple features representation and relevance vector machine is presented in this paper. Moreover, this method can be well applied to quickly solve nonlinear problems, which makes RVM have a wide application prospect in solving classification and prediction problems. The retrieval experiment is conducted on COREL image library which is collected from internet and has about 68040 photo images belonging to various categories. The experimental results show that it can significantly improve the accuracy for image retrieval to obtain parameters with cross validation, so MF-RVM presented in this paper has very high practicability in image retrieval [10-11].

Here we briefly summary the advantages of relevance vector machine, which accounts for reason that uses relevance vector machine to multiple feature based image retrieval. First, relevance vector machine introduces the sparseness inducing prior over the weight. Consequently, relevance vector machine has the ability to simultaneously perform feature selection and classification. On the other hand, because the proposed MF-RVM uses multiple image features and produces a very high feature space, it is important to fish out the informative features and remove the noise features, from the perspective of both computational efficiency and model performance [12]. By integrating RVM, MR-RVM is able to benefit from multiple features while improving its efficiency and performance. Moreover, it is very flexible due to its adaption ability from probabilistic formulation.

The step of the proposed MF-RVM is summarized as follows. (1) Remove the noise data so as to minimize the ambiguity and highlight the useful information. (2) Model the distribution of image features and extract the feature based on the model. (3) Choose the relevance vector machine as the retrieval model, and learn relevance vector machine based on the feature space or equally the similarity defined on the feature space. (4) We perform image retrieval using the learned MF-RVM and extracted features for test data.

The contributions of this paper are three aspects. (1), the feature extraction method proposed in this paper could highlight the informative components while reducing the data variance by means of data distribution. (2), relevance vector machine is used as retrieval model, by which MF-RVM benefits from its satisfied generalization ability and data adaption ability for image retrieval. (3), the results of our comprehensive experiments validates the advantages of MF-RVM in image retrieval, i.e., beats other methods on different experimental conditions.

The remainder part of the paper is organized as follows. We report the formulation details of the proposed method MF-RVM in Section 2. The empirically experiments of MF-RVM and related algorithms are reported in Section 3. We draw a conclusion in Section 4.

## II. THE PROPOSED SCHEMA

On the basis of the above discussion, in this part, we propose an approach, MF-RVM for image retrieval. Our approach is based on the relevance vector machine, which is able to conquer the defects of previous algorithm [13]. Our developed algorithm is graphically illustrated in Figure. There are three main steps. First, collect data. Second, data modeling and feature extraction [14]. Third, train the model described in Section 2 and perform test.

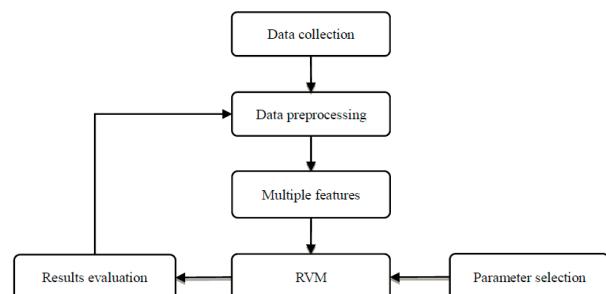


Figure 1. The experiment framework for MF-RVM

### A. Multiple Image Features

Image feature is one of the most important factors for image retrieval because it captures low-level and local cues such as gradient and texture [15]. However, it has been validated that, different features have different abilities and capture different information. Therefore, single image feature is not sufficient to represent the images. To overcome this limitation, we in this paper propose to use multiple features for image representation. Specifically, we use scale invariant feature (SIFT), HOG and HMAX. These features are combined and feed into the following relevance vector machine for image retrieval.

### B. Relevance Vector Machine

There are many machine learning problems belonging to the heading of supervised learning. Here we consider a set of input vectors  $X = \{x_n\}_{n=1}^N$  that combined with corresponding target values  $T = \{t_n\}_{n=1}^N$ . Our goal of that is to employ this training data, together with arbitrary relevant prior knowledge, to predict T for new values of

$X$ . We can tell apart two distinct cases: regression, in that  $T$  is a continuous variable, and classification, in which  $T$  belongs to a discrete set.

Here we make a consideration of models, in which the prediction label  $y(X, W)$  can be expressed as the linear combination of basis function  $\phi_m(x)$  of the form,

$$y(x, w) = \sum_{m=0}^M w_m \phi_m(x) = w^T \phi \quad (1)$$

where  $\{w_m\}$  are the weight parameters of the model. Relevance Vector Machine (RVM) makes the probabilistic predictions and yet that keeps the excellent predictive performance of the support vector machine. It also can make the preservation of the sparseness property of the SVM. In fact, for a wide variety of test problems it actually results in models which are dramatically sparser than the corresponding SVM, while sacrificing little if anything in the accuracy of prediction.

RVM models the conditional distribution of the target variable, given an input vector  $x$ , as a Gaussian distribution of the form,

$$P(t|x, w, \tau) = N(t | y(x, w), \tau^{-1})$$

where  $N(z | m, S)$  denotes a multivariate Gaussian distribution over  $z$  with mean  $m$ . The conditional probability of  $y(x, w)$  is given through Equation (1). The parameters  $w$  are given a Gaussian prior,

$$P(w|\alpha) = \prod_{m=0}^N N(w_m | 0, \alpha_m^{-1})$$

where  $\alpha = \{\alpha_m\}$  is a vector of hyper-parameters, with a hyper-parameter  $\alpha_m$  assigned to each model parameter. These hyper-parameters can be estimated through utilizing type-II maximum likelihood in which the marginal likelihood  $P(T|X, \alpha, \tau)$  is maximized with consideration of  $\alpha$  and  $\tau$  making an assessment of this marginal likelihood require integration over the model parameters

$$P(T|X, \alpha, \tau) = \int P(T|X, w, \tau) P(w|\alpha) dw.$$

We can neglect the relevant conditions from the trained model, the representation is Equation(1), and the training data are used that is connected with the remaining kernel functions, called relevance vectors. A step is applied to improve the  $m$  parameters simultaneously, in the relevance vector machine's identification version. The targets' conditional distribution is applied by

$$P(t|x, w) = \sigma(y)^t [1 - \sigma(y)]^{1-t}$$

where  $\sigma(y) = (1 + \exp(-y))^{-1}$  and  $y(x, w)$  is given by Equation (1). Please pay attention to the case  $t \in \{0, 1\}$ . Make an assumption which it is independent example which is equivalently distributed. The above equation requires the integration to compute marginal likelihood which cannot be conducted analytically arbitrary more. Therefore, a local Gaussian approximation is used to the weights' posterior distribution. Then optimizing the

hyper parameters can be made with a re-estimation framework, re-verifying the posterior's mode in alternation till convergence.

### C. Retrieval as Identification

Nevertheless, categorization is more complicated than the regression case. Note that we do no have a completely conjugate hierarchical structure. How to settle this problem, think about the log marginal probability of the target data again, offer the input instance which can be written like this,

$$\ln P(T|X) = \ln \int \int P(T|X, w) P(w|\alpha) P(\alpha) dw d\alpha$$

The same as ever, we bring in a factorized variational posterior of the form  $Q_w(w)Q_\alpha(\alpha)$ , and acquire the lower bound as follow on the log marginal probability

$$\ln P(T|X) \geq Q_w(w)Q_\alpha(\alpha) \ln \int \left\{ \frac{P(T|X, w) P(w|\alpha) P(\alpha)}{Q_w(w)Q_\alpha(\alpha)} \right\} dw d\alpha$$

We can get the predictions from the trained model for new inputs by replacing the posterior suggest weights,. You can see the predictive distribution in the form of  $P(t | x, E[w])$ . To make an accurate estimation, we should take the uncertainty of weight into consideration by means of marginalizing over the posterior distribution over weights.

## III. EXPERIMENTAL RESULTS

In this part, we will evaluate our proposed MF-RVM approach for image retrieval [16]. The experimental steps of our proposed MF-RVM algorithm are graphically presented in the above section and Figure 2. It contains the following procedure: (1) data collection; (2) feature extraction via approach in Section 2; (3) train the model and evaluate its performance.

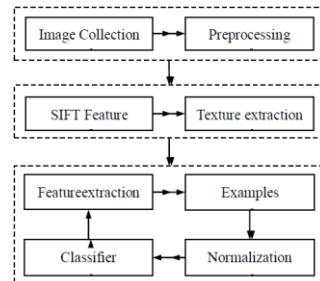


Figure 2. The flowchart of the proposed MF-RVM method

### A. Experimental Database

The Corel Image database used in our experiments was collected by means of Michael Ortega-Binderberger who is with University of California at Irvine. The original image collection was got from Corel. There are 68040 photo images from various categories. Each configured of features is stored in a divide file. For each file, a line corresponds to a single image. The first value in a line is the image ID and the subsequent values are the feature vector (e.g. color histogram, etc.) of the image. The same image has the same ID in all files whereas the image ID is not the same as the image filename. The database includes 68040 samples with 89 attributes, these

attributes describes as each image's four configures of features were extracted: Color Histogram, Color Histogram Layout, Color Moments, Co-occurrence Texture. Color Histogram: 32 dimensions ( $8 \times 4 = H \times S$ ); Color Histogram Layout: 32 dimensions ( $4 \times 2 \times 4 = H \times S \times$  sub-images); Color Moments: 9 dimensions ( $3 \times 3$ ); Co-occurrence Texture: 16 dimensions ( $4 \times 4$ ). The distribution of datum is presented in Table 1.

TABLE I. CLASS DISTRIBUTION OF SAMPLES

Classes	Number of samples
Car	40000
Bridge	5000
People	18040
Scenery	5000
Total number of samples	68040

### B. Assessment Standard

To assess the advantage of our proposed MF-RVM approach for image retrieval, and other compared algorithms for image retrieval, we in this paper select some of classification accuracy, recognition precision and recognition recall as the assessment criterions. The definitions can be determined in Table. TP denotes true positive that is the correct result; TP (true positive) represents items correctly labeled as belonging to the positive category; TN (true negative) represents items correctly labeled as belonging to the negative category; FP (false positive) denotes items incorrectly labeled as belonging to the category; and FN (false negative) represents items that were not labeled as belonging to the positive category whereas should have been. These assessment criterions can be directly for two classes or multiple class classification problem of image retrieval.

TABLE II. THE ASSESSMENT CRITERION FOR IMAGE RETRIEVAL

Evaluation standard	Definition
Precision	$TP / (TP + FP)$
Recall	$TP / (FN + TP)$

TABLE III. THE PERFORMANCE COMPARISON OF DIFFERENT ALGORITHM

Experiment	Approach	Precision (%)	Recall (%)
Trial 1	SIFT-SVM	75.29	77.83
	MF-RVM (ours)	79.81	82.39
Trial 2	SIFT-SVM	76.49	75.26
	MF-RVM (ours)	83.43	82.85
Trial 3	SIFT-SVM	75.56	76.55
	MF-RVM (ours)	80.55	80.26
Trial 4	SIFT-SVM	73.97	76.24
	MF-RVM (ours)	81.48	80.81
Trial 5	SIFT-SVM	73.98	75.79
	MF-RVM (ours)	79.60	81.40
Average	SIFT-SVM	75.12	75.81
	MF-RVM (ours)	80.38	81.29

### C. Main Results

In the first experiment, we assess our proposed MF-RVM method for image retrieval, over the Corel image dataset. We make use of two comprehensive criterions, Precision and Recall, for experimental verification. Identification accuracy and recall are two typical and popular measures for the correctness of the

identification model. The experimental step is summarized in the experiment part. The pre-processing procedure and feature extraction procedure are important due to the capture discriminant information. Our developed approach MF-RVM is trained utilizing above described algorithm, and some parameters of MF-RVM are got through cross-validation strategy. We do the test for multiple trials, where in each trial we randomly divide the dataset to training set and test set.

The Precision and Recall is utilized as the evaluation standard for the image retrieval. We conduct the experiment for 20 times and present the experimental results of partial time are in Table 3 and Figure 3. As report in Table 3, by means of utilizing our algorithm to learn parameter, MF-RVM for image retrieval reach the highest performance of 83.43% under the standard of Precision, while MF-RVM reach the highest performance of 82.85% under the criterion of Recall. Additionally, the average Precision of MF-RVM is 80.38% which outperforms that of SIFT-SVM (75.12%). The potential reasons for these results are mainly threefold. Firstly, The MF-RVM is capable to adapt complexly distributed data and deal with it well, where the adaptability essentially comes from the flexibility of the model parameters. Secondly, the parameter selection method is according to the distribution information of the input data to select the model parameters of the MF-RVM, which makes the MF-RVM has better adaptability. Thirdly, the processing procedure for data is able to remove noise and keep useful information effectively, and the element steps of our method could cooperate.

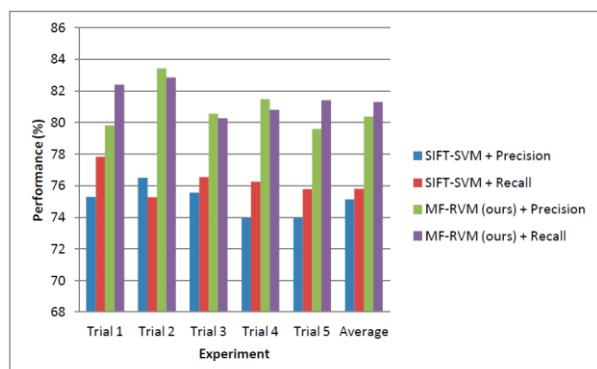


Figure 3. The comparison of experimented approach over two standard

This experiment is run over Corel image dataset. The dataset, collected by Michael, includes 68040 samples with 89 attributes for each image. This experiment aims to validate the ability of our proposed MF-RVM as well as its solution method in the task of image retrieval. The experiment procedure can be found in the above part of this paper, where the parameters of MF-RVM are found using the solution algorithm variational inference and cross-validation. We also compare MF-RVM with other related algorithms. The verification standards make use of here are Precision and Recall where Identification accuracy and recall are two typical and popular measures for the correctness of the identification model.

We conduct experiments over Corel image dataset. The dataset, collected by Michael Ortega-Binderberger, includes 68040 samples with 89 attributes for each image. This experiment will assess the capability of MF-RVM in image retrieval, and optimization. It employs the method show in above part to learn MF-RVM and cross-validation approach to select the parameters. The assessment criterions are precision and recall respectively where identification accuracy and recall are two typical and popular measures for the correctness of the identification model. The test was performed for 10 trials on this method, and the overall results of varying experimental configuration are present in Table 4 and Figure 4. As present in Table 4 and Figure 4, the value of Precision is around 77.77%, consistently beating the compared approach PLSA. Additionally, for varying experimental rounds, the Precision of our proposed approach also outperform other compared approach. These results are consistent with the previous work, which demonstrates which Precision is a reliable measure for image retrieval and MF-RVM. The reasons are from the following three aspects. Firstly, the MF-RVM has the ability to map the nonlinear data in the low dimensional space to the high dimensional space by a Kernel function, which makes the classification problem easy. Secondly, the parameter selection method is according to the distribution information of the input data to select the model parameters of the MF-RVM, which makes the MF-RVM has better adaptability. Thirdly, the framework of the proposed method MF-RVM contains a group of comprehensive procedures which sequentially maximize the performance.

TABLE IV. THE IDENTIFICATION RESULTS OF IMAGE RETRIEVAL ADOPTING MF-RVM

Training data	Verification standard	MF-RVM (ours)	PLSA
30%	Precision	68.59	65.35
	Recall	67.89	68.94
40%	Precision	75.30	73.14
	Recall	79.23	74.99
50%	Precision	77.77	77.82
	Recall	82.18	78.14
60%	Precision	81.59	81.33
	Recall	85.06	81.28
70%	Precision	85.04	78.91
	Recall	86.03	83.79

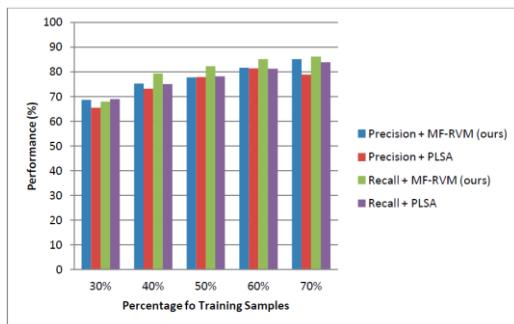


Figure 4. The classification results of image retrieval employing MF-RVM

In the third experiment, we verify the capability of the proposed MF-RVM method in image retrieval, using

comparison experiment. Two popular criterions Precision and Recall are employed for evaluation. Identification accuracy and recall are two typical and popular measures for the correctness of the identification model. The dataset is Corel image dataset. The experimental procedures are summarized in above part. Our approach MF-RVM is learnt by the approach in above section, where some parameters of MF-RVM are configured to defaults. The test is repeatedly done for several rounds over stochastically split database.

TABLE V. IDENTIFICATION PERFORMANCE COMPARISON OF FOUR ALGORITHM FOR IMAGE RETRIEVAL

Experiment trial	Approach	Precision (%)	Recall (%)
Trial 1	SIFT-SVM	75.29	77.83
	PLSA	78.30	80.29
	SC-SVM	77.45	81.03
	MF-RVM (ours)	79.81	82.39
Trial 2	SIFT-SVM	76.49	75.26
	PLSA	75.92	77.64
	SC-SVM	80.92	81.19
	MF-RVM (ours)	83.43	82.85
Trial 3	SIFT-SVM	75.56	76.55
	PLSA	78.26	77.27
	SC-SVM	77.34	78.55
	MF-RVM (ours)	80.38	81.29

We extensively compare our developed MF-RVM approach for image retrieval with three algorithm, SIFT-SVM, PLSA and SC-SVM. The classification results are show in Table 5 and Figure 5. These experimental results indicate which: (1) our proposed MF-RVM significantly beats all three compared approach, under varying experimental configurations, varying number of training sample, and different assessment criterion. (2) The proposed algorithm exhibit robustness against the trial of experiments, and the assessment criterion, that no wonder mean which the proposed algorithm could be used to a number of tasks. The reasons are three folds. (1) The MF-RVM method can be applied to the dataset of large scale and high dimension, and containing a large number of heterogeneous information. (2) The parameter selection method is according to the distribution information of the input data to select the model parameters of the MF-RVM, which makes the MF-RVM has better adaptability. (3) The framework of the proposed method MF-RVM contains a group of comprehensive procedures which sequentially maximize the performance.

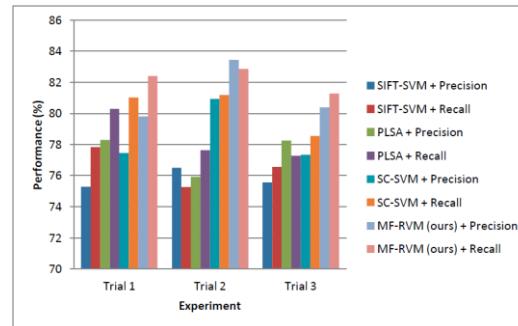


Figure 5. The performance comparison of experimented approach for image retrieval

#### IV. DISCUSSION

One image retrieval method based on multiple features and RVM is presented in this paper. Since RVM is a convex optimization problem, seen from the properties of convex optimization, local optimum solution solved with convex optimization must be global optimum solution, which is not included in other classification methods. In addition, RVM shows its effectiveness on image retrieval research in the aspects of independent study classification and automatic processing [1]. However, the application of RVM is affected by the parameters and its classification effect highly depends on parameters. For this reason, selection method for cross validation, which is an expansion of typical RVM, is presented in this paper. Moreover, this method can be well applied to quickly solve nonlinear problems, which makes RVM have a wide application prospect in solving recognition and retrieval problems. Retrieval experiment is conducted on COREL collected from internet.

Experimental results show that it can significantly improve the accuracy for image retrieval to obtain parameters with cross validation, so MF-RVM presented in this paper has very high practicability in image retrieval. Moreover, MF-RVM learning method can be well applied to conditions of large-scale sample data, complex dimension and large amount of isomerism information. MF-RVM learning method provides machine learning with wide application prospect and rich design thought in fields of feature extraction, multi-class objective detection and pattern recognition. Retrieval accuracy is highly dependent on selection the best wide y of parameter and penalty parameter c, so study on selection for kernel parameter is still the key point of the next study.

#### REFERENCES

- [1] XieCheng-Wang, Comparative study on various SVM algorithms, *Mini-MicroSystems*, Vol. 1, No. 1, pp. 56-58, 2008
- [2] Gong Zhi-Le, Zhang De-Xian. An improved text classification algorithm for SVM, *Computer simulation*, Vol. 7, No. 26, pp. 89-90, 2009
- [3] Gao Jin, Image classification based on SVM, *Master thesis from Northwest University*, Vol. 2, No. 10, pp. 125-1262010
- [4] Zhang Shu-Ya, Zhao Yi-Ming, Algorithm and implementation for image classification based on SVM. *Computer engineering and application*, Vol. 43, No. 25, pp. 156-157, 2007.
- [5] Zhang Bao-Hua, Application of decision rule classifier in network intrusion detection. *Computer engineering and application*, Vol48, No. 26, pp. 93-95, 2012
- [6] Haralick R M. Statistical and Structural Approaches to Texture, *Proceedings of IEEE*, Vol. 67, No. 5, 1957
- [7] Xiang Li, Xuan Zhan, A New EPMA Image Fusion Algorithm based on Contourlet-lifting Wavelet Transform and Regional Variance, *Journal of Software*, Vol. 5, No. 11, pp. 1200-1207, 2010
- [8] Tipping, Michael E, Sparse Bayesian learning and the relevance vector machine, *The Journal of Machine Learning Research*, No. 1, pp. 221-244, 2001.
- [9] Li Jing, Chao Shao, Image Copy-Move Forgery Detecting Based on Local Invariant Feature, *Journal of Multimedia*, Vol. 7, No. 1, PP. 90-97, 2012
- [10] Bobo Wang, Hong Bao, Shan Yang, Haitao Lou, Crowd Density Estimation Based on Texture Feature Extraction, *Journal of Multimedia*, Vol. 8, No. 4, pp. 331-337, 2013
- [11] Yin Bo, XiaJing-Bo, FuKai, The prediction research of the Network traffic based on the IPSO Chaos Support Vector Machine (SVM), *Computer application research*, Vol. 29, No. 11, pp. 4293-4299, 2011
- [12] Linhao Li, Content-Based Digital Image Retrieval based on Multi-Feature Amalgamation, *Journal of Multimedia*, Vol. 8, No. 6, pp. 739-746, 2013
- [13] Zemin Liu, Zong Wei, Image Classification Optimization Algorithm based on SVM, *Journal of Multimedia*, Vol 8, No. 5, pp. 496-502, 2013
- [14] Shoujia Wang, Wenhui Li, Ying Wang, Yuanyuan Jiang, Shan Jiang, Ruilin Zhao, An Improved Difference of Gaussian Filter in Face Recognition, *Journal of Multimedia*, Vol. 7, No. 6, pp. 429-433, 2012
- [15] Xiao-wei ZHU, Research on Automatic Classification Technology of Flash Animations based on Content Analysis, *Journal of Multimedia*, Vol. 8, No. 6, pp. 693-698, 2013
- [16] Xiang Zhang, Changjiang Zhang, Satellite Cloud Image Registration by Combining Curvature Shape Representation with Particle Swarm Optimization, *Journal of Software*, Vol. 6, No. 3, pp. 483-489, 2011

**Zemin Liu** is currently an associate professor in the College of Mathematics and Computer Science at the Panzhihua University of P. R. China. He received a bachelor degree of Engineering from Xi'an Jiao Tong University in 1987 and a master's degree from University of Electronic Science and Technology of China in 2007. He is in charge of teaching compiler principles, operating system. He hosts and participates in a number of projects and research topics. At present his research interests include: machine intelligence, computer graphic, digital image processing.

**Wei Zong** received the B.S. degree in Electrical Engineering and Computer Science, the M.S. and the Ph.D. degrees in Biomedical Engineering, from Xi'an Jiaotong University (XJTU), Xi'an, P. R. China, in 1983, 1986, and 1993, respectively. From 1993 to 1997, he was an associate professor at the Institute of Biomedical Engineering, XJTU. Since 1997, he has been affiliated with the Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology (MIT). His research interests include signal and image processing, pattern recognition, machine learning, data mining, and medical informatics.

# Low SNR Speech Recognition using SMKL

Qin Yuan

Suzhou University, Anhui, 234000, China

**Abstract**—While traditional speech recognition methods have achieved great success in a number of real word applications, their further applications to some difficult situations, such as Signal-to-Noise Ratio (SNR) signal and local languages, are still limited by their shortcomings in adaption ability. In particular, their robustness to pronunciation level noise is not satisfied enough. To overcome these limitations, in this paper, we propose a novel speech recognition approach for low signal-to-noise ratio signal. The general steps for our speech recognition approach are composed of signal preprocessing, feature extraction and recognition with simple multiple kernel learning (SMKL) method. Then the application of SMKL in speech recognition with low SNR is presented. We evaluate the proposed approach over a standard data set. The experimental results show that the performance of SMKL method for low SNR speech recognition is significantly higher than that of the method based on other popular approaches. Further, SMKL based method can be straightforwardly applied to recognition problem of large scale dataset, high dimension data, and a large amount of isomerism information.

**Index Terms**—SMKL; SNR; Speech Recognition

## I. INTRODUCTION

Computer based automatic speech recognition started from 1950s. Given the human speech signal for certain language, the target of computer based speech recognition is to transform speech signal into the corresponding text or command through recognition and comprehension process [1, 4]. By means of computer based automatic speech recognition system, human and computer are able to communicate with each other in speech, namely, make computer accurately recognize the content of speech in a variety of situations and circumstances, even understand human language, in order to be applied by human. With the development of computer science and technique, artificial intelligence techniques, computer speech recognition technique has developed into an independent field attracting a number of researches from different fields [2, 3, 5]. Speech recognition and its related techniques can be widely used in automatic control, communication and electronic system, information processing and so on. In a word, the ultimate goal of computer based speech recognition technique is to make machine understand human language, realizing language communication between human and computer.

The first speech recognition system that recognizes isolate English and number was set up by the staff of Bell laboratory in 1952 [3, 6]. In 1960s, computer application

promoted the development of speech recognition. Moreover, linear predictive coding (LPC) and digital processing (DP) techniques were developed for speech signal. 1970s have seen the development and success of LPC technique and hidden Markov model (HMM) [7, 9]. In 1980s, a number of speech recognition algorithms for continuous signal were developed, where the speech recognition algorithm had an evolution from template matching technique to statistics model based technique [10, 11]. More specifically, HMM, as a breakthrough, achieved the state-of-the-art performance for continuous speech recognition. Then, parameter extraction and optimization, elaboration of model design and the system's adaptive technique got great progress, making the speech recognition technique matured and beginning bring products to market [2]. In recent years, due to the development of computer technique and the wide application of internet, speech recognition greatly affects the industrial manufacture and daily life [12].

Usually, current speech recognition approaches are able to reach satisfied performance for signal with high SNR. However, for speech signal with low SNR, those methods are easy to be failed. It is worth noting that, in real application, speech signal mostly with low SNR which means that these methods are difficult to satisfy the real world applications [9, 11]. For this reason, how to improve speech recognition accuracy of low SNR signal to meet users' needs has become increasingly important. Recently, research methods of recognition under low SNR mainly focus on traditional speech recognition model, such as artificially neural network (ANN) and deep learning, which are both based on statistical theory, have very high requirement of data regularity and shall be conducted under the assumption of infinite sample size. However, they show good performance when the training data is enough, and are still not robust enough to deal with the real world speech signal with low SNR, small samples.

Researches show that it cannot only get better mapping performance but also process isomerism or various data often existing in typical learning problem to combine different kernel functions. At this moment, multiple kernel method can show more flexible and effective processing ability. In addition, it can be also used as one ingenious analytical method to explain learning results, so that application problem can be understood more deeply and accurately. Moreover, synthetic kernel method is just a typical learning method in multiple kernel learning. This type of multiple kernel learning method is mainly achieved through linear combination of

various kernel functions. The structure diagram of its component is shown as Figure 2. The detailed description of multiple kernel learning can be found in Section 2.

For this reason, on the basis of analyzing relevant research achievement, on account of limitations and shortcomings of traditional speech recognition model, a deep searching about speech recognition under low SNR is conducted in this paper. What's more, according to the functions and features for speech recognition, the general method step for speech recognition is analyzed and the application in speech recognition under low SNR of simple multiple kernel (SMKL) method is presented to improve speech recognition accuracy, so that methods for speech recognition under low SNR can be enriched in theory. Further, SMKL method can be well applied to conditions of large-scale sample data, complicated dimension and a great deal of isomerism information.

Our proposed SMKL method provides machine learning with widespread application prospect and rich design thought in the field of feature extraction, multi-class object detection and pattern recognition. In addition, SMKL method can get more effective recognition performance through combining different kernel functions according to certain rules. Meanwhile, classical machine learning approaches usually encounter problems, such as isomerism or data with complicated kinds, so it will be more reasonable selection to consider SMKL method under complicated conditions [1-3]. The proposed approach in this paper contains three procedures. (1) Preprocess the input speech signal. (2) Extract features for the preprocessed speech, and then match the extracted speech signal with the speech model in computer. (3). Output the matching results or transform them into specific instructions. Experimental results show that: (1) the accuracy of our SMKL method recognizing speech under low SNR is obviously higher than that in the case of SVM. (2) Our SMKL method can be well applied to the conditions of large-scale sample dataset, high dimension and a large amount of isomerism information. SMKL method provides widespread application prospect and rich design thought in the field of feature extraction, multi-class object detection and pattern recognition.

The main contributions of the proposed approaches are threefold. (1) It extracts multiples features from the speech signal for speech recognition, which improve the robustness significantly. (2) It seamlessly integrates the speech recognition problem with SMKL through using multiple kernels for multiple features. (3) This approach is computationally effective. These advantages will be verified in the experiment section. The remainder part of this paper is organized as follows. We first present the SMKL based low SNR speech recognition approach in Section 2, and then conduct the experiments in Section 3. Section 4 draws a conclusion of the whole paper.

## II. THE PROPOSED SCHEME

In this section, we will propose the SMKL based speech recognition approach. This approach is composed of five main procedures, signal collection and processing, feature extraction, model training, recognition and

evaluation, followed by a feedback procedure. The framework of the proposed approach can be found in Figure 1. In this section, we mainly introduce the main model, which contains training model and performing recognition. The rest procedures will be introduced in section 3.

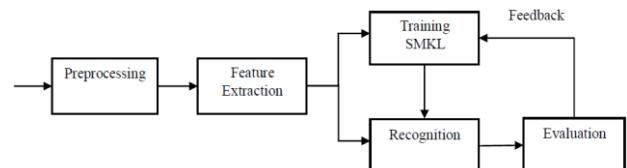


Figure 1. The framework of the proposed speech recognition approach

Now we will propose the idea of MKL and derive its dual form. Let  $i$  and  $j$  index the samples and  $m$  index the kernel. To reduce the use of symbol, we simply set that values of  $i$  and  $j$  are both from 1 to  $l$ , and the value of  $m$  is from 1 to  $M$ . Under linear separable circumstance, SVM directly uses hyper-plane for classification. But it is nonlinear when processing most problems, then kernel function shall be used for phenotype switching, transforming primal data  $j$  from low-dimensional space to high-dimensional space, so that the linear separable goal can be obtained.

Primal problem for SMKL and its solution. The alternating optimization algorithm presented by Grandvaletis possible to solve problem the simple form of multiple kernel learning. First, for the optimization of problem,  $f_m, b, \xi$ , and  $d$  are fixed. Next, weight vector  $d$  is the objective function updated to decrease problem,  $f_m, b, \xi$ , and  $d$  are fixed. In this section, we show that the second step can be achieved with closed form. However, this method is short of convergence guarantee and possible to cause number problem, especially some elements in  $d$  are closed to 0. Note that these number problems can be solved through importing substitute algorithm of disturbance processing version, as shown by Argyriouin 2008.

We consider the following constraint optimization problems if another optimization algorithm are used:

$$J(d) = \begin{cases} \min_{\{f_m\}, b, \xi} \frac{1}{2} \sum_m \frac{1}{d_m} f_m^2 + C \sum_i \xi_i & \forall i \\ s.t. y_i \sum_m f_m(x_i) + y_i b \geq 1 - \xi_i & \\ \xi_i \geq 0 & \forall i. \end{cases} \quad (1)$$

Then we will show that how to solve Equation (1) with simple gradient method. First, we note that objective function  $J(d)$  is actually the objective value for an optimum SVM. We can discuss and calculate  $J(*)$  with gradient and its kernel is just this method. We have to process Equation (1), which is a nonlinear object function constraint of simplex. Once  $J(d)$  is solved out,  $d$  is updated according to decrease direction, ensuring that this equality constraint and non-negative constraint  $d$  are

met. Equality constraint for gradient treatment can be decreased according to Luenberger's work in 1984. Let  $d_\mu$  be a non-zero entrance of  $d$ ,  $J(d)$  and  $\nabla_{\text{red}} J$  whose gradients are simplified have components:

$$\begin{aligned} [\nabla_{\text{red}} J]_m &= \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu}, \quad \forall m \neq \mu \\ [\nabla_{\text{red}} J]_\mu &= \sum_{m \neq \mu} \left( \frac{\partial J}{\partial d_\mu} - \frac{\partial J}{\partial d_m} \right) \end{aligned} \quad (2)$$

To get better numerical stability, we select  $\mu$  as the index of vector  $d$  of the largest component. This positive constraint is also considered at decrease direction. Because we want to reduce  $J(\cdot)$ ,  $-\nabla_{\text{red}} J$  is a decrease direction, if there is an index  $m$  making  $d_m = 0$  and  $[\nabla_{\text{red}} J]_m > 0$ , it will violate positive  $d_m$  of the constraint to use this direction. For this reason, the component is set as 0 at decrease direction.

$$D_m = \begin{cases} 0 & \text{if } d_m = 0 \text{ and } \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu} > 0 \\ -\frac{\partial J}{\partial d_m} + \frac{\partial J}{\partial d_\mu} & \text{if } d_m > 0 \text{ and } m \neq \mu \\ \sum_{g \neq \mu, d_g > 0} \left( \frac{\partial J}{\partial d_g} - \frac{\partial J}{\partial d_\mu} \right) & \text{for } m = \mu \end{cases} \quad (3)$$

The general update plan is  $d \leftarrow d + \gamma D$ , where  $\gamma$  is step length. Once  $D$  at one decrease direction has been solved out, we first need to look for the maximum allowable step length at that direction; then check whether the objective value is reduced. If the objective value is reduced and  $d$  is updated, we can set  $D_v = 0$  and normalize  $D$  to observe equality constraint. This process will be repeated until the objective value stops reducing. At this point, we can look for the optimum step length  $\gamma$ , one dimensional is used to ensure the global convergence through suitable stopping criterion, such as Armijo.

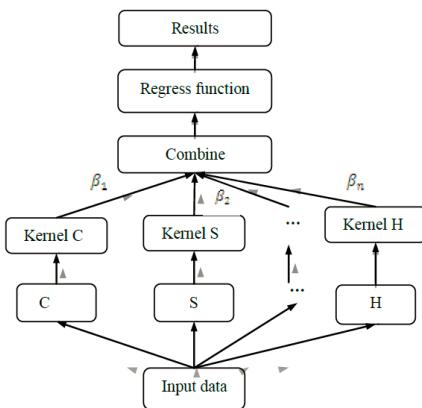


Figure 2. Sketch map for composition of kernel function

In this algorithm, it's the objective function  $J(\cdot)$  based on evaluation to calculate decrease direction and linear

search, it needs to solve a problem of SVM. It seems to be time consumed, but it's very fast to learn for small change; meanwhile, the previous value  $\alpha^*$  is used when solution of SVM is initialized. Please note that cost function for gradient is not the calculating weight vector  $d$  after each updating. Instead, as long as the objective value is decreased, it will be easier to take use of a updated decrease direction for us. Comparing with general updating plan, this method can save a large amount of calculating time, and the decrease direction is  $d$  after each updating. Note that we also investigate gradient project algorithms, the results show that efficiency for gradient project algorithms are slightly lower than that of proposed method, so we will not report these results.

The algorithm will stop when meeting stopping criterion. Our operation is based on dual gap, which will be introduced in remaining part in detail. The process for SMKL algorithm is shown as below: first,  $n \leftarrow 0$ ; then randomly initialize  $d^0$ ; repeat the above two steps. Let  $K \leftarrow k(d^n)$ , so that SVM can be selected to solve single kernel problem with kernels  $K$  and  $\alpha^*$ , achieving

$$d^{n+1} \leftarrow d^n - s^n \left( \frac{\partial r}{\partial d_k} - \frac{1}{2} \alpha^* \frac{\partial H}{\partial d_k} \alpha^* \right).$$

Repeat  $n \leftarrow n+1$  until it converges.

### III. EXPERIMENT

Flow diagram of experiment is shown as below: (1). Preprocess the input speech. (2). Extract features of the preprocessed speech, and then match the extracted speech signal with speech model in computer. (3). Output the matching results or transform them into specific instructions, as shown in the following figure. Preprocessing for speech signal mainly includes filtering, weighting, short-time windowing processing and terminal detection. Then extract then features of speech signal. Then save these feature data as specific feature file, being the basic data of multiple kernel learning. In this experiment, we choose two popular kernels. (1) Summation kernel  $k(x, z) = \sum_{j=1}^M k_j(x, z)$ . (2) Extended weighting polynomial  $k(x, z) = \alpha k_1^p(x, z) + (1-\alpha)k_2^q(x, q)$

#### A. Data Sources and Preprocessing

Training data in experiment of this paper is taken from 8 people's pronunciations in different SNR (0,5,10,15,20,25 db), and the collected speech samples are English pronunciations of words respectively consisted of 5, 10, 15, 20 English letters, in addition, each person pronounces different words for 4 times in various SNR. Under each type, there are respectively 192( $8 \times 4 \times 6 = 192$ ) corresponding training samples. 10.75Hz is selected as sampling rate of the signal during experimental process. Also, pronunciations of 5 persons in various SNR are selected as test recognition sample. Data preprocessing mainly includes pre-weighting of collected speech signal, in which filtering processing of

data is finished with one transfer function. Then operations, such as windowing and framing, shall be conducted to data, which are finished with Hamming window. For data, feature extraction shall be conducted after preprocessing, in order to better classify data.

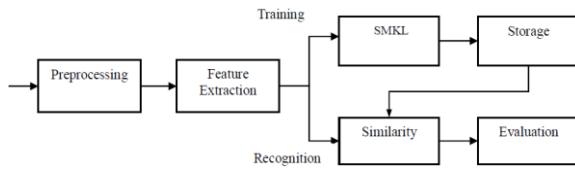


Figure 3. Flow diagram of experiment

### B. Feature Extraction

It is mainly that extract feature parameters reflecting speech essence from primal speech signal for data feature extraction, forming proper vector sequence. The selectable speech feature parameters include time domain parameters, frequency domain parameters. Time domain parameters consist of short-time average zero-crossing rate, short-time average energy, and pitch period and so on. Short-time average zero-crossing rate and short-time average energy are usually used to detect speech terminals while pitch period is used to distinguish tone and unvoiced or voiced sound for Chinese characters. Frequency domain parameters include spectrum (MFCC and LPCC), short-time frequency spectrum (DFT and average spectrum of 10-30 channel filter banks)and the first three formants.

People's auditory system is a specific nonlinear system and has different responds to signals with various frequencies, mostly logarithm relationship. LPCC factors are based on synthetic parameters and do not take use of hearing characteristics of people's ears, MFCC parameters combine generation mechanism of speech with hearing perception characteristic. For this reason, MFCC with better performance is selected as feature parameters of speech. The solution of MFCC is shown in the following Figure 4.

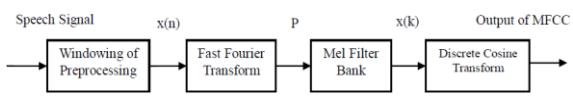


Figure 4. Solving process of MFCC

Speech recognition can be grouped into narrow and broad speech recognition, the former means one technique extracting text content from speech signal while the latter means one technique extracting any interesting content. There are many classification methods for speech recognition, mainly including classification methods according to vocabulary, manner of articulation, speaker and recognition method; classification method according to vocabulary is applied in this paper. The vocabulary can be divided into small vocabulary, middle vocabulary and large vocabulary. Generally speaking, 10-100 entries is small vocabulary, that 100-500 entries is middle vocabulary and that more than 500 entries is large vocabulary. Usually, with the increase of vocabulary, recognition accuracy of speech

recognition will decrease, so with the increase of vocabulary, the degree of difficulty for the study on speech recognition also gradually increases.

### C. Selection of Kernel Parameters

We have to select proper kernel functions, kernel parameters and high-dimensional mapping spaces when classifying, so that we can get the separator with excellent learning and generalization ability. Error penalty factors  $C$  and  $\sigma$  are the key factors for SMKL [6], so these parameters have great effect on classification precision and generalization ability for SMKL.

TABLE I. SELECTION FOR KERNEL PARAMETERS

1: Select the scope of parameters, such as $C$ and $\sigma$
2:A coarse search is conducted for the grid
3:Construct two-dimensional grid
4:Calculate mean values of accuracy rate for parameters in various group when predicting
5:A subtle search is conducted for the grid

Grid search method is introduced to optimize and select kernel parameters  $C$  and  $\sigma$  when selecting SMKL parameters. This method can be simply operated and easily understood. The procedure of kernel parameters selection is shown in the following Table 1. Note that, during step 1,  $C$  and  $\sigma$  are generally selected; during step 2, the step length is usually set 2 when a coarse search being conducted. The goal of step 5 is to make the obtained search result be more accurate. Generally speaking, areas with high predicting accuracy are selected, which is equivalent to decrease step length to conduct a binary search.

### D. Experimental Result and Analysis

Features are input SMKL classifier to conduct speech recognition experiment under low SNR after sample feature being extracted. In this paper, grid search method is used to determine kernel parameters, for details, see section 3.3. SVM method and SMKL method are respectively used to conduct experiment, under each condition. The experiment is repeated for 20 times, then average recognition accuracy for each experiment as the final experimental results.

In the first experiment, SVM method and SMKL method are compared to verify the effectiveness of SMKL method in speech recognition under low SNR. In each experiment, 2/3 of each kind of data samples can be randomly selected as training sets, and remaining 1/3 can be test sets. Grid search method is used to determine kernel parameters during experimental process. The experiment is conducted under SNR of 0.5, 10, 15, 20, 25 and word number of 5, 10, 15, 20, and experimental results are shown as Table 2. Experimental parameters are error penalty factors  $C$  and  $\sigma$ .

The average recognition accuracy of various word numbers under the same SNR. Seen from Table3, recognition rates of methods proposed in this paper under different SNR are obviously higher than that of SVM method. The recognition accuracy of the proposed methods in this paper are between 84.72%~88.30%,

TABLE II. LOW SNR SPEECH RECOGNITION RESULTS

SNR/dB	Method	Letter number of words				Average recognition (%)
		5	10	15	20	
0	SVM	81.23	83.21	79.36	76.32	80.03
	SMKL	83.26	84.36	89.23	89.23	86.52
5	SVM	83.45	81.65	81.36	84.23	82.67
	SMKL	86.14	82.12	82.16	94.62	86.26
10	SVM	84.12	83.45	79.65	79.32	81.65
	SMKL	85.05	86.45	81.23	86.15	84.72
15	SVM	81.36	82.13	84.36	74.12	80.49
	SMKL	84.23	85.56	85.26	98.15	88.30
20	SVM	81.96	82.16	79.36	85.63	82.28
	SMKL	82.01	86.23	81.56	98.36	87.04
25	SVM	84.32	83.56	78.56	82.14	82.15
	SMKL	85023	86.21	81.23	94.32	86.75

TABLE III. PERFORMANCE OF LOW SNR SPEECH RECOGNITION USING SMKL

Experiment	Accuracy	Area under curve (AUC)
Run 1	86.24%	0.9184
Run 2	85.21%	0.9270
Run 3	86.12%	0.9459
Run 4	81.32%	0.9672
Run 5	88.69%	0.9187
Run 6	94.69%	0.9542
Run 7	89.17%	0.9564
Run 8	90.85%	0.9225
Run 9	88.14%	0.9433
Run 10	86.79%	0.9455
Average	87.72%	0.9395
Std	1.59%	0.0460

mostly around 86%, while recognition rates of SVM method are mostly around 81%. Reasons for the above results include the following three aspects: First, SMKL exploits the ability of the nonlinear feature mapping through learning the implicit feature space. By means of learning nonlinear feature mapping, the classifier could adapt to data distribution well. The adaptability of SMKL comes from the flexibility and the adaptability of the parameters of kernels of SMKL. Second, the selection method for kernel parameters can adapt to dataset much better, in comparison with empirical parameter selection. Third, SMKL can be applied to the condition that sample data is large-scale, high dimension data, and data containing a large number of heterogeneous information.

The second experiment is to test the effectiveness of the SMKL algorithm. This research divided the data set into 10 groups by using the 10 groups cross validation ways. Every group took turns as the test samples, and the others are as the training samples. Then we can get 10 groups different training samples and test samples. The experiment tool ten times, then we can get ten groups classification accuracy. At last we can get the average classification accuracy of the ten experiments. Experimental parameters are error penalty factors  $C$  and  $\sigma$ . During the experiment it put the recognition accuracy and AUC as the evaluation standards. The two indicators have monotonic relations. Although variability of AUC is greater than that of accuracy, AUC is more reliable than accuracy; the main reason is that accuracy is just the performance pointer when  $TH = 0.5$  while AUC is the average performance pointer of all possible TH.

From the experiment results in Table 3 and Figure 4, we can see the recognition accuracy of each experiment is lower than that of AUC. The average recognition accuracy of the ten experiments is 87%, but in AUC it can be as high as 92%. The reasons are mainly from the following two aspects:(1) compared with traditional machine learning methods, SMKL can be applied to the condition that the dataset is highdimension, containing a large number of heterogeneous information. (2) The adaptability of SMKL comes from the flexibility and the adaptability of the parameters of kernels of SMKL, which can make get the more effective mapping performance. However, for the classical machine learning problems, it can often encounter the heterogeneous or the complex data, so it showed the high AUC and little standard.

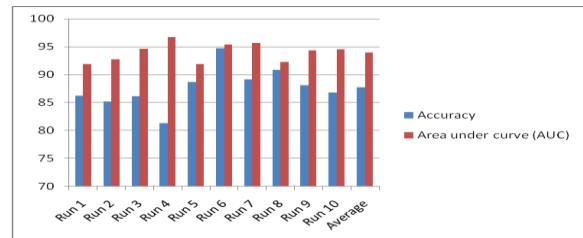


Figure 5. Recognition performances for low SNR signal

In the third experiment, we evaluate the kernel parameter selection approaches. Grid search method is used to determine kernel parameters in this paper when recognizing with SMKL method. In above experiments, kernel parameters are obtained by experience and determined based on grid search method, so that the effectiveness of grid search method determining the kernel parameters can be verified. The fixed parameter determination method and grid search method can be used to determine kernel parameters for conducting experiment, in each experiment, 2/3 of images can be randomly selected as training sets, and remaining 1/3 can be test sets. The average training times in Table 4 are the average times under different SNR. The experimental parameters are error penalty factors  $C$  and  $\sigma$ . In this experiment, average relative error, error of mean square and average training time are used as criteria evaluating experimental results.

From the experimental results, average relative error, error of mean square and average training time under grid

TABLE IV. CLASSIFICATION PERFORMANCE COMPARISON FOR DIFFERENT KERNEL PARAMETERS

Parameters selection Assessment criteria	Experience based kernel parameters selection	Determine kernel parameters with grid search
MAPE	9.3875	4.1902
MSE	429.12	386.91
Training time (second)	371.12	296.45

search method are respectively 4.1902 seconds, 386.91 seconds and 296.45 seconds, and the corresponding values under the fixed kernel parameter method are respectively 9.3875 seconds, 429.12 seconds and 371.12 seconds. So we can see that it expresses huge superiority in aspects of average relative error, mean square error and average training time to conduct experiment based on determining kernel parameters with grid search method. Reasons for the above results are twofold. (1)The kernel parameters for SVM are selected according to distributed information of input data, which can make kernel parameters for SVM have better adaptability than default kernel parameters. (2) Determine parameters of kernel function with grid search method, which can better map features from low-dimensional space to high-dimensional space, being propitious to improve recognition accuracy.

#### IV. CONCLUSION

On the basis of analyzing relevant research works, on account of limitations and shortcomings of traditional speech recognition model, a deep searching about speech recognition under low SNR is conducted in this paper. What's more, according to the functions and features for speech recognition, the general method step for speech recognition is analyzed and the SMKL method is presented to improve speech recognition accuracy for low SNR speech recognition, so that methods for speech recognition under low SNR can be enriched in theory. SMKL method can be well applied to conditions of large-scale sample data, complicated dimension and a great deal of isomerism information. SMKL can be widely applied to the field of feature extraction, object detection and recognition. In addition, SMKL method can archive higher recognition performance by combining different kernels according to certain regulations. Meanwhile, classical approaches can usually encounter problems, such as isomerism or data with complicated kinds, so it will be more reasonable selection to consider MKL method under complicated conditions [1-3].

Experimental results show that: (1) the accuracy of SMKL method under low SNR is obviously higher than that in the case of SVM. (2) SMKL method can be well applied to the conditions of large-scale sample data, complicated dimension and a large amount of isomerism information. SMKL method provides machine learning with widespread application prospect and rich design thought in the field of feature extraction, multi-class object detection and pattern recognition. Moreover, it will take more calculated amount and long training time, so it's the next research direction of this paper to improve the model's training efficiency and the detection

performance of algorithm under the circumstance of small training sample.

#### REFERENCES

- [1] Qin Yuan, "Efficient Audio Recognition Algorithm based on Simple Multiple Kernel Learning", *Journal of Multimedia*, Vol. 9, No. 1, pp. 59-66, 2014
- [2] Yang Shien, Chen Chunmei, "Study on network attack detection system based on SVM", *Yangtze University Journal (Science version)*, Vol. 8, No. 8, pp. 89-92, 2011
- [3] Yang Xiaofeng, Sun Mingming, Hu Xuelei, "Network attack detection method based improved hidden Markov model", *Journal On Communications*, Vol. 31, No. 3, pp. 126-129, 2010
- [4] Zhaofeng Li, Xiaoyan Feng, "Near Duplicate Image Detecting Algorithm based on Bag of Visual Word Model", *Journal of Multimedia*, Vol. 8, No. 5, pp. 557-564, 2013
- [5] Mika S, Ratsch G, Weston J, Scholkopf B, Mullers K R, "Fisher discriminant analysis with kernels. In: Proceedings of the Conference on Neural Networks for Signal Processing", Washington D. C., USA: IEEE, pp. 41-48, 1999.
- [6] ManikVarma, BodlaRakeshBabu, "More Generality in Efficient Multiple Kernel Learning", *the 26th International Conference on Machine Learning*, Canada, 2009.
- [7] Guo Chao, Zhang Xueying, Liu Xiaofeng, "Application of SVM in speech recognition with low SNR", *Computer engineering and application*, Vol. 49, No. 5, pp. 213-215, 2013
- [8] Deng Naiyang, TianYingjie, "Theory, algorithm and prolongation of SVM", Beijing: Science Press. 2009.
- [9] Zhou Shuge, "Study on speech recognition algorithm", Nanjing: Nanjing University of Science and Technology, 2004.
- [10] Vapnik V N, "Statistical Learning Theory", New York: John Wiley and Sons, 1998.
- [11] Xueying Zhang, Jing Bai, Wuzhou Liang, "The Speech Recognition System Based On Bark Wavelet MFCC", *8th International Conference on Signal Processing*, pp. 16-20, 2006
- [12] Du Chengjin, Sun Dawen, Patrick Jachman, et al, "Development of a hybrid image processing algorithm for automatic evaluation of intramuscular fat in beef M. longissimusdorsi", *Meat science*, Vol. 80, No. 4, pp. 1231-1237, 2008
- [13] Guangming Zhang, Zhiming Cui, Pengpeng Zhao, JianWu, "A Novel De-noising Model Based on Independent Component Analysis and Beamlet Transform", *Journal of Multimedia*, Vol. 7, No. 3, 247-253, 2012
- [14] Gang Liu, Jing Liu, Quan Wang, Wenjuan He, "The Translation Invariant Wavelet-based Contourlet Transform for Image Denoising", *Journal of Multimedia*, Vol. 7, No. 3, pp. 254-261, 2012
- [15] Yan Zhao, Hexin Chen, Shigang Wang, Moncef Gabbouj, "An Improved Method of Detecting Edge Direction for Spatial Error Concealment", *Journal of Multimedia*, Vol. 7, No. 3, pp. 262-268, 2012

# Electronic Commerce Data Mining using Rough Set and Logistic Regression

Li Xiuli, Zhao Rui, and Xiao Yan

Hebei Normal University of Science & Technology, Finance & Economics College, Qinhuangdao, Hebei Province, 066004

**Abstract**—Electronic commerce (E-commerce) has gradually been the mainstream of business. There may be some unpredictable but frequent problems such as delay in shipment, shipping errors caused by E-commerce participants' low efficiency. These problems will have negative impact on the business of participants eventually. Correct evaluation of the efficiency of E-commerce is an important way to improve operations. This paper introduces the knowledge discovery theory of data mining-based on Rough Set Theory (RST) to deal with the vague and inaccurate information about the evaluation of supplier and mine the law knowledge that exists between input variables and adverse position. The output of RST is then used as the feature and is delivered to the Logistic Regression (LR) to rank the product of electronic commerce website. The proposed approach, termed as RST-LR, is composed of the procedure of attribute values discretization; filtration processing of minimum attributes sets; evaluation rule; calculating the ranking accuracy and the establishment of evaluation systems. We evaluated the proposed approach on a real world dataset. The experimental results show that it achieves a high accuracy, and the rule has met the requirements of application.

**Index Terms**—E-Commerce; Service Quality; Rough Set Theory (RST); Logistic Regression; Efficiency Evaluation

## I. INTRODUCTION

According to the fact that wide application of computer networks and the rapid development of the Internet, more and more people access the internet and the great improvement of internet transaction security technology, E-commerce has gradually been accepted by the public and it has become a mainstream business model [1, 2]. At the same time, through the application of Business-to-Business (B2B) and (Business-to-consumer) B2C in e-business practices, many companies realize their dreams about opening a store on the internet, by real-time online payments, supply chain management (SCM) and other mechanisms [1, 3, 4]. They can manage their logistics and funds efficiency, and provide a safe and substantial Internet trading environment and in turn attracts more people to shopping online. Global E-commerce online shopping market is rising year by year [1]. The application and research of Web data mining technique in E-commerce attracts widespread attention, and more and more scholars begin to apply Web data mining techniques to E-commerce. Using data mining techniques to mine E-commerce data has tremendous

commercial value. It benefits for customers' segment and can realize personalized services to customers. This in turn can increase overall management chain, returning investment rate of industry chain, business models, products and service innovation. Web Data Mining aims to dig out hidden, unknown and potentially useful information from a lot of materials [5,6]. Web data mining has different functions, such as identification, prediction, association, clustering and so on. Many researchers use data mining in different fields to solve different problems. In particular, there are also many researchers use Web data mining to solve quality issues. For example, work [2] uses the two algorithms, self-organization map (SOM) neural networks and rule induction for data mining. The results indicate that a process step and the quality of materials used are the major factor that decreases the excellent rate and significantly improve the quality of transistor collector and emitter when the leakage current is excessive. Work [3] uses the methods of Kruskal-Wallis verification and decisiontree to develop manufacturing diagnostic data mining according to electronic materials testing and related process information. There are also many studies on the security of Web data mining about electronic transactions [7, 8]. So far, there is very few related research on Data Mining in E-commerce service quality diagnostic systems. This paper attempts to use rough set theory for Web data mining.

Data mining is an interdisciplinary topic. According to the different requirements of specific tasks, we can use statistical methods, neural networks, online analytical processing, genetic algorithms and decision trees, rough sets and so on [1, 3, 6, 10]. (1) Statistical analysis methods. Statistical analysis methods mainly refer to the relevant statistical methods for data correlation analysis, regression analysis, cluster analysis and principal component analysis. In practice, we usually use SPSS for Web data mining. The generally procedure is normalizing the data firstly. Standardized data can be operated directly in SPSS, such as cluster analysis can use Classify menu. We can make many customizations on the output results, including analysis, graphics, and text etc. (2) Neural network based methods. Neural network is similar to the data processing method which caused by simulating human brain neurons process information, which is more common Back Propagation (BP) neural network. Through the network learning function, we can find a suitable

weight value to obtain the best results. BP is the more common neural network. BP neural network algorithm was proposed in 1986 [11]. Neural networks have different network modes. In general, according to the division of the network structure, it can be divided into forward neural networks and feedback neural network [12]. (3) Online processing method. Online processing is different with traditional Online Transaction Processing (OLTP) which is primarily used to deal with transaction processing, such as civil aviation calibration system, the bank's storage systems and so on [13]. These operations mainly meet the requirements of high-volume, simultaneous and quick response. However, it is mainly used for data analysis and decision supporting. (4) The genetic algorithm (GA) [14] is an evolution algorithm, its basic principle is simulate the biological evolution laws, encode the parameters of the problem, label them as chromosomes and then use an iterative approach to selection, crossover and mutation, and ultimately make the generating chromosomes meet optimization goals. (5) Decision tree algorithm [15] is a kind of inductive learning methods. Decision tree method can predict unknown outputs. The decision tree is similar with the conventional tree. That is, it has nodes and leaves, and each node can be arranged a suitable test, and according to the test results, we can determine which node as the identification criteria can be used for further decisions until we achieve our target. In essence, the decision tree can be considered as a Boolean function, where the inputs of the function are a set of objectives and conditions which has the same property, and its output is binary. In a decision tree, each internal node is mapped to a corresponding property's test, and the branch emanated by node represents the possible values.

Rough Set Theory (RST) was proposed to deal with the ambiguity in the 1980s [4]. The main idea of RST is using data to collect the corresponding rules and provide effective assistance for decision [6]. The rough set theory overcomes the limitations of the previous approaches and can potentially be applied to our task. Further logistic regression is a popular approach for both identification and regression. We in this paper cast the product ranking problem as a identification problem.

The main purpose of this paper is for the service quality and efficiency evaluation of E-commerce, and we introduce RST theory in the course [7-8]. By using the effective mining knowledge rule of RST and logistic regression, we can provide appropriate advice for the service quality and efficiency of E-commerce eventually. In the experiment, there are many procedures as follows: data collection, data preprocessing, discrimination, attribute reduction, reduction filtering, rule generation, rule filtering, identification prediction, accuracy calculation and diagnostic system. And the results show that web data mining can greatly improve the service quality and efficiency when used in r-commerce.

The contributions of this work are threefold: (1) to our best knowledge, the rough set approach is firstly applied to the efficiency evaluation of EE-commerce; (2) we modify the rough set approach so that it can well adapt to

our application; (3) the proposed approach shows very compleutive performance and reasonable results [9-12].

## II. PROPOSED SCHEME

The main purpose of this paper is to evaluate the service quality and the service efficiency of E-commerce, and we introduce RST theory in the course. By using the effective mining knowledge rule of RST, we can provide appropriate advice for the service quality and efficiency of E-commerce eventually [13]. In the identification predict ion problems, rough set first provides a set of data with training results, these data are stored in relational tables, and the decision tree can obtain identification rules according to the data, and then guide and predict future data results by using the identification rules. Web data mining mainly include the following aspects: data base, server, knowledge base, Web data mining algorithms, model assessment, and user interface and so on. Our general framework for Web data mining is shown in Figure 1.

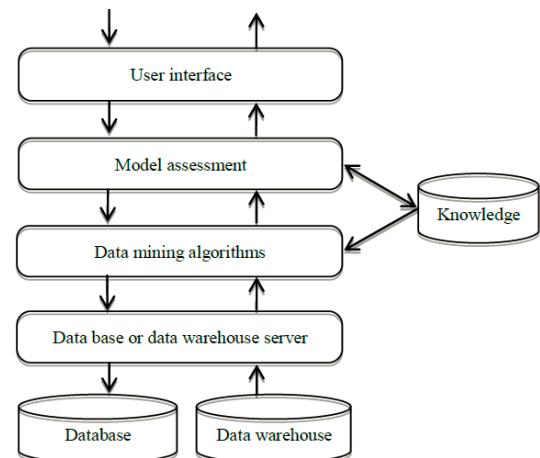


Figure 1. The general framework of Web data mining

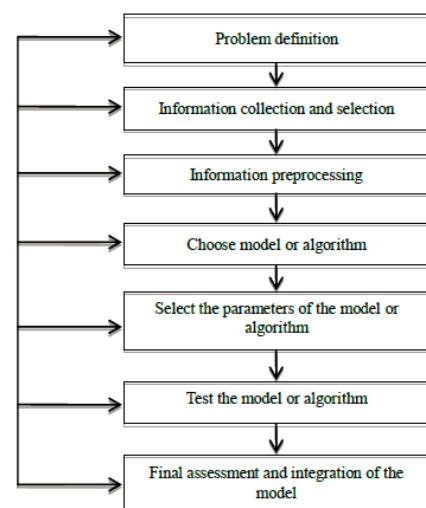


Figure 2. The framework of the proposed approach for Web data mining

Web data mining uses a series of automated or semi-automated process, conversion, and machining process to selects appropriate information that stored in the

historical database according to the purpose, and find the association that hidden in the data and meanwhile assist users to solve issues related [14]. In general, web data mining mainly includes the following steps as shown in Figure 2.

#### A. Data Collection and Preprocessing

UCI database is maintained by the University of California at Irvine California Irvine, which are widely used in machine learning. The database currently has a total of 264datasets, and the number is still growing. The UCI data sets are commonly used standard test data sets. It fully refers to the international famous design standard. Besides, it also combined the actual situation of Chinese financial markets, and it put the empirical research as the guide to finish the whole design.

In this paper, we used the Amazon Commerce reviews data set of the UCI as the main data resource. The dataset is used for authorship identification in online writepring which is a new research field of pattern recognition. The features of the data set are multivariate. The dataset is the customer reviews from Amazon Commerce Website for authorship identification. Most previous researches conducted the identification experiments for two to ten authors. But in the online context, reviews to be identified usually have more potential authors, and normally identification algorithms are not adapted to large number of target classes. To examine the robustness of identification algorithms, we identified 50 of the most active users (denoted by a unique ID and username) who frequently posted reviews in these newsgroups. The number of reviews we collected for each author is 30. As the data set is from different angles, so it is strong representative and challenging. This paper used the database and by using the Rough Set it can deal with the vague and imprecise information during the process of evaluation of electricity, which can achieve the correct, reliable and comprehensive evaluation for the efficiency.

#### B. Rough Set for Feature Representation

When the rough set approach deals with data, it defines the table of information as below  $S = \langle U, A, V, F \rangle$  where  $U = \{x_1, x_2, \dots, x_{10}\}$  is there search set, for example, {weight, gender,..., blood type};  $A = \{a_1, a_2, \dots, a_5\}$  is the attribute set, such as{heavy, medium, light}.The information function expression is  $f : U \times A \rightarrow V$ . It also can be described as  $f(x, a) \in V_a$ . Let  $S = \langle U, A, V, F \rangle$ , and  $P \subseteq A$  is the subset [9]. So the relationship without distinction must meet the following formula:

$$f(x, a) \in f(y, a), \quad \forall a \in P \quad (1)$$

Based on the above definition, the lower limit can be defined as  $PX_{down} = \{x_i \in U | [x_i]_{Ind(P)} \subset X\}$  ; and the upper limit is  $PX_{down} = \{x_i \in U | [x_i]_{Ind(P)} \cap X \neq \emptyset\}$  . The critical value is described as:

$$PNX = PX - PX_{down} \quad (2)$$

There are four comments relating to this formula. (1) If  $X$  meets the formula  $PX_{down} \neq \emptyset$ ,  $PX \neq U$  ,  $X$  can be considered as RST element in the data set. (2) If  $X$  meets the expression  $PX_{down} \neq \emptyset$  and  $PX = U$  ,  $X$  cannot be affirmed as RST element out of the data set. (3) If  $X$  meets the expression  $PX_{down} = \emptyset$  and  $PX \neq U$  ,  $X$  cannot be affirmed as RST element in the data set. (4) If  $X$  meets the expression  $PX_{down} = \emptyset$  and  $PX = U$  ,  $X$  can be affirmed as RST element out of the data set.

The calculation of identification error rate is:

$$\mu_p(X) = \frac{\text{card}(PX_{down})}{\text{card}(PX)}, \quad 0 \leq \mu_p(X) \leq 1 \quad (3)$$

### III. EXPERIMENTS RESULTS

This section will empirically validate our proposed RST-LR based on rough set theory (RST) and logistic regression (LR) for product ranking. The experiment procedures are as follows. First, collect data according to experiment design; second, extract feature from the processed data; third, model training and test. The experimental steps are shown in Figure 3. This section will sequentially present the dataset, verification criterion and experimental results. Note that, we integrate RST and LR in the same framework where RST extracts feature from the raw data while LR performs product ranking based on the chosen dataset.

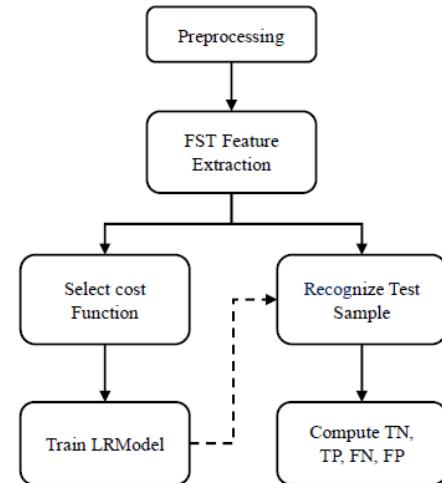


Figure 3. The flowchart of the proposed method

#### A. Experimental Dataset

The Amazon Commerce reviews data used in our experiments was collected by National Engineering Research Center for E-Learning. The dataset are from the customers' reviews in Amazon Commerce Website for authorship identification. It was used to identify the review authors from Amazon Commerce Website. Most previous works conducted the identification experiments for two to ten authors. But in the online context, reviews to be identified usually have more potential authors, and normally identification approaches are not adapted to large number of target classes. To examine the robustness of identification algorithms, we identified 50 of the most

active users (a unique ID and username) who frequently posted reviews in these newsgroups. The number of reviews we collected for each author is 30. The dataset includes 1500 samples with 10000 attributes that includes authors' linguistic style such as usage of digit, punctuation, words and sentences' length and usage frequency of words and so on. The categories are shown in Table 1.

TABLE I. CATEGORY DISTRIBUTION OF SAMPLES

Categories	Number of samples
News	500
Books	500
Video	500
Total number of samples	1500

The Data in our experiment were collected by ZhiLiu of National Engineering Research Center for E-Learning. The dataset are derived from the customer's reviews in Amazon commerce website for authorship identification. The dataset includes have 1500 samples with 10000 attributes, which used to identify the review authors from Amazon Commerce Website. Through which, we divided the dataset into two part, training data set with 1000 samples and forecasting data set with 500 samples.

### B. Evaluation Criterion

To validate the advantages of the proposed approach, we employ the identification accuracy as the evaluation criterion which can be defined as follows:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}$$

where TP indicates the true positive; TN indicates the true negative; FP denotes false positive; FN represents false negative. Four parameters of the confusion matrix are presented in Table 2. It is worth noting that, the parameters use a partition threshold between 1 and 2.

TABLE II. THE CONFUSION MATRIX

		Actual class	
		Positive	Negative
Predicted class	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

### C. Main Results

In the first experiment, we take advantage of the RST-LR approach for product ranking, and use the Amazon Commerce reviews collected from Amazon Commerce Website for authorship identification as the samples to run experiment. The dataset in this experiment were collected by ZhiLiu from National Engineering Research Center for E-Learning. The dataset are collected from the customer's reviews in Amazon Commerce Website for authorship identification. It uses accuracy and precision as the assessment standard to verify the effectiveness of RST-LR for product ranking. During the experiment, it adopts the standard method to determine the parameters of RST-LR. Then it utilizes the trained RST-LR to run the identification. In the experiment, the parameters of RST-LR are configured to the default ones.

The accuracy and precision is employed as the assessment criterion for the product ranking. We run the experiment for 20 rounds and report the experimental results of partial rounds are in Table 3. As show in Table 3, by using our algorithm to train parameter, RST-LRfor product ranking reach the highest performance of 87.49% under the criterion of accuracy, while RST-LReach the highest performance of 87.97% under the standard of precision. Further, the average accuracy of RST-LRis 84.55% which outperforms that of NN (79.82%). The potential reasons for these results are mainly threefold. Firstly, the RST-LRhas the ability to map the nonlinear data in the low dimensional space to the high dimensional space by a Kernel function, which makes the identification problem easy. Secondly, the parameter selection method can be adaptive to different dataset, in comparison with empirical parameter selection method. Thirdly, the processing procedure for data is able to remove noise and keep useful information effectively, and the element steps of our method could cooperate.

TABLE III. THE PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS

Experiment	Method	Evaluation standard	
		Accuracy	Precision
Round 1	RST	82.21	80.63
	RST-LR	87.49	87.24
Round 2	RST	81.03	82.18
	RST-LR	85.48	87.97
Round 3	RST	79.43	79.83
	RST-LR	84.01	86.32
Round 4	RST	81.72	83.72
	RST-LR	87.30	86.34
Round 5	RST	80.39	83.13
	RST-LR	82.48	85.55
Average	RST	79.82	81.72
	RST-LR	84.55	86.24

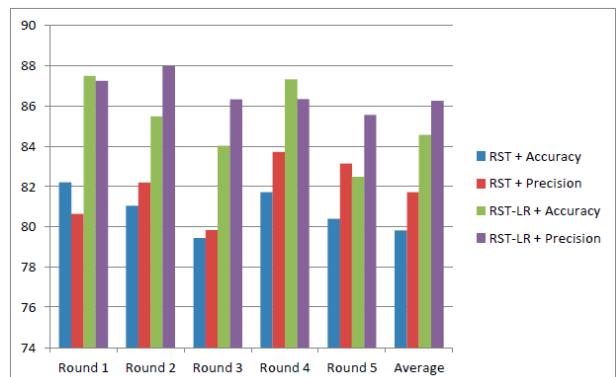


Figure 4. The performance comparison of two methods

In the second experiment, we further evaluate our proposed RST-LR method for product ranking, over the Amazon Commerce reviews. We use two comprehensive criterions, accuracy and precision, for the experimental evaluation. Identification accuracy and recall are two typical and popular measures for the correctness of the identification model. The experimental procedure is shown in the experiment section. The preprocessing step and feature extraction procedure are important to the experiment because they encode discriminant information.

The proposed approach RST-LR is learnt using the above approach, and some parameters of RST-LR are obtained using two schemes: hand-specification scheme and cross-validation strategy. We do the test multiple rounds, where in each round we randomly separate the dataset to training set and test set.

We seek to verify the advantage of the proposed approach for product ranking, through extensively comparing our method with three state-of-the-art algorithms, that is, RST, NN and BP-NN. Further, to verify the robustness of our method to the number of training sample, we also vary the number of training examples. The experimental results are report in Table 4. As present in Table 4, the proposed method consistently outperforms all three compared approach, with the range of 75.72-89.56% on accuracy and 76.86-96.12% on precision. Further, our RST-LR exhibit strong robustness against the percentage of training data from 30% to 70%. These results indicate that, our algorithm can be straightforwardly utilized to a number of applications. We here provide some explanations. (1) The RST-LR is capable to adapt and deal with complexly distributed data well, where the adaptability essentially comes from the flexibility of the parameters. (2) In comparison with empirical parameter selection approach, the selection algorithm for model parameters can adapt to the dataset. (3) The experimental procedure of the proposed approach could provide informative features and could maximize the discrimination ability.

TABLE IV. PEFFORMANCE COMPARISION FOR PRODUCT RANKING

Training samples	approach	Evaluation Criterion	
		Accuracy	Precision
30%	RST	72.40	74.16
	NN	74.06	69.39
	BP-NN	70.32	68.11
	RST-LR	75.72	76.86
50%	RST	79.28	83.56
	NN	83.68	83.32
	BP-NN	82.30	84.44
	RST-LR	82.85	88.03
70%	RST	89.04	88.98
	NN	89.90	87.81
	BP-NN	85.75	91.92
	RST-LR	89.56	96.12

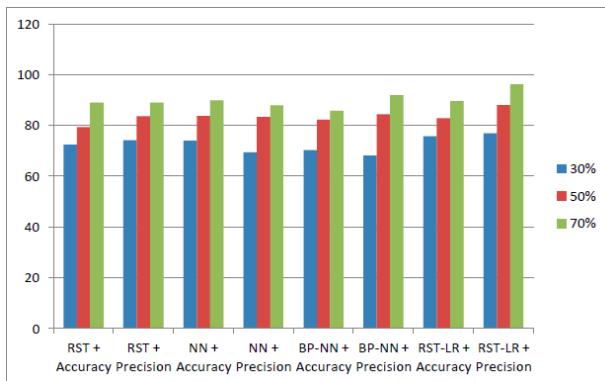


Figure 5. Pefformance comparision of 4 methods for product ranking

In the third experiment, it targets to validate the advantage and robustness of RST-LR method for product

ranking, in comparison with others. The Amazon Commerce reviews is collected from Amazon Commerce Website for authorship identification and is randomly partition to training set and test set. The dataset are collected from the customer's reviews in Amazon Commerce Website for authorship identification.. We take advantage of accuracy and precision as the assessment criterion for assessment. As show in previous section of this paper, the parameters of RST-LR is solve using the standard algorithm. The learn RST-LR to is then employ to run identification. The test is performed for multiple rounds, with default set.

The experimental results of our proposed algorithm RST-LR and the compared algorithm NN are respectively summarized in Table 5. These results are getting through the selection approach of cross validation algorithm under varying parameter setting. Identification accuracy and precision are two typical and popular measures for the correctness of the identification model. From the results of Table 5, we can see that, for different experimental rounds, under the verification criterion of accuracy or precision, the proposed algorithm for product ranking is significantly higher than that of the compared NN, respectively about 84.55-80.51% and 86.24-79.75%. The reasons for these results are mainly three aspects. (1) The RST-LR method can be applied to the conditions that sample data is large scale, complex dimension, containing a large number of heterogeneous information. (2) The learning method is according to the data distribution of the input data to select the model parameters of the RST-LR, which makes the RST-LR having better adaptability. (3) The framework of the proposed algorithm is composed of some comprehensive procedures which sequentially maximizes the identification ability.

TABLE V. THE COMARED EXPERIMENTAL RESULTS OF NN AND RST-LR

Experimental round	Algorithm	Evaluation standard	
		accuracy	precision
1	NN	80.52	78.27
	RST-LR (ours)	87.49	87.24
2	NN	78.42	82.19
	RST-LR (ours)	85.48	87.97
3	NN	82.77	78.34
	RST-LR (ours)	84.01	86.32
4	NN	81.85	78.95
	RST-LR (ours)	87.30	86.34
5	NN	82.31	83.28
	RST-LR (ours)	82.48	85.55

#### IV. CONCLUSION

The quality of the E-commerce service efficiency has relationship with the survival of E-commerce, so this paper is about RST based E-commerce service efficiency evaluation system. But in the application of the model, there are some aspects should be considered. There are several simplifying attribute methods. Here, we adopt the latter genetic algorithm, the design principles for the fitness function of the genetic algorithm is the rate of the cases included in the attribute set. The answer got from the simplified attribute is not unique; we can obtain many minimal set of attributes. But which set should be

adopted, should not only compare the accuracy of the identification results but also refer to the advice of experts in the field. We can also filter rules if it is necessary. The principles whose ability of predicting is less should be filtered, and the others are kept.

Generally, this step should refer to experts' knowledge so as to determine the reasonableness of these principles, and decide whether to delete or retain. When there are new objects that should be classified, we should on the basis of the following conditions: the new object exactly in line with a deterministic logical rule. The new object accords with many rules, but their decision attribute value are same. So there will be no confusing problems (n-to-1). But if the decision attribute values are not the same (n-to-n), it is necessary to compute which decision attribute value's probability is higher, and then select the principle that has more samples. New object exactly in line with a non-deterministic logical rule, namely, in line with an attribute value with the same premise, but it has several decision attribute values (1-to-n). This principle must calculate the possibility of decision attribute value is relatively high, and then select the principle that has more samples. If new object doesn't in line with any principle, it is necessary to find a similar near principle to utilize.

#### REFERENCES

- [1] Ahna, B. S., Cho, S. S. and Kim, C. Y., "The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction", *Expert Systems with Applications*, Vol. 18, pp. 65-74, 2010.
- [2] Chen, F. -L. and Liu, S. -F., "A Neural-Network Approach To Recognize Defect Spatial Pattern In Semiconductor Fabrication", *IEEE transactions on semiconductor manufacturing*, Vol. 13, No. 3, pp. 366-373, 2010
- [3] Dimitras, A. I., "Business Failure Prediction Using Rough Sets, " *European Journal of Operational Research*, Vol. 114, pp. 263-279, 2009
- [4] Gardner, R. and Bieker, I., "Solving Tough Semiconductor Manufacturing Problems Using Data Mining", *IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, pp. 46-55, 2008
- [5] Han, J., and Kamber, M., *Data Mining: Concepts and Techniques*, San Francisco, CA: Morgan Kaufmann Publishers, 2001.
- [6] Johnson, D. S., "Approximation Algorithrns for Combinatorial Problerns." *Journal of Computer and System Sciences*, Vol. 9, pp. 258-278, 1974
- [7] Kusiak, A., "Decomposition in Data Mining: An Industrial Case Study", *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. 23, No. 4, pp. 345-353, 2009
- [8] Kusiak, A., "Rough Set Theory: A Data Mining Tool for Serniculator Manufacturing", *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. 24, No. 1, pp. 44-50, 2011
- [9] Lee, I. -H., Yu, S. -J. and Park S. -c., "Design of Intelligent Data Sampling Methodology Based on Data Mining", *IEEE transactions on semiconductor manufacturing*, Vol. 17, No. 5, pp. 637-649, 2008
- [10] Zhang Ji, Bo Ning, "Research on Web Information Retrieval based on Vector Space Model", *Journal of Networks*, Vol. 8, No. 3, pp. 688-695, 2013
- [11] Fang Dong, Mengchi Liu, Kun Ma, "Web Entities Extraction Based on Semi-Structured Semantic Database", *Journal of Networks*, Vol. 8, No. 7, pp. 1640-1646, 2013
- [12] Qiang Deng, YupinLuo, "Saliency Detection by Selective Strategy for Salient Object Segmentation", *Journal of Multimedia*, Vol. 7, No. 6, pp. 420-428, 2012
- [13] Hongzeng He, Jingbo Sun, Yue Wang, Qi Liu, Jian Yuan, "Cluster Analysis Based Switching-off Scheme of Base Stations for Energy Saving", *Journal of Networks*, Vol. 7, No. 3, pp. 494-501, 2012
- [14] Jing Ma, Shuli Sun, "Centralized Fusion Estimators for Multi-sensor Systems with Multiplicative Noises and Missing Measurements", *Journal of Networks*, pp. 1538-1545, Vol. 7, No. 10, 2012
- [15] Yang Hongyu, Wang Sen, Zhang Yang, "Web Service Replacement Algorithm based on Complex Network", *Journal of Networks*, Vol. 8, No. 4, pp. 829-835, 2013

# Eye State Recognition Algorithm GAHMM of Web-based Learning Fatigue

Qiufen Yang<sup>1</sup>, Zhenjun Li<sup>1</sup>, Canjun Li<sup>1</sup>, Xianlin Yang<sup>1</sup>, and Can Zhu<sup>2</sup>

1. Department of Computer Science, Hunan Radio & TV University, Changsha, China

2. School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha, China

Email: yqf5569@sohu.com; Catorljzj@163.com; li\_canjun@163.com; yxlhnrtu@126.com; zhc@csust.edu.cn

**Abstract**—With the increasingly deeper application of multimedia and network technology in education, the proportion of web-based learning presents a trend of continuous growth in people's daily study, and the web-based learners tend to have learning fatigue during their learning process of distance education. Under the 4 learning states of normal study, fatigue and doubt, the eye opening degrees of web-based learner have certain difference. Through preprocessing of the color images of web-based learners' eye state, the 2-dimensional kernel function is selected to build 48 optimal filters and obtain 48 eigenvalues, and then, through HMM, training is conducted to identify the eye state and feature classification. The experiment results show that this algorithm has very high identification ability of the web-based learning fatigue.

**Index Terms**—Learning Fatigue; Internet Learning; Hidden Markov Model

## I. INTRODUCTION

The 21st century is a century of knowledge as well as a century of continuous reform in education. During the process in which the Internet and information technology have promoted the human society to comprehensively enter the informationization age, it has also significantly expanded the boundary of time and space and increased people's interest in learning! Efficiency and initiation, “the traditional teaching and learning patterns are experiencing a big breakthrough, the education is facing profound reform”, and in the new learning environment, the “web-based learning” has been endowed with rich connotations.

Modern web-based learning is a necessary result of the global informationization process, and informationization represented by the information superhighway has not only brought brand new technology and media to education, but also comprehensively influenced the concept, system, model and content of education. Its status in the era of knowledge-based economy is like words in the agricultural era and printing in the industrial era, which has become the third milestone in the human educational history. The emergence of network communication has triggered a profound educational reform, which has generated an important leap to current educational system and model [1].

In accordance with the proportion of students getting education at school in China's higher education system,

the modern distance education, adult education and self-taught examination account for nearly 50% of the entire scale of higher education, and the distance education accounts for 50% of the overall adult education and continuing education [2]. At present, the approaches of distance education have been gradually transformed from the correspondence college and radio & television university to the networking academy and virtual university. Therefore, the society will attach increasingly more importance to the digital and web-based learning. In addition, with the deepening of Ministry of Education's “teaching quality and teaching reform program for colleges and universities” and the continuous promotion of the construction of excellent courses at various levels, the web-based learning environment, digital learning resources and network-based teaching method have become important components of full-time college teaching. Web-based learning is not the exclusive learning activity for students in colleges of distance education anymore, and it has become one of the basic learning methods for students in full-time colleges.

During the process of web-based learning, separation from the teachers and students makes the learners stay in a separated state of time and space from the teachers and students, they cannot have face-to-face communication, and attention are generally not paid to the learners' learning moods and state [3]. This kind of learning does not have the regular cultural atmosphere of campus or timely reminding from the teacher, plus long-time continuous study, the web-based learner tends to have learning fatigue, which might cause decrease in the learning efficiency, and the learner may even reach the state of being unable to continue learning. There are many manifestation patterns of learning fatigue. Learning fatigue mainly consists of the two types of “physiological fatigue” (i.e., physical fatigue) and “psychological fatigue” (i.e., mental fatigue).

At present, people are thinking of using certain methods to conduct identification and intervention with fatigue, in order to prevent traffic accidents or improve the efficiency of web-based learning. In the field of fatigue driving detection, some learners try to use the sensor technology, signal detection and collection technology or identification technology of facial expression to conduct identification, prediction and fatigue intervention of the driver's fatigue state, which is

mainly the detection of physiological fatigue. In addition, in the field of web-based learning, some scholars are trying to promote the emotional communication under the web-based learning environment from the perspective to improve the design of network course teaching, which can in a certain degree reduce the possibility of the learner to get into learning fatigue; some scholars are using the multimedia technology, BBS and even instant messaging tools like QQ to promote the emotional communication under the web-based learning environment, which can prevent learning fatigue in a certain degree. These studies mainly concentrate on the identification or detection of physiological fatigue, while ignoring the complexity of learning fatigue, it includes two levels of physiological fatigue and psychological fatigue, and there are not many researches which initiatively conduct identification and intervention of the state of learning fatigue. Based on the theories of psychology, cognitive science and affective computing, 3 basic learning moods are defined, the eye state identification technology is used to identify the eye state of distance learner, determine whether the learner is in a state of learning fatigue and discriminatively treat the physiological fatigue and psychological fatigue, and in this way, real-time intervention with learning fatigue can be implemented to guide the learner to avoid learning fatigue and increase the efficiency of web-based learning.

Obvious difference of ocular shape exists in 3 different kinds of states – normal learning, fatigue and focus [2]. In normal learning, eyes open largely, full of energy and spirit; in fatigue, eyes diminish unnaturally, in sagging spirit; in confusion, eyes open up, upper eyelids too high and lower eyelids slumping. At present, there remains few research on fatigue detection technique in field of education. Liu Jinyi from Shanghai Jiaotong University has conducted sleepiness detection research and realization on message image processing and fuzzy logic in intelligent classroom. He analyzes the actual requirement of doze scene and detection technical scheme of intelligent classroom, adopts digital image processing and fuzzy logic as final mode of research and realization, and performs realization on MATLAB platform. Nevertheless, he merely sets the eye opening state as the standard of being sober and the eye closing state as the standard of being dozing. Borrowing mature driver fatigue techniques, we carry out research work on the learning fatigue of Internet learners.

Based on the Gabor wavelet and HMM, this paper proposes an algorithm to identify the eye state of learning fatigue, which is embedded into the distance intelligence teaching system. First of all, based on the binarization pretreatment, this method is used to obtain the image of eye state; then, the 2-dimensional selection Gabor kernel function is built, 48 optimal filters are built, 48 eigenvalues are obtained, these 48 eigenvalues generate 48 eigenvectors, and HMM is used to conduct eye state identification of an observation sequence  $O$  formed by the eigenvectors of eye state image.

The recognition result of this paper will be sent into remote intelligent tutoring system as feedback

information, which can provide a basis for teachers to adjust teaching progress, re-arrange teaching content and correct teaching method in time, and can provide Internet learners with individualized learning environment, so as to make up for the problem of emotional deficiency of Internet learners.

## II. IMAGE PREPROCESSING

In our method, the  $YCbCr$  color space is obtained from  $RGB$  color space using

$$\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ Cr = R - Y \\ Cb = B - Y \end{cases} \quad (1)$$

where  $R$ ,  $G$ , and  $B$  are the red, green, and blue components of the color image, respectively. The  $Y$ ,  $Cr$ , and  $Cb$  values are normalized to the range  $[0, 255]$ .

The eye regions have low intensity  $Y$ , low red chrominance ( $Cr$ ), and high blue chrominance ( $Cb$ ), when compared to the forehead region of the face. Using this fact, we will preprocess the input image to a gray level image. The PDF of gray-level differences between neighboring pixels is well approximated by a generalized Laplacian [3]. The procedure of gray level is shown in Fig. 1.

$$f_L(\Delta I) = \frac{1}{Z_L} \exp\left(-\left|\frac{\Delta I}{\lambda}\right|^\beta\right) \quad (2)$$

where  $\Delta I$  is the gray level difference,  $\lambda$  depends on the distance between the two sampled image locations,  $\beta$  is a parameter approximately equal to 0.5 and  $ZL$  is a normalization constant. In the following we assume that  $\beta = 0.5$ , which implies  $ZL = 4\lambda$ .

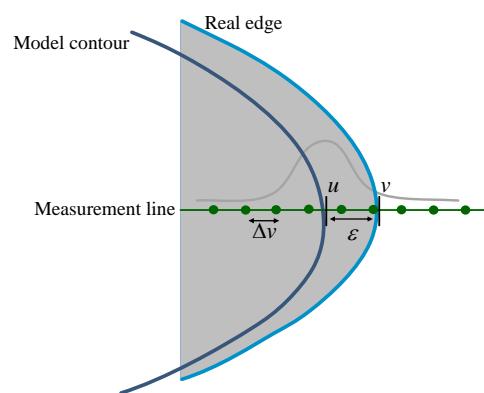


Figure 1. Marginalized contour definitions

In Fig. 2, the distribution of gray level differences over images of eyes is shown. The gray level differences, calculated over different scales of skin, are shown in the figure. It can be seen that the distribution of gray level differences for the class of eye images is well approximated by the  $YCbCr$  color space. Further, it can be seen that the width  $k$  of the distribution increases with the scale.

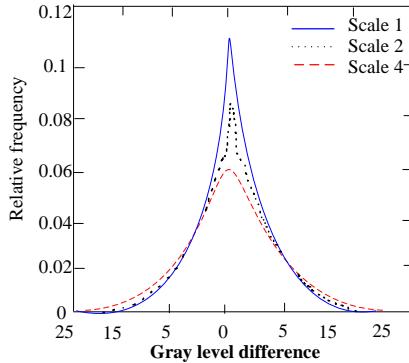


Figure 2. Distribution of gray level differences over scales for images of eyes. The scale number refers to the distance between gray level measurements: “scale 1” refers to 1 pixel distance and “scale 4” refers to 4 pixels distance between the measurement points.

In Fig. 2, the distribution of gray level differences over images of eyes is shown. The gray level differences, calculated over different scales of  $\Delta v$ , are shown in the figure. It can be seen that the distribution of gray level differences for the class of eye images is well approximated by the generalized Laplacian, defined in Equation (2). Further, it can be seen that the width  $\lambda$  of the distribution increases with the scale. This is caused by de-correlation of pixel values as  $\Delta v$  increases.

Then, binarizes the gray level image to a “binary image” by simple global threshold with threshold  $T$  [3]. After linearization, the next task is to get 4-connected components, label them, and then find the center of each block. We label two eyes, mouth, ear etc. The details of connected component finding can be found in [4].

The image of ocular region after processing is shown in Fig. 3.

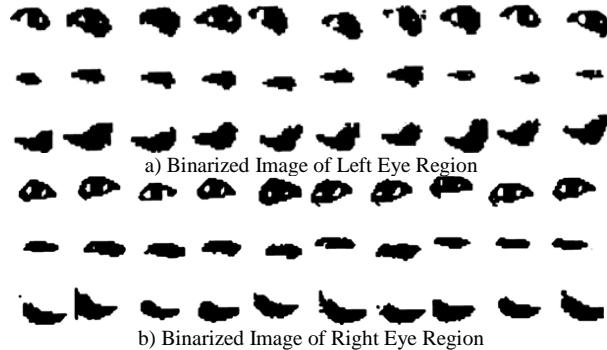


Figure 3. Ocular region image preprocessing

### III. FEATURE EXTRACTION OF GABOR FILTERING

Two-dimensional Gabor wavelet transformation is an important tool for the time-frequency domain to make signal analysis and processing, and its transformation coefficient has good visual features and biology background, thus, it is widely used in image processing, mode recognition, etc.,

Gabor filter of different parameters can capture local feature information in images, which is corresponding to different spatial frequencies, spatial positions and directions. Because of the features of Gabor filter, it is

not sensitive to changes to the brightness and face postures, thus, Gabor filter is widely used for image coding, handwritten numeral recognition, face recognition and edge detection, etc.

Then, make two-dimensional Gabor wavelet transformation in the rectangle feature region obtained via the features on the above, to obtain the expression features of the face region of drivers.

Two-dimensional Gabor wavelet kernel function is:

$$\psi_j(\vec{k}, \vec{x}) = \frac{\vec{k}_j^2}{\sigma^2} \exp\left(-\frac{\vec{k}_j^2 \bullet \vec{x}^2}{2\sigma^2}\right) [\exp(\vec{k}_j^2 \bullet \vec{x}) - \exp(-\frac{\sigma^2}{2})] \quad (3)$$

Of which,  $\vec{k}_j^2$  ensures that filters of different frequency bandwidths have almost the same energy.  $\exp(-\frac{\sigma^2}{2})$  is to offset the DC shunt of the images, thus, the filter will not be sensitive to the overall lamination. Its advantage is that it can maintain the relation of spatial relation while allowing to describe the spatial frequency structure.

$$\vec{k}_j = \begin{pmatrix} k_v \cos \varphi_u \\ k_v \sin \varphi_u \end{pmatrix}, k_v = 2^{-\frac{v+2}{2}} \pi, \varphi_u = \mu \frac{\pi}{8} \quad (4)$$

$\vec{k}_j$  constitutes different wavelets for sides with different values, and 4 changes to size and directions are adopted in the paper.

$$k_v = 2^{-\frac{v+2}{2}} \pi \quad (v=1,2,3,4),$$

Six directions  $\varphi: 0-6/8\pi$  ( $\mu=1,2,\dots,6$ ), with interval of  $\pi/8$ .  $\sigma$  is the length of the filter, take  $\sigma=\pi$ , suppose it is 1 time frequency. Input the image  $I(x, y)$  and the wavelet for convolution, then,

$$g(\vec{k}_j, \vec{x}) = \iint I(\vec{x}, \vec{y}) \psi_j(\vec{k}, \vec{x}) d\vec{x} d\vec{y} \quad (5)$$

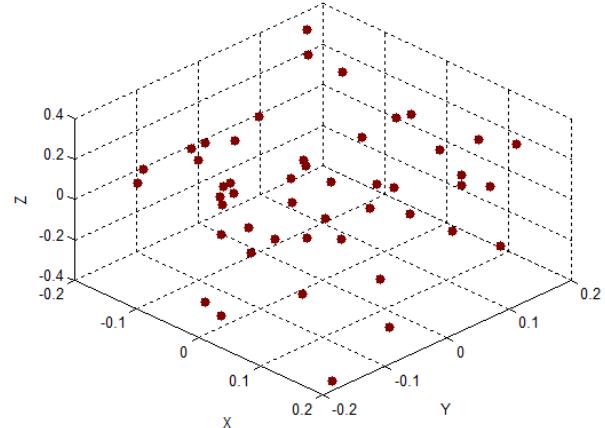


Figure 4. 48 feature points distribution drawing obtained by two-dimensional Gabor filtering

Of which,  $g(\vec{k}_j, \vec{x})$  is the amplitude. Thus, the number of Gabor filter is 48, which constitute a group of optimal filters that represent the target features. These filters constitute wavelet subspace, project the images to the wavelet subspace to get the coefficient of the wavelet, and extract the mean value and variance to represent the statistical features of the facial expression images of drivers.

Figure 5 shows the ocular image after Gabor reconstruction, extracting four sets of feature basis vectors, of which a) is original ocular state image and b) is Gabor reconstruction state image.

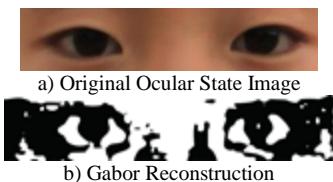


Figure 5. Ocular state feature extraction

#### IV. HMM MODEL

##### A. Basic Concept

HMM is a kind of very mature matching technique that uses parameter to show the probability model of random process statistical feature [14], which consists of two parts: on part is Hidden Layer, Hidden Markov Chain; the other part is Observation Layer, actual observed quantity.

One HMM can be marked as Formula (8):

$$\lambda = (N, M, \pi, A, B) \quad (6)$$

where  $N$  refers to the state number of Markov chain in the model, the states of which are represented as  $\theta_1, \theta_2, \dots, \theta_N$ , and the state of Markov chain at the time of  $t$  is  $q_t$ ;  $M$  refers to the number of observed value corresponding to each state, the observed value of which is represented as  $V_1, V_2, \dots, V_M$ , and the observed value at the time of  $t$  is  $N$ ;  $\pi$  refers to the probability distribution vector in the initial state,  $O$ , i.e.

$$\lambda = (N, M, \pi, A, B), \pi = (\pi_1, \pi_2, \dots, \pi_N);$$

$A$  refers to the state transition probability matrix,  $B = \{b_j(O_i)\}$ , i.e.

$$b_j(O_i) = \frac{1}{\sqrt{(2\pi)^n |\sum_j|}} \exp \left\{ -\frac{1}{2} (o_i - \mu_j)^T \sum_j^{-1} (o_i - \mu_j) \right\},$$

$$\sum_j$$

$B$  refers to the observation probability matrix,  $\mu_j$ , i.e.

$$\mu_j = \frac{1}{T_j} \sum_{i=1}^{T_j} o_i, \sum_j = \frac{1}{T} \sum_{i=1}^{T_j} (o_i - \mu_j)(o_i - \mu_j)^T.$$

As the Gaussian mixture model is adopted to describe the probability distribution matrix  $B$  of HMM parameter and matrix  $B$  describes the probability statistic properties of and each state, the moderate increase of  $M$

could promote the freedom degree of Gaussian mixture model, which is conducive to the performance of the statistical properties of each state, so as to improve the recognition rate of expressions. However, the blind increase of  $M$  will result in the decrease of effect by HMM model, leading to the over-reliance of the training process on the data set. As can be seen from Figure 5, the state number of HMM shall be 4 in order to achieve better recognition effect. When the mixed element number of Gaussian probability is 12, the system is considered to be relatively more reasonable, which is initially defined as  $\lambda = (4, 12, \pi, A, B)$ . Generally, the impact of the parameter  $\pi$  and initial value selection of  $A$  is insignificant, which can be selected randomly or uniformly under the premise of satisfying the condition of sum as 1. However, the initial value of  $B$  has great influence on the trained HMM, so a method that can match the observation sequence shall be required. Gaussian probability density function is selected by this paper, of which,  $\pi = \{1, 0, \dots, 0\}$ . As for the state transition probability matrix  $A = \{A_{ij}\}$ , it can be made that  $a_{ij} = 0$ , of which,  $j > i+1$  or  $j < i$ . The probability distribution matrix can be calculated in accordance with the formula below:

$$b_j(o_i) = \frac{1}{\sqrt{(2\pi)^n |\sum_j|}} \exp \left\{ -\frac{1}{2} (o_i - \mu_j)^T \sum_j^{-1} (o_i - \mu_j) \right\} \quad (7)$$

where,  $\mu_j$  and  $\sum_j$  represent the mean and covariance matrix of Gaussian probability density function respectively,

$$\mu_j = \frac{1}{T_j} \sum_{i=1}^{T_j} o_i \quad (8)$$

$$\sum_j = \frac{1}{T} \sum_{i=1}^{T_j} (o_i - \mu_j)(o_i - \mu_j)^T \quad (9)$$

##### B. HMM Training

The fatigue state can be classified into three most basic categories, and every state needs to design one HMM, besides, similar to the characteristics of voice sequence, fatigue state image sequence is an irreversible process with time sequence, thus, every HMM adopts a left-to-right model, and three such HMMs constitute one face expression classifier.

The determination of HMM model parameter mainly refers to the determination of various parameters in  $\lambda = (N, M, \pi, A, B)$ , in the paper, the selection of  $N$  and  $M$  is determined as per the experiment result, the value range of  $N$  is 2~8, and the value range of  $M$  is 5~21. Petar et al. made automatic expression recognition for HMM with direct definition status number of 3 when making HMM modeling for face animation parameter, however, the HMM initial parameter obtained actually is

random and does not have a fixed value. Thus, it can be further understood that the sequence that we have observed is generated by several states, which are abstract, do not have specific meaning, and can only be estimated via observing sequence. When HMM id adopted to recognize the fatigue state, the length of fatigue state sequence is not limited theoretically.

HMM training can be carried out either with one single image or multiple images. The purpose is to optimize HMM parameter. Training procedure is as follows:

Perform FastICA processing on the ocular state image that is to be classified. A set of observation sequence  $O_i$  is formed. The obtained feature value is regarded as the feature vector of ocular state image.

Build a universal model  $\lambda = (N, M, \pi, A, B)$ , specify the state transition permitted by this model, the number of states and the number of Gaussian mixed probability components.

Equally segment training data, and calculate the initial parameter of model, corresponding to  $N_t$ -th (moment t) state. State transition matrix  $A = \{a_{ij}\}$ , we make  $a_{ij}=0$ , when  $j < i$  or  $j > i+1$ . Initial probability distribution  $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ , HMM starts with the first state, suppose  $\pi_1=1$ . If  $\pi_i=0(i \neq 1)$ , probability distribution matrix B adopts Gaussian Probability-Density Function,  $B = \{b_j(O_i)\}$ , calculation is made according to Formula (8):

$$b_j(o_i) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} \exp \left\{ -\frac{1}{2} (o_i - \mu_j)^T \Sigma_j^{-1} (o_i - \mu_j) \right\} \quad (10)$$

Among which,  $\Sigma_j$  and  $\mu_j$  respectively refer to covariance matrix and mean value of Gaussian Probability-Density Function.

$$\mu_j = \frac{1}{T_j} \sum_{i=1}^{T_j} o_i \quad (11)$$

$$\Sigma_j = \frac{1}{T_j} \sum_{i=1}^{T_j} (o_i - \mu_j)(o_i - \mu_j)^T \quad (12)$$

Uniform Segmentation [14] is replaced with Viterbi Segmentation, the parameter of Gaussian Mixed Model is determined with subsection K mean value clustering method, then optimal state sequence of HMM is obtained.

Re-evaluation on parameters adopts Baum-Welch Algorithm [14]. Specify one  $\lambda = (N, M, \pi, A, B)$ , make  $P(O|\lambda)$  the largest and optimize parameter with the model finally obtained.  $P(O|\lambda)$  belongs to a certain class in ocular fatigue state.

### C. HMM Eye Fatigue State Recognition

We firstly perform processing on eye image with FastICA classification. The obtained feature value is regarded as the feature vector of ocular state image so as to form a set of observation sequence. Forward-Backward

Algorithm is adopted to calculate the probability  $P(O|\lambda_i)$  produced by each training model  $\lambda_i (1 \leq i \leq 3)$ . Shown in Formula (11), the model that the largest value responds to is the class that the ocular state to be recognized belongs to.

$$\phi_n = \arg \max_i P(O|\lambda_i) \quad (13)$$

If the probability of the sequence  $O$  produced by  $n$ -th model  $\phi_n$ , then we classify ocular state as  $n$ -th class.

## V. EXPERIMENTAL RESULT AND ANALYSIS

Firstly, we have totally shot 410 images in the laboratory. Each image is 256×256 large (pixel). The experiment is carried out in two parts: on one part, 80 images are used for HMM training; on the other part, 330 images are used for testing, 110 images per state.

When implementing Matlab simulation on PC machine, this algorithm has its processing speed reach 12fps, being able to trace and recognize the fatigue state of testing personnel in real time. The eye image detection and recognition rate of testing personnel is shown in TABLE I,

TABLE I. OCULAR STATE RECOGNITION RATEBASED ON GA HMM

	Normal learning (%)	Fatigue learning (%)	Dozing learning (%)
Normal Learning	84.87%		
Fatigue		90.23%	
Confusion			92.5%

Then, in the face recognition database CAS-Peal-R1, samples of the numbers of 20000, 10000 and 5000 are selected respectively to conduct experiment. The experiment results are as shown in TABLE II.

TABLE II. SPEED COMPARISON BETWEEN THE IMPROVED GAHMM ALGORITHM AND COMMON HMM EYE STATE ALGORITHM(S)

Number of image experiments	Improved GAHMM algorithm	Common HMM algorithm
20000	2.87	4.32
10000	1.67	2.26
5000	1.42	1.23

In accordance with the table, we can see that when the tested number of images is big (20000 and 10000), the improved GAHMM algorithm has a higher recognition speed than the common HMM algorithm; when the number is small (5000), the improved GAHMM algorithm has a lower speed than the common HMM algorithm, because before comparing the facial images, it takes a long time for the GAHMM algorithm to conduct classification of facial images. Therefore, we can reach this conclusion: the algorithm that needs to conduct classification of facial images before the recognition applies to large-scale face database, which generally has a number higher than 10000.

The results are shown in fig. 6 and fig. 7.

## VI. CONCLUSION

Among three kinds of states, recognition rate of normal learning is relatively low, because the expression feature of focus and fatigue is more obvious than that of normal learning, which can be improved by increasing the number and variety of training samples. In remote intelligent education system, we raise recognition rate on two ways: first, require camera to shoot color image, and update Gaussian model parameter in time according to the change of component; second, further refine the classification of learning fatigue, and make clearer judgment standard on eye opening and closing state, so as to improve the accuracy rate and stability of recognition and tracing.

Therefore, this study tries to analyze the basic process and principles of web-based learning from the perspective of web-based learning. Considering at present the web-based learners are mainly adults, this study plans to select various college students as the research objects, and with the excellent course *Engineering Mechanics* as the platform of web-based learning, the research method and experiment equipment of cognitive psychology are used to conduct related experiment of web-based learning and analyze the web-based learners' behavior characteristics in order to disclose the general process and rules of web-based learning. It can help improve the design of various current support platforms of web-based learning to build the support platform more consistent with the characteristics and rules of web-based learning, which can further promote various platforms of web-based learning to better carry out their positive effect in the future.

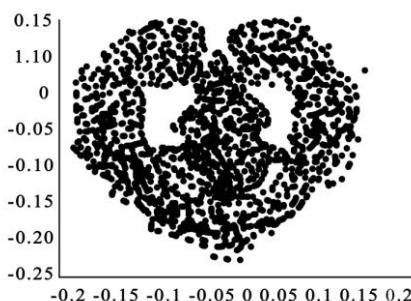


Figure 6. Recognition effect of GAHMM algorithms

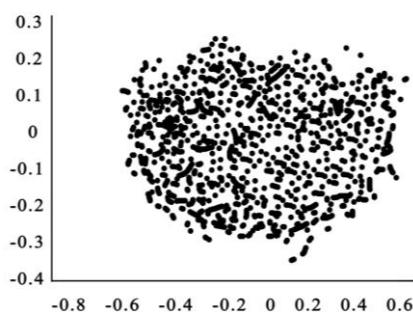


Figure 7. Recognition effect of general HMM algorithm

## ACKNOWLEDGMENT

This research is supported by the 12th five-year plan project in hunan province (No. XJK013BXX006), the Science and Technology Planning Project of Hunan Province (No. 2012GK3096, No. 2012FJ4334), Scientific Research Fund of Hunan Provincial Education Department (No. 12C1158, No. 10C0374, No. 13C613 and No. 13C614).

## REFERENCES

- [1] Wang Ting-ting, WU Yan-wen, AI Xue-yi, Learning fatigue recognition and intervention based on facial expression recognition, *Computer Engineering and Design*, 2010, 31 (8), pp. 1764-1767.
- [2] Zhang Jiahua, Research on the Information Processing Model of web based Learning and Its Application, *PhD thesis, Southwest University*, 2010.
- [3] M. Balasubtamanian, S. Palanivel, V. Ramalingam, Real time face and mouth recognition using radial basis function neural networks. *Expert System with Applications* (2008), doi:10. 1016/j.eswa. 2008. 08. 001.
- [4] R. C. Gonzalez, R. E. Woods. Digital Image Processing. Addison-Wesley Publishing Company, *Reading, MA, USA*, 1992.
- [5] Alldrin N, Smith A, Turnbull D. Classifying Facial Expression with Radial Basis Function Networks, *Using Gradient Descent and K-means*. CSE253, 2003.
- [6] Thomas Plötz, Gernot A. Fink. Pattern recognition methods for advanced stochastic protein sequence analysis using HMMs. *Pattern Recognition* 39 (2006) 2267 – 2280
- [7] Tian Ying li, Takeo Kanade, Cohn Jeffrey F. Eye-State Action Unit Detection by Gabor Wavelet. *ICMI 2000*: 143-150
- [8] GUO Ke-you, CHU Jiang-wei, WANG Rong-ben. Study on the Recognition of the Driver's Eye States. *Systems Engineering and Electronics*, 2003, 10:1186-1228
- [9] W. A. Fellner, J. G. Taylor, N. Tsapatsoulis and S. Kollias, "Comparing Template-Based, Feature-Based and Supervised Classification of Facial Expressions from Static Images, " *In Proceedings of Circuits, Systems, Communications and Computers*, pp. 5331-5336, 1999.
- [10] Chao-Fa Chuang, Frank Y. Shih. Recognizing facial action units independent component analysis and support vector machine. *Pattern Recognition*. 39 (2006) 1795~1798
- [11] Y. Wang, H. Ai, B. Wu and C. Huang, "Real Time Facial Expression Recognition with Adaboost, " *In Proceedings of 17th IEEE International Conference on Pattern Recognition*, vol. 3, pp. 926-929, 2004.
- [12] Jing, X. Y., Tang, Y. Y., Zhang, D., 2005. A Fourier-LDA approach for image recognition. *Pattern Recognition* 38 (2), pp. 453–457.
- [13] Wang Rong-ben, Guo Ke-you, Chu Jiang-wei. Study on Method of Recognizing Driver's Eye State Based on Gabor Wavelet in Driver Behavior Surveillance. *Journal of Image and graphic*, 2003, 9(8), pp. 1043-1047.
- [14] Y. Wang, H. Ai, B. Wu and C. Huang, "Real Time Facial Expression Recognition with Adaboost, " *In Proceedings of 17th IEEE International Conference on Pattern Recognition*, vol. 3, pp. 926-929, 2004.
- [15] Thomas Plötz, Gernot A. Fink. Pattern recognition methods for advanced stochastic protein sequence analysis using HMMs. *Pattern Recognition* 39 (2006) 2267 – 2280.
- [16] J. -H. Lee, T. -W. Lee, H. -Y. Jung, S. -Y. Lee, On the efficient speech feature extraction based on independent

- component analysis, *Neural Process. Lett.* 15 (3) (2002) 235–245.
- [17] C. F. Beckmann, S. M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging, *IEEE Trans. Med. Imaging* 23 (2004) 137–152.
- [18] L. E. Baum, T. Petrie, Statistical inference for probabilistic functions of finite state Markov chains, *Ann. Math. Stat.* 37 (1966) 1554–1563.
- [19] Yang Liu, Nitesh V. Chawla, Mary P. Harper, Elizabeth Shriberg, Andreas Stolcke, A study in machine learning from imbalanced data for sentence boundary detection in speech, *Computer Speech and Language* 20 (2006), pp. 468–494.
- [20] Bing Feng, Xiao Qing Ding, Off-line handwritten Chinese character recognition with hidden Markov models, in: *5th International Conference on Signal Processing Proceedings*, WCCC-ICSP 2000, vol. 3, pp. 1542–1545.
- [21] A. V. Nefian, Embedded Bayesian networks for face recognition, *Proc. of the IEEE International Conference on Multimedia and Expo*, Vol. 2, 26-29 August 2002, Lusanne, Switzerland, pp. 133-136
- [22] A. V. Nefian. Embedded Bayesian networks for face recognition. *Proc. of the IEEE International Conference on Multimedia and Expo*, Vol. 2, 26-29 August 2002, Lusanne, Switzerland, pp. 133-136.
- [23] Chao Tang; Changle Zhou; Wei Pan; Lidong Xie; Huosheng Hu. Fusing mixed visual features for human action recognition. *Int. J. of Modelling, Identification and Control*, 2013 Vol. 19, No. 1, pp. 13 – 22.
- [24] Changxi Ma; Yinchen Li; Ruichun He; Bo Qi; Fang Wu; Aixia Diao; Cunrui Ma. Latent ringlike road traffic control system based on compound mechanism particle swarm optimisation algorithm. *Int. J. of Modelling, Identification and Control*, 2013 Vol. 18, No. 1, pp. 47 – 53.
- [25] Yi Zhang; Jiao Liu; Yuan Luo; Huosheng Hu. Hybrid lip shape feature extraction and recognition for human-machine interaction. *Int. J. of Modelling, Identification and Control*, 2013 Vol. 18, No. 3, pp. 191 – 198.
- [26] Y. Zhang, J. Liu, Y. Luo and H. Hu, Hybrid lip shape feature extraction and recognition for human-machine interaction, *International Journal of Modelling, Identification & Control*, Vol. 18, No. 3, 2012, pp. 191–198.
- Qiufen Yang** received her B.Eng. and M. Eng. degrees degree from Nanjing University, Jiangsu, China and National Defense Science and Technology University, ChangSha, Hunan, China in 1996 and 2004, respectively. She is currently pursuing her Ph.D. degree in the School of software, Central South University, Changsha, Hunan, China. Her current research interests are in the areas of data fusion and computer vision.
- Zhenjun Li** received the degree of the B.Eng.( Electronic Engineering) and the M. Eng. (Electronic Science and technology) from Xiangtan University, Hunan, China and Hunan University, Changsha, China in1999 and 2009, respectively. His current research interests are in the areas of the key technology of mobile Internet.
- Canjun Li** received the degree of the B.Eng.(Industrial Automation) and the M.Eng. (Control Science and Engineering) from Zhengzhou University, Henan, China and Central South University, Changsha, China in 2000 and 2007, respectively. She is a visiting scholar at Central South University, Changsha, Hunan, China. Her current research interests are in the areas of modeling and control of complex system.
- Xianlin Yang** received the degree of Doctor of Engineering in General Mechanics and Basis of Mechanics in Hunan University, in 2008, and the degree of the B.S. and the M.S.in Physics from Hunan Normal University, in1987 and 1992, respectively, His current research interests are in the areas of Nonlinear dynamics.
- Can Zhu** received his Ph.D. in Computer Application from Central South University, China, in 2009. He is engaged as a lectorate of Changsha University of Science & Technology, China.His research interests involve modeling and optimization concerned transportation planning and management.

# Restoration Technique of Video Motion Image Estimation Based on Wavelet

Ruibin Chen

Department of Information Engineering, Binzhou University, Binzhou City Shandong Province 256603, China

**Abstract**—Video compression technology is the research field of video compression coding attention. It mainly includes the elimination of temporal redundancy and spatial redundancy elimination algorithm, this paper mainly focuses on the study of motion estimation and compensation algorithm to eliminate the temporal redundancy. The starting point of this article is how to improve the precision and the subjective quality of the reconstructed image motion estimation and compensation, to reduce the computation complexity of motion estimation algorithms, to improve the efficiency of motion estimation. This paper makes some studies on the redundant wavelet domain block matching motion estimation and compensation, then the video image for non translational motion, the DT triangular mesh motion estimation and compensation in the redundant wavelet domain to do related research. Experimental verification, the algorithm can effectively extract the feature points of video image stabilization and improve the matching, motion estimation and compensation accuracy and efficiency, and the subjective quality of reconstructed image is good, especially for non translational motion video sequence motion estimation effect..

**Index Terms**—Estimation and Compensation; Video Image; Block Matching Algorithm; Reconstruction; Adaptive; DT Mesh; Improved SIFT

## I. INTRODUCTION

With the rapid development of information technology and Internet, multimedia communication services grow with each passing day, multimedia information has become the most important source for human to obtain information, research and development of multimedia technology has become a hot topic in the field of electronic information. Digitized multimedia easy encryption and high reliability; and can relay transmission and strong anti-interference ability; easy to very large scale integrated circuit (VLSI) implementation and low cost; convenient and computer networking and powerful interactive capabilities; easy to integrate the advantages in a public multimedia platform, so it has been widely applied in various fields. But the multimedia data generated is massive, it puts forward higher requirement for the information storage devices and the network communication, are beyond the capability of the existing communication technology [1].

In recent years, the new video business continues to rise, such as video conference, video on demand, remote monitoring, remote medical treatment, put forward higher requirements of real-time transmission and video quality

of video data, but the inherent limitations of Internet and wireless network poses significant problems for video applications, but also to the video data compression technology has put forward higher requirements. Efficient video compression is the video data compression is a problem that must be solved technology facing [2]. Data compression can remove redundant information in huge data, namely the removal of the correlation between data, information and keep the independent of each other, so the video compression is possible. In order to ensure the quality of the video information, compression efficiency is higher, the storage space or bandwidth requirement is lower. Therefore, the compressed video data is more important, how to efficiently implement video compression has become a research hotspot of video compression and communication field [3].

Because the overall continuity and moving objects in video motion, the current frame image can be thought of as the displacement, the movement of a previous frame image of objects in space. Therefore, displacement current motion frame objects in the image can be launched by the previous image frame [4]. So in the video image storage or transmission, all information does not need to be stored or transmitted each frame image, but only the motion information of moving objects are stored or transmitted. By using the motion information of the current frame image and the previous frame image can be restored from the current frame image. This paper presents the analysis of blind de-convolution, G function and slow evolutionary constraint algorithm, the approximate point spread function detection algorithm are studied and improved, and studies the problem of non uniqueness of solution of blind convolution in the video image, the experiments are carried out, a satisfactory effect.

From the point of view of information theory, video compression is achieved by removing redundant information in video information, with a more close to the essence of video information description to replace the original with redundant description. Statistical redundancy is redundant coding, derived from the encoded the probability density distribution of video image non-uniformity, can be removed by entropy coding. Structural redundancy and temporal redundancy and spatial redundancy, mainly from the video image data space (intra) and temporal (inter) correlation [5, 6]. The adjacent pixels in the same frame video image spatial correlation between sources, and the similarity between

the adjacent rows, can eliminate the spatial redundancy using spatial transform and vector quantization. The similarity between time correlation derived from video images of adjacent frames, namely the change between before and after the video sequence frame in most of the region is slow, and the background is almost unchanged, so the difference between adjacent frames is small, with very high similarity. The motion estimation and compensation, can effectively remove the temporal redundancy in video image [7].

In video sequence, the inter temporal redundancy is far greater than the intra frame spatial redundancy and the coding redundancy, therefore, motion estimation is very important in the system of video compression coding module, it directly affects the efficiency and the quality of video data compression coding. Motion estimation is more accurate, the coding efficiency is higher, decoded video quality is better [8-10]. At the same time, the computational complexity of motion estimation in video compression coding system in the largest, accounted for the entire system to calculate the volume of more than 50%. Therefore, motion estimation performance not only determines the quality of video image compression coding, but also fundamentally determines the real time performance of the whole video compression coding system [11]. Study on the efficient motion estimation algorithm has very important practical significance to improve the efficiency of video compression coding, is an effective way to solve the problem of video data compression. Therefore, research on motion estimation research has become the focus of many years of video coding field [12].

The starting point of this article is how to improve the precision and the subjective quality of the reconstructed image motion estimation, to reduce the computation complexity of motion estimation algorithms, to improve the efficiency of motion estimation [13]. This paper makes some studies on the redundant wavelet domain block matching motion estimation and compensation, then the video image for non translational motion, the DT mesh motion estimation and compensation in the redundant wavelet domain to do related research. The airspace prediction of initial search point method and search method is introduced in the redundant wavelet domain motion estimation, and made the improvement, proposed a redundant wavelet transform fast adaptive block matching motion estimation algorithm based on. The experimental results show that, the algorithm can keep good motion estimation precision, effectively reduce the motion estimation time, improve the efficiency of motion estimation, and image subjective quality is very good. With the existing redundant wavelet domain block matching motion estimation algorithm, this algorithm has obvious advantages in the quality and efficiency of motion estimation. Special video image sequences of different motion characteristics of motion estimation, the algorithm has a strong ability to adapt.

## II. RELATED WORK

### A. Motion Parameter Estimation

Block matching motion estimation algorithm is based on the translational motion mechanism to realize the motion estimation based on the. The basic principle is: first, each frame of the sequences is divided into many non overlapping sub blocks, and assuming the pixel sub block with motion consistency, and only used for translational motion, which ignores the amplification, rotation, occlusion secondary factor, and that the moving object in the image pixel value is not changes over time [14].

The matching block in the reference frame position is that position block before displacement current, block matching with the relative displacement of the current block called the motion vector, the matching block and the current block of pixels is called the residual difference. In the motion compensation, only matching block and the residual pixel values can fully recover the pixel value of the current block by block motion vector [15]. Block matching method between the macro block, the search area and motion vector as shown in Figure 1.

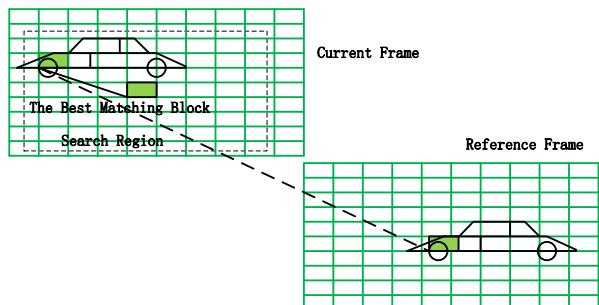


Figure 1. Block matching relationship between sub block, the search area and motion vector

### B. Optical Flow Method

Motion parameters refer to the reference image and the image to be matched in the corresponding image spatial displacement vector between the point, through a certain method to estimate the motion parameters of point like called motion parameter estimation [16-18]. Motion estimation is an important work of multi frame video image super-resolution reconstruction in displacement, provide accurate information for the low resolution image pixel interpolation to the corresponding position of high resolution grid, the reconstructed image quality depends on the degree of accuracy in the estimation of motion parameters [19].

Motion between images can be divided into two categories: global motion and non global motion [20]. Global motion estimation can be used to solve the general affine transformation model, rather than the global motion estimation problem solving approach are numerous, such as methods based on optical flow equation, based on the block matching method, and block matching and optical flow method combination algorithm.

$$S_c(x_1, x_2, t) = S(x_1 + v_1 \Delta t, x_2 + v_2 \Delta t, t + \Delta t) \quad (1)$$

Using the differential chain method, can be expressed as:

$$\frac{\partial S_c(x_1, x_2, t)}{\partial x_1} v_1(x_1, x_2, t) + \frac{\partial S_c(x_1, x_2, t)}{\partial x_2} v_2(x_1, x_2, t) + \frac{\partial S_c(x_1, x_2, t)}{\partial t} = 0 \quad (2)$$

### C. Local Polynomial Fitting

As the motion estimation of direct and indirect influence the coding efficiency and image quality, so the research of motion estimation algorithm in recent years become a hot research spot in the field of video compression. The main research contents of the motion estimation are how to quickly and effectively obtain more accurate motion vector. At present, the block matching algorithm (Block Matching Algorithm) is an effective and practical motion estimation algorithm, has been used widely and tremendous success in various coding standards.

Another alternative partial differential estimation method is by X, and the linear combination. Where low order polynomial of T from local approximation.

$$S_c(x_1, x_2, t) \approx \sum_{i=0}^{N-1} a_i \phi_i(x_1, x_2, t) \quad (3)$$

If the estimated coefficient  $a_1$ , can be obtained by simple calculation of gradient component [13]:

$$\frac{\partial S_c(x_1, x_2, t)}{\partial x_1} \approx a_1 + 2a_4 x_1 + a_6 x_2 + a_7 t \Big|_{x_1=x_2=t=0} = a_1 \quad (4)$$

$$\frac{\partial S_c(x_1, x_2, t)}{\partial x_2} \approx a_2 + 2a_5 x_2 + a_6 x_1 + a_8 t \Big|_{x_1=x_2=t=0} = a_2 \quad (5)$$

$$\frac{\partial S_c(x_1, x_2, t)}{\partial t} \approx a_3 + a_7 x_1 + a_8 x_2 \Big|_{x_1=x_2=t=0} = a_3 \quad (6)$$

Similarly, also easily from the coefficient of  $a_4$  to  $a_8$  estimate two orders and mixed partial derivative.

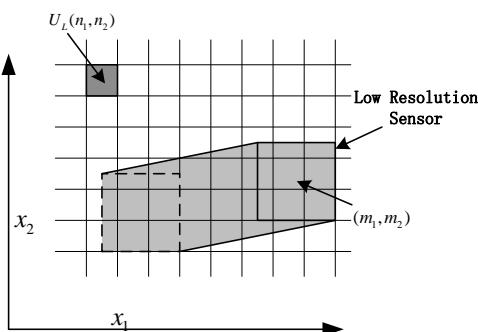


Figure 2. Block matching diagram

### D. Block Matching Method

Block matching is to find the best motion vector estimation through the pixel domain search procedures, the basic idea was shown in Figure 2, by considering a central location in K frame (the frame) pixel ( $N_1, N_2$ )  $N_1 \times N_2$  block, at the same time, search the  $K+1$  frame

(search frame) to obtain the best matching block position the same size, so as to determine the K frame ( $N_1, N_2$ ) [14-16]. Considering the displacement from a computational perspective, search is usually limited within the scope of the size of a  $M_1 \times M_2$ , the range is called the search window

## III. MOTION BLUR FOR SUPER RESOLUTION RECONSTRUCTION

### A. Video Imaging Model

In the exposure time in motion blur is usually in space and time changes, will not be able to eliminate motion blur from the SR restoration problem in separate as pretreatment process alone, so it is necessary to establish the initial model will consider the effects of motion blur. Continuous SR image is defined as a spatial density distribution on the focal plane of the camera (sensor is placed in the focal plane), it exist without any degradation factors in the ideal case. In fact, due to the impact of the LR camera system and other factors, the camera output for LR image. In the establishment of the LR imaging model, consider the following factors:

- (1) The motion compensation (caused by LR camera movement or scene content change);
- (2) The motion blur (by exposure time objects relative movement caused by);
- (3) The based on fuzzy (caused by the sensor unit physical size);
- (4) The optical blur (caused by the imaging optics);
- (5) The sensor noise

### B. Spatial Domain Block Matching Motion Estimation Theory and Wavelet Transform

#### 1) The Principle of Block Matching Motion Estimation

In order to reduce the effect of noise on image edge detection, Prewitt operator for edge detection template to expand, the Prewitt operator, it will direction difference operation and the average operation combination. The expression formula (7) and (8) are shown.

$$f_x(x, y) = f(x-1, y+1) + f(x, y+1) + f(x+1, y+1) - f(x-1, y-1) - f(x, y-1) - f(x+1, y-1) \quad (7)$$

$$f_y(x, y) = f(x+1, y-1) + f(x+1, y) + f(x+1, y+1) - f(x-1, y-1) - f(x-1, y) - f(x-1, y+1) \quad (8)$$

Evaluation of the effect of motion estimation algorithm, mainly see the matching calculation results and search complexity. Specifically, the matching effect can be obtained by subjective assessment of quality of the reconstructed image of the eye, but it has great randomness, and is not easy to make quantitative comparisons.

The corresponding convolution template is:

$$f_x(x, y) = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad (9)$$

$$f_y(x, y) = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (10)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (11)$$

$$\nabla^2 G(x, y, \sigma) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} = \frac{1}{\pi\sigma^2} \left(\frac{x^2 + y^2}{2\sigma^2} - 1\right) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (12)$$

Sobel based on Prewitt operator Sobel operator is proposed, which is the direction difference algorithm and weighted average arithmetic are combined, the expression formula (13) and (14) are shown. The operator combines the weighted average arithmetic, it can detect the image edge and further reduce noise, and easy to calculate, but it has larger value of edge gradient.

But  $\frac{\partial G}{\partial \sigma}$  can be approximately expressed as:

$$\frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \quad (13)$$

Comprehensive the above equation, we can get:

$$F(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \nabla^2 G \quad (14)$$

$$E(x, y, \sigma) \approx ((k-1)\sigma^2 \nabla^2 G) * I(x, y) \quad (15)$$

## 2) The Key Technical Indexes

Block matching criterion to judge the similarity degree is block basis, so the matching criteria will directly influence the accuracy of motion estimation; on the other hand, matching the computational complexity depends largely on the matching criterion. At present, the matching criteria commonly used has the minimum absolute difference of MAD (Minimum Absolute Difference), minimum mean square error MSE (Mean Square Error) and the normalized mutual function NCCFF (Normalized Cross Correlation Function).

Minimum absolute difference MAD

$$MAD(i, j) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N |f_k(m, n) - f_{k-1}(m+i, n+j)| \quad (16)$$

MAD (mean square error)

Minimum mean square error MSE:

$$MSE(i, j) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N [f_k(m, n) - f_{k-1}(m+i, n+j)]^2 \quad (17)$$

SAD, defined as follow:

$$SAD(i, j) = \sum_{m=1}^M \sum_{n=1}^N |f_k(m, n) - f_{k-1}(m+i, n+j)| \quad (18)$$

## 3) The Commonly Used Block Matching Motion Estimation Algorithm

At present, the most commonly used block matching motion estimation search algorithm includes the full

search algorithm FS (Full Search) and fast search algorithm. The full search algorithm, is to search all the search window, so as to find the best matching block, get the best motion vector, so it is optimal search algorithm. Therefore, the full search algorithm has the best peak signal to noise ratio, the quality of the reconstructed image is best, and the algorithm of rule, easily implemented by VLSI, but has a great computation and real-time performance is not very good.

In order to improve the real-time image compression system, generally use the fast search algorithm for motion estimation, which satisfy certain performance requirements, to improve the search speed by reducing the number of search points. Fast motion estimation algorithm using single peak characteristic of SAD, with the search away from the global minimum point, block distortion are monotonically increasing. Therefore, the assumption that the movement towards the global minimum matching error is monotone decreasing, along the SAD decreases in the direction of the search, can greatly reduce the motion estimation search points, improves the search speed. Fast search algorithm of the common main three step search method, the new TSS three step NTSS, four step search method, gradient method, 4SS BBGDS algorithm, HEXBS algorithm and DS hexagon diamond, motion vector field adaptive algorithm MVFAST and predictive motion vector field adaptive algorithm PMVFAST fast search algorithm.

### 4) The Basic Theory of Wavelet Transform

If the functions satisfy the admissibility condition:

$$C_\psi = \int_R |\hat{\psi}(\omega)|^2 |\omega|^{-1} d\omega < \infty \quad (19)$$

Called a admissible wavelet, the chemical signal  $f(x)$  with continuous wavelet transform as the base for the

$$N_\psi f(a, b) = \frac{1}{\sqrt{|a|}} \int f(x) \overline{\psi\left(\frac{x-b}{a}\right)} dx \quad (20)$$

$$(f \in L^2(R), (a, b) \in R, a \neq 0)$$

Wavelet transform is a multi-resolution analysis, the resolution can be adjusted by the value of  $a$ , just to meet the actual needs analysis. Therefore, wavelet transform has played a mathematical microscope role. Wavelet is the good property of time-frequency localization, can approximate the original signal is more accurate; in addition, the wavelet coefficient between different resolution correlation exists, can further improve the compression effect by using this correlation. Therefore, the wavelet transform in image and video compression applications are very good.

The symbol is defined as:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right) \quad (a, b \in R, a \neq 0) \quad (21)$$

$$N(x) = e^{-\frac{x^2}{2}} - e^{-\frac{x^2}{8}} / 2 \quad (22)$$

The frequency formula

$$\hat{N}(\omega) = \frac{1}{\sqrt{2\pi}} \int W(x) e^{-i\omega x} dx = e^{-\frac{\omega^2}{2}} - e^{-2\omega^2} \quad (23)$$

Select a certain time interval T, DOG continuous wavelet function can be divided into:

$$N(a, iT) = T \frac{1}{\sqrt{a}} \sum_n \psi \left[ \frac{(n-i)T}{a} \right] f(nT) \quad (24)$$

The biggest characteristic of wavelet transform is capable of multi resolution analysis of image signal, which laid the foundation for the motion estimation and motion compensation in wavelet domain video image. Motion estimation in the wavelet domain, the wavelet transform is performed on all of the current frame and the reference frame, and then the motion estimation and compensation directly in the wavelet domain. This method significantly reduces the motion estimation time, eliminate the block effect in the reconstructed image, can get a better peak signal to noise ratio and subjective quality.

#### IV. EXPERIMENTAL RESULTS

##### A. Fast Adaptive Motion Estimation and Compensation Algorithm Based on Redundant Wavelet Transform

This paper presents a fast adaptive block matching motion estimation and compensation algorithm based on redundant wavelet transform. The algorithm steps are as follows:

(1) The input video image of redundant discrete wavelet transform, wavelet coefficients and the sub-band block.

(2) If the image as the reference frame (I frame), then the decomposed sub-band images are stored and sampling, discrete wavelet transform DWT corresponding to the image, and go to step (5); otherwise, go to step (3).

(3) If the image is predicted frame (P frame), using adaptive prediction of initial search point method and fast adaptive block matching motion estimation search algorithm for motion estimation, the best motion vector of image blocks in the redundant wavelet domain to get. According to the error of each candidate vector and the final block motion vector prediction coefficients, each candidate vector method in the corresponding adaptive adjustment.

(4) The motion vectors obtained by the RDWT coefficients and motion estimation of I frames, each frame of P with image motion compensation prediction image respectively, P frame in the redundant wavelet domain, and calculate the residual image and the corresponding. Residual image of redundancy in wavelet domain and then down sampled, obtain the residual image based on DWT coefficients.

(5) Finally, image representation of DWT coefficient can be obtained by quantization coding

Figure 3 is the redundant wavelet transform fast adaptive block matching motion estimation and compensation algorithm based on the flow chart.

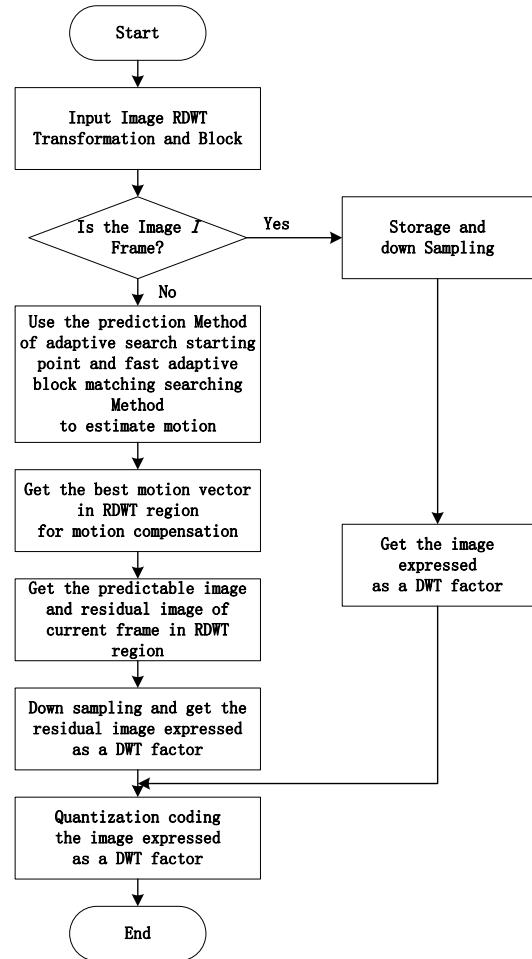


Figure 3. The redundant wavelet transform fast adaptive block matching motion estimation and compensation algorithm

##### B. SIFT Algorithm for Feature Point Extraction

In order to verify the validity of this method, the Matlab simulation experiment on standard video test sequences with different characteristics. The test sequence experiment used include: 120 frame "Akiyo" gray image sequence, 90 frame "Bus" gray image sequence and the 60 frame "Football" gray image sequence. Images are 256 gray level image of 352 \* 288 size, frame rate to 30fps (frames per second), which includes the motion information of the color video image almost all. "Akiyo" image sequences belongs to the head shoulder sequence, typical of slow motion, change, and the spatial details are not rich; "Bus" motion image sequence changes in medium and spatial detail is very rich; "Football" image sequence movement speed, variation.

Each frame of the peak signal to noise ratio (PSNR) and each frame for each macro-block average search points (ANSP) respectively was shown in Figure 4a, and Figure 4b.

In order to verify the proposed redundant wavelet domain adaptive prediction of initial search point method for motion estimation to improve the effectiveness of the test video sequence, simulation experiments. The specific method is: the median forecast method and adaptive

prediction method (referred to as the starting point prediction method), were used to search method based on RDWT diamond (RDWT-DS) and the RDWT fast adaptive block matching search method based on (referred to in this search method) for motion estimation, and then the motion estimation of each frame in the peak signal to noise ratio (PSNR) and each frame for each macroblock average search points (ANSP) are compared.

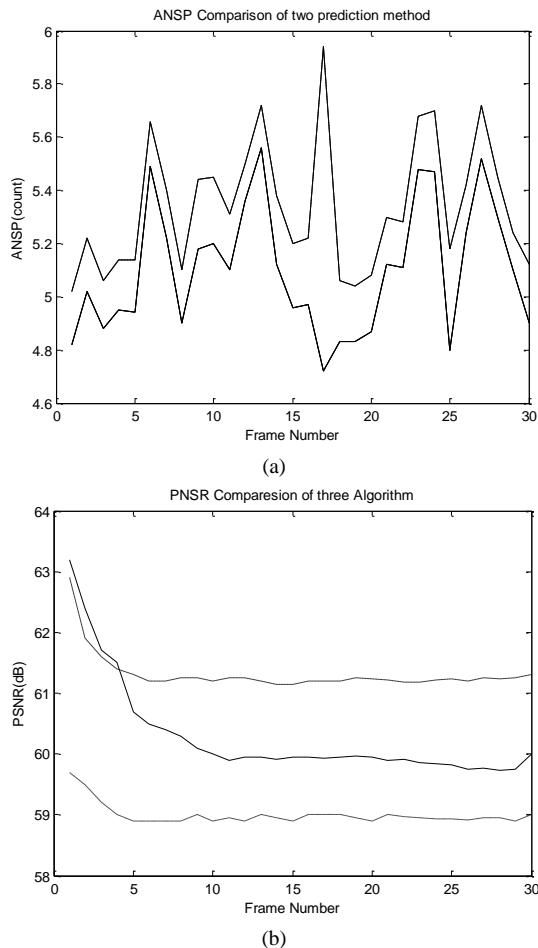


Figure 4. Each frame of the peak signal to noise ratio (PSNR) and each frame for each macro-block average search points (ANSP) respectively

TABLE I. COMBINED WITH PREDICTION METHODS AND TWO DIFFERENT SEARCH METHODS ON TWO DIFFERENT SEARCH POINT AVERAGE PEAK SIGNAL-TO-NOISE (APSNR) AND THE MEAN OF EACH MACRO BLOCK OF THE AVERAGE SEARCH POINTS (AANSR) COMPARISON

Sequence types	Comparative category	RDWT-DS	RDWT-DS	This method
Akiyo	APSNR	63	64	78
Akiyo	AANSR	5	5	1
Bus	APSNR	34	23	23
Bus	AANSR	11	11	13
Basketball	APSNR	58	59	59
Basketball	AANSR	13	14	10

In order to verify the proposed adaptive threshold prediction method and fast adaptive block matching search method which combines the improved algorithm of motion estimation in the overall performance of the

test video sequence, simulation experiments. The specific method is: the median forecast method and diamond search algorithm based on RDWT combining (RDWT-DS algorithm), the method and the starting point prediction based fast adaptive block matching search method based on RDWT combining

### C. Simulation Results and Analysis

In this paper, three series of experiments were carried out to verify the performance of the improved POCS algorithm for video image motion blur of super resolution reconstruction. Each group of experimental data are acquired by digital imaging equipment, image motion parameters according to the actual movement situation to select the appropriate method to calculate the. According to the character of Butterworth filter, to better maintain the edge details, select the parameter D=0.9, n=3, its frequency response as shown in Figure 5.

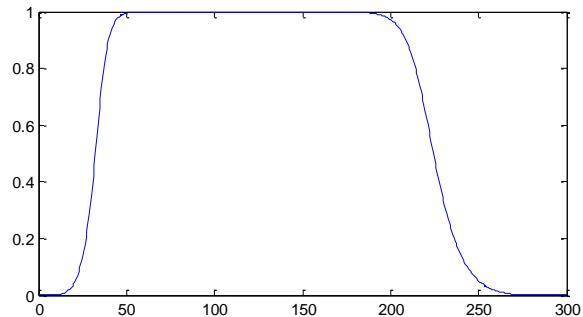


Figure 5. The six order Butterworth low-pass filter frequency response

In order to realize the multi-scale feature point detection, must build a DOG linear scale space. Therefore, Lowe build image Pyramid through the method of continuous convolution and decimation, in each layer of the Pyramid adjacent to the Gauss scale function convolution the image subtraction to construct DOG linear scale space. Figure 6 is a linear DOG scale space. The specific construction process: first, the original image is preprocessed to eliminate the influence of noise, a feature point extraction for stable; secondly the image of a linear interpolation by a factor of 2, sampling, in order to obtain more scale and feature point; the image sampling after, selected the scale factor K and initial scale sigma, constantly with scale of K,  $\sigma$  Gauss kernel convolution and won the first set, Gauss Pyramid in Figure 6.

In the scale space, extreme feature is the DOG scale space. In order to detect differential extremum points

Gauss function, each sampling point and its neighboring points to all comparison, each point should be compared with around the same scale image in 8 adjacent points and two adjacent scales in the image of each 9 adjacent point total of 26 points. At this point DOG values were less than or are higher than that of other points, said the point of extreme point.

The outdoor scene video sequence experimental data for the image size is 550x661 pixels. In the video, is composed of two parts, one part is in shadow, part of it is in natural light irradiation, the ability to detect moving

target detection algorithm in outdoor scene changes in natural light intensity. The training phase in Figure 7, (a) as the reference background, (b) image includes the foreground object is known. This paper introduces the GF space, a new polynomial kernel function GF space is proposed. The kernel function is not good, like other kernel learning, requires a predefined kernel functions. According to the actual situation of sample, the Gauss kernel function, the conventional polynomial kernel function or other forms.

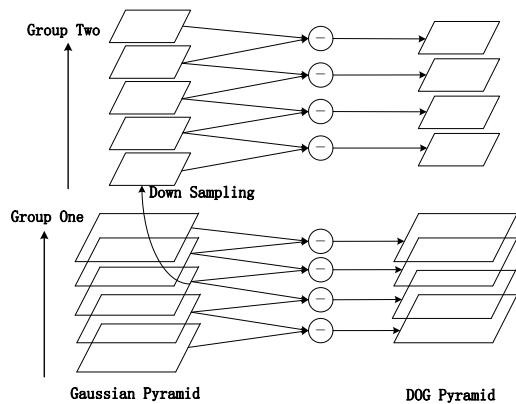


Figure 6. A linear DOG scale space



Figure 7. The training phase image

Deformation can be obtained by DT grid mesh nodes motion vector and the affine transformation, namely computational grid within each pixel motion vector, and then complete the motion estimation. But in the process of motion estimation, most of image pixels in adjacent frames are static, so the pixels of motion only affine transformation, in order to improve the efficiency of motion estimation. First, determine the grid within each

pixel is in the PMA. If it is, then the affine transformation to obtain the motion vector, otherwise, the provisions of the motion vector points is 0, which can effectively reduce the number of pixels and motion estimation time, improved the speed of motion estimation.

## V. CONCLUSIONS

Theory and technology of this method on the video image super-resolution restoration were studied and compared in depth, on the motion estimation, three aspects of video image motion blur and improvement of POCS algorithm for super resolution reconstruction and approximate point spread function APEX algorithm in-depth study based on, some deficiencies are effective improved, the meaningful research results have been obtained. This paper makes some discussion and Research on the redundant wavelet domain block matching motion estimation and compensation, then the redundant wavelet domain is triangulated irregular mesh motion estimation and compensation to do related research. This paper mainly discusses the value of algorithm, aimed at the theoretical research on the field of play certain reference significance.

The airspace prediction of initial search point method and search method is introduced in the redundant wavelet domain motion estimation, and made the improvement, proposed a redundant wavelet transform fast adaptive block matching motion estimation algorithm based on. Adaptive block matching criterion is proposed in this algorithm, can get smaller error block matching in motion estimation; the extraction potential moving blocks and termination of adaptive threshold search, it can overcome the shortcoming of the fixed threshold, can be well suited to the different characteristics of the image sequence and image block motion estimation. By dividing the block motion condition, only the potential motion block motion estimation, can effectively reduce the number of blocks in motion estimation, thus reducing the time required for motion estimation.

## ACKNOWLEDGMENT

The work was supported by the Science Foundation of Binzhou University, China (Grant No. BZXYG1101).

## REFERENCES

- [1] M. Sadek, A. Tarighat, A. H. Sayed, "A Leakage-based Precoding Scheme for Downlink multi-user MIMO Channels", *IEEE Transactions on Wireless Communications*, vol. 26, no. 8, pp. 1505-1515, 2008.
- [2] A. Tarighat, M. Sadek, A. H. Sayed, "A multi User Beamforming Scheme for Downlink MIMO Channels based on Maximizing Signal-to-Leakage Ratios", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1129-1132, 2005.
- [3] J. van de Beek, O. Edfors, M. Sandell, S. Wilson, P. Borjesson, "On Channel Estimation in OFDM System", in *Proceedings of the IEEE Vehicular Technology Conference*, pp. 815-819, 1995.
- [4] K. Wong, R. Cheng, K. B. Letaief, R. D. Murch, "Adaptive antennas at the mobile and base stations in an

- OFDM/TDMA system", *IEEE Transactions on Communications*, vol. 49, no. 1, pp. 195-206, 2001.
- [5] M. Sadek, A. Tarighat, A. H. Sayed, "Active Antenna Selection in multi-user MIMO Communications," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1498-1510, 2007.
- [6] Zhu Yazhou, Zheng Guoxin, Rui Yun., Li Mingqi, "A Novel Distributed Precoding Scheme Based on THP for Downlink Multi-Cell Multi-User OFDMA Wireless Systems", *IJACT: International Journal of Advancements in Computing Technology*, vol. 5, no. 9, pp. 213-220, 2011.
- [7] Raore Soungalo, Li Renfa and Zeng Fanzi, "Evaluating and Improving Wireless Local Area Networks Performance", *IJACT: International Journal of Advancements in Computing Technology*, vol. 3, no. 2, pp. 156-164, 2011.
- [8] 3GPP TS 36. 300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved UTRA (E-UTRA)", Dec. 2008. V8. 0. 0.
- [9] L Tong, G Xu, B Hassibi, and T Kailath. Blind channel estimation based on second-order statistics: a frequency-domain approach. *IEEE Trans. Inform. Theory*, Vol. 41. Jan 1995. pp. 329-334.
- [10] Y Zhao, A Huang. A novel channel estimation methods for OFDM mobile communication systems based on pilot signals and transform domain processing. *inPro. IEEE 47th Vehicular Technology Conference, Phoenix, USA*, 1997, pp. 2089-2093.
- [11] Theodore S. Rappaport etc. "Wireless Communications Principles and Practice. Publishing House of Electronics Industry. Mar. 2003.
- [12] Sinern Estimation Coleri, Mustafa Ergen, Anuj Puri, Ahmad Bahai. A Study of Channel in OFDM Systems. *IEEE VTC, Vancouver, Canada. September*, 2002.
- [13] D. B. Van, O. Edfors, and M. Sandle. On channel estimation in OFDM systems, Proc. *IEEE Vehic. Tech. Conf*, 1999. pp. 815-819.
- [14] Huang Kai, Zhou Yong-quan. Chaotic Artificial Glowworm Swarm Optimization Algorithm with Mating Behavior. *Computer Science*, 2012, 39(3):231-234.
- [15] Yang Xinshe, Hosseini S, Gandomi A H. Firefly algorithm for solving non-convex economic dispatch problems with valve loading effect. *Applied Soft Computing*, 2012, 12(3) pp. 1180-1186.
- [16] Deep K, Bansal J C. Mean Particle Swarm Optimization for Function Optimization. *International Journal of Computational Intelligence Studies*, 2009, 1(1) pp. 72-92.
- [17] Rahimi-Vahed A, Mirzaei A H, "A Hybrid Multi-Objective Shuffled Frog-Leaping Algorithm For A Mixed Model Assembly Line Sequencing Problem", *Computers and Industrial Engineering*, vol. 53, no. 4, pp. 642-666, 2007.
- [18] Zhang X C, Hu X M, Cui G Z, "An Improved Shuffled Frog Leaping Algorithm with Cognitive Behavior ", *7th WCICA Conf*, pp. 6197-6202, 2008.
- [19] Huynh T H, "A Modified Shuffled Frog Leaping Algorithm For Optimal Tuning of Multivariable PID Controllers", *IEEE Int Conf on Industrial Technology*, pp. 1-6, 2008.
- [20] Elbctagi E, Hegazy T, Griemon D, "Comparison Among Five Evolutionary-Based Optimization Algorithms", *Advanced engineering Informatics*, vol. 19, no. 1, pp. 43-53, 2005.

**Ruimin Chen** was born in the city of Binzhou, Shandong Province, China, in 1977. He received his master's degree in Kunming University of Science and Technology in 2005. He is currently a lecturer of Department of Information Engineering, Binzhou University, Shandong Province, China. His research interests are in image processing algorithm, software structure and application, information processing of the internet of things.

# New Video Target Tracking Algorithm Based on KNN

Ding Ma

College of software, Jilin University, Changchun, 130012, China

Email: martin3436@yeah.net

Zhezhou Yu\*

College of Computer Science and Technology, Jilin University, Changchun 130012, China

\*Corresponding author, Email: yuzz@jlu.edu.cn

**Abstract**—Compressed sensing is an advanced technique used in visual tracking problem. However, current research uses too many pre-knowledge, and cannot deal with quick object sensitively. Based on the problems, the paper proposes a new real-time visual tracking algorithm based on compressive sensing. On one hand, the algorithm uses the characteristics of compressive sensing, and make sure of the time constraint of the tracking. On the other hand, KNN classifier is used to differentiate the target and the background. Results show that the algorithm can provide the accuracy of the real-time tracking.

**Index Terms**—Video; Target Tracking; KNN

## I. INTRODUCTION

Applying computer vision to realize recognition and tracking of moving objects is widely applied to inspect factory products, monitor traffic flow and violation of laws. There are the following methods detecting moving objects, optical flow, frame difference method and background difference method. But most tracking methods of moving objects need to use models and matches of adjacent video frames, which not only make complexity of computation process high, but also is difficult to realize real-time tracking for objects. Compressed sensing can extract information from small amounts of data, which greatly reduces computing space tracking objects. So it is widely applied to real-time tracking on moving objects. However, tracking algorithm based on compressed sensing still has the problems that many video frames are used, it is not sensitive to objects with quick movement and it is not enough to model background information, which makes the performance of the algorithm not meet the requirements. Based on the problems, the paper proposes an improved real-time tracking algorithm based on comprehensive sensing. The algorithm uses the features of compressed sensing to extract features in the compressed domain, which ensures real time of track. On the other hand, KNN classifier is used to differentiate objects and background. The experimental results indicate that the algorithm can ensure improving accuracy of track.

In the second section, the achievements of real-time tracking are analyzed. In the third section, the concept of compressed sensing and KNN classifier is proposed. In the fourth section, the algorithm is described in detail. In the fifth section, the experiments verify the algorithm. Lastly, the conclusion is made.

## II. RELATED WORK

Real-time tracking is a popular research filed. In order to improve the performance of tracking to achieve the effect of real-time tracking, the researchers propose many algorithms [1-4]. The computation of relevant feature vectors needs to be made on video frame scale, which makes it difficult to improve the performance of tracking algorithm. The introduction of compressed sensing guides correct direction for the problem. Compressed sensing comes from the field of digital processing, and the essence is to reduce the cost measuring signals mathematically. The methods was introduced into signal processing field in 2006 [5, 6], and then it was applied to video processing. Gurbuz [7] and Cevher [8] applied compressed sensing to pattern recognition and create background model, which achieves good effect. Li [9] applied the method to track moving objects, used the method of compressed sensing sparsening signal reconstruction to reduce complexity, and proposed a tracking algorithm. And the performance of the algorithm improves 5000 times compared with common algorithms. The algorithm applies compressed sensing to moving target tracking. But the algorithm still needs to be improved for performance and accuracy. Based on the method, the research organizations make deep researches. Literature [10] proposes a multi-feature tracking algorithm based on compressed sensing. The algorithm not only is suitable for the track with great target texture and illumination change, but also ensures real time. Literature [11] proposes a video coding scheme based on compressed sensing. While ensuring real time, the scheme improves the security and stability of code. Moving object tracking based on compressed sensing has made progress, but there are deficiencies as follows. The algorithm needs to many videos besides the current video frame, which depends on the text. The research considers

texture change and light intensity, but there are problems for tracking objects with strenuous exercise. Under-utilization of background information makes the performance of tracking algorithm not ideal.

Based on the problem, the paper proposes a real-time tracking algorithm based on compressed sensing. The algorithm uses compressed sensing and KNN classifier, which not only ensures real time of tracking, but also can differentiate objects and background.

### III. COMPRESSED SENSING AND KNN CLASSIFIER

#### A. Introduction of Compressed Sensing Tracking Algorithm

From compressed sensing [12, 13], we can see that the vector  $x \in R^n$  with dispersion  $\eta$  can be restored by a measurement matrix with a great probability, and the formula is as follows.

$$y_i = \omega_i x, \quad x = 1, \dots, m \ll n \quad (1)$$

In the formula,  $\omega \in R^{m \times n}$ , is the measurement matrix.  $x$  can get the optimal solution by solving the following optimization problems after transformation. .

$$\min_x \|x\|_1, \quad s.t. \|\omega x - y\|_2 \leq \varepsilon \quad (2)$$

In the formula,  $\varepsilon$  is a pre-set constant. When the optimization problem is solved, greedy algorithm or convex-optimum method can be used to solve  $x$ .

#### B. KNN Classifier

KNN (K-Nearest Neighbor) [14, 15] algorithm is a typical machine learning algorithm. The core of the algorithm is that if a most of  $k$  adjacent samples of a sample in feature space belong to a category, the sample belongs to the category. In practice, KNN algorithm depends on limited adjacent samples for classification. Applying KNN algorithm to object tracking has long history [16]. The reasons which the paper uses KNN as the classifier are as follows. Firstly, constraint conditions of KNN are consistent with that of the algorithm in the paper, that is, the classifier needs some pre-classified samples. Secondly, KNN classifier has better performance than other classifiers. The classifier doesn't need training set for training, and the time complexity of training is a constant.

### IV. ALGORITHM DESCRIPTION

#### A. Using Improved Compressed Sensing Method for Feature Extraction

Matrix  $R \in R^{n \times m}$ , and the matrix can reduce the graphic from high-dimensional space to low-dimensional space. And the transformed formula is as follows.

$$v = Rx \quad (3)$$

In the formula,  $v \in R^m$ , which means the low-dimensional space, and  $x \in R^m$ , which means high-dimensional space,  $n \ll m$ . It is general that

measurement matrix is a random matrix satisfying Gaussian distribution. And  $r_{ij} \sim N(0,1)$ . The sparsity of the random matrix is not enough, so the computation needs great space. Therefore, the paper transforms the random matrix, and uses a similar matrix to compute dimensionality reduction. The generation type of improved random matrix  $R$  needs to obey the following rules.

$$r_{ij} \begin{cases} \ln k & \text{with probability } 1/2k \\ 0 & \text{with probability } 1-1/2k \\ -\ln k & \text{with probability } 1/2k \end{cases} \quad (4)$$

When  $k$  is 2 or 3, the matrix can be used for dimensionality reduction operation, which can retain information [17]. The transformation of 0-1 matrix can make that only non-0 elements and corresponding position of  $R$  is retained for storage and computation, which greatly reduces storage and computation space.

After computing a group of random matrixes  $R$  by off-line method,  $R$  can be applied for dimensionality reduction operation. At the beginning of tracking algorithm, the position of tracking objects is preset, and the position is used as benchmark to acquire a series of positive samples and negative samples. The positive samples make acquisition in preset area, and negative samples make acquisition in area which is far from preset areas. And dimensionality reduction operation is made for each sample, which can get low-dimensional feature vector.

Above all, the feature extraction methods in the paper are as follows.

Converting the image into gray level image, and the gray information of image is a two-dimensional matrix  $x$ .

Using (4) to get sparse matrix  $R$ .

Using  $R$  achieved in step b for sparsification (3).  $V$  is the classification feature.

#### B. Application and update of KNN Classifier

##### Initialization of KNN classifier

The first frame of video sequence needs to give object region, based on which positive samples and negative samples with the same number generate. The features of the sample set are used for exercising KNN, which can get KNN classifier used for the second frame.

##### Determinating location of objectives

Several samples are collected randomly around the location of the first frame of objectives. The features are extracted and entered into KNN classifier for classification. The number of background samples is  $N(n)=k_1$ , and the number of objective samples is  $N(p)=k_2$ . For each sample  $s(t)$  of video frame  $f(t)$  to be recognized, the number of samples  $Count(p)$  and  $Count(n)$  in constraint range is calculated by computing Euclidean distance. And the following formula is used to get the optimal solution.

$$\max_s \left( \frac{Count(p)+1}{Count(n)+1} \right) \quad (5)$$

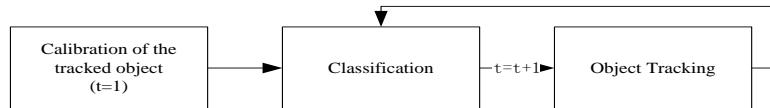


Figure 1. Algorithm framework

TABLE I. EXPERIMENTAL VIDEO INFORMATION

Figure caption	Resolution	Bit rate (kbit/s)	Frame rate (Hz)	Coded format
Figure 3	320*240	18.59	10	MPEG4
Figure 4	720*576	108.40	25	MPEG4
Figure 5	720*640	107.38	30	MPEG4

The center of positive samples is used as object position of the second frame.

In order to avoid the situation without match, under the condition without influencing results, 1 is added to member and denominator for calculation.

#### Update of KNN classifier

The objects have great light change and shape change in video sequence. Therefore, with the operation of track, KNN classifier needs to be updated. There is little change between adjacent frames. Therefore, after computing ( $t$ ), some results of the achieved non-optimal solutions are added to samples of classifiers, and KNN classifier is trained again, which completes classifier update.

The selection rules of adding samples to training sample set of classifiers are as follows.

$$top_s(\frac{N(p) + N(n)}{2}, Count(n)) \quad (6)$$

It means that the distance from the selected samples to the original classification center is less than the sample of the original average.

#### C. Algorithm Description

Based on dimensionality reduction strategy and classifier, the tracking algorithm is supposed as object recognition, and the specific tracking algorithm is given. The algorithm framework is as follows.

The algorithm description is as follows.

Initial condition: in the first video frame  $f(1)$ , the position of target object to be tracked is given.

Initial operation: the improved compressed sensing method is used to extract features of goal objects, and the method in chapter 4.2 is used to initialize KNN classifier.

Input: video frame  $f(t)$

Sampling  $f(t)$ , reducing each sample  $s$  from high-dimensional computation space to low-dimensional computation space, and extracting feature vectors.

Using KNN classifier to find the optimal tracking location  $loc(t)$  of the current frame.

Outputting  $loc(t)$  and updating KNN classifier.

$t=t+1$ . Judging if video ends. If it doesn't end, it returns to step 1. If it ends, the program exits.

Lastly, the motion location  $loc$  of objects is achieved in the video.

In initial state, hand-annotated way is used to get the location information of goal object to be tracked. The algorithm uses rectangular region as tracking area. In the beginning of the algorithm, dimension-reduction strategy

is used to sample the video frames, and the features are extracted, which can get sample set and establish KNN classifier. Figure 2 is the classification diagram. The diagram describes the operational principle of classifier. For the recognition of the  $t+1$  video frame, the optimal solution is computed according to the algorithm of classifier in Chapter 4.2. Computing samples of 20 areas can get the largest ratio, 0.87, which is the optimal solution. After solving the optimal solution, the optimal solution is returned, and the results are added to classifier, and the classifier is updated.

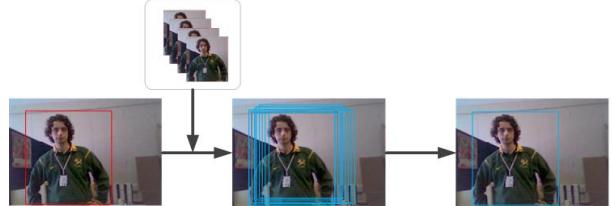


Figure 2. Working sketch map of classifier

#### V. PERFORMANCE ANALYSIS

In order to verify the performance and effect of the algorithm, the paper develops a real-time tracking model based on the algorithm in the paper. The prototype uses open video data as experimental data and evaluates tracking effect. The computer with CPU Intel i5 processor, 3.2GHz master frequency, and 4GB of memory is used for evaluation. The paper selects video data set under three conditions. Figure 3 is the recognition result with normal illumination. The video has 550 frames, and the initial state is (40, 40, 170, 200). In the video, tracking objects move or turn around, Figure 4 is the recognition results with light change, and the video has 275 frames. The initial state is (120, 100, 290, 480). In the video, light goes through the process from shadow to light. And in the process, tracking objects may be overlapped by other objects. Figure 5 is the recognition results under complex background. The video has 725 frames, and the initial state is (190, 210, 40, 110). The video shows a basketball match. In the video, tracking objects may move rapidly or be sheltered. The information of four videos is shown in Table 1.

The experimental results indicate that for video data in experiment, the algorithm can return recognition results in real time. The performance of the algorithm is similar to that of literature [9], which meets the requirements including real-time recognition and tracking.

We can get from experimental results that in most cases, the algorithm proposed in the paper can accurately track moving objects. From Figure 3 and Figure 4, we can see that the algorithm is not influenced by moving objects in video frame. From Figure 3, we can get that light change has little influence on the algorithm. From Figure 3 and Figure 4, we can see that the algorithm can process when the objects are overlapped. In three experiments, all recognition and track can be completed in real time, and the recognition results are feedback. In order to compare recognition results, we use the method of literature [9] and get some comparison results, and

uses tracking error to compare two recognition algorithms. In the paper, tracking error is defined as pixel value differencing between object position and recognition results. The comparison results indicate that for processing objects with light change and rapid movement, the algorithm proposed in the paper has good performance. Figure (6) and Figure (8) shows the comparison between the algorithm proposed in the paper and the algorithm in Literature [9]. For tracking error, we can see that the performance of the algorithm in the paper is good under the situation with light change and rapid movement.



Figure 3. Normal illumination ( Initial state ,1, 43, 128, 139, 236 frame)

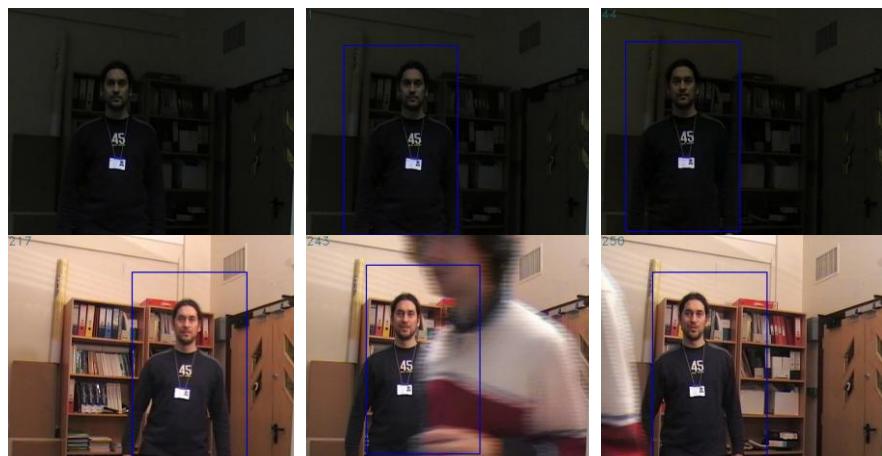


Figure 4. Light change and overlapping condition ( Initial state ,1, 44, 217, 243, 250 frame)



Figure 5. Complicated situation ( Initial state ,1, 8, 102, 114, 236 frame)

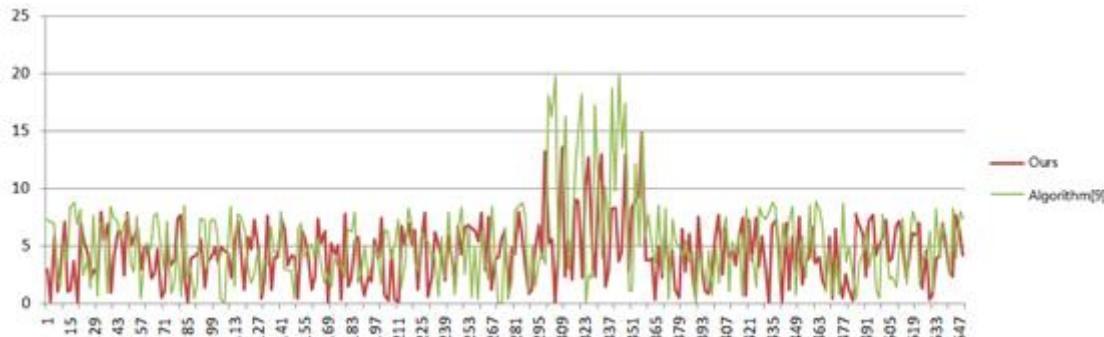


Figure 6. Comparison of video tracking error in Figure 3

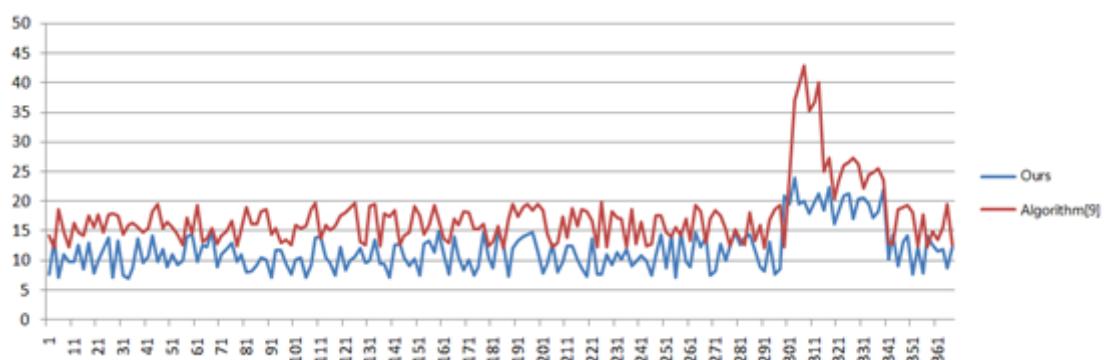


Figure 7. Comparison of video tracking error in Figure 4

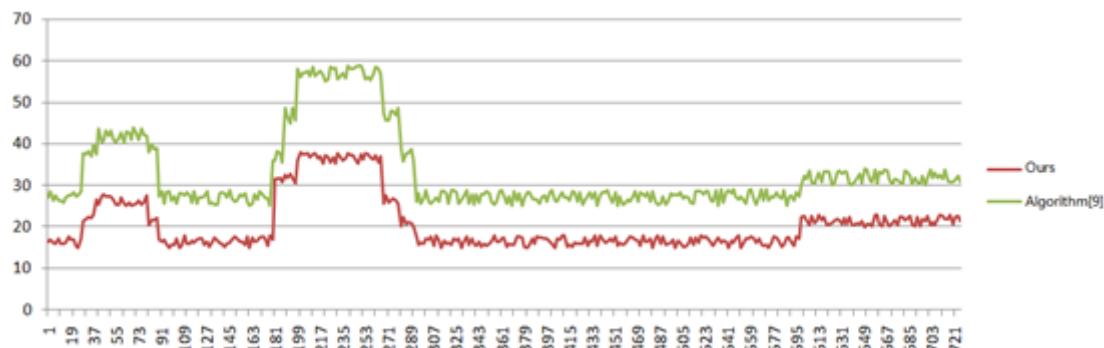


Figure 8. Comparison of video tracking error in Figure 5

## VI. CONCLUSION

Based on the problems of the algorithm based on compressed sensing, the paper proposes an improved real-time tracking algorithm based on compressed sensing. The algorithm uses compressed sensing, which greatly reduces computation space. And the algorithm uses KNN classifier to classify objects and background. The experimental results indicate that under the premise of ensuring real-time tracking, the algorithm improves tracking effect. But the algorithm still has some problems, which needs to be developed.

1. The algorithm can improve tracking performance of objects with rapid movement, but there are problems for too rapid objects. A self-adaption method needs to be designed to improve tracking effect.

2. The algorithm uses KNN classifier. In the subsequent work, other classifiers can be used for the algorithm. But exercise set still needs to be solved.

3. The experimental results indicate that the algorithm makes the size of the tracked objects change greatly because of the size of recognition frame in tracking process, and the recognition effect is not good. On the other hand, the algorithm doesn't consider multi-object tracking, which needs to be researched.

## ACKNOWLEDGMENT

Project supported by the Ph.D. Programs Foundation of Ministry of Education of China under Grant No. 20120061110045, and the Natural Science Foundation of Jilin province of China under Grant No. 201115022, and Graduate Innovation Fund of Jilin University No. 20121103.

## REFERENCE

- [1] Kalal, Z., K. Mikolajczyk, and J. Matas, Tracking-learning-detection. Pattern Analysis and Machine

- Intelligence, *IEEE Transactions on*, 2012. 34(7) pp. 1409-1422.
- [2] Bai Xiangfeng, Li Aihua, Li Xilai, Cai Yanping, Bandwidth self-adaption Mean-Shift tracking algorithm, *Computer Application*, 2011. 31(1) pp. 254-257.
- [3] Dang Xiaojun, Yin Junwen, Moving object tracking method based on template matching, *Computer Engineering and Applications*, 2010. 46(005) pp. 173-176.
- [4] Kaihua Zhang, Lei Zhang, Ming-Hsuan Yang, Real-Time Compressive Tracking, Proceedings of the IEEE International Conference on Computer, Florence, Italy, 2012 pp. 864-877.
- [5] Donoho, D.L., Compressed sensing. *Information Theory, IEEE Transactions on*, 2006. 52(4) pp. 1289-1306.
- [6] Candes, E.J., J.K. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 2006. 59(8) pp. 1207-1223.
- [7] Gurbuz Ali Cafer, McClellan James H, Romberg Justin, Scott Waymond R. Compressive sensing of parameterized shapes in images. in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. 2008. IEEE.
- [8] Cevher Volkan, Sankaranarayanan Aswin, Duarte Marco F, Reddy Dikpal, Baraniuk Richard G, Chellappa Rama, Compressive sensing for background subtraction, in *Computer Vision-ECCV 2008. 2008, Springer*. pp. 155-168.
- [9] Li, H., C. Shen, and Q. Shi. Real-time visual tracking using compressive sensing. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 2011. IEEE.
- [10] Zhu Qiuping, Yan Jia, Zhang Hu, Fan Cien, Deng Dexiang, Real-time tracking based on compressed sensing and multiple features, *Optics and Precision Engineering*, 2013. 21(2) pp. 437.
- [11] Guo Yansong, Yang Aiping, Hou Zhengxin, He Yuqing, Target tracking based on compressed sensing, *Computer Engineering and Applications*, 2011. 47(32) pp. 4-6.
- [12] Jiao Licheng, Yang Shuyuan, Liu Fang, Hou Biao, Prospect and retrospection of compressed sensing, *Journal of Electronics*, 2011. 39(7) pp. 1651-1662.
- [13] Wang Xinyue, Zhou Dequan, Research on image compressing technique based on compressed sensing theory, *Microprocessor*, 2013. 04 pp. 48-50.
- [14] Zhang Zhuying, Huang Yulong, Wang Hanhu, An efficient KNN classification algorithm, *Computer Science*, 2008. 35(3) pp. 170-172.
- [15] Pan Lifang, Yang Bingru, Research on KNN classification algorithm based on clusters, *Computer Engineering and Design*, 2009. 30(18) pp. 4260-4262.
- [16] Carbonell, J., et al. CMU report on TDT-2: Segmentation, detection and tracking. in *Proceedings of the DARPA Broadcast News Workshop*. 1999.
- [17] Achlioptas, D., Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of computer and System Sciences*, 2003. 66(4) pp. 671-687.

# Face Detection and Location System Based on Software and Hardware Co-design

Hua Cai, Yong Yang\*, Fuheng Qu, and Jianfei Wu  
 Chang Chun University of Science and Technology, Chang Chun, P. R. China  
 \*Corresponding author

**Abstract**—Face localization was an important research direction of face recognition technology. Its aim was to segment the face from the background of the detecting image. This technology was widely used in many areas of research, such as identity verification, HMI, visual communication, virtual reality, public files, etc. In this paper, we firstly constructed skin model and calculated the similarity of the image data, then calculated the face boundaries based space projection. Importantly, in order to improve the real-time, we utilized Xilinx high-level synthesis tool Vivado HLS (AutoESL) to achieve a hardware from a C program, which was based on Zynq platform. And the hardware module is used to realize the face location greatly improved the computing speed. The simulation results show that the proposed method worked well and the efficiency increased by 80%.

**Index Terms**—Face Location; Zynq; Software/Hardware Co-design

## I. INTRODUCTION

With the development of science & technology and the continuous progress of human society, computer vision had entered every aspect of social life. The biometric feature of a face was widely used in identity verification, HMI, visual communication, virtual reality, etc based on the visual analyzing. Face location was the first step in face feature analysis, which was responsible for the face detection and localization in a digital image. Face location was an important research content in face recognition and face expression analysis. Viola and Jones proposed the rapid real-time detection method in 2001[1] firstly, many face location algorithm was presented, such as region segmentation, gray projection, edge detection and template matching, etc. In [2], the region segmentation method was applied, but it couldn't find the effective position when the man wore the glasses. In [3], the gray projection method was utilized; however, the accuracy was impacted by the face gestures changing. In [4], the edge detection was adopted, nevertheless, a lot of data preprocessing was required and couldn't be implemented completely in real time. Meanwhile, with the increase of the image resolution, a single CPU platform had been unable to meet the demand for rapid detection. Parallel computing based on multi-core GPU was widely used in compute-intensive application. Because of the inherent parallel characteristics in face location, more and more people considered parallel

processing to accelerate the detection. NVIDIA launched CUDA computing platform in 2007 and CPU and GPU co-processing was presented in [5]. Although the ability of computing was powerful, it couldn't reduce its volume and power consumption. At the same time, due to heavy demand for mobile detection terminals, embedded face detection system was provoked attention. In [6], implement a parallel accelerating face detection system, but when the image size was changed, the system needed be revised and overhead was larger. In [7], GPU-accelerated face detection method achieved a certain effect, but its GPU load was not balance and impacted the accelerating effect.

In this paper, we presented an embedded face location system based on Zynq-7000.

Zynq-7000 series integrated a dual-core ARM Cortex-A9 (PS) and Xilinx 7 series FPGA (PL) on a single chip [8], which was the first integrated high-performance ARM CortexA series processors and high-performance FPGA products. Each Zynq-7000 chip contained the same PS, however, the PL and I/O resources were different. Compared to a single ARM Cortex A9 board or a single Xilinx FPGA board, Zynq series produced not only integrated different technology characteristic processor and FPAG on a single chip, but also designed the high-performance chip interconnection path between the processors and FPGA, meanwhile, Xilinx provided vivado HLS tools to convert the high-level language (such as C, Matlab, etc) into a hardware module, which could accelerate the software to improve the system's real time performance [9].

In this paper, we provided a face location algorithm in C language. In the face location algorithm, we adopted skin color segmentation method to achieve face region and projected image after edge detection to locate face region. The algorithm was fast and simple to synthesize the C program into a hardware module based on HLS tools and Zynq platform. The experimental results indicate that the hardware module worked well and efficiency increased greatly.

## II. THE STRUCTURE OF ZYNQ

### A. Internal Structure of Zynq

The internal structure of Zynq contained processor system (Processing System, PS) and a programmable logic (Programmable Logic, PL) the two parts.

Application processing unit (APU) located in the PS part. APU contained two ARM Cortex A9 dual-core processors and two Neon coprocessors. They both shared 512KB L2 cache. Each processor had a high-performance, low-power cores, and independently owned L1 level 32KB cache of instruction and data. ARM Cortex A9 dual-core processor was based on the ARMv7-A architecture and supported for virtual memory, which could run 32-bit ARM instructions, 16-bit and 32-bit Thumb instructions, and Jazelle state 8 Java byte code. Neon coprocessor aimed to process media and signal, which was optimized for increased audio, video, image and voice processing and 3D graphics instruction. Zynq-7000 AP SoC architecture diagram [10] was shown as Figure1.

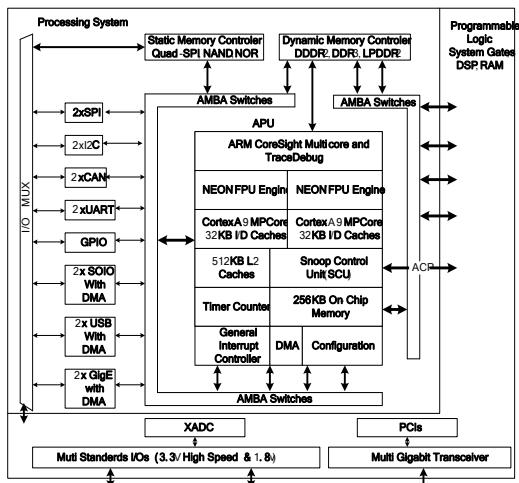


Figure 1. Zynq-7000 AP SoC architecture diagram

### B. SCU Module

In Figure1, Snoop control unit (SCU) module [11] was connected with the two ARM Cortex A9 processor and memory subsystem. It could intelligently managed data consistency between of two processors and L2 shared cache. This module was responsible for the data transmission, arbitration, memory interconnection and the cache coherence. The SCU module communicated with each ARM Cortex A9 processor by cache coherency bus (Cache Coherency Bus, CCB). SCU supported MESI (Modified, Exclusive, Shared, Invalid) monitoring system by avoiding unnecessary access to achieve higher efficiency and performance. SCU module implemented a four-way repeated association RAM to label local directory in the L1 data cache, which checked the relevance of the data. SCU could copy a clean data from one processor cache to another processor cache without participating of the main memory. In addition, it could move the dirty data between processors, while avoiding the delays caused by the sharing or writing back. Consistency management system provided by the independent hardware management unit not only reduced the complexity of the software, but also maintained the consistency across different operating systems and software drivers.

### C. Accelerator Coherence Port

Accelerator Coherence Port(ACP) was is 64-bit AXI slave interface, which provided an asynchronous cache coherent access points and could be accessed directly from PL SCU to ARM Cortex A9 MP-Core processor. PL could use this interface to access the entire APU cache and memory system, which could simplify software designing and improve system performance. As a standard AXI slave interface, ACP supported all standard reading or writing operation without adding any additional PL component. Therefore, ACP provided a cache memory interface to maintain cache consistency from PL end to ARM Cortex. When any read operation from ACP interface to a continuous memory region, SCU module checked whether the required information had been stored in the L1 cache. If it already existed in the L1 cache, the data would be returned directly. If the L1 cache missed, it still had the opportunity to L2 cache hit. If L2 cache was not hit, SCU would eventually be forwarded to the main memory.

### D. I / O Peripheral and Programmable Logic Peripheral

Zynq-7000 AP SoC included a lot of common internal I / O peripherals and memory interfaces and it was an important part of PS [12]. These I/O peripherals included GPIO, Gigabit Ethernet controller, USB controller, SPI controller and UART controller etc. These I/O peripherals in addition to complete common I/O functionality, but also made some changes for the Zynq-7000 AP SoC, so that it could be a good support for PS + PL architecture and flexibly used of PL. Storage interface on Zynq was also very rich, which included DDR controller, Quad-SPI controller and Nand / Nor / SRAM controller etc.

Programmable logic "peripherals" (PL) was Xilinx FPGA essentially. To facilitate understanding, on the Zynq platform, PL firstly could be seen as a reconfigurable "peripheral", which could be used as a part of the PS controlled by the ARM processor. Actually, PL could be reconfigurable to a variety of extensional peripherals, such as extended serial, Ethernet or video interface. Secondly, PL could be seen as a master device without ARM controlling. In this case, PL could actively complete data exchange through the interface with external chip, more even it could also be used as the master device to get data from the main memory of the APU, data storage and could control the ARM processor computing. PL part was added to the traditional ARM based SoC chip brought more flexibility.

### E. Connection between PL and PS

PL and PS used multiple interfaces and signals to achieve tighter or looser coupling and there were more than 3000 connections between these interfaces and signals. This ensured that the designers could efficiently complete hardware accelerators or other PL logic integration. PL could be designed to a module accessed by the processor or be designed to a module to access the main memory resources within the PS. When start-up the system, PS always started firstly and then used a software-centric approach to configure PL. PL could be

reconfigurable in the boot process or also be arranged at a future time. Particularly, PL could be fully reconfigurable or partly reconfigurable by dynamic configuration.

The bus between PS and PL data exchange was based on AXI bus protocol [13]. AXI bus, which was inside the PS pathway, had been designed to follow AMBA bus specification. As long as PL module designed by user complied with AXI bus protocol, it could communicate through these pathways with PS, such as providing data to the processor, accessed DDR, OCM and L2 Cache and so on. There were two types interface signals between PL and PS:

**Functional interface:** It included AXI, EMIO, interrupts, DMA flow control, clock, and debugging interface. When designers developed FPGA in the PL, the designers could use these interface signals to exchange data with PS. Signal interface for different purposes were not the same, should according to the need to use.

**Configuration Interface:** It included PCAP, SEU, configuration status signals and Program / Done / Init signal interface. These signals were connected to the built-in module within PL and provided PS with the ability to control PL.

In the hardware designing, we mainly used AXI, EMIO, DMA, PCAP interfaces.

Here we introduced two important concepts-MIO and EMIO in the Zynq platform. MIO meant Multi-function I / O, and EMIO stood for extend Multiplexed I / O. MIO was a I / O interfaces of PS, which had 54 pins. These pins could be used on GPIO, SPI, UART, TIMER, Ethernet, USB and other functions. Meanwhile, each pin had several functions simultaneously, so called multi-function I / O. Because of the number of MIO was limited, when it was not enough, Zynq provided us a method that was to use the extended MIO. The main feature was that when the MIO was not enough and at the same time wanted to use the I / O interface to connect the PL, from PL pin is connected to the outside of the chip. Another function using EMIO was to connect the PL module as a control signal, such as using the GPIO to control several PL modules. Using this method, we could skip the AXI bus to connect the modules, but we needed to consume CPU resources to control PL modules. The difference between MIO and EMIO was shown in Figure 2.

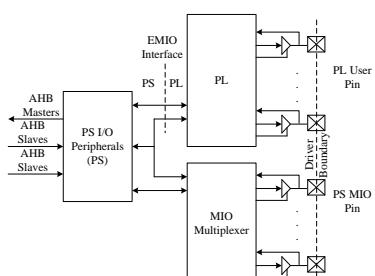


Figure 2. The difference between MIO and EMIO

#### F. Advanced eXtensible Interface

AXI (Advanced eXtensible Interface) protocol described the master devices and slave devices data

transmission. The master equipment and slave equipment was connected through the handshaking signal. When the data of master devices was ready, AXI maintained the VALID signal, which meant the data was valid. When the data of slave devices was ready, AXI maintained the READY signal. Both the VALID signal and READY signal were valid, the data transmission began. When these two signals continued to maintain valid, the master device would continue to transmit the next data. The transmission would be terminated when master devices canceled the VALID signal or slave devices canceled the READY signal. In essential, AXI was the upgraded version of AMBA (Advanced Microcontroller Bus Architecture) bus, which was proposed by ARM company. It was a high-performance, high-bandwidth, low-latency on-chip bus and could be used to replace the previous AHB and APB bus. AXI4.0 included three interface standards, namely AXI4 (also called AXI-Memory Map or AXI Full), AXI4-Stream and AXI4-Lite. Which AXI4-Stream was also defined by Xilinx and ARM together, designed for large data path FPGA design application preparation. AXI 4.0 protocol structure was shown as Figure 3.

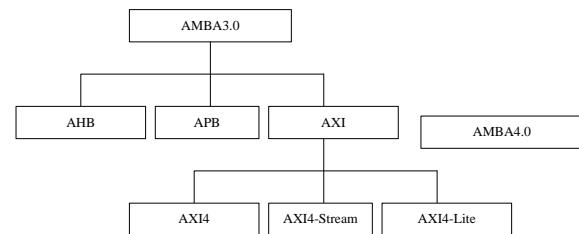


Figure 3. AXI 4.0 protocol structure

AXI4.0 protocol was equivalent to the AHB bus standard, which provided high-speed interconnect channel and supported burst mode. AXI4.0 mainly used access high-speed data storage. AXI4-Lite provided a single data transmission to peripheral, equivalent to the original APB protocol, which mainly used to access some of the low-speed peripherals. AXI4-Stream Interface was similar to FIFO. When data transmitted, it didn't need address and continuously read or wrote data directly from the main device to slave device. It mainly used in the case of high-speed data transmission, such as video, high-speed AD, PCIe, DMA. AXI4-Stream was similar to Xilinx's Local Link.

#### III. SOFTWARE/HARDWARE CO-DESIGN BASED ON ZYNQ

Software/Hardware co-design was defined as designing and coordination of hardware and software at the same time, which based on the definition of the whole system. Software/Hardware co-design included hardware and software division, hardware and software system development and joint debugging. Software/hardware co-design procedure was shown in Figure 4

First, the system was described in a high-level language, such as C language, Matlab, SystemC, etc. Then according to the system requirements to divide hardware and software, the system described in the

original C program was decomposed into a software part executed on the processor and a hardware part which was transformed from a C program to the hardware module. Third, if the hardware part required a special interface, designed it through the Zynq AXI interface and independently verified function. Fourth, in the software part, we should design software interface for the hardware and if the development involved in an OS then a driver would be designed for the coprocessor hardware. Finally, we needed to be combined hardware and software simulation and debugging together.

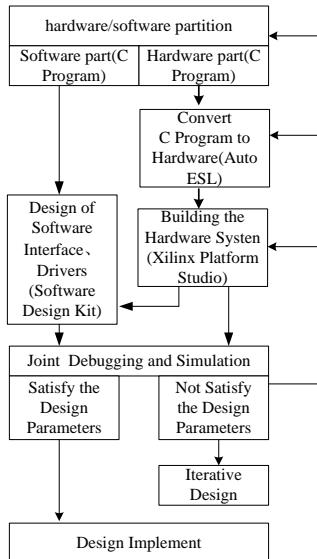


Figure 4. Software/hardware co-design procedure

#### IV. FACE LOCATION

Face localization [14] [15] was an important research direction of face recognition [18] [19] [20] technology and the process was actually human face detection. It was firstly determined whether someone's face in the input image, if where was a face in the image, then to determine the further information [21] on the human face. Skin color didn't depend on other features of the face and was not sensitive to the posture and facial expression changing. So it had the good stability and significantly different from the background color of most objects.

If we wanted to extract face region from the image, we must have a variety of skin color model under all kinds of light. So we needed an appropriate color space firstly. Because RGB tricolor color space tricolor represented not only color but also brightness, so it was not suitable for skin models. Due to changing in illumination, brightness makes it more difficult to locate the face. In this case, we could utilize clustering in color space by means of the skin and transform the RGB model into the chroma and brightness independent color space. Namely convert RGB space to  $Y\text{C}_b\text{C}_r$  color space, in which the components were unrelated.

Face color in the  $Y\text{C}_b\text{C}_r$  color space was relatively concentrated (referred to as color clustering features) so choose  $Y\text{C}_b\text{C}_r$  color space for face detection, RGB space was needed to complete the mapping  $Y\text{C}_b\text{C}_r$  color space.

$$Y = 0.29990R + 0.5870G + 0.1140B$$

$$C_b = -0.1687R - 0.3313G + 0.5000B + 128 \quad (1)$$

$$C_r = 0.5000R - 0.4187G - 0.0813B + 128$$

In the equation (1), R, G, B were the red, green and blue color values in the three color space. In the  $Y\text{C}_b\text{C}_r$  color space, Y was in the range of 16-235 and  $C_bC_r$  was in the range of 16-240. However, due to the Y and  $C_bC_r$  might exceed the range of 16-235 and 16-240 (video processing and noise), the following equation would be more convenient.

$$Y = \lfloor 0.257R + 0.504G + 0.098B + 16 \rfloor$$

$$C_b = \lfloor -0.148R - 0.291G + 0.439B + 128 \rfloor \quad (2)$$

$$C_r = \lfloor 0.439R - 0.368G - 0.071B + 128 \rfloor$$

Considering the hardware computing floating-point arithmetic was more complexity, we floored  $Y\text{C}_b\text{C}_r$  to an integer value, which denoted as  $\lfloor \cdot \rfloor$ .

In the  $Y\text{C}_b\text{C}_r$  color space, after normalized chromaticity histogram, through large amounts of data, we performed statistical computations to multiple samples facial skin pigmentation. From different color face data,  $C_r$  and  $C_b$  values projected to 3D graph as Figure 5 shown.

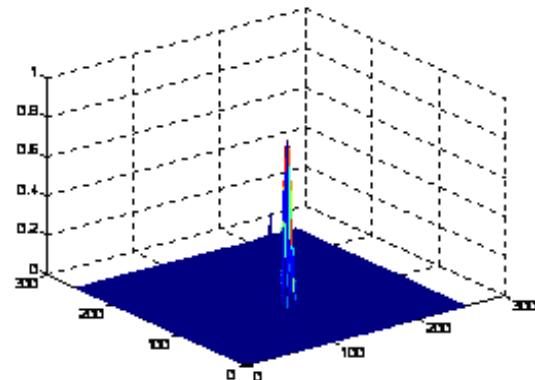


Figure 5.  $C_b$  and  $C_r$  projected to 3D graph

From Figure 5, we could see that skin color distribution was very similar to a two-dimensional Gaussian distribution. So we defined skin color met a two-dimensional Gaussian model [16]  $M = (m, C)$ , in which, m was the mean value

$$m = E(x), \quad x = (C_b, C_r)^T \quad (3)$$

C was the covariance matrix,

$$C = E((x - m)^T(x - m)) \quad (4)$$

Through the color model to detect the probability of an arbitrary pixel color was skin.

$$P(C_b, C_r) = \exp[-0.5(x - m)^T C^{-1}(x - m)] \quad (5)$$

After calculated every pixel in the image, we found out the maximum similarity  $P_{\max}(C_b, C_r)$  and then

normalized all pixel similarity in the range [0, 1]. The higher the value, the greater of the likelihood belonged to skin, whereas the smaller.

Assuming only two level gray scale in the image, one was the representative objects and the other was the background. There was significant difference between the two objects in the gray histogram. After statistical calculation, histogram statistics was bimodal and the simplest way was to select a valley point of the histogram of the gray as the threshold for image binaryzation. Shown as Figure 6.

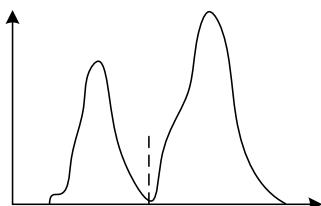


Figure 6. Bimodal gray histogram

After binarization processing, there was a lot of noise in the binaryzation image. Sometimes the noise would interfere the detection and location of the face region.

In this paper, we adopted morphology to remove the noise and smooth the boundary.

Efficiency of edge detection based on the grayscale morphology erode transform was proved by experimental results.

$$f \Theta b = \{x \mid b_x \subseteq f\} \quad (6)$$

In which,  $f$  meant original image and  $b$  meant a structure component. Erode transform was to find the region, where  $b$  could be put its in  $f$ . Erode transform was a shrinkage change. It always removed outermost layer of the original image, so it utilized to remove noise and thinning the image.

The other basic operation in morphology was dilation transform.

$$f \oplus b = \{x \mid \hat{b}_x \cap f \neq \emptyset\} \quad (7)$$

Dilation was a expansion procedure, namely connected the outer boundary and expanded the original image. Dilation transform aimed to eliminate holes and connect the cracks.

Erode transform and dilation transform were not inverse operation but they could be cascade connection. Using a same structure component to one original, firstly used the erode transform then dealt with dilation transform. This procedure was defined open operation, which was a fundamental operation in mathematical morphology.

$$f \circ b = (f \Theta b) \oplus b \quad (8)$$

Aimed to above mentioned binarization image, we used open operation to smooth the boundary and remove the burr. Make the outline of the face more clearly.

On the basis of analyzing the above-mentioned, we proposed projection method to locate the face region. Statistics the number of white pixels in each column, when the maximum value was obtained, using the values normalized of the other columns to determine the left and right boundaries. Statistics of the number of white pixels per line, when obtained the maximum value, using the max value normalized of the other rows to determine the upper and lower bounds. Finally, mark the boundary in the color image. The algorithm flow diagram was shown in Figure 7.

The face location results were shown in Figure 8.

Xilinx provided design tools to solve the traditional problem of multi-core heterogeneous [17]. Xilinx utilized high-level synthesis tool Vivado HLS (AutoESL) to achieve a C program into hardware. Xilinx utilized Platform Studio to achieve the hardware system design and construction. Xilinx Software DesignKit achieved the design software interfaces and drivers.

The whole system structure and system frame were shown as Figure 9.

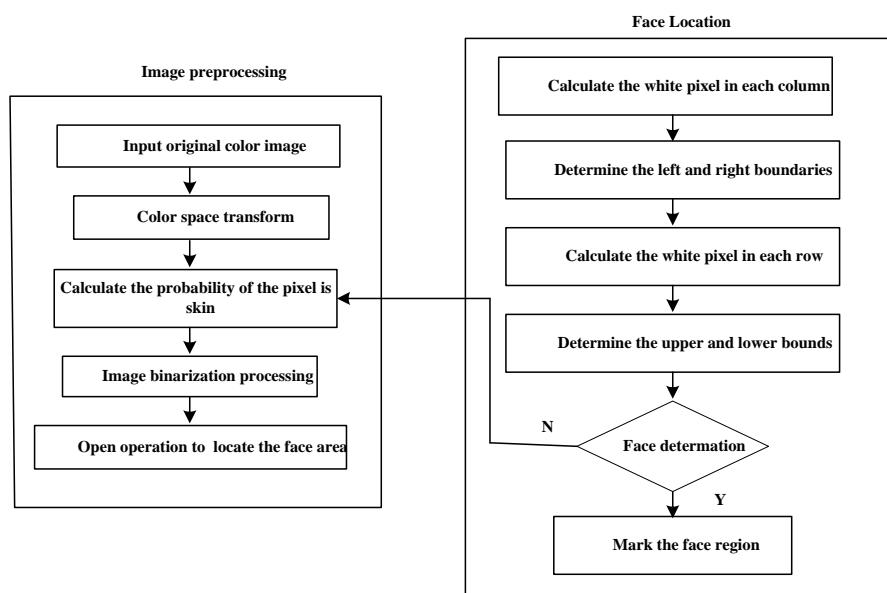


Figure 7. Algorithm flow diagram

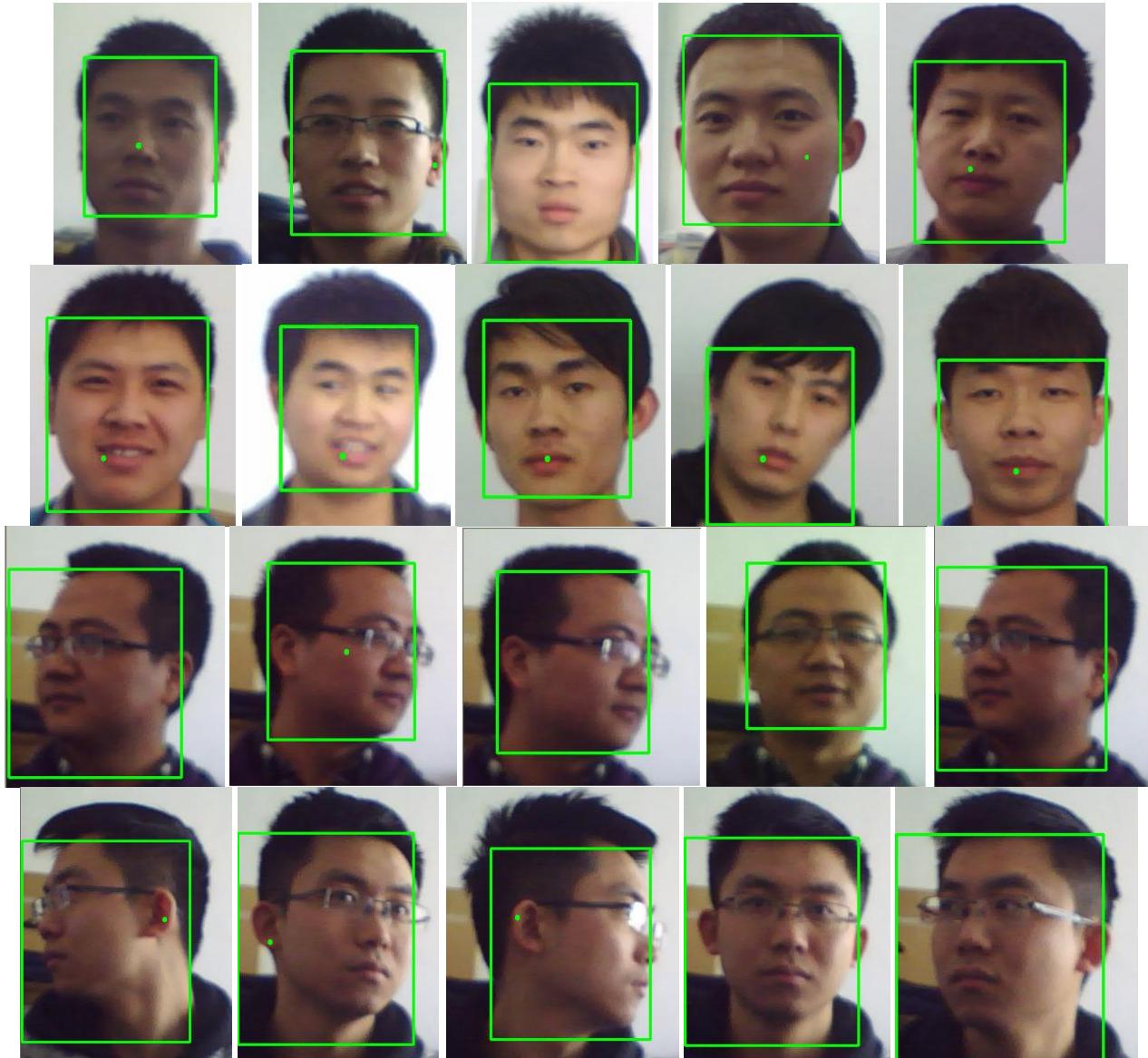


Figure 8. The results of face location

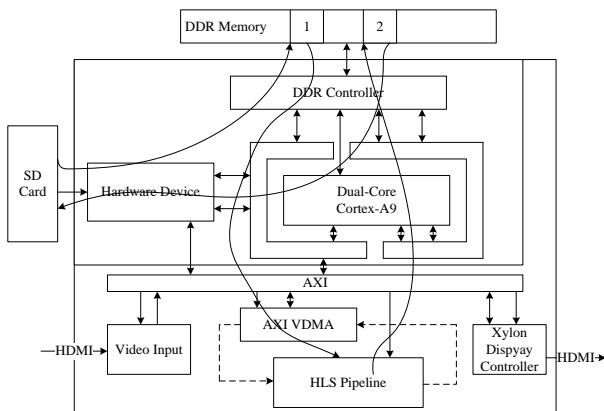


Figure 9. System frame of face location

## V. CONCLUSION

In this paper, we utilized xilinx Zynq 7000 platform to design a face location system, which applied hardware accelerator method converting a C program to a hardware

module. Through face location testing, the results demonstrated that using hardware module to complete a face location algorithm had very high performance and real time.

The experimental results had also shown that the design encouraged the decrease of the time occupied greatly and reduces the design cost. The result of FPGA verification shown that the system generated a good image quality, all this proved that the proposed hardware structure was valid and worked well, it could be integrated into the location, monitoring system well and used as the geometric normalizing operation of license plate recognition system. These technologies will broad application prospects in high-speed face recognition system in the future.

## REFERENCES

- [1] Viola P, Jones M. "Rapid object detection using a boosted cascade of simple feature", *Proceedings of the 2001 IEEE*

- Computer Society Conference on Computer Vision and Pattern Recognition.* 2001 pp. 511-518.
- [2] ZHU Shuliang, WANG Zengcai, WANG Shuliang et al. "Eye localization and state analysis of head multiposition in color image". *Journal of Chongqing University*, 2010, 33(11) pp. 20-26.
- [3] Wang WC, Chang F L, Zhang G Q, et al. "A precise eye localization method based on ratio local binary pattern". *Journal of Convergence Information Technology*, 2011, 6(1) pp. 126-134.
- [4] Sun W, Tang H Q. "Driver fatigue driving detection based on eye state". *International Journal of Digital Content Technology and its Applications*, 2011, 10(5) pp. 307-314
- [5] NVIDIA. CUDA\_C "Programming Guide [EB/OL]". [Http://docs.nvidia.com/cuda/cuda-c-programming-guide/](http://docs.nvidia.com/cuda/cuda-c-programming-guide/), 2013-04-02.
- [6] Junguk Cho, Shahnam Mirzaei, Jason Oberg, et al. "FPGA based face detection system using Haar classifiers", *Proceedings of the ACM/SIGDA International Symposium on Field Programmable Gate Arrays*. 2009: pp. 103-112.
- [7] Kong Jiangang, Deng Yangdong. "GPU accelerated face detection", *International Conference on Intelligent Control and Information Processing*. 2010: pp. 584-588.
- [8] Xilinx Inc. DS190, "Zynq-7000 Extensible Processing Platform Overview", 2012.
- [9] Xilinx Inc. UG702, "Partial Reconfiguration user guide" 2012.
- [10] Xilinx Inc. UG585, "Zynq-7000 Extensible Processing Platform Technical Reference Manual", 2012.
- [11] "ZedBoard-RevC. 1\_120824" Diligent Inc. 2012.
- [12] Xilinx Inc. UG744, "Partial Reconfiguration of a processor peripheral tutorial" 2012.
- [13] Xilinx Inc. DS768, "LogiCORE AXI Interconnect IP (v1. 0. 01a)", 2012".
- [14] Hsu R L. Mohamed A M, Jain A K. "Face detection in color images". *IEEE Transactions on PAMI*, 2002, 24 (5) pp. 696-706.
- [15] Erik Hjelmas. "Face Detection: A Survey", *Computer Vision and Image Understanding* 2001, 8(3), pp. 236-274.
- [16] XiYa Chen, Ying Wang, "A Novel Method of Face Features Location Based on the Skin-Color and Geometry Features", *Computer Engineering*, 2006, 32(3) pp. 212-222.
- [17] Xilinx Inc. Xapp890, "Zynq all programmable socsobel filter implementation using the vivado HLS tool".
- [18] Yunqi Lei, Haibin Lai, Qingmin Li, "Geometric Features of 3D Face and Recognition of It by PCA", *Journal of multimedia*, vol. 6 no. 2, April. 2011, pp. 207-216.
- [19] Jiadong Song, Xiaojuan Li, Pengfei Xu, Mingquan Zhou, " A New Face Recognition Framework: Symmetrical Bilateral 2DPLS plus LDA", *Journal of multimedia*, Vol 6, No 1 (2011), pp. 74-82.
- [20] Di Wu, Jie Cao, Huajin Wang, Wei Li, "Multi-feature Fusion Face Recognition Based on Kernel Discriminate Local Preserve Projection Algorithm under Smart Environment", *Journal of computer*, vol. 7, no. 10, October 2012, pp. 2479-2487.
- [21] Wenxing Li, Liming Lian. "Multiple Faces Tracking Based on Relevance Vector Machine". *Journal of software*. Vol. 7, No. 4. 2012. pp. 810-813.

# Video Image Object Tracking Algorithm based on Improved Principal Component Analysis

Wang Liping<sup>1,2</sup>

1. Engineering Technology Research Center of Optoelectronic Technology Appliance, AnHui Tongling Anhui 244000, China

2. Hefei University of Technology, Hefei Anhui 230009, China

**Abstract**—Since the existing object tracking algorithms are very difficult to adapt the object appearance changes caused by illumination changes, large pose variations, and partial or full occlusions, an object tracking algorithm based on two-dimensional principal component analysis (2DPCA) and sparse-representation is proposed in this paper. The tracking algorithm adopts 2DPCA and sparse-representation to establish object appearance model. In order to reduce dimension of object template, incremental subspace updating algorithm is introduced to online update the object template, reduce the requirement of memory space and enhance the precision of object appearance description. Experimental results show the proposed algorithm is robust for image illumination variance and object partial occlusion.

**Index Terms**—Object Tracking; Sparse-Representation; Incremental Learning; Subspace Updating Algorithm; Object Appearance Description

## I. INTRODUCTION

Object tracking [1] in video is one of the hot topics in the field of computer vision. It has a wide variety of applications, such as video conference, robotics navigation, virtual reality etc, which has important theoretical significance and practical value. However, object tracking result is usually affected by the background, brightness, objects posture, occlusion or lighting condition and so on, which makes the tracking process more complicated and becomes a challenging task. In order to overcome these difficulties, a large collection of algorithms are proposed to improve and refine object tracking. One category of existing algorithms is to detect and track each individual object using an independent tracker where multiple object tracking is performed by using multiple independent trackers [2]. Although tracking isolated objects is relatively simple, especially in simple backgrounds, tracking multiple objects in complex backgrounds remains a challenging task [3], especially in cases of long-term partial occlusions, object intersections, object pose changes, deformations, abrupt motion speed changes, cluttered background or dynamic complex background [4]. Therefore, it is worthy of studying the topic deeply that object tracking should exactly track the selected objects [5].

The existing object tracking algorithms can be classified into three categories: Correlated-template matching based tracking methods, Adaptive filter theory tracking approaches, Classification-based tracking algorithms [6]. The tracking algorithms based on correlated-template matching can achieve a higher accuracy with simple scheme, but the model is sensitive with the situation of partial occlusion, non-rigid image deformation and natural appearance changes [7]. The tracking algorithms based on adaptive filter theory can mainly be classified into two categories: extended Kalman filter and particle filter [8]. Kalman filter can only handle the scenario with linear, Gaussian and single mode, while particle filter adapts to nonlinearity and non-Gauss object tracking [9]. Classification-based tracking algorithms are regarded as a two-classification problem, where the object is tracked through distinguishing foreground from background using two thresholds [10]. It mainly utilizes the prior knowledge of each classifier to locate the moving or still object accurately, but two groups of samples are needed to construct classifier, which it is not appropriate for real-time scenario [11]. And in the process of tracking, there must be many sudden changes in the object image, and then the reasonable updating of the reference image will be the key to the tracking problem [12].

Recently, sparse representation theory can significantly improve the robustness of tracking in cases of long-term partial occlusions, object intersections, object pose changes, deformations, abrupt motion speed changes, cluttered background or dynamic complex background and reduce the computational complexity [13], which this technology is gradually applied and spread in object tracking. Scholars have proposed a robust visual tracking algorithm by viewing tracking as a sparse approximation problem in a particle filter framework [14] [15]. In this framework, occlusion, corruption and other challenging issues are addressed seamlessly through a set of trivial templates. Specifically, in order to find the tracking target at a new frame, each target candidate is sparsely represented in the space spanned by target templates and trivial templates. The sparsity is achieved by solving a  $\ell_1$ -regularized least squares problem. Then the candidate with the smallest projection error is taken as the tracking target. After that, tracking is continued by using a Bayesian state inference framework in which a particle

filter is used for propagating sample distributions over time. Although the proposed approach shows excellent performance in comparison with three previously proposed trackers, it is always unstable in the case of small objects or large deformation because of its high computation complexity and complicated calculation. Based on the above analysis, an object tracking algorithm based on two-dimensional principal component analysis and sparse-representation is proposed in this paper. The tracking algorithm adopts 2DPCA and sparse-representation to establish object appearance model. In order to reduce dimension of object template, incremental subspace updating algorithm is introduced to online update the object template, reduce the requirement of memory space and enhance the precision of object appearance description.

## II. OBJECT TRACKING BASED ON SPARSE-REPRESENTATION

### A. Feature Subspace

Assume  $\mathbf{Y}_t = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$  denotes an observed sample set of object on the  $t$  frame, where  $\mathbf{y}_i \in \mathbf{R}^{d \times d}$  is the  $i$ th observed sample,  $\mathbf{I}_t = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$  denotes the means of observed sample set  $\mathbf{Y}_t$ . In order to reduce storage costs and computation time of computer, an object tracking algorithm based on two-dimensional principal component analysis (2DPCA) is introduced to analyze the image. In comparison with PCA, 2DPCA is based on image method, which doesn't have to be translated into vector [16] [17]. Computing covariance matrix is simple and direct, and computing feature vector has less time. 2DPCA algorithm is described as follows:

Computing the covariance matrix of observed sample set  $\mathbf{Y}_t$ ,

$$\mathbf{S}_t = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{I}_t)(\mathbf{y}_i - \mathbf{I}_t)^T = \frac{1}{n} \mathbf{Y}_t \mathbf{Y}_t^T \quad (1)$$

Computing the eigenvalue of covariance matrix [18]  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_n$ ;

Computing the unit orthogonal eigenvectors relative to eigenvalues [19]  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ ;

Retain the first  $k$  eigenvectors to constitute feature subspace  $\mathbf{U} = \{\mathbf{u}_i\}_{i=1}^k$ .

### B. Sparse-representation

Assume the object template set and feature subspace can be denoted as  $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\} \in \mathbf{R}^{d \times k}$  ( $d \gg k$ ), where includes  $k$  object templates [20]. Therefore, the tracking results can use the feature subspace  $\mathbf{U} = \{\mathbf{u}_i\}_{i=1}^k$  to express approximately as the nonlinear curve [21]. Let us consider the discrete signal  $x$  of size  $N$ , which can be viewed as an  $N \times 1$  column vector in  $\mathbf{R}^N$  with elements  $x[n] \in \mathbf{R}^N$ . Specifically, using the

basis matrix  $\Psi = [\Psi_1, \Psi_2, \dots, \Psi_N]$  with the vectors  $\{\Psi_i\}$  as columns, a signal  $x$  can be sparsely expressed as

$$x = \sum_{i=1}^N s_i \Psi_i = \Psi s \quad (2)$$

$$\mathbf{y} \approx \mathbf{U}\mathbf{a} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_k \mathbf{u}_k \quad (3)$$

where,  $s$  is the  $N \times 1$  column vector of weighting coefficients in the  $\Psi$  domain,  $\mathbf{a} = (a_1, a_2, \dots, a_k)^T \in \mathbf{R}^k$  is a vector of object coefficient. Due to the interference, such as noise or occlusions in the process of object tracking, the error term is introduced, as shown as follows [22]:

$$\mathbf{y} = \mathbf{U}\mathbf{a} + \boldsymbol{\varepsilon} \quad (4)$$

where,  $\boldsymbol{\varepsilon} \in \mathbf{R}^d$  is the error term caused by noise or occlusions and its non-zero element denotes the object has noise or the object is occluded. Therefore, the identity matrix  $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d]^T \in \mathbf{R}^{d \times d}$  is introduced to the interference location [23]. So Equation (4) can be rewritten as follows:

$$\mathbf{y} = [\mathbf{U}, \mathbf{E}] \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \mathbf{D}\mathbf{c} \quad (5)$$

where,  $\mathbf{b} = (b_1, b_2, \dots, b_d)^T \in \mathbf{R}^d$  is noise coefficient,  $\mathbf{D} = [\mathbf{U}, \mathbf{E}]$  is an over-complete dictionary, and  $\mathbf{c}^T = [\mathbf{a} \ \mathbf{b}]$  is coefficient vector [24].

The signal  $x$  is K-sparse if it is a linear combination of only K basis vectors; namely, only K elements of the coefficients in (1) are nonzero and (N-K) are zero [25]. The measurement value of a signal  $x$  can be obtained by adopting nonlinear measurement, so  $y$  can be written as

$$y = \Phi x = \Phi \Psi s = \Theta s \quad (6)$$

where,  $\Phi$  is an  $M \times N$  measurement matrix with  $M < N$ . Because  $y$  is the  $M \times 1$  vector and  $x$  is the  $N \times 1$  column vector, so it is an underdetermined equation to recover  $x$  from only measurements  $y$ , which its exact solution cannot be obtained. If, however,  $x$  is K-sparse and the K locations of the nonzero coefficients in  $s$  are known, then the problem can be solved. According to the compressed sensing theory [26] [27], a sufficient condition for a stable solution for both K-sparse and compressible signals is that the matrix  $\Phi$  satisfies RIP criterion and is incoherent with the basis matrix  $\Psi$ . So the original signal  $x$  can be precisely reconstructed by solving the following  $l_1$  norm optimization problem [28].

$$s = \arg \min_s \|s\|_1 \text{ s.t. } y = \Phi \Psi s \quad (7)$$

Applying the relaxation to the problem (5) yields a new optimization problem as follows

$$\min \|y - Dc\|_2^2 + \lambda \|c\|_1 \quad (8)$$

The linear programming technology is adopted to solve the Equation (7), so the reconstruction methods include some greedy algorithms, such as OMP [29], StOMP [11] and CP [12] have recently been proposed. Thus, we can get the sparse solution of coefficient vector  $\mathbf{c}^T = [\mathbf{a} \ \mathbf{b}]$

$$\mathbf{c}^* = \arg \min_c \|y - \mathbf{D}\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 \quad (9)$$

By the solution  $\mathbf{c}^T = [\mathbf{a} \ \mathbf{b}]$  in above equations, reconstruction error between sample and template can be defined as

$$RE = \|y - \mathbf{U}\mathbf{a}\|_2^2 \quad (10)$$

### C. Incremental Subspace Updating Algorithm

Given the first  $n$  frames  $\mathbf{A} = [\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_n]$ , and the corresponding mean value  $\bar{\mathbf{I}}_A = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_i$ , we can get the centered matrix  $\bar{\mathbf{A}} = [\mathbf{I}_1 - \bar{\mathbf{I}}_A, \dots, \mathbf{I}_n - \bar{\mathbf{I}}_A]$ . If we denote the singular value decomposition of data matrix  $\bar{\mathbf{A}}$  as  $\bar{\mathbf{A}} = \mathbf{U}_A \Sigma_A \mathbf{V}_A^T$ , where  $\Sigma_i = \text{diag}(\lambda_{i,1}, \dots, \lambda_{i,r})$ ,  $r = \min(m, n)$  and  $\lambda_{i,j}$  is singular value of data matrix  $S_i$  (including zero), and  $\mathbf{U}_A$  and  $\Sigma_A$  are unitary matrix and diagonal matrix, respectively. In addition, each column in matrix  $\bar{\mathbf{A}}$  is basis vector of subspace. Assume  $\mathbf{B} = [\mathbf{I}_{n+1}, \mathbf{I}_{n+2}, \dots, \mathbf{I}_{n+m}]$  denotes  $m$  frames image, the corresponding mean value  $\bar{\mathbf{I}}_B = \frac{1}{m} \sum_{i=n+1}^{n+m} \mathbf{I}_i$ , so we can get  $\mathbf{C} = [\mathbf{A}, \mathbf{B}] = [\mathbf{I}_1, \dots, \mathbf{I}_n, \dots, \mathbf{I}_{n+m}]$ . In order to the unitary matrix  $\mathbf{U}_A$  and diagonal matrix  $\Sigma_A$  of matrix  $\mathbf{C}$ , the concrete steps are described as follows:

(1) Compute the mean value

$$\bar{\mathbf{I}}_c = \frac{fn}{fn+m} \bar{\mathbf{I}}_A + \frac{m}{fn+m} \bar{\mathbf{I}}_B \text{ of matrix } \mathbf{C}, \text{ where } f \text{ is a forgetting factor and is also a non-negative number (no more than 1)}$$

(2) Compute incremental centered matrix of  $\mathbf{B}$ :

$$\mathbf{B}^+ = [(\mathbf{I}_{n+1} - \bar{\mathbf{I}}_B), \dots, (\mathbf{I}_{n+m} - \bar{\mathbf{I}}_B), \sqrt{\frac{nm}{n+m}}(\bar{\mathbf{I}}_B - \bar{\mathbf{I}}_A)]$$

(3) Compute orthogonalization  $\tilde{\mathbf{B}}$  of  $(\mathbf{B}^+ - \mathbf{U}\mathbf{U}^T\mathbf{B}^+)$

and matrix  $\mathbf{R} = \begin{bmatrix} f \ \Sigma & \mathbf{U}^T \mathbf{B}^+ \\ \mathbf{0} & \tilde{\mathbf{B}}(\mathbf{B}^+ - \mathbf{U}\mathbf{U}^T\mathbf{B}^+) \end{bmatrix}$ .

(4) Decompose  $\mathbf{R}$  in singular value decomposition to get  $\mathbf{U}_R$  and  $\Sigma_R$ , thus  $\mathbf{U}_C = [\mathbf{U}_A \ \tilde{\mathbf{B}}]\mathbf{U}_R$ ,  $\Sigma_C = \Sigma_R$ .

### III. PARTICLE FILTER

The particle filter is an effective technique for the state estimation in non-linear and non-Gaussian dynamic systems. A novel method for object robust tracking based on particle filter has been proposed. Under the theory

framework of particle filter, the posterior distribution of the object is approximated by a set of weighted samples, while object tracking is implemented by the Bayesian propagation of the sample set. The state transition model is chosen as the simple second order auto-regressive model, and the system noise variance is adaptively determined in object tracking.

The objective of the particle filter is to estimate the values of the hidden states  $x_t$ , given the values of the observation process  $y_t$  at time  $t$ . According to the state transferring probability  $p(x_t | x_{t-1})$  and the observation probability  $p(y_t | x_t)$ , the posterior probability  $p(x_t | y_{1:t})$  can be derived and obtained by the following two steps: predict and update.

$$p(x_t | y_{1:t}) = \int p(x_t | y_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1} \quad (11)$$

$$p(x_t | y_{1:t}) = \frac{p(y_t | x_t) p(x_t | y_{1:t-1})}{p(y_t | y_{1:t-1})} \quad (12)$$

The optimal Bayes estimator can be constituted by prediction equation in (10) and updating equation in (11). Each particle has an assigned weight to represent the probability of that particle being sampled from the probability density function. Weight disparity leading to weight collapse is a common issue encountered in these filtering algorithms; however it can be mitigated by introducing a re-sampling step before the weights become too uneven. In the resampling step, the particles with negligible weights are replaced by new particles in the proximity of the particles with higher weights.

Assume the particle is  $\{x_t^i\}_{i=1}^N$  at time  $t$ , the normalized weight is denoted as  $\{w_t^i\}_{i=1}^N$ , and  $\sum_{i=1}^N w_t^i = 1$ . In other words, we use  $\{(x_t^i, w_t^i)\}_{i=1}^N$  to describe the posterior probability.

$$p(x_t | y_{1:t}) \approx \sum_{i=1}^N w_t^i \delta(x_t - x_t^i) \quad (13)$$

In addition, the weight  $w_t^i$  of particle is updated by using the following equation:

$$w_t^i = w_{t-1}^i \frac{p(y_t | x_t^i) p(x_t^i | x_{t-1}^i)}{q(x_t^i | x_{t-1}^i, y_{1:t})} \quad (14)$$

where  $\delta(\bullet)$ ,  $q(\bullet)$  are Dirac function and importance density function, respectively. Generally,  $q(x_t | x_{t-1}, y_{1:t}) = p(x_t | x_{t-1})$  is viewed as importance density function, where observations likelihood  $p(y_t | x_t)$  is weight. The maximum posterior estimation can be acquired by the optimal state estimation problem of object:

$$\hat{x}_t = \arg \max p(x_t | y_{1:t}) \quad (15)$$

### A. Motion Model

In the tracking framework based on particle filter, motion model is adopted to predict the available states of object between adjacent frames. Especially, rectangular box is used to locate object. The affine transformations of rectangular box are used to describe image scaling, translation, and rotation. In this paper, state vector of object is defined as

$$X_t = [x_t, y_t, \alpha_t, \beta_t, \varphi_t, \gamma_t] \quad (16)$$

where six parameters correspond to six affine transformation parameters of rectangular box, and denote as the translation between two images in x, y direction, scale variation, the aspect ratio, rotation and gradient, respectively. Given the probability model of State Transition is called the normal distribution or the Gaussian distribution, namely,

$$p(X_t | X_{t-1}) = N(X_t; X_{t-1}, \Psi) \quad (17)$$

where,  $N(\bullet)$  is Gaussian distribution,  $\Psi$  is a covariance matrix. In general, these parameters are independent of each other, and  $\Psi$  is a diagonal matrix where its elements are variance of affine parameters  $\delta_x^2, \delta_y^2, \delta_\alpha^2, \delta_\beta^2, \delta_\varphi^2, \delta_\gamma^2$ , respectively.

### B. Observation Model

Observation model measures the degree of similarity between candidate object region and object model. According to the Bayesian inference framework, observation model plays an important role in processing unknown scenario. Using the reconstruction error, the object observations likelihoods function can be defined as follows:

$$p(y|x) = \exp(-RE) \quad (18)$$

As we can see from Equation (18), the smaller the reconstruction error between sample and object template, so the larger its corresponding weight and the more reliable the sample will be. Our proposed tracking algorithm is given in Table 1.

### IV. EXPERIMENT AND ANALYSIS

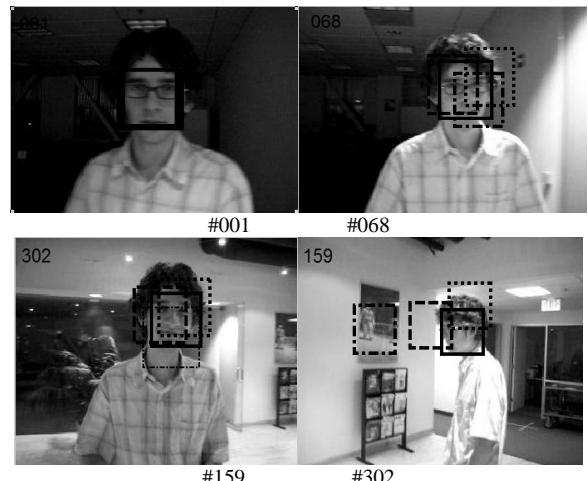
In order to validate the performance of our proposed algorithm, four standard video image sequences are selected to perform the tracking experiment. All of the experiments are run under MATLAB v7.8 (R2012a) on PCs with an Intel Pentium Dual Core CPU at 3.06 GHz and 4 GB memory. The proposed method is compared with three state-of-the-art methods which include Incremental Visual Tracking (IVT), L1tracker (L1), Multiple Instance Learning (MIL). In the actual experiment, the object position is manually selected in the first frame, the number of particle samples is 600, the dimension of object subspace is 16, and the update frequency of object subspace is set as 5. Throughout this study, the size of observatory image in our implementation is set as  $32 \times 32$ .

TABLE I. OUR PROPOSED TRACKING ALGORITHM

Our proposed tracking algorithm
Input: video image sequence $\{F_t\} (t=1, \dots, T)$ , the current number of particles $N$ .
Output: object location tracked in each frame $\{x_t\} (t=1, \dots, T)$ .
1. Initialization: manually choose the initial object template and disturb some pixels to generate $n$ templates; using the affine transformation, the $n$ templates are converted into image with the same size $32 \times 32$ , then 2DPCA algorithm is adopted to generate the feature subspace $U$ for object template set, namely, object templates set.
2. Construct over-complete dictionary: use feature subspace $U$ and $E$ to construct over-complete dictionary $D = [U \ E]$ .
3. Generate particles: according to Gauss distribution, use the initialized affine transformation matrix to generate $N$ particles (affine transformation parameters of candidate object templates).
4. Compute coefficient of sparse representation: minimize the Equation(8) in $l_1$ norm to solve sparse representation coefficient of each particle in template space, so as to get the coefficient-based affine coordinate of tracking object in the current frame.
5. Particle resample: according to the actual size of the weight, particles set is re-sampled for getting $N$ particles in the next frame.
6. Update object template: use Maximum A Posteriori Probability to estimate the particle $\hat{x}_t$ with maximal weigh, save the corresponding observation sample; at 5-frame intervals, use the updating algorithm to update object template.
7. Output result: show the tracking result in the current frame, go back to step 3.

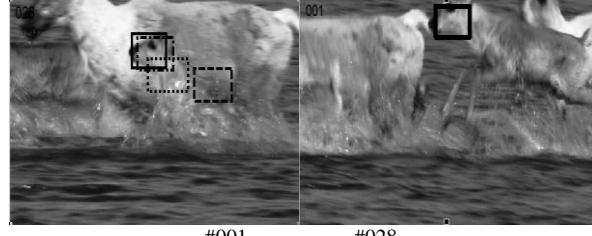
Figure 1(a) demonstrates David Indoor video sequences to evaluate the performance of tracking algorithm under the condition of illumination and posture change. In David Indoor sequence, there are two obvious light-intensity changes, and posture changes, such as picking glass and wearing glass. Throughout the whole tracking process, our proposed algorithm is not very sensitive to changes of illumination and posture, while L1, IVT and MIL algorithm show some deviation from the original object to some extent, where the MIL is more sensitive to all these influence.

In Deer video sequences shown in Figure 1(b), our proposed algorithm and MIL algorithm show a good performance. Although error tracking is appeared when the occlusion is occurred in #052 frames, MIL algorithm soon returns to the exact position. In addition, aiming at fast-moving object, L1 and IVT algorithm has poor tracking performance.

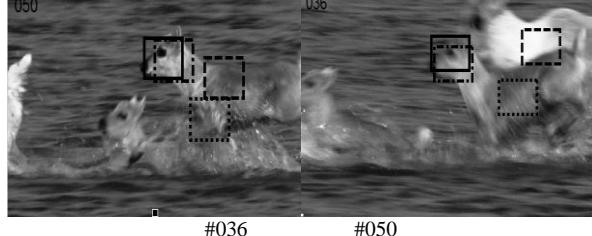




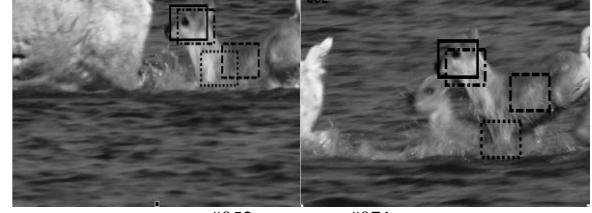
(a) Tracking comparisons of David Indoor in different algorithms



(b) Tracking comparisons of Deer in different algorithms



(c) Tracking comparisons of Car4 in different algorithms



(d) Tracking comparisons of Caviar1 in different algorithms



#001 #090



#115 #125



#150 #182

(d) Tracking comparisons of Caviar1 in different algorithms

Figure 1. Tracking comparisons of video sequences among our algorithm, IVT, L1 and MIL

As shown in Figure 1(c), our proposed algorithm and IVT algorithm show a good performance. Nevertheless, the tracking result of L1 algorithm has partial excursion after the vehicles pass over the bridge, while MIL algorithm has a larger excursion so that the tracking algorithm loses object.

Video sequences Caviar1 display the woman is walking down the aisle, as shown in Figure 1(d). The tracking result of MIL algorithm has the worst performance when appearing similar occlusion, and L1 and IVT algorithms have also some deviation from the original object to some extent, while our proposed algorithm shows a good performance in condition of partial occlusion or similar interference.

In addition, we adopt the location error to analyze quantitatively the performance of our tracker and the referenced trackers, where the location error is Euclidean distance between object location and tracking location, as shown in Figure 2. The maximum, average and standard deviation of each algorithm are listed in Table 2.

The underlined text shows the optimal result of each algorithm in Table 2, as we can see that our algorithm and L1 algorithm have the similar optimal result in David Indoor; our tracking algorithm and MIL algorithm have steady tracking performance when appearing the occlusion in Deer; comparing with the other three algorithms, our tracking algorithm shows a good performance when processing Deer and Car4 video sequences; our tracking algorithm acquires the least means and standard deviation in Caviar1 sequences. According to the overall performance, our proposed

algorithm yields the best performance in the process of object tracking.

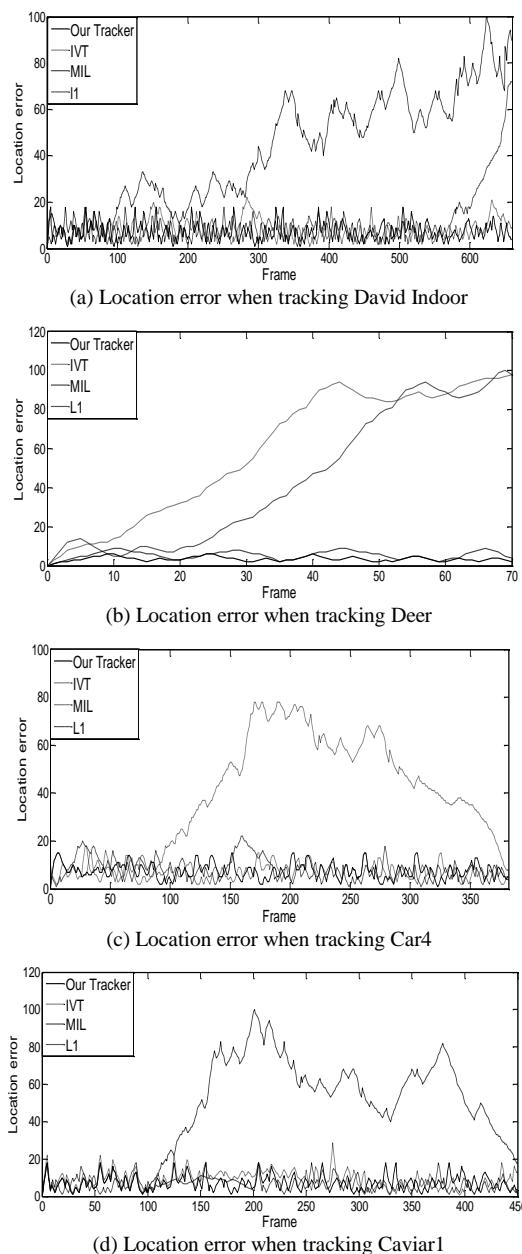


Figure 2. Quantitative evaluations in terms of center location error (in pixel). The proposed algorithm is compared with, IVT, L1 and MIL on four challenging image sequences

TABLE II. LOCATION ERROR

		David Indoor	Deer	Car4	Caviar1
IVT Tracker	Max	28.55	102.34	22.36	15.20
	Mean	9.18	59.10	8.39	7.50
	Std	4.45	32.82	4.22	3.16
MIL Tracker	Max	98.88	9.24	103.21	78.43
	Mean	45.25	5.46	42.46	38.61
	Std	27.22	2.09	24.38	21.82
L1 Tracker	Max	20.02	97.88	72.44	22.58
	Mean	6.77	44.01	10.72	8.15
	Std	3.35	34.63	10.91	4.48
Our Tracker	Max	21.23	6.35	18.21	15.32
	Mean	7.30	3.57	7.37	7.48
	Std	3.64	1.28	3.45	3.13

## V. CONCLUSION

An object tracking algorithm based on two-dimensional principal component analysis (2DPCA) and sparse-representation is proposed in this paper, which adopts 2DPCA and sparse-representation to establish object appearance model and avoid to process high dimensional data. In order to reduce dimension of object template, incremental subspace updating algorithm is introduced to online update the object template, reduce the requirement of memory space and enhance the precision of object appearance description. Experimental results show comparing with IVT, MIL and L1 algorithm, the proposed algorithm can track the object accurately and is also robust for image illumination variance and target partial occlusion. However, our algorithm only uses the global feature, which makes full occlusion hasn't been effectively solved. Therefore, local features have reasonably been combined to better describe object and the improvement of our proposed algorithm will be the next research priority.

## REFERENCES

- [1] Wagner, Andrew, et al. "Toward a practical face recognition system: Robust alignment and illumination by sparse representation." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34.2 (2012) pp. 372-386.
- [2] Rhodes, Matthew G., and Jeffrey S. Anastasi. "The own-age bias in face recognition: a meta-analytic and theoretical review." *Psychological bulletin* 138.1 (2012) pp. 146.
- [3] de Heering, Ad da ile, Bruno Rossion, and Daphne Maurer. "Developmental changes in face recognition during childhood: Evidence from upright and inverted faces." *Cognitive Development* 27.1 (2012) pp. 17-27.
- [4] Gui, Jie, et al. "Discriminant sparse neighborhood preserving embedding for face recognition." *Pattern Recognition* 45.8 (2012) pp. 2884-2893.
- [5] Tanaka, James W., Bonnie Heptonstall, and Simen Hagen. "Perceptual expertise and the plasticity of other-race face recognition." *Visual Cognition* ahead-of-print (2013) pp. 1-19.
- [6] Zou, Wilman WW, and Pong C. Yuen. "Very low resolution face recognition problem." *Image Processing, IEEE Transactions on* 21.1 (2012) pp. 327-340.
- [7] Zhong S, Xia K, Yin X, et al. The representation and simulation for reasoning about action based on Colored Petri Net//*Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*. IEEE, 2010 pp. 480-483.
- [8] Le, D., Jin, Y., Xia, K., & Bai, G. (2010, March). Adaptive error control mechanism based on link layer frame importance valuation for wireless multimedia sensor networks. In *Advanced Computer Control (ICACC), 2010 2nd International Conference on* (Vol. 1, pp. 465-470). IEEE.
- [9] Xia K, Cai J, Wu Y. Research on Improved Network Data Fault-Tolerant Transmission Optimization Algorithm. *Journal of Convergence Information Technology*, 2012, 7(19).
- [10] XIA, K., WU, Y., REN, X., & JIN, Y. (2013). Research in Clustering Algorithm for Diseases Analysis. *Journal of Networks*, 8(7), 1632-1639.

- [11] Yao, Yufeng, Jinyi Chang, and Kaijian Xia. "A case of parallel eeg data processing upon a beowulf cluster." *Parallel and Distributed Systems (ICPADS), 2009 15th International Conference on. IEEE, 2009.*
- [12] Kai-jian, Xia, et al. "An edge detection improved algorithm based on morphology and wavelet transform." *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on. Vol. 1. IEEE, 2010.*
- [13] Hofmann, Martin, et al. "Combined face and gait recognition using alpha matte preprocessing." *Biometrics (ICB), 2012 5th IAPR International Conference on. IEEE, 2012.*
- [14] Connolly, Jean-François, Eric Granger, and Robert Sabourin. "An adaptive classification system for video-based face recognition." *Information Sciences192* (2012) pp. 50-70.
- [15] Li, Annan, Shiguang Shan, and Wen Gao. "Coupled bias-variance tradeoff for cross-pose face recognition." *Image Processing, IEEE Transactions on 21.1* (2012) pp. 305-315.
- [16] Sugiura, Motoaki, et al. "Self - face recognition in social context." *Human brain mapping 33.6* (2012) pp. 1364-1374.
- [17] Mikic, Ivana, et al. "Moving shadow and object detection in traffic scenes." *Pattern Recognition, 2000. Proceedings. 15th International Conference on. Vol. 1. IEEE, 2000.*
- [18] Prati, Andrea, et al. "Detecting moving shadows: algorithms and evaluation." *Pattern Analysis and Machine Intelligence, IEEE Transactions on 25. 7* (2003) pp. 918-923.
- [19] Shotton, Jamie, et al. "Real-time human pose recognition in parts from single depth images." *Communications of the ACM 56. 1* (2013) pp. 116-124.
- [20] Shotton, Jamie, et al. "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context." *International Journal of Computer Vision 81. 1* (2009) pp. 2-23.
- [21] Gall, Juergen, et al. "Hough forests for object detection, tracking, and action recognition." *Pattern Analysis and Machine Intelligence, IEEE Transactions on 33. 11* (2011) pp. 2188-2202.
- [22] Gould, Stephen, Tianshi Gao, and Daphne Koller. "Region-based segmentation and object detection." *Advances in neural information processing systems. 2009.*
- [23] Wang, Yang. "Real-time moving vehicle detection with cast shadow removal in video based on conditional random field." *Circuits and Systems for Video Technology, IEEE Transactions on 19. 3* (2009) pp. 437-441.
- [24] Ladicky, Lubor, et al. "Associative hierarchical crfs for object class image segmentation." *Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009.*
- [25] Ommer, Björn, Theodor Mader, and Joachim M. Buhmann. "Seeing the objects behind the dots: Recognition in videos from a moving camera." *International journal of computer vision 83. 1* (2009) pp. 57-71.
- [26] Rhodes, Matthew G., and Jeffrey S. Anastasi. "The own-age bias in face recognition: a meta-analytic and theoretical review." *Psychological bulletin138.1* (2012) pp. 146.
- [27] De Heering, Adélaïde, Bruno Rossion, and Daphne Maurer. "Developmental changes in face recognition during childhood: Evidence from upright and inverted faces." *Cognitive Development 27.1* (2012) pp. 17-27.
- [28] Gui, Jie, et al. "Discriminant sparse neighborhood preserving embedding for face recognition." *Pattern Recognition 45.8* (2012) pp. 2884-2893.
- [29] Tanaka, James W., Bonnie Heptonstall, and Simen Hagen. "Perceptual expertise and the plasticity of other-race face recognition." *Visual Cognition ahead-of-print* (2013) pp. 1-19.

# Data Interpretation Technology for Continuous Measurement Production Profile Logging

Junfeng Liu

Key Laboratory of Exploration Technologies for Oil & Gas Resources (Yangtze University), Ministry of Education,  
Wuhan 430100, Hubei, China

Heng Li

College of Geophysics and Oil Resources, Yangtze University, Wuhan 430100, Hubei, China

Yingming Liu

Research Institute of Petroleum Exploration and Development, Beijing 100083, China

**Abstract**—Up till now, there is no production logging data interpretation module in CIFLog, which is the 3rd generation well-logging software platform in China. So the situation has a strong impact on its promotion and utilization. In this paper, firstly, the authors introduce the characteristics of the existing and mature logging interpretation software, and design the data interpretation module functions for continuous measurement production profile logging based on JAVA-NetBeans. Secondly, the calculation methods of apparent fluid velocity, holdup, superficial velocity and flow rate of each phase are presented. Thirdly, eight module functions including wellbore message, curve value, physical parameters, and parameter settings are described. Finally, the authors has analyzed three-phase flow production profile logging data of X well using this module, which includes seven parameters of continuous measurement, and provided the result chart and table. In a word, the practice has proved that the module application effect is good.

**Index Terms**—Java-NetBeans; CIFLog; Production Logging; Multiphase Flow; Continuous Measurement; Production Profile

## I. INTRODUCTION

The second generation of logging interpretation software in domestic, such as FORWARD, WATCH and LEAD, etc., is developed in Windows and Visual C++ environment [1-4]. FORWARD is open hole logging evaluation software platform in exploration stage, and it provides data management, preprocessing, interpretation evaluation, result output module. What's more, it has integrated domestic logging software results for many years, for example, conventional processing, purpose processing, and imaging logging data processing, and so on [5]. WATCH is a production logging data interpretation software in domestic. It combines the advanced technology of production logging evaluation and interpretation experience, considering the complexity and the oilfield production logging situation evaluation habits, and includes injection profile, production profile, engineering logging and reservoir parameter evaluation

[6]. Forward.NET is a web log data processing platform, and it is based on Visual Studio.Net development environment [7]. LEAD from CNPC provides open hole and cased hole data interpretation modules [8]. Smart is specifically applicable to Daqing oilfield in production logging interpretation software [9]. However, GeoFrame, Techlog, Express and LandMark are mainly used in the workstation. At the same time, they also have the function of logging data processing and interpretation [10-14]. Emeraude is the industry standard software platform for production log interpretation, and it offers Multiple Probe Tool (MPT) data processing, Pulsed Neutron Log (PNL) interpretation [15]. Research Institute of Petroleum Exploration Development has released the third generation of logging interpretation software platform named CIFLog in 2011, and it can work under three major operating systems, i.e. Windows, Linux and Unix, with characteristics of exploration well logging and production logging integration, data acquisition and processing network [16-18]. At present, there is no the data interpretation functions for production profile logging of continuous measurement in this platform. So it has affected the extension and application of the platform. In this paper, by using data reading, writing and management interface in main body frame data layer of CIFLog platform [19-24], the authors have compiled CMPPL (Continuous Measurement Production Profile Logging) module using JAVA language in the NetBeans environment and the module is suitable for data interpretation in production profile logging of continuous measurement. What is more, it has eight functions including wellbore information editing, curve values, cross-plot analysis, results view, physical property parameter conversion, interpretation parameter setting, data processing and results table mapping. The module can deal with of data interpretation for one-phase flow, two-phase flow and three-phase flow in borehole. On the whole, it has been tested and verified by Oil field examples well, and the results show that interpretation function and data accuracy are consistent with the similar software, so it has a good application potential.

TABLE I. SOME LOGGING SOFTWARE FEATURES

NO.	Software name	Developer	Environment	Application Range
1	GeoFrame	Schlumberger	workstation	open hole
2	Techlog	Schlumberger	workstation, PC	open hole
3	EXpress	Baker Atlas	Workstation	open hole, cased hole
4	LandMark	Halliburton	Workstation	open hole
5	Emeraude	Kappa	PC	cased hole
6	Smart	Daqing oil field	PC	cased hole
7	LEAD	CNPC Logging	PC	open hole, cased hole
8	FORWARD, FORWARD.NET, WATCH	Start Petrossoft Technique Ltd	workstation, PC	open hole, cased hole

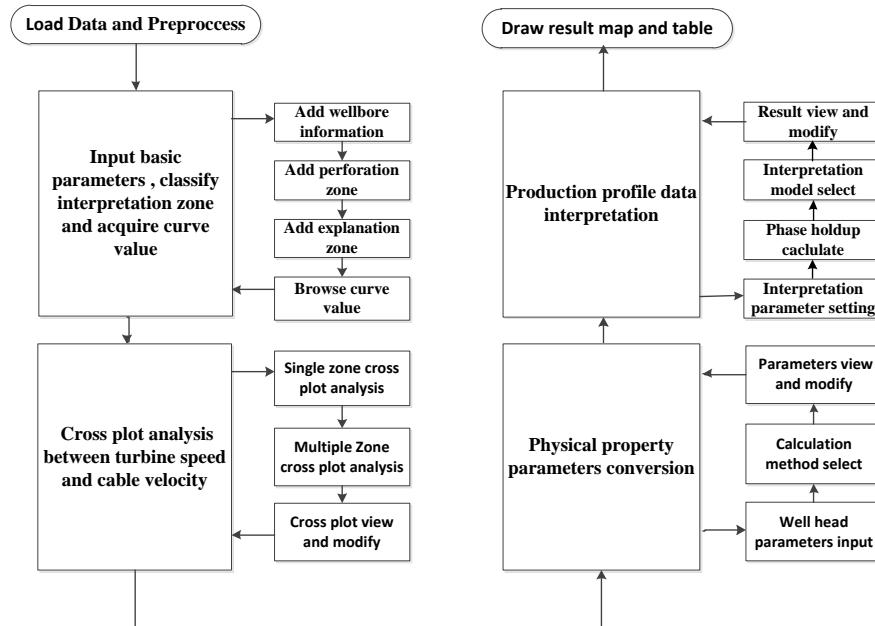


Figure 1. The overall design chart

## II. PRODUCTION PROFILE LOGGING TECHNOLOGY OF CONTINUOUS MEASUREMENT

Production profile logging of continuous measurement is a kind of complicated technology. That is to say, many instruments are connected to the cable, including temperature, pressure, casing collar, gamma, radioactive fluid density, capacitance water holdup and turbine flowmeter, and so on. When the instrument is working, the logging engineer will lift and lower cable for many times at a constant speed, and then obtain many kinds of information such as fluid types, flow pattern and flow rate for multiple production zones in the wellbore. So this logging technology can guide the well to maximize production [25]. Nowadays, the technology has been widely used in various fields at home and abroad. As shown in Table I, the corresponding production profile logging data interpretation technology mainly is represented by platform module from WATCH, FORWARD.NET, and LEAD in domestic, as well as Emeraude and PLATO professional software in overseas.

## III. INTERPRETATION MODULE DESIGN

The module design is shown in Figure 1, mainly including four parts, i.e. basic parameter, curve value, cross plot analysis between cable velocity and turbine

rotate speed, physical property parameter conversion and production profile data interpretation.

## IV. CMPPL MODULE ALGORITHMS

### A. Apparent Fluid Velocity

The authors use the least square method to calculate apparent fluid velocity of the whole flow layer or each zone with measurement by mixing up and down [26]. And the formulas are as follows

$$V_b = \frac{\sum_{i=1}^N LSPD_i * \sum_{i=1}^N CFB_i^2 - \sum_{i=1}^N (LSPD_i * CFB_i) \sum_{i=1}^N CFB_i}{N \sum_{i=1}^N CFB_i^2 - (\sum_{i=1}^N CFB_i)^2} \quad (1)$$

$$V_a = |V_b| + V_{off} \quad (2)$$

where,  $N$  is the number of measurements, dimensionless.  $i$  is from 1 to  $N$ , dimensionless.  $LSPD$ ,  $V_b$ ,  $V_{off}$  are cable speed, intercept of cable speed axle and turbine toggle speed respectively,  $m/min$ .  $CFB$  is flow frequency,  $r/s$ .

### B. Holdup

In the case of gas-liquid two-phase fluid, each phase holdup is obtained by the data of fluid density.

$$\begin{cases} Y_w = \frac{\rho_m - \rho_g}{\rho_w - \rho_g} \\ Y_g = 1 - Y_w \end{cases} \quad (3)$$

In the case of oil-water two-phase fluid, each phase holdup is obtained by the data of capacitance water holdup tool.

$$\begin{cases} Y_w = \frac{CPS - CPS_o}{CPS_w - CPS_o} \\ Y_o = 1 - Y_w \end{cases} \quad (4)$$

where,  $Y_o$ ,  $Y_g$ ,  $Y_w$  is oil holdup, gas holdup and water holdup respectively, dimensionless.  $\rho_m$ ,  $\rho_g$ ,  $\rho_w$  is the measuring response value of densimeter, and the scale value in total gas or total water, respectively,  $g/cm^3$ .  $CPS$ ,  $CPS_w$ ,  $CPS_o$  is the measuring response value of capacitance water holdup tool, and the scale value in total water or total oil, respectively, dimensionless.

Oil, gas and water three-phase fluid is considered to be water-hydrocarbon two phase flow. Firstly,  $Y_{w1}$  is calculated by using the data of capacitance water holdup tool. Secondly,  $Y_{w2}$  is calculated by using the data of fluid density in gas-water two-phase fluid. And then, after their average, you can get water holdup in three-phase fluid.

Where,  $Y_{w1}, Y_{w2}$  is water holdup.

### C. Flow Model

Interpretation model includes slip velocity model, drift flux model and special processing methods. In the case of oil-water two phase, the drift flux model is

$$\begin{cases} V_{so} = Y_o [C_o V_m + V_t (1 - Y_o)^n] \\ V_{sw} = V_m - V_{so} \end{cases} \quad (5)$$

$$V_m = C_v \times V_a \quad (6)$$

When flow pattern is bubble flow and slug flow in oil and water two-phase fluid,

$$\begin{cases} V_t = 1.53 \left[ \frac{g \delta (\rho_w - \rho_o)}{\rho_w^2} \right]^{0.25} \\ C_o = 1.2 \end{cases} \quad (7)$$

When flow pattern is bubble flow and slug flow in oil and water two-phase fluid, emulsion flow and foam flow,

$$\begin{cases} V_t = 0 \\ C_o = 1 \end{cases} \quad (8)$$

where,  $V_{so}$ ,  $V_{sw}$ ,  $V_m$  are oil superficial velocity, water superficial velocity and the average velocity of fluid,

respectively,  $m/min$ . And  $V_t$  is average drift velocity,  $m/min$ .  $C_o$  is the phase distribution coefficient.  $C_v$  is the velocity profile correction factor.  $g$  is gravity acceleration.  $\delta$  is the surface tension of oil and water.

### D. Multiphase Flow Rate

In the case of oil-water two-phase, each phase fluid flow rate in wellbore respectively are

$$\begin{cases} Q_o = V_{so} \times PC \\ Q_w = V_{sw} \times PC \end{cases} \quad (9)$$

where,  $PC$  is the pipe constant, which is calculated by wellbore internal diameter and instrument outside diameter.  $Q_o$ ,  $Q_w$  are oil and water flow rate respectively.

## V. CMPPL MODULE FUNCTION

### A. CMPPL Module Setting

Before the module running, it is necessary to set up form template with common table under the CIFlog platform, such as perforation data table, curve value table, cross plot data table, PVT parameter conversion data table, parameter table and interpretation result table, and so on.

### B. Module Features

After the module work, firstly, we need to load the well data and the corresponding graphics card template for data analysis. As the Figure 2 shows, it is important to note that templates can be used for fast graphics display with the same measurement curves in different well data. Secondly, after loading wellbore information and setting interpretation zone depth section for multiple perforation zones, the module identifies channel name and curve automatically. At the same time, as shown in Figure 3, we can choose the name of taking part in curve value and get curve value automatically with the mean value method, and then the result will be saved in “Value” table.

Thirdly, it may achieve cross plot analysis of each interpretation zone between cable velocity and turbine rotate speed. As shown in Figure 4, if the place of spot is inappropriate, we may move or delete that point and get apparent fluid velocity of each interpretation zone.

After that, as shown in Figure 5, we may import wellhead metering data to physical property parameters calculating (“Property”), and then we can calculate the parameters of each interpretation zone for crude oil, natural gas and formation water in high temperature and high pressure condition, such as bubble point pressure, density, viscosity and volume factor, and so on.

Finally, as shown in Figure 6 and Figure 7, considering the relationship of bubble point pressure and flowing pressure, setting interpretation parameters, importing geometrical parameters, and selecting the appropriate algorithm, for example, fluid phase, holdup computing method and interpretation model of multi-phase flow, we can achieve the total flow rate and each phase flow rate for each perforation zone.

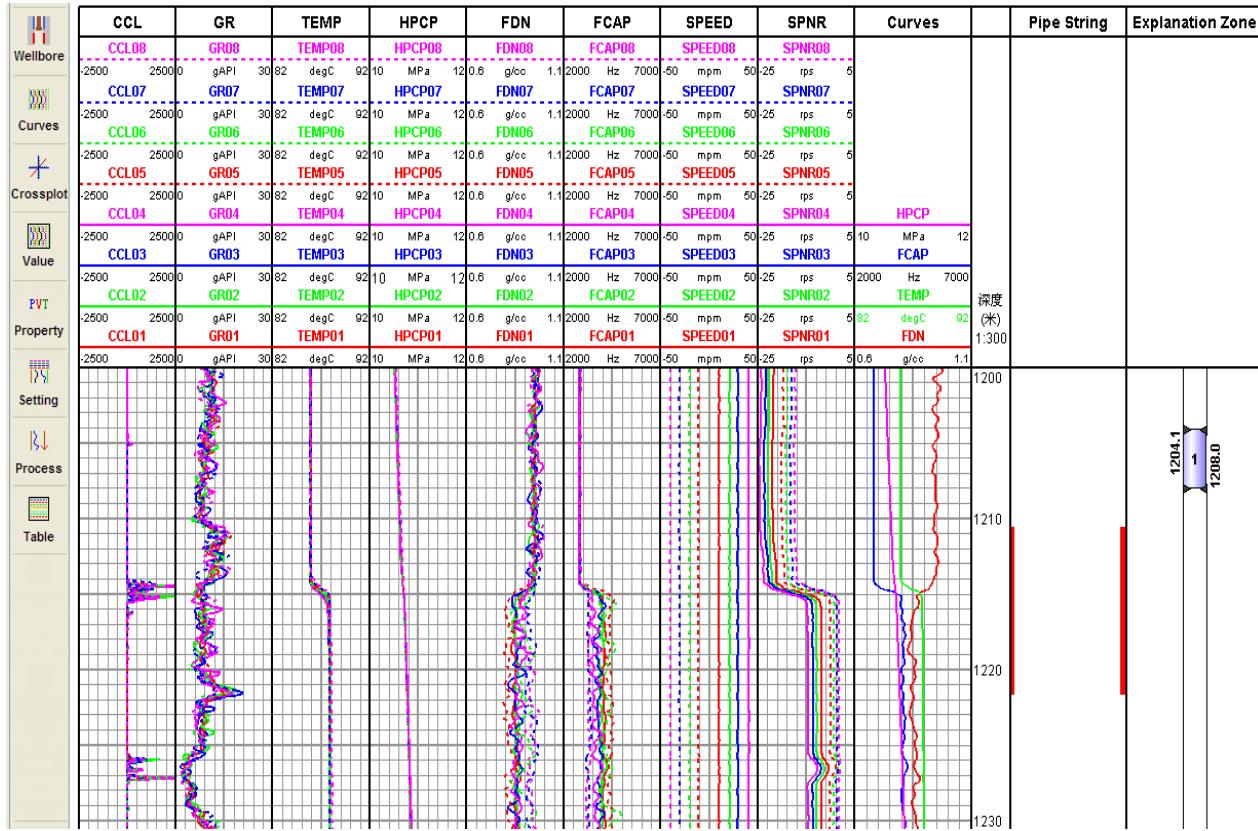


Figure 2. Measuring curve playback and explain zone setting (Depth: 1200-1230m)

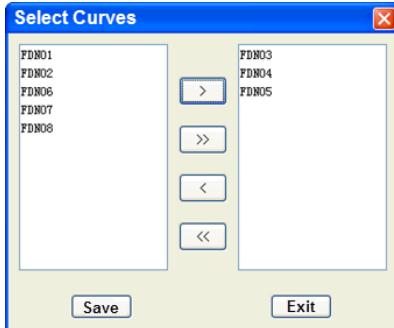


Figure 3. Acquire curve value setting

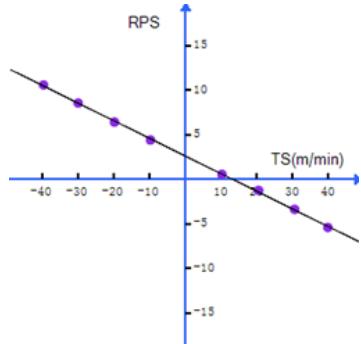


Figure 4. Cross plot analysis

## VI. EXAMPLE ANALYSES

As shown in Figure 8, X well from Xinjiang district is three-phase fluid flow in wellbore. Through production profile logging of continuous measurement, we can get

nine parameters as follow: natural gamma (GR), casing collar (CCL), fluid density, Pressure, temperature (TEMP), differential temperature (DTEM), Water holdup, turbine rotate speed (Flowmeter) and cable velocity(Logging velocity). Loading these curves to the CMPPL module, we can obtain two results as below: (1) From 4572 to 4575 meter well depth, comparing with the second zone, we can find that the curves of Flowmeter and Water holdup change lager significantly, but the Fluid Density curve changes smaller. Through the analysis above, we can conclude that this zone is the main oil production layer. (2) From 4575 to 4577 meter well depth, we can see the Flowmeter curve changes lager slightly, Fluid Density curve is constant on the whole and Water holdup curve has no significant change. Through the analysis above, we can conclude that this zone is the secondary production layer, and the fluid is almost water.

At the same time, Table II shows flow rate data of oil, gas and water in two perforation zones respectively. The wellbore produces whole fluid 92.53 t/d, oil is 56.98 t/d, water is 35.54 t/d and produces gas 4310.53m<sup>3</sup>/d, then we can get total water cut is 38.41%.

## VII. CONCLUSIONS

Based on the above research, main conclusions could be drawn as follows. In this paper, we applied multiphase flow model to compile the CMMPL module in java-NetBeans environment. And the module can deal with production profile logging data of continuous measurement. What is more, the module fills gaps in relevant fields. However, it is noteworthy that we test

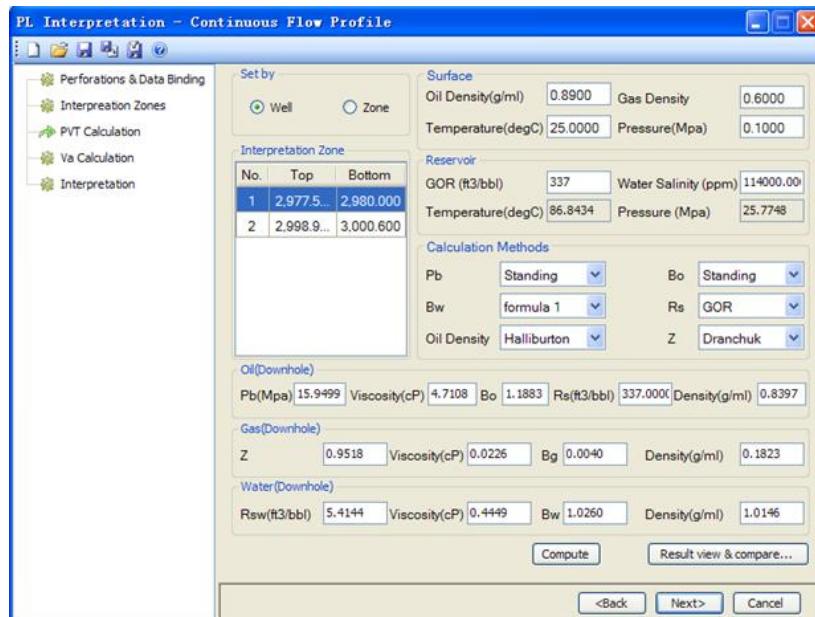


Figure 5. Physical property parameters conversion

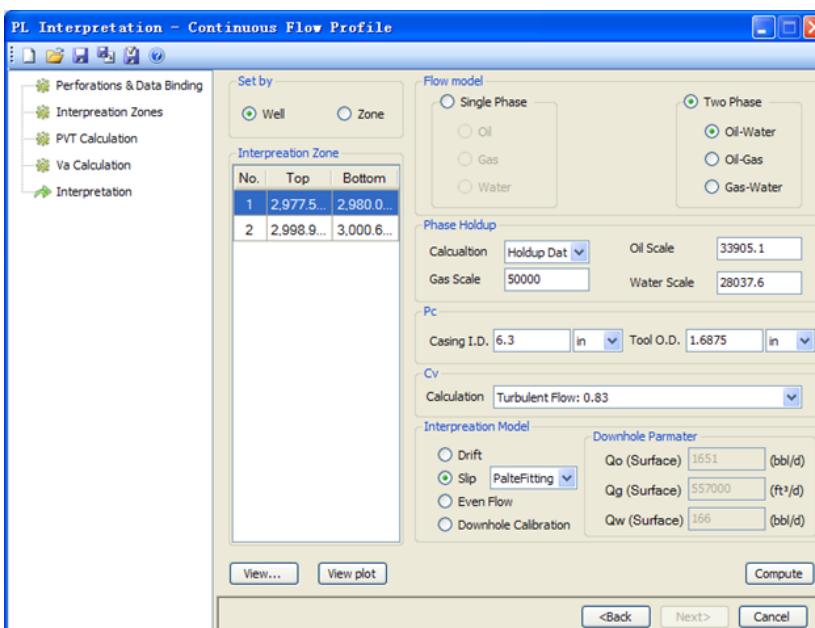


Figure 6. Interpretation parameters setting

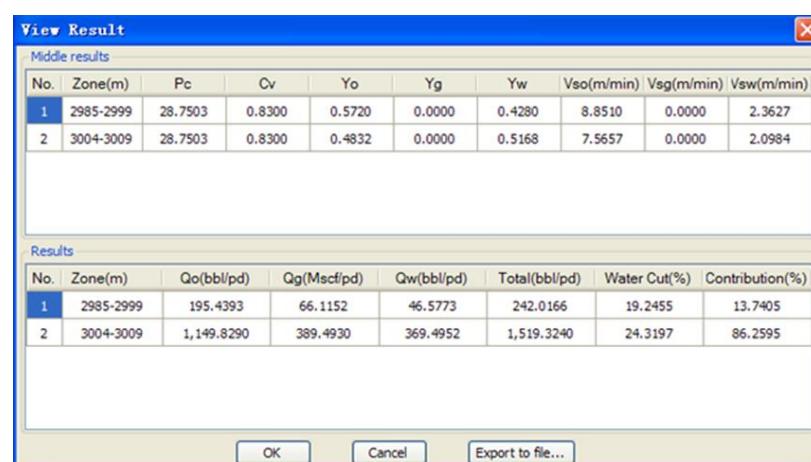


Figure 7. Total flow rate and each phase flow rate

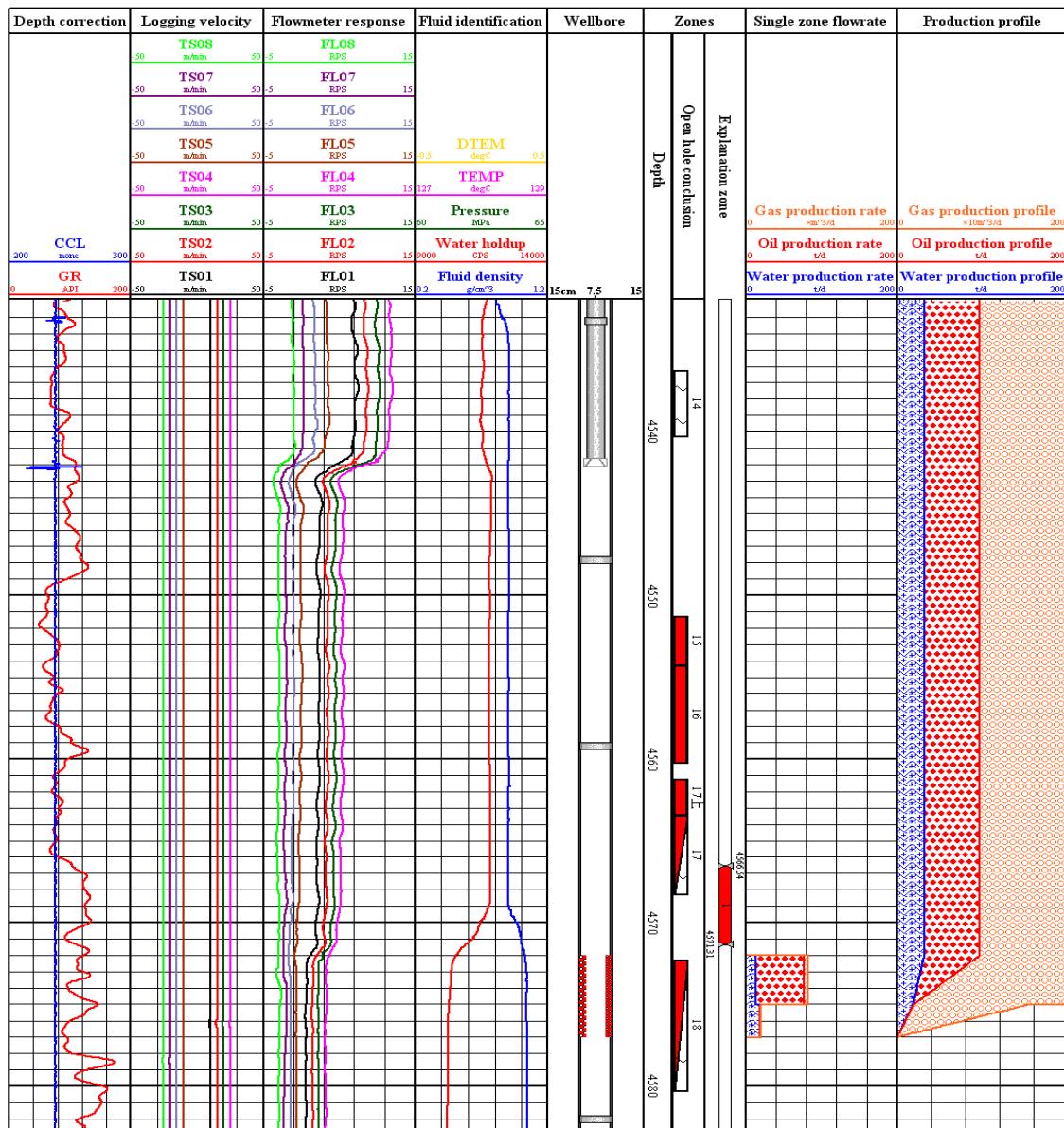


Figure 8. X well interpretation result chart

TABLE II. X WELL INTERPRETATION RESULT

Perforation zone (m)	Interpretation zone (m)	Production fluid (t/d)	Production oil (t/d)	Production water (t/d)	Production gas (m³/d)	Water cut (%)	Relative production fluid (%)
4572-4577	4572-4575	71.26	56.63	14.63	4174.89	20.53	77.02
	4575-4577	21.27	0.35	20.91	135.64	98.34	22.98
Total		92.53	56.98	35.54	4310.53	38.41	100.00

only one Well data, so we will need a large amount of test data in oil field to improve its function. In addition, in future we can consider production profile logging data interpretation of point measurement in low production well with packer flowmeter.

#### ACKNOWLEDGMENT

The authors are grateful to production logging laboratory of Yangtze University. This paper is supported by PetroChina Innovation Foundation (Grant No. 2012D-5006-0211), National Science and Technology Major

Project of the Ministry of Science and Technology of China (Grant No. 2011ZX05020-006).

#### REFERENCES

- [1] Lu Dawei, Zhang Shigang. "On Integration of Well Logging Software". *Well Logging Technology*, pp. 5-11, 20 (01), 1996.
- [2] Wei Danong, Kuang Bibo, Dai Changlin, Li Muyuan. "Development of Data Processing Software for Neutron Lifetime Logging and Its Application". *Petroleum Instruments*, pp. 13-16, 17 (6), 2003.

- [3] Yu Yalou. "Log Data Processing and Interpretation Software for Development Wells". *Well Logging Technology*, pp. 404-406, 29(5), 2005.
- [4] Liu Haihe, Zhang Defeng, Li Shaoxia. "Software Development for Multi Well Logging Interpretation". *Computing Techniques for Geophysical and Geochemical Exploration*, pp.104-108, 29(2), 2007.
- [5] Wang Qinhua, Zhang Yong, Li Shuying, and Zhao Chenguang. "A Recipe for Development and Application of the FORWARD Platform". *Well Logging Technology*, pp. 261-264, 27(3), 2003.
- [6] CHEN Fuli, Jin Yong, Zhang Shigang. "Depth Unit System Conversion in FORWARD and WATCH Platform". *Well Logging Technology*, pp. 153-155, 31(2), 2007.
- [7] Chen Meng. "Drawing The Profile of Oil and Water Relationship Based on FORWARD.Net". *Computer Applications of Petroleum*, pp. 42-43, (03), 2011.
- [8] Tan Bolei, Li Wei, Yu Chunhao, et al. "Development of Multipole Acoustic Analysis System Based on LEAD Platform and Its Application". *Well Logging Technology*, pp. 442-448, 34(05), 2010.
- [9] Tong Zhenghui, Yang Jinghai, Zhao Yipeng, et al. "Smart 2004 Production Logging Interpretation System". *Oil Forum*, pp. 64-66, (01), 2011.
- [10] Dou Yumin. *Study on Plug-In Log Interpretation of Software Systems*. Chengdu University of Technology. 2010.
- [11] Guo Haimin, Liu Junfeng, Dai Jiacai, et al. "Interpretation Models and Charts of Production Profiles in Horizontal Wells". *Science in China (Series D: Earth Sciences)*, pp. 161-166, 52(S1), 2009.
- [12] Fu Yongmei, Wang Xiufeng, Gu Lihua, et al. "On Huance 3000 Field Logging Software and PPI Interpretation Software". *Journal of Jianghan Petroleum University of Staff and Workers*, pp. 61-63, 23 (5), 2010.
- [13] Zhao Peng. "On the Development and Application of Elog Logging Integrated Interpretation Platform". *Journal of Jianghan Petroleum University of Staff and Workers*, pp. 53-54, 24(3), 2011.
- [14] Ge Jike, Li Zushu, Li Taifu. "A Novel Chinese Domain Ontology Construction Method for Petroleum Exploration Information". *Journal of Computers*, pp. 1445-1452, 7(6), 2012.
- [15] Liu Junfeng, Ding Yu. "The Research of Array Holdup Tools Response Characteristic for Oil-Water Two-Phase in Near Horizontal Well". *Journal of Chemical and Pharmaceutical Research*, pp. 31-37, 6(2), 2014.
- [16] Liu Yingming, Li Ning, Zhang Ling and Zhubo. "Integration of the C++ well logging application based on JNI technology". *Acta Petrolei Sinica*, pp. 980-984, 31(6), 2010.
- [17] Xia Shouji, Li Ning, Li Weizhong, and Liu Xiaobai. "Design of integration framework for multi-language applications on logging platform in Java". *Acta Petrolei Sinica*, pp. 810-814, 31(5), 2010.
- [18] Liu Yingming, Li Ning, Xia Shouji, et al. "Design and realization of the high-efficiency data operation cache on the Java well logging software platform." *Petroleum Exploration and Development*, pp. 328-331, 38(3), 2011.
- [19] Yuan Ye. *The Research and Realization on Key Technologies in the Water Flooded Layer Interpretation Platform*. Northeast Petroleum University, 2012.
- [20] Li Ning, Wang Caizhi, and Liu Yingming, et al. "CIFLog: the 3rd generation logging software based on Java-NetBeans". *Acta Petrolei Sinica*, pp. 192-200, 34(01), 2013.
- [21] Li Ning, Wang Caizhi, Liu Yingming, et al. "CIFLog - Integrated Network Logging Processing and Interpretation Software Platform". *Oil Forum*, pp. 6-10, (03), 2013.
- [22] Feng Qingfu, He Zongbin, Zhang Gong, and Fan He. "Development and Application of CIFLog-CMR Nuclear Magnetic Resonance Data Processing and Interpretation Software". *Well Logging Technology*, pp. 513-516, 37(05), 2013.
- [23] Fan He. *The Study of Production Logging Data Processing Method and Software Development*. Yangtze University, 2013.
- [24] Zhang Gong. *The Study of NMR Logging Data Processing Method and Software Development*. Yangtze university, 2013.
- [25] Kittiphong J, Robert Peter K. "Flow Model Selection for Production Logging Interpretation". SPE Asia Pacific Oil and Gas Conference and Exhibition. Society of Petroleum Engineers: Jakarta, Indonesia, 2011.
- [26] Guo Haimin, Dai Jiacai, Chen Kegui. *Production Logging Principle and Data Interpretation*. Beijing Petroleum Industry Press, 2010.

**Junfeng Liu** received his B.E degree and M.E degree in geophysical well logging from Yangtze University, Jingzhou, China, in 2004 and 2007, respectively, and Ph.D. degree in geodetection and information technology, Yangtze University, China, in 2010. His current research interests include production logging data processing and interpretation.

**Heng Li** received her B.E degree in geophysical well logging from Yangtze University, Jingzhou, China, in 2012. Currently, she is a master graduate student at Yangtze University, mainly does researches on software development of production logging data interpretation.

**Yingming Liu** received his M.E degree in geodetection and information technology from China University of Geosciences (Beijing), in 2008. Currently, he is an on-job doctoral student at Research Institute of Petroleum Exploration and Development, mainly does researches on the new method and technology of well-logging interpretation.

# Rival Penalized Image Segmentation

Shaojun Zhu, Jieyu Zhao, Lijun Guo

Institute of Computer Science and Technology, Ningbo University, Ningbo, 315211, China

Email: shj.zhu@gmail.com, {zhao.jieyu, guolijun}@nbu.edu.cn

**Abstract**—Image segmentation plays an important role in computer vision and image analysis. In this paper, we cast natural image segmentation into a problem of feature clustering. We extract local homogeneity, textures and color features from images and describe them with Gaussian Mixture Models. Unlike most existing clustering based segmentation methods, our method is capable of model selection automatically by de-learning redundant segments (clusters) during the clustering process. Thus, our method does not need to specify the exact number of segments in advance. Comprehensive experiments are conducted to measure the performance of the proposed algorithm in terms of visual evaluation and a variety of quantitative indices for image segmentation. The proposed algorithm compares favorably against other well-known image segmentation methods on the BSDS500 image database.

**Index Terms**—Image Segmentation, Rival Penalized Competitive Learning, Gaussian Mixture Model, Gabor Filter, Local Homogeneity

## I. INTRODUCTION

Image segmentation refers to the process of partitioning a digital image into multiple homogenous segments that share certain visual characteristics. This problem has received extensive attention since the early years of computer vision research [1]. It has been known that image segmentation plays an important role in human visual system [2]. Many computer vision problems, such as stereo vision [3], motion estimation [4], image retrieval [5], and object recognition [6], can be solved better with reliable results of image segmentation. Although the problem of image segmentation has been studied for more than three decades, great challenges still remain in this area.

In recent years, much attention has been paid to spectral clustering algorithms [7], in particular the Normalized Cuts criterion [8], which provides a way of integrating global image information into the grouping process. The original NCuts criterion is concerned on the 2-way situation, which aims at partitioning image into two parts. Two recent variants extent the NCuts criterion to the  $k$ -way multi-class and multi-scale situation, known as the Multi-class NCuts [9] and Multi-scale NCuts [10]. These progress address segmentation in a  $k$ -way global optimization framework and guarantee a globally optimal

solution in the relaxed continuous solution space. However, the  $k$ -way NCuts criterion can not automatically select the number of segments,  $k$ , since the objective function of  $k$ -way NCuts increases monotonically as  $k$  is varied. In many applications such as natural image segmentation, due to the diversity and complexity of image contents and semantics, the optimal number of segments  $k$  may be different for varying images. For unsupervised segmentation, how to adaptively select the number of segments  $k$  for varying images is a fundamental open problem left in [9] and [10]. We also note that how to construct affinity matrix is another important issue in NCuts framework, which significantly influences the segmentation performance [11], [12].

Due to the efficiency and intuitive, clustering based image segmentation algorithms have been one of the most popular image segmentation methods. First, image features, such as color or texture, are extracted on small windows centered around the pixel to be classified. Then, the extracted image features, handled as vectors, are grouped in compact but well-separated classes by clustering algorithms. The method such as  $k$ -means [13] and its variations [14]–[16] are the most commonly used techniques in the clustering-based segmentation area.

Nevertheless, despite its popularity and advantages, clustering based image segmentation algorithms suffers from several difficulties in unsupervised segmentation of natural images:

- The most challenging problem is how to determine the segment number [17], [18]. Traditional methods often end with wrong results due to the fixed segment number because different images usually have different numbers of segments. The algorithm is able to correctly find out the clustering centers only if the preassigned segment number is exactly equal to the components number. Otherwise, it will typically lead to an incorrect clustering result.
- The other problem is that the clustering result is highly sensitive to the initial centers [19]. The quality of the clustering result depends largely on the initial centers, and may, in practice, be much worse than the global optimum.

In the view of problems mentioned above, a plenty of approaches and their corresponding improvements have been proposed to ensure the accuracy and rapidity of natural image segmentation. The Minimum Description Length (MDL) principle [20] is used with the EM algorithm to estimate the parameters of Gaussian Mixture

Manuscript received 2013; revised 2013; accepted 2013. © 2014

This work is supported by NSF of China (No. 61175026), international science and technology cooperation special programme (No. S2013GR0113), NSF of Zhejiang province (No. D1080807), Zhejiang Province Technology Innovation Team of China (New Generation of Mobile Internet Client Software : 2010R50009), and Ningbo Municipal Natural Science Foundation of China (2011A610193,2011A610191).

Models, which is the joint distribution of pixel color and texture features, but the segmentation process needs to be carried out many times with different region numbers to find the best result. This takes a large amount of computation, and the segmentation result may not accord with human segmentation. Recently, an objective criterion based on the notion of lossy minimum description length (LMDL) has been proposed for evaluating segmentation of images [21]. The optimal segmentation is defined as the one that minimizes the overall coding length of the feature vectors extracted from images. The most recent progress based on LMDL [17], [22] have shown that this criterion is highly consistent with human segmentation of natural images. However, as the minimization problem is NP hard, a suboptimal solution is found by iteratively merging regions to reduce the coding length [23]. However it still need a lot of time.

**Contributions** These algorithms above have been proven to be successful in many applications, but there is still much work to be done to overcome their drawbacks. In this paper, we proposed an unsupervised natural image segmentation method based on Rival Penalized Competitive Learning (RPCL) [24], [25] theory. We construct image feature vectors on super pixels, model the extracted vectors as a Finite Mixture Model, and cluster them by RPCL strategy. We conduct extensive experiments to compare the results with human segmentation on the Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500) [18]. Although our method is conceptually simple, the segmentation results match extremely well with those by human, exceeding or competing with the best segmentation algorithms. The main novelty and the special contributions of this paper are as follows:

- 1) We propose a method to automatically select the number of components for unsupervised natural image segmentation, using RPCL theory. The optimal number of components is obtained during the learning procedure, no additional criterion is needed. As long as the initial number of components is large enough, the algorithm will end at an appropriate value of component number.
- 2) The proposed method is not sensitive to initial centers. Provided by large enough number of components, the method will drive away redundant components and preserve good components automatically during the clustering procedure.
- 3) The experimental results demonstrate that the proposed algorithm has addressed the problems of most clustering based natural image segmentation methods have, thus achieves comparable or even better segmentation results compared with other well-known methods in the BSDS500 database.

The rest of this paper is organized as follows. In Section 2, we briefly introduce Rival Penalized Competitive Learning theory. Section 3 specified the features we extracted to segment images. In section 4, we derive the clustering algorithm used in this paper. Section 4 give our experimental results and Section 5 concludes the paper.

## II. RIVAL PENALIZED COMPETITIVE LEARNING

Rival Penalized Competitive Learning is a development of competitive learning in help of an appropriate balance between two opposite mechanisms (namely a participating mechanism and a leaving mechanism), such that an appropriate number of components will be allocated to learn multiple structures underlying observations. The basic idea in RPCL is that, for each sample, not only the winner is updated to adapt to the sample, but also its nearest rival (i.e., the second winner) is de-learned by a much smaller fixed learning rate. The empirical studies have shown that the RPCL can indeed automatically select the correct cluster number by gradually driving extra components far away from the input data set [26], [27].

Suppose we have  $N$  samples,  $x_1, x_2, \dots, x_N$ , need to be clustered, which come from  $k^*$  components, here  $k^*$  is unknown. The RPCL randomly choose  $k$  ( $k$  is large enough so that we are sure of  $k > k^*$ ) samples,  $m_1, m_2, \dots, m_k$ , as initial centers, and adaptively updates them so that those  $x_t$ s can be correctly classified.

As a sample  $x_t$  comes, each  $m_j$  evaluates a degree of its suitability on representing  $x_t$  by an individual criterion

$$\varepsilon_t(\theta_j) = \alpha_j \|x_t - m_j\|^2 \quad (1)$$

where  $\alpha_j$  is the frequency that the  $j$ -th component won in past.

Then, each competes to be allocated to represent  $x_t$ . The competition is guided globally by

$$p_{j,t} = \begin{cases} 1, & \text{if } j = c, \\ -\gamma, & \text{if } j = r, \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where

$$c = \arg \min_j \varepsilon_t(\theta_j), \quad (3)$$

$$r = \arg \min_{j \neq c} \varepsilon_t(\theta_j) \quad (4)$$

and  $\gamma$  is a parameter for controlling the penalizing strength.

Then, each  $m_j$  is updated by

$$m_j^{new} = m_j^{old} + \eta p_{j,t}(x_t - m_j^{old}). \quad (5)$$

where  $\eta$  is learning rate.

This rival penalized mechanism makes extra components driven away from data as long as the number of agents is initially given to be larger than the number of components to be detected. In other words, RPCL can push away needless components to detect a correct number of components automatically during learning. RPCL has been successfully utilized in many applications in the last decade.

## III. FEATURE EXTRACTION

In this paper, each pixel of an image is identified as belonging to a homogenous region corresponding to an object or part of an object. The problem of image

segmentation is regarded as a classification task, and the goal of segmentation is to assign a label to individual pixel or a region. So, it is very important to extract the effective pixel-level image feature. In this paper, our features include local homogeneity, Gabor filter and CIELAB color space. We select the CIELAB color space to extract all features, because CIELAB space has metric color difference sensitivity to a good approximation and is very convenient to measure small color difference, while the RGB color space does not.

Homogeneity measures the spatial closeness of the distribution color is one of the most dominant and distinguishable low-level visual features in describing image. It plays an important role in image segmentation since the result of image segmentation would be several homogeneous regions. Intensity inhomogeneity always appears in nature images. It is difficult to extract the objects accurately from the images with intensity inhomogeneity. However, it can often still be recognized by human eyes. The reason is that the pixels in the object are still different from these in their neighborhoods. We use the definition of homogeneity introduced in [28].

Suppose  $p_{mn}$  is the intensity of a pixel at the location  $(m, n)$  in an  $M \times N$  image. The standard deviation is calculated as

$$v_{ij} = \sqrt{\frac{1}{d^2} \sum_{m=i-\frac{d-1}{2}}^{i+\frac{d-1}{2}} \sum_{n=j-\frac{d-1}{2}}^{j+\frac{d-1}{2}} (p_{mn} - \mu_{ij})^2} \quad (6)$$

where  $0 \leq i, m \leq M - 1, 0 \leq j, n \leq N - 1$ .  $\mu_{ij}$  is the mean of pixel intensities in window  $w_{ij}$ , and is calculated by

$$\mu_{ij} = \frac{1}{d^2} \sum_{m=i-\frac{d-1}{2}}^{i+\frac{d-1}{2}} \sum_{n=j-\frac{d-1}{2}}^{j+\frac{d-1}{2}} p_{mn} \quad (7)$$

The discontinuity is described by edge value. We employ Sobel operator to calculate the discontinuity and use the magnitude  $e_{ij}$  of the gradient at location  $(i, j)$  as the measurement

$$e_{ij} = \sqrt{G_x^2 + G_y^2} \quad (8)$$

where  $G_x$  and  $G_y$  are the components of the gradient of intensities  $p_{ij}$  in the  $x$  and  $y$  directions, respectively. In our experiments, we use the  $3 \times 3$  window for computing the edge measure.

The standard deviation and discontinuity values are normalized in order to achieve computational consistence

$$v_{ij} = \frac{v_{ij}}{v_{max}}, e_{ij} = \frac{e_{ij}}{e_{max}} \quad (9)$$

where  $v_{max} = \max_{i,j} v_{ij}$ ,  $e_{max} = \max_{i,j} e_{ij}$ ,  $0 \leq i \leq M - 1, 0 \leq j \leq N - 1$ .

The local homogeneity is represented as

$$H_{ij} = 1 - e_{ij} \times v_{ij} \quad (10)$$

where  $0 \leq i \leq M - 1, 0 \leq j \leq N - 1$ .

The value of the local homogeneity at each location of an image has a range from 0 to 1. The more uniform

the local region surrounding a pixel is, the larger the local homogeneity value the pixel has. Weighing the pros and cons, we choose a  $5 \times 5$  window for computing the standard deviation and discontinuity.

Let  $H_{ij}^L, H_{ij}^a, H_{ij}^b$  represent the three homogeneity components in the CIELAB color space at location  $(i, j)$ , we define our homogeneity as

$$H_{ij} = \{H_{ij}^L, H_{ij}^a, H_{ij}^b\} \quad (11)$$

Figure 1 shows an example of local homogeneity.

Gabor filters [29]–[31] have been successfully applied to many image processing applications. Generally speaking, people use Gabor filters when trying to solve problems involving complicated images comprised of textured regions.

A two dimensional Gabor filter  $g(x, y)$  is an oriented sinusoidal grating, which is modulated by a two dimensional Gaussian function  $h(x, y)$  as follows:

$$g(x, y) = h(x, y) \exp(2\pi j W x) \\ = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \exp(2\pi j W x) \quad (12)$$

and the Fourier transform  $G(u, v)$  of  $g(x, y)$  is given by

$$G(u, v) = H(u-W, v) = \exp\left\{-\frac{1}{2}\left[-\frac{1}{2}\left(\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right)\right]\right\} \quad (13)$$

Here  $h(x, y)$  is a two dimensional Gaussian function with its center at the origin, and  $\sigma_x$  and  $\sigma_y$  denote its variances in  $x$  and  $y$  directions, respectively. The variances of the Fourier transform function  $G(u, v)$  are  $\sigma_u = \frac{1}{2}\pi\sigma_x$  and  $\sigma_v = \frac{1}{2}\pi\sigma_y$ . And  $W$  is the modulation frequency.

For the mother Gabor filter  $g(x, y)$ , its children Gabor filters  $g_{mn}(x, y)$  are defined to be its scaled and rotated versions:

$$g_{mn}(x, y) = a^{-2m} g(x', y'), a \geq 1 \quad (14)$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = a^{-m} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (15)$$

$$\theta = \frac{n\pi}{L}; m = 0, 1, \dots, K-1; n = 0, 1, \dots, L-1 \quad (16)$$

where  $a$  is a fixed scale factor,  $m$  is the scale parameter,  $n$  is the orientation parameter,  $K$  is the total number of scales, and  $L$  is the total number of orientations. In this paper, we set the Gabor function parameters as follows:  $W = 5, a = 2, \sigma_x = \sigma_y = \frac{1}{2}\pi, K = 3, L = 4$ .

Let  $I(x, y)$  denote an image with size  $w \times h$ . The Gabor-filtered output  $G_{mn}(x, y)$  of the image  $I(x, y)$  is defined as its convolution with the Gabor filter  $g_{mn}(x, y)$ :

$$G_{mn}(x, y) = I(x, y) \times g_{mn}(x, y) = \sum_{x_1=1}^w \sum_{y_1=1}^h I(x-x_1) \quad (17)$$

Here,  $G_{mn}(x, y)$  is the Gabor filtered image at the scale parameter  $m$  and the orientation parameter  $n$ .

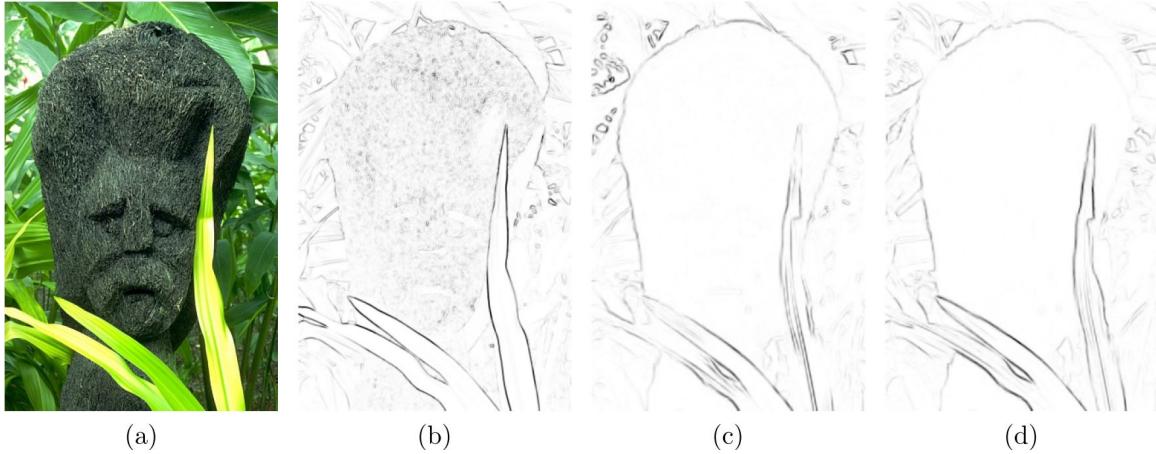


Figure 1. Local homogeneity. (a) is the original image, (b)-(d) are homogeneities in L, a, b color component respectively.

The Fourier transform function  $\hat{G}_{mn}(u, v)$  of the Gabor filter output  $G_{mn}(x, y)$  is given by

$$\hat{G}_{mn}(u, v) = \hat{I}(u, v) \times \hat{g}_{mn}(u, v) \quad (18)$$

where  $\hat{I}(u, v)$  is the Fourier transform function of the image  $I(x, y)$ , and  $\hat{G}_{mn}(u, v)$  is the Fourier transform function of the Gabor filter  $G_{mn}(x, y)$ .

In this paper, we apply gabor filters to the L component of CIELAB color space to extract texture features. Let  $B_{ij}^{m,n}$  represent the Gabor subband coefficient at location  $(i, j)$  that corresponds to the  $m$  ( $m = 1, 2, 3$ ) scale and  $n$  ( $n = 0, 45, 90, 135$ ) orientation. We will use  $GB_{ij}$  to denote the maximum (in absolute value) of the 12 coefficients at location  $(i, j)$ , which is the pixel texture feature at location  $(i, j)$ . Figure 2 shows an example.

#### IV. CLUSTERING VIA RIVAL PENALIZED COMPETITIVE LEARNING

##### A. Assumptions

Suppose  $N$  feature vectors  $x_1, x_2, \dots, x_N$  extracted from superpixels are independently and identically distributed from a mixture density of Gaussian, and all dimensions of  $x_t$  are independently:

$$p^*(x; \Theta^*) = \sum_{j=1}^{k^*} \pi_j^* G_j(x; \mu_j^*, \Sigma_j^*) \quad (19)$$

with

$$\sum_{j=1}^{k^*} \pi_j^* = 1, \quad \pi_j^* \geq 0 \quad (20)$$

where  $k^*$  is the components number,  $\Theta^* = \{(\pi_j^*, \mu_j^*, \Sigma_j^*) | 1 \leq j \leq k^*\}$  is the true parameter set, and  $G(x|\mu, \Sigma)$  denotes a multivariate Gaussian density of  $x$  with mean  $\mu$  and covariance matrix  $\Sigma$ . Since all dimensions of  $x_t$  are independently, so that  $\Sigma_j$  is a diagonal matrix.  $k^*$  and  $\Theta^*$  are unknown and need to be estimated.

##### B. The Clustering Algorithm

In Equation (19), the mixture components  $G_j(x; \mu_j^*, \Sigma_j^*)$  can be interpreted as the pdf of those feature vectors that form the corresponding image segment  $S_j$ , and  $\pi_j^*$  is the proportion of feature vectors that belong to  $S_j$ . Consequently, if the parameter set  $\Theta^*$  is already known, clustering becomes direct and simple, which classifies every feature vectors  $x_t$  into a complement according to the posterior probability of the density that  $x_t$  comes from, the posterior probability can be written as:

$$h(j|x_t, \Theta^*) = \frac{\pi_j^* G_j(x_t; \mu_j^*, \Sigma_j^*)}{\sum_{i=1}^{k^*} \pi_i^* G_i(x_t; \mu_i^*, \Sigma_i^*)} \quad (21)$$

To obtain the estimate of  $\Theta^*$ , we model the feature vectors by

$$p(x; \Theta) = \sum_{j=1}^k \alpha_j G(x|\mu_j, \Sigma_j) \quad (22)$$

with

$$\sum_{j=1}^k \alpha_j = 1, \quad \alpha_j \geq 0 \quad (23)$$

$$G(x|\mu_j, \Sigma_j) = \frac{1}{(2\pi)^{dim/2} |\Sigma_j|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_j) \Sigma_j^{-1} (x - \mu_j) \right\} \quad (24)$$

where  $k$  is components number,  $dim$  is the dimension of feature vectors, and  $\Theta = \{(\alpha_j, \mu_j, \Sigma_j) | 1 \leq j \leq k\}$  is an estimate of  $\Theta^*$ , and all  $\Sigma_j$  are diagonal matrices. It is assumed in [32] that every sample  $x_t$  is accompanied by its hidden label

$$x_{t,h} = [x_{t,h}^{(1)}, x_{t,h}^{(2)}, \dots, x_{t,h}^{(k)}] \quad (25)$$

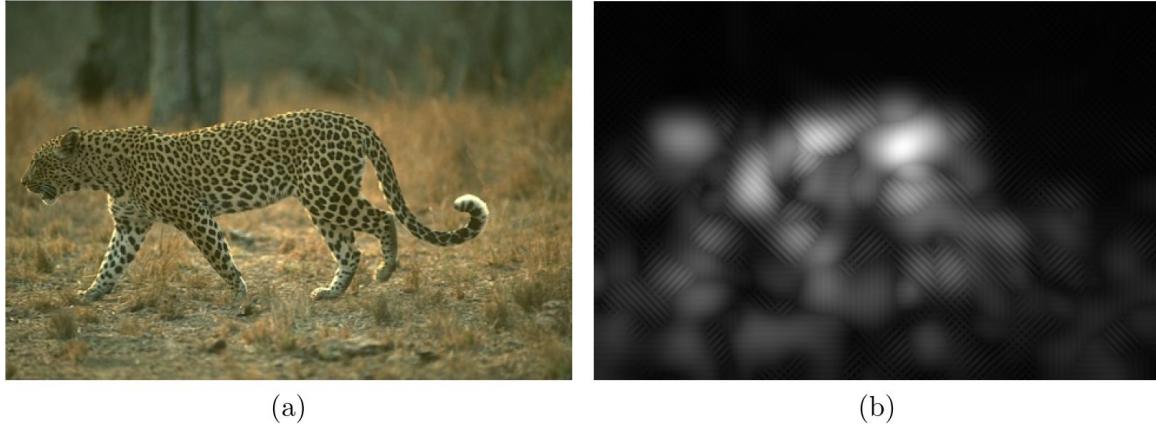


Figure 2. Gabor feature. (a) is the original image, (b) is the gabor feature extracted.

with

$$x_{t,h}^{(j)} = \begin{cases} 1, & \text{if } x_t \text{ is drawn from } G(x|\mu_j, \Sigma_j) \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

$$\sum_{j=1}^k x_{t,h}^{(j)} = 1$$

shows which component  $x_t$  drawn from. Since all  $x_t$  are randomly chose from one of  $k$  components,  $x_{1,h}, x_{2,h}, \dots, x_{N,h}$  are also iid. Authors of [32] further assumed that every  $x_{t,h}$  is from a multinomial distribution consisting of one draw on  $k$  densities with probabilities  $\alpha_1, \alpha_2, \dots, \alpha_k$  respectively. which means the marginal density distribution of  $x_{t,h}$  is

$$p(x_{t,h}|\Theta) = \prod_{j=1}^k (\alpha_j)^{x_{t,h}^{(j)}} \quad (27)$$

Therefor, the joint pdf of  $x_{1,h}, x_{2,h}, \dots, x_{N,h}$  is

$$p(x_{1,h}, x_{2,h}, \dots, x_{N,h}|\Theta) = \prod_{t=1}^N \prod_{j=1}^k (\alpha_j)^{x_{t,h}^{(j)}} \quad (28)$$

Suppose  $x_1, x_2, \dots, x_N$  are conditionally independent as given  $x_{1,h}, x_{2,h}, \dots, x_{N,h}$ , respectively, we have

$$p(X_N|X_{n,h}, \Theta) = \prod_{t=1}^N p(x_t|x_{t,h}, \Theta) \quad (29)$$

where

$$X_N = (x_1^T, x_2^T, \dots, x_N^T)^T$$

$$X_{N,h} = (x_{1,h}^T, x_{2,h}^T, \dots, x_{N,h}^T)^T \quad (30)$$

$$p(x_t|x_{t,h}, \Theta) = \prod_{j=1}^k G(x_t|\mu_j, \Sigma_j)^{x_{t,h}^{(j)}}$$

upon the fact that all  $x_t$  exclusively depend on  $x_{t,h}$ . Hence, the joint pdf of the complete data is

$$p(X_N, X_{n,h}|\Theta) = \prod_{t=1}^N \prod_{j=1}^k [\alpha_j G(x_t|\mu_j, \Sigma_j)]^{x_{t,h}^{(j)}} \quad (31)$$

Consequently, the empirical log likelihood function of  $\Theta$  is

$$L(\Theta; X_N) = \frac{1}{N} \sum_{t=1}^N \sum_{j=1}^k x_{t,h}^{(j)} [\ln \alpha_j + \ln G(x_t|\mu_j, \Sigma_j)] \quad (32)$$

Since all hidden labels  $x_{t,h}$  are unknown, we then replace them with their expected value conditioned on  $x_t$ , which is  $h(j|x_t, \Theta)$  the posterior density probability. In conclusion, given a specific  $k$ ,  $\Theta$  can be calculated by maximizing the following log likelihood function:

$$L(\Theta; X_N) = \frac{1}{N} \sum_{t=1}^N \sum_{j=1}^k h(j|x_t, \Theta) [\ln \alpha_j + \ln G(x_t|\mu_j, \Sigma_j)] \quad (33)$$

In implementation, maximizing Equation (33) can be achieved by the Expectation-Maximization (EM) algorithm [32]

E-Step Fix the parameter set  $\Theta^{old} = \{\alpha_j^{old}, \mu_j^{old}, \Sigma_j^{old}\}_{j=1}^k$ , update posterior probabilities

$$h(j|x_t, \Theta^{old}) = \frac{\alpha_j^{old} G_j(x_t, \mu_j^{old}, \Sigma_j^{old})}{\sum_{i=1}^k \alpha_i^{old} G_i(x_t, \mu_i^{old}, \Sigma_i^{old})} \quad (34)$$

where  $j = 1, 2, \dots, k$ .

M-Step Fix posterior probabilities, update the parameter set

$$\Theta^{new} = \Theta^{old} + \eta \Delta \Theta \quad (35)$$

where  $\eta$  is a small positive constant parameter for learning rate.

The EM algorithm are iteratively performed for every sample until  $\Theta$  converges. However, this algorithm does not contain a mechanism for automatic model section, i.e., a mechanism to select an appropriate value of  $k$ . As a consequence, we can get a good estimate of  $\Theta^*$  only when  $k$  is exactly equal to the true  $k^*$  which is unknown. Otherwise, the EM algorithm will probably get a bad result. Through integrating RPCL into the EM algorithm, we can make the EM algorithm has the automatic model section ability.

As a feature vector  $x^{(t)}$  comes, the first step, namely E-Step, is compute the posterior probability with fixed parameter set  $\Theta^{(t)} = \left\{ \alpha_j^{(t)}, \mu_j^{(t)}, \Sigma_j^{(t)} \right\}_{j=1}^k$  for each component  $\mathcal{S}_j$

$$h(j|x^{(t)}, \Theta^{(t)}) = \frac{\alpha_j^{(t)} G_j(x^{(t)}, \mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{i=1}^k \alpha_i^{(t)} G_i(x^{(t)}, \mu_i^{(t)}, \Sigma_i^{(t)})} \quad (36)$$

where  $\alpha_j^{(t)}$  is the frequency that the  $j$ -th component  $\mathcal{S}_j$  won in past, that is

$$\alpha_j^{(t)} = \frac{n_j^{(t)}}{\sum_{i=1}^k n_i^{(t)}} \quad (37)$$

where  $n_j^{(t)}$  is the cumulative number of the winning occurrences of  $\mathcal{S}_j$  in the past.

Then, each competes to be allocated to represent  $x^{(t)}$ . The competition is guided globally by

$$p_{j,t} = \begin{cases} 1, & \text{if } j = c, \\ -1, & \text{if } j = r, \\ 0, & \text{otherwise} \end{cases} \quad (38)$$

where

$$\begin{aligned} c &= \arg \max_j h(j|x^{(t)}, \Theta^{(t)}) \\ r &= \arg \max_{j \neq c} h(j|x^{(t)}, \Theta^{(t)}) \end{aligned} \quad (39)$$

Since in Equation (36), all of denominators are the same, Equation (39) can be simplified as

$$\begin{aligned} c &= \arg \max_j \alpha_j^{(t)} G_j(x^{(t)}, \mu_j^{(t)}, \Sigma_j^{(t)}) \\ r &= \arg \max_{j \neq c} \alpha_j^{(t)} G_j(x^{(t)}, \mu_j^{(t)}, \Sigma_j^{(t)}) \end{aligned} \quad (40)$$

we denote that

$$\begin{aligned} \psi_j(x^{(t)}) &= \ln \left[ \alpha_j^{(t)} G_j(x^{(t)}, \mu_j^{(t)}, \Sigma_j^{(t)}) \right] \\ &= \ln \alpha_j^{(t)} - \frac{1}{2} (x^{(t)} - \mu_j^{(t)}) \Sigma_j^{(t)^{-1}} (x^{(t)} - \mu_j) \\ &\quad - \frac{1}{2} \ln |\Sigma_j^{(t)}| - \frac{\dim}{2} \ln(2\pi) \end{aligned} \quad (41)$$

Eliminating the last common item,  $\psi_j(x^{(t)})$  is simplified to be

$$\begin{aligned} \psi_j(x^{(t)}) &= \ln \alpha_j^{(t)} - \frac{1}{2} (x^{(t)} - \mu_j^{(t)}) \Sigma_j^{(t)^{-1}} (x^{(t)} - \mu_j^{(t)}) \\ &\quad - \frac{1}{2} \ln |\Sigma_j^{(t)}| \end{aligned} \quad (42)$$

As a result, Equation (39) is changed to be

$$\begin{aligned} c &= \arg \max_j \psi_j(x^{(t)}) \\ r &= \arg \max_{j \neq c} \psi_j(x^{(t)}) \end{aligned} \quad (43)$$

The second step, namely M-Step, is updating the parameters. Only the parameters of winner and rival are modified. For the winner  $\mathcal{S}_c$ , we update the parameters by

$$\mu_c^{(t+1)} = \mu_c^{(t)} + p_{c,t} \eta_c (x^{(t)} - \mu_c^{(t)}) \quad (44)$$

$$\Sigma_c^{(t+1)} = \Sigma_c^{(t)} + p_{c,t} \eta_c \left[ (x^{(t)} - \mu_c^{(t)})^T (x^{(t)} - \mu_c^{(t)}) - \Sigma_c^{(t)} \right] \quad (45)$$

where  $\eta$  is a small constant value for learning rate. Since all dimensions in  $x^{(t)}$  are independent, the covariance matrices are diagonal matrices, Equation (45) should be

$$\Sigma_c^{(t+1)} = \Sigma_c^{(t)} + p_{c,t} \eta_c \left\{ \text{diag} \left[ (x^{(t)} - \mu_c^{(t)})^T (x^{(t)} - \mu_c^{(t)}) \right] - \Sigma_c^{(t)} \right\} \quad (46)$$

where  $\text{diag}(M)$  is the diagonal matrix consists of the diagonal elements of  $M$ .

For the rival  $\mathcal{S}_r$ , we update the parameters by

$$\mu_r^{(t+1)} = \mu_r^{(t)} + p_{r,t} \eta_c \eta_r (x^{(t)}; c, r) (x_t - \mu_r^{(t)}) \quad (47)$$

$$\begin{aligned} \Sigma_r^{(t+1)} &= \Sigma_r^{(t)} + p_{r,t} \eta_c \eta_r (x^{(t)}; c, r) \\ &\quad \left\{ \text{diag} \left[ (x^{(t)} - \mu_r^{(t)})^T (x_t - \mu_r^{(t)}) \right] - \Sigma_r^{(t)} \right\} \end{aligned} \quad (48)$$

where  $\eta_r(x^{(t)}; c, r)$  is rival penalization rate, [33] suggested the controlled rival penalization mechanism

$$\eta_r(x^{(t)}; c, r) = \frac{\min \{ f(\mathcal{S}_c, \mathcal{S}_r), f(x^{(t)}, \mathcal{S}_c) \}}{f(\mathcal{S}_c, \mathcal{S}_r)} \quad (49)$$

where  $f(\cdot, \cdot)$  is a certain distance measuring function. We adopt this mechanism into our method, and get our rival penalization rate

$$\eta_r(x^{(t)}; c, r) = \min \left\{ 1, \frac{G_r(x^{(t)}, \mu_r^{(t)}, \Sigma_r^{(t)})}{G_c(\mu_r^{(t)}, \mu_c^{(t)}, \Sigma_c^{(t)})} \right\} \quad (50)$$

## V. EXPERIMENTS

The proposed method has been used to segment an image into different regions on the Berkeley segmentation database BSDS500 [18]. This database is an extension of the BSDS300 and comprises of various images from the Corel dataset and contains ground truth of 500 images for benchmarking image segmentation and boundary detection algorithms. The content of the images includes landscapes, animals, portraits and various objects.

We show six representative segmentation results Figure 3, compared with two well-known image segmentation algorithms [21], [34]. Note that all the small regions with the numbers of pixels less than 200 are removed in our results. The results show that our algorithm is feasible. More segmentation results are shown in Figure 4.

To quantitatively evaluate the performance of our method, we use four metrics for comparing pairs of image segmentation:

- The **Probabilistic Rand index** (PRI) [35] is a classical metric that measures the probability that an arbitrary pair of samples have consistent labels in the two partitions.
- The **Global Consistency Error** (GCE) [36] measures the extent to which one segmentation can be viewed as a refinement of the other. Segmentations which are related in this manner are considered to be consistent, since they could represent the same natural image segmented at different scales.

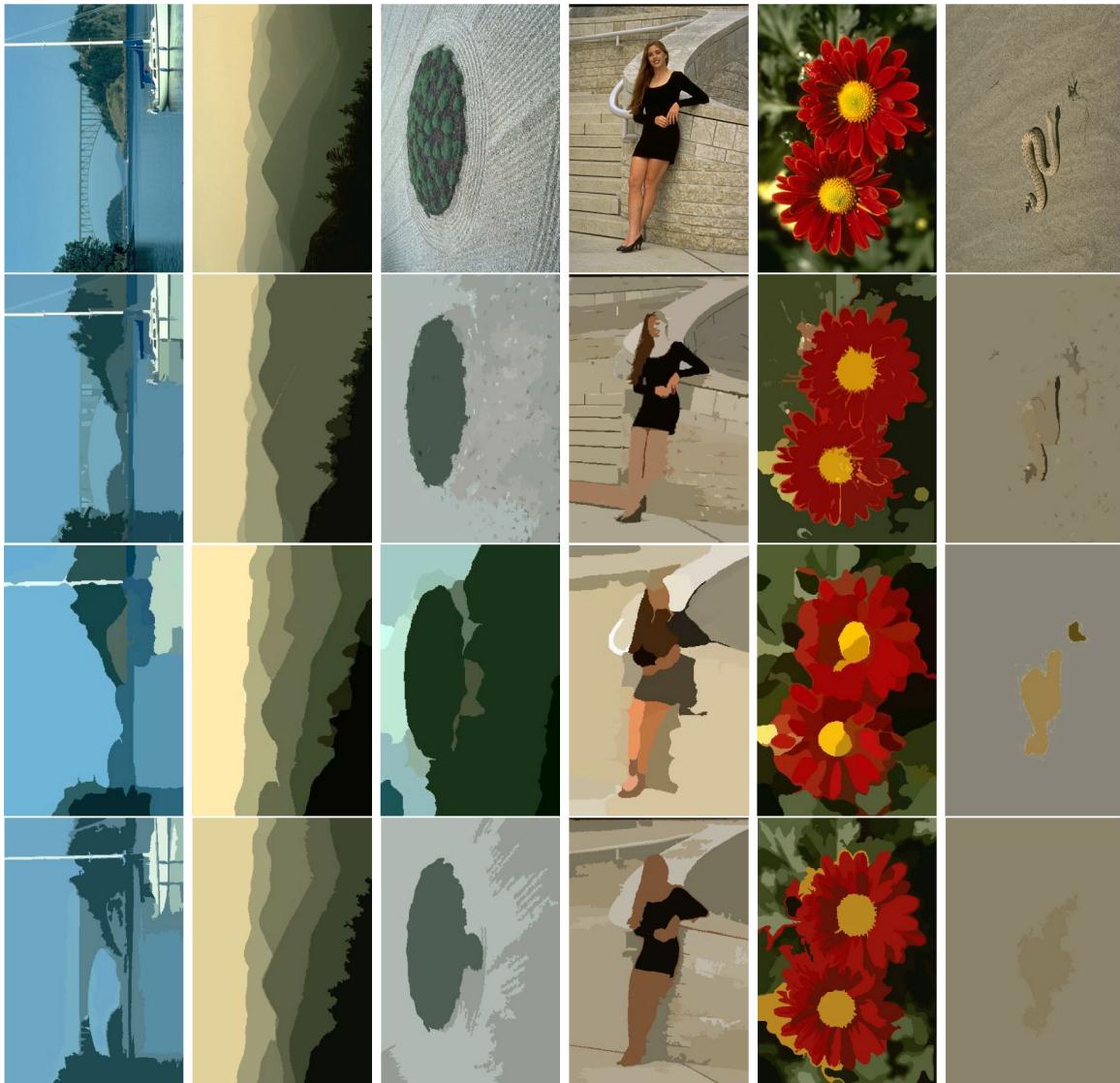


Figure 3. Comparison with [21], [34]. The first row is the input images, the second row is the results obtained by [34], the third row is the results obtained by [21] and the last row is the results obtained by our method.

- The **Variation of Information** (VOI) [37] measures the sum of information loss and information gain between the two clusterings, and thus it roughly measures the extent to which one clustering can explain the other.
- The **Boundary Displacement Error** (BDE) [38] measures the average displacement error of boundary pixels between two segmented images. Particularly, it defines the error of one boundary pixel as the distance between the pixel and the closest pixel in the other boundary image.

In cases where we have multiple ground-truth segmentations, we simply average the results of the metric between the test segmentation and each ground truth segmentation. Table I shows the average performances of these four measures over the 500 color images in BSDS500.

TABLE I.  
AVERAGE PERFORMANCE ON BSDS500. CTM REPRESENTS  
COMPRESSION-BASED TEXTURE MERGING ALGORITHM PROPOSED  
BY YANG ET.AL. [21], GRAPHBASED REPRESENTS THE METHOD  
PROPOSED BY FELZENZWALB AND HUTTENLOCHER [34].

	PRI	GCE	VOI	BDE
Humans	0.8854	0.0767	1.1020	4.9942
RPCL	0.7649	0.2176	2.6882	8.8980
CTM	0.7628	0.2093	2.0788	9.4038
GraphBased	0.6930	0.3255	2.4456	15.5592

## VI. CONCLUSION

Automatic image segmentation is one of the fundamental problems in computer vision. To segment the pixels in image space, the most straightforward idea is first to obtain a coherent or robust clustering result on the pixels' feature space, then each pixel is labeled with the cluster that contains its feature vector. In this paper, we propose a novel approach that is capable of determine the seg-

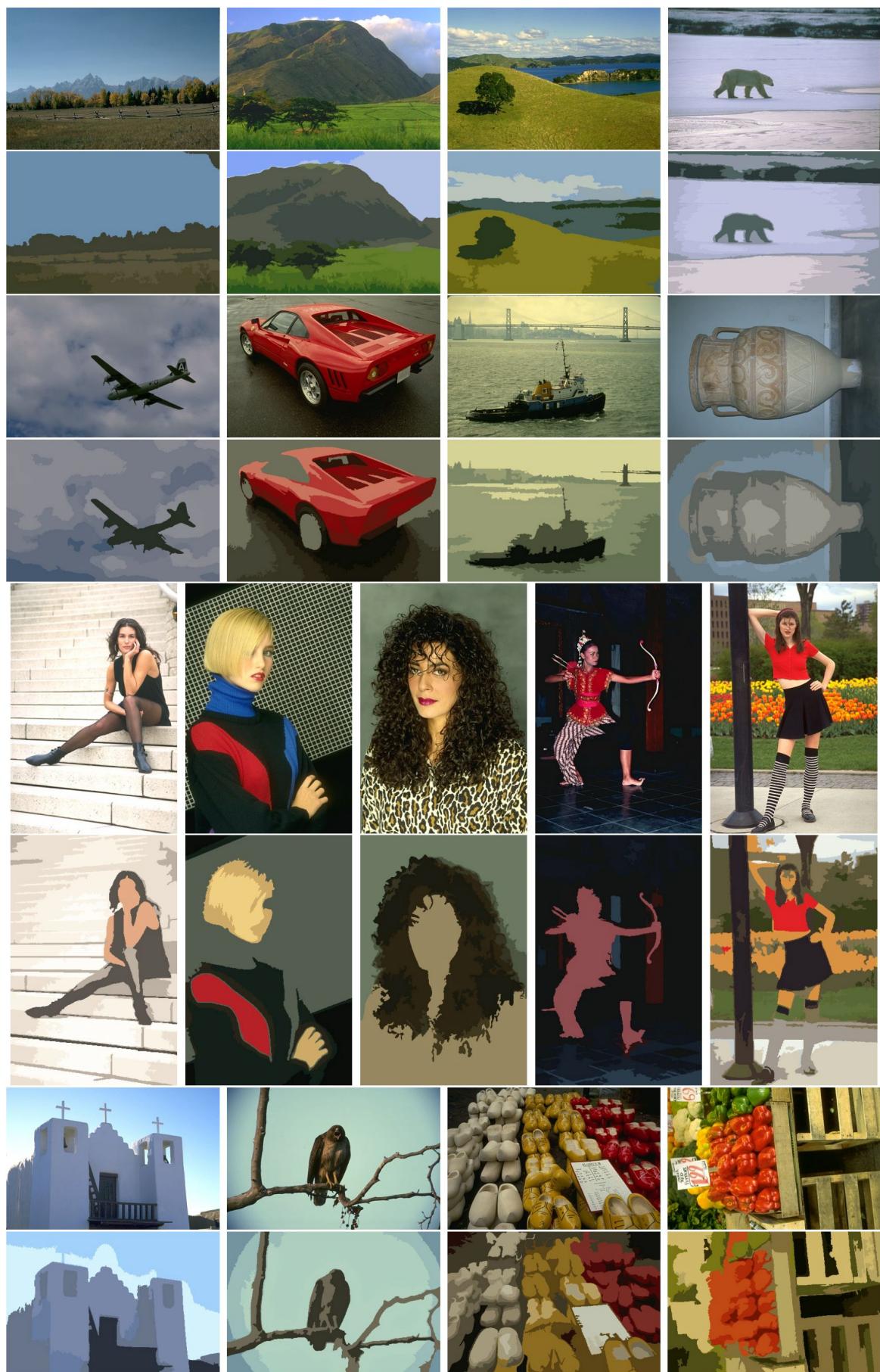


Figure 4. Image segmentation results using proposed approach.

**Algorithm 1** Clustering

**Require:** Feature vector set  $X = \{x_1, x_2, \dots, x_N\}$ , initial components number  $k$ .

**Ensure:**  $y = x^n$

- 1: Initialization. Randomly pick  $k$  samples  $x^1, x^2, \dots, x^k$  from the feature vector set  $X$ , set

$$\mu_j^{(1)} = x^j, j = 1, 2, \dots, k \quad (51)$$

Set covariance matrices of  $k$  components empirically, i.e.,

$$\Sigma_j^{(1)} = \sigma I, j = 1, 2, \dots, k \quad (52)$$

where  $\sigma$  is a small positive number, and  $I$  is the identity matrix has same dimension with feature vectors.

Set  $n_j^{(0)} = 1, j = 1, 2, \dots, k$  and  $t = 0$ .

- 2:  $t = t + 1$ . Randomly take an input  $x^{(t)}$  from the feature vector set;
- 3: Calculate  $\psi_j(x^{(t)})$  for each component with Equation (42);
- 4: Compute  $\alpha_j^{(t)}$  for each component with Equation (37);
- 5: Determine winner component  $c$  and rival component  $r$  using Equation (43);
- 6: Update the parameters of winner component by Equation (44) and Equation (45)
- 7: Calculate the rival penalization rate with Equation (50);
- 8: Update the parameters of rival component by Equation (47) and Equation (48)
- 9: If  $t \% N == 0$ ,

$$n_j^{(t)} = \frac{n_j^{(t)}}{2}, j = 1, 2, \dots, k \quad (53)$$

- 10: Iterate through step 2 to 9 for each input until a stop criterion is satisfied.

ment number automatically. Our approach is formulated under Gaussian Mixture Models and with Rival Penalized Competitive Learning. Experimental results demonstrate the efficacy. Our future work includes investigate other separable features, combine them into our approach and improve the experimental results.

## ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for their valuable comments and suggestions to improve the presentation of this paper.

## REFERENCES

- [1] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2010.
- [2] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, 2nd ed. Prentice Hall, 2012.
- [3] Y. Wang, K. Liu, Q. Hao, X. Wang, D. Lau, and L. Hassbrook, "Robust active stereo vision using kullback-leibler divergence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 548–563, 2012.
- [4] S. Ricco and C. Tomasi, "Dense lagrangian motion estimation with occlusions," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1800–1807.
- [5] L. Zhang, L. Wang, and W. Lin, "Semisupervised biased maximum margin analysis for interactive image retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2294–2308, 2012.
- [6] M. Choi, A. Torralba, and A. Willsky, "A tree-based context model for object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 240–252, 2012.
- [7] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.
- [8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [9] S. X. Yu and J. Shi, "Multiclass spectral clustering," in *IEEE International Conference on Computer Vision*. IEEE, 2003, pp. 313–319.
- [10] T. Cour, F. Benzeit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2005, pp. 1124–1131.
- [11] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *International journal of computer vision*, vol. 43, no. 1, pp. 7–27, 2001.
- [12] T. H. Kim, K. M. Lee, and S. U. Lee, "Learning full pairwise affinities for spectral segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1690–1703, 2013.
- [13] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar 1982.
- [14] A. Irani and B. Belaton, "An automated adaption of k-means based hybrid segmentation system into direct volume rendering object distinction mode for enhanced visualization effect," in *International Conference on Computer Graphics, Imaging and Visualization*. IEEE, 2012, pp. 62–69.
- [15] K. Tan, N. Isa, and W. Lim, "Color image segmentation using adaptive unsupervised clustering approach," *Applied Soft Computing*, 2012.
- [16] M. Yang, L. Xue-bin, and Q. Zhou, "Image segmentation algorithm based on incomplete k-means clustering and category optimization," *Journal of Computer Applications*, vol. 1, p. 061, 2012.
- [17] H. Mobahi, S. Rao, A. Yang, S. Sastry, and Y. Ma, "Segmentation of natural images by texture and boundary compression," *International Journal of Computer Vision*, vol. 95, no. 1, pp. 86–98, 2011.
- [18] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.
- [19] M. Mignotte, "Mds-based segmentation model for the fusion of contour and texture cues in natural images," *Computer Vision and Image Understanding*, vol. 116, no. 9, pp. 981–990, 2012.
- [20] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, 2002.
- [21] A. Yang, J. Wright, Y. Ma, and S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 212–225, 2008.
- [22] S. Rao, H. Mobahi, A. Yang, S. Sastry, and Y. Ma, "Natural image segmentation with adaptive texture and boundary

- encoding,” *Asian Conference on Computer Vision*, pp. 135–146, 2010.
- [23] M. Jiang, C. Li, J. Feng, and L. Wang, “Segmentation via ncuts and lossy minimum description length: a unified approach,” *Asian Conference on Computer Vision*, pp. 213–224, 2011.
- [24] L. Xu, A. Krzyzak, and E. Oja, “Rival penalized competitive learning for clustering analysis, rbf net, and curve detection,” *IEEE Transactions on Neural Networks*, vol. 4, no. 4, pp. 636–649, 1993.
- [25] L. Xu, “Rival penalized competitive learning, finite mixture, and multisets clustering,” in *IEEE International Joint Conference on Neural Networks Proceedings*, vol. 3, 1998, pp. 2525–2530.
- [26] C. Wang and J. Lai, “Energy based competitive learning,” *Neurocomputing*, vol. 74, no. 12-13, pp. 2265–2275, Jun 2011.
- [27] Y. Cheung, “ $k^*$ -means: A new generalized k-means clustering algorithm,” *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2883–2893, Nov 2003.
- [28] H. Cheng and Y. Sun, “A hierarchical approach to color image segmentation using homogeneity,” *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2071–2082, 2000.
- [29] T. Lee, “Image representation using 2d gabor wavelets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, 1996.
- [30] T. Weldon, W. Higgins, and D. Dunn, “Efficient gabor filter design for texture segmentation,” *Pattern Recognition*, vol. 29, no. 12, pp. 2005–2015, 1996.
- [31] J. Khan, R. Adhami, and S. Bhuiyan, “A customized gabor filter for unsupervised color image segmentation,” *Image and Vision Computing*, vol. 27, no. 4, pp. 489–501, 2009.
- [32] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [33] Y. Cheung, “On rival penalization controlled competitive learning for clustering with automatic cluster number selection,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 11, pp. 1583–1588, Nov 2005.
- [34] P. Felzenszwalb and D. Huttenlocher, “Efficient graph-based image segmentation,” *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [35] W. M. Rand, “Objective criteria for the evaluation of clustering methods,” *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, Dec 1971.
- [36] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *IEEE International Conference on Computer Vision*, vol. 2, 2001, pp. 416–423.
- [37] M. Meila, *Comparing clusterings*. ACM Press, 2005, pp. 577–584.
- [38] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cuffí, “Yet another survey on image segmentation: Region and boundary information integration,” *European Conference on Computer Vision*, pp. 21–25, 2002.



# Instructions for Authors

## Manuscript Submission

All paper submissions will be handled electronically in EDAS via the JMM Submission Page (URL: <http://edas.info/newPaper.php?c=7325>). After login EDAS, you will first register the paper. Afterwards, you will be able to add authors and submit the manuscript (file). If you do not have an EDAS account, you can obtain one. Along with the submission, Authors should select up to 3 topics from the EDICS (URL: <http://www.academypublisher.com/jmm/jmmedics.html>), and clearly state them during the registration of the submission.

JMM invites original, previously unpublished, research papers, review, survey and tutorial papers, application papers, plus case studies, short research notes and letters, on both applied and theoretical aspects. Submission implies that the manuscript has not been published previously, and is not currently submitted for publication elsewhere. Submission also implies that the corresponding author has consent of all authors. Upon acceptance for publication transfer of copyright will be made to Academy Publisher, article submission implies author agreement with this policy. Manuscripts should be written in English. Paper submissions are accepted only in PDF. Other formats will not be accepted. Papers should be formatted into A4-size (8.27" x 11.69") pages, with main text of 10-point Times New Roman, in single-spaced two-column format. Authors are advised to follow the format of the final version at this stage. All the papers, except survey, should ideally not exceed 12,000 words (14 pages) in length. Whenever applicable, submissions must include the following elements: title, authors, affiliations, contacts, abstract, index terms, introduction, main text, conclusions, appendixes, acknowledgement, references, and biographies.

## Conference Version

Submissions previously published in conference proceedings are eligible for consideration provided that the author informs the Editors at the time of submission and that the submission has undergone substantial revision. In the new submission, authors are required to cite the previous publication and very clearly indicate how the new submission offers substantively novel or different contributions beyond those of the previously published work. The appropriate way to indicate that your paper has been revised substantially is for the new paper to have a new title. Author should supply a copy of the previous version to the Editor, and provide a brief description of the differences between the submitted manuscript and the previous version.

If the authors provide a previously published conference submission, Editors will check the submission to determine whether there has been sufficient new material added to warrant publication in the Journal. The Academy Publisher's guidelines are that the submission should contain a significant amount of new material, that is, material that has not been published elsewhere. New results are not required; however, the submission should contain expansions of key ideas, examples, elaborations, and so on, of the conference submission. The paper submitting to the journal should differ from the previously published material by at least 30 percent.

## Review Process

Submissions are accepted for review with the understanding that the same work has been neither submitted to, nor published in, another publication. Concurrent submission to other publications will result in immediate rejection of the submission.

All manuscripts will be subject to a well established, fair, unbiased peer review and refereeing procedure, and are considered on the basis of their significance, novelty and usefulness to the Journals readership. The reviewing structure will always ensure the anonymity of the referees. The review output will be one of the following decisions: Accept, Accept with minor changes, Accept with major changes, or Reject.

The review process may take approximately three months to be completed. Should authors be requested by the editor to revise the text, the revised version should be submitted within three months for a major revision or one month for a minor revision. Authors who need more time are kindly requested to contact the Editor. The Editor reserves the right to reject a paper if it does not meet the aims and scope of the journal, it is not technically sound, it is not revised satisfactorily, or if it is inadequate in presentation.

## Revised and Final Version Submission

Revised version should follow the same requirements as for the final version to format the paper, plus a short summary about the modifications authors have made and author's response to reviewer's comments.

Authors are requested to use the Academy Publisher Journal Style for preparing the final camera-ready version. A template in PDF and an MS word template can be downloaded from the web site. Authors are requested to strictly follow the guidelines specified in the templates. Only PDF format is acceptable. The PDF document should be sent as an open file, i.e. without any data protection. Authors should submit their paper electronically through email to the Journal's submission address. Please always refer to the paper ID in the submissions and any further enquiries.

Please do not use the Adobe Acrobat PDFWriter to generate the PDF file. Use the Adobe Acrobat Distiller instead, which is contained in the same package as the Acrobat PDFWriter. Make sure that you have used Type 1 or True Type Fonts (check with the Acrobat Reader or Acrobat Writer by clicking on File>Document Properties>Fonts to see the list of fonts and their type used in the PDF document).

## Copyright

Submission of your paper to this journal implies that the paper is not under submission for publication elsewhere. Material which has been previously copyrighted, published, or accepted for publication will not be considered for publication in this journal. Submission of a manuscript is interpreted as a statement of certification that no part of the manuscript is copyrighted by any other publisher nor is under review by any other formal publication.

Submitted papers are assumed to contain no proprietary material unprotected by patent or patent application; responsibility for technical content and for protection of proprietary material rests solely with the author(s) and their organizations and is not the responsibility of the Academy Publisher or its editorial staff. The main author is responsible for ensuring that the article has been seen and approved by all the other authors. It is the responsibility of the author to obtain all necessary copyright release permissions for the use of any copyrighted materials in the manuscript prior to the submission. More information about permission request can be found at the web site.

Authors are asked to sign a warranty and copyright agreement upon acceptance of their manuscript, before the manuscript can be published. The Copyright Transfer Agreement can be downloaded from the web site.

## Publication Charges and Re-print

The author's company or institution will be requested to pay a flat publication fee of EUR 360 for an accepted manuscript regardless of the length of the paper. The page charges are mandatory. Authors are entitled to a 30% discount on the journal, which is EUR 100 per copy. Reprints of the paper can be ordered with a price of EUR 100 per 20 copies. An allowance of 50% discount may be granted for individuals without a host institution and from less developed countries, upon application. Such application however will be handled case by case.

*(Contents Continued from Back Cover)*

---

New Video Target Tracking Algorithm Based on KNN <i>Ding Ma and Zhezhou Yu</i>	709
Face Detection and Location System Based on Software and Hardware Co-design <i>Hua Cai, Yong Yang, Fuheng Qu, and Jianfei Wu</i>	715
Video Image Object Tracking Algorithm based on Improved Principal Component Analysis <i>Wang Liping</i>	722
Data Interpretation Technology for Continuous Measurement Production Profile Logging <i>Junfeng Liu, Heng Li, and Yingming Liu</i>	729
Rival Penalized Image Segmentation <i>Shaojun Zhu, Jieyu Zhao, and Lijun Guo</i>	736

---