Real-time TV Logo Detection based on Color and HOG Features

Fei Ye[1, *], Chongyang Zhang[1, 2], Ya Zhang[1, 2], and Chao Ma[3]
[1]Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Shanghai Key Labs of Digital Media Processing and Communication, Shanghai 200240, China
Email: sleepingco@163.com, {sunny_zhang, ya.zhang, chaoma}@sjtu.edu.cn

***Abstract*—This paper proposes a real-time TV logo detection algorithm that can detect logos embedded in TV videos or in the real-world videos/images. Unlike most existing TV logo detection methods, the proposed algorithm makes no assumption on temporal motion, spatial location, or any other visual view constrains on TV logos. The detection process consists of three stages: in the first stage, a color based region segmentation and candidate selection strategy is developed, which can narrow down the candidate search space and reduce computation cost significantly; at the second stage, SVM based classifier is trained, where geometric correction based on minimum rectangle bounding is used to improve the accuracy of the classifier, and affine transformation is adopted to construct a robust sample database; finally, the candidates are recognized by the trained SVM Classifiers using their HOG features. Experiments on several video sequences and logotypes have been carried out to verify the robustness and effectiveness of the proposed method.**

I. INTRODUCTION

In conventional video production, logotypes are used to convey information about content originator or the actual video content. Logotypes contain information that is critical to infer genre, class and other important semantic features of video [1]. Thus, detection of TV logos is essential for video content understanding, video semantic annotation, and display protection [1, 2].

Many works related to logo detection in video sequences have been reported in the literatures [1]—[4]. The existing TV logo detection techniques can be mainly classified into the following three types: 1) Difference based detection. Meisinger *et al*. used the image difference between consecutive frames to extract the logo mask with an assumption that the video content changes over time except the logo, and the frequency selective extrapolation technique was employed for logo in-painting [3]. In [4], average pixel gradient is computed over multiple key frames, and then high gradient regions are extracted to detect logos. In addition to static logos, background is assumed to be non-textured. 2) Spatial information based detection. Ahmet Ekin*et al*. use purely spatial information to detect TV logo robustly [2]. In [5], the detection accuracy is improved by assuming that the probability of the logos appearing in the four corners of the video frames is higher than that in the center. 3) Feature based detection. In [6], a neural network is trained using two sets of logo and non-logo examples to detect a transparent logo. It obtains a good detection rate at the expense of a rather large training set. In [7], from each logo, three sets of 2D HAAR coefficients are computed as the detection features. The logo's feature vector is formed by selecting the coefficients representing the averages and the low frequency coefficients of the RGB channels.

Most of the existing logo detectors have made remarkable achievement under the following assumptions: 1) Logos are static; or 2) logos have specific spatial location, or 3) logos have invariant features.

Fig.1. Examples of TV logos detection. The logos may appear in any arbitrary location, with different scale, different viewing angles and different lighting condition.

However, there are several problems brought about inevitably by the assumptions above.1) The difference based methods, which assume that logos are static, will fail when the scene itself is mostly stationary. 2) The spatial information based methods, which assume that logos have specific spatial location, will fail when logo appears in other place of a picture. For the examples shown in Fig. 1, the logo may appear in any location in the pictures. 3) The Feature based detections, which assume that logos have invariant features, will fail when the logo's features change with different view angles and different lighting conditions (see the logo on the microphone in left picture of Fig.1).

To solve these mentioned problems, a real-time TV logo detection algorithm with no assumption on temporal motion,

[*]This work was supported in part by the National Science Foundation of China (61001147) and the China National Key Technology R&D Program (2012BAH07B01), and by the High Technology Research and Development Program of China (2011AA01A107), the STCSM of Shanghai (12DZ2272600 and 10DZ2253200).

spatial location, or any other visual view constrains on the TV logos, is proposed in this work. Thus, the proposed method can detect logos not only from the TV video sequences but also from the life videos/images in the real-world, as shown in Fig. 2.

Fig. 2 Detection results on some sample images

In order to recognize TV logos appeared in each place of the video frames, a color features based candidate selection strategy is developed to narrow down the candidate search space.The color feature analysis is conducted in HSV color space to make it more robust to the illumination. Based on the fact that each TV logo mostly has a unique shape, HOG algorithm [10] is employed to extract the logo's contour feature. Moreover, a database that covers logo samples under 200 view angles is constructed to make the SVM classifier, which is based on HOG features, to be robust to the view angle changes. To solve the over learning of the SVM classifier, a geometric correction is also employed to simplify the classifier's decision hyper plane to improve its accuracy.

II. Real-time TV Logo Detection Using Color and HOG Features

The TV logo detection is performed through three stages. At the first stage, considered that the bright and unique color features owned by most TV logos, segmentation through color domain is employed to reduce the candidate region and hence speed up the recognition. At the second stage, in order to simplify the construction of the training data set, which is the key to build robust SVM Classifiers, affine transformation is applied. To solve the over learning of the SVM classifier, geometric correction is adopted. At the third stage, based on the obvious contour features of TV logo, HOG+SVM algorithm is employed to recognize the logo in the candidate region. The three stages will be described in detail in the following subsections.

A. Color based region segmentation and candidate selection

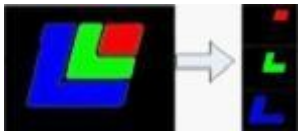

Fig.3. Example of extracting color features by selecting the first three dominant colors with the largest proportion.

The candidate selection starts with analyzing the sample logos and extracting the color features in the HSV Space. HSV(Hue, Saturation, Value) [11] is shown to have better results for image segmentation than RGB color space [12],[13], and it is also capable of emphasizing human visual perception in hues. Moreover, it shows high stability in hues under different illumination. The HSV space is divided manually into 12 areas by its H component. Each area stands for a kind of color. Three dominant colors with the largest proportion in the logo are extracted and kept as the logo's color features, as shown in Fig.3.

With the above selected domain colors, the candidate regions are obtained by low-cost color processing. As shown in Fig.4, the test image is divided into three sub-images by the three dominant colors. Each color lump in each sub-image is located by its bounding rectangle.

Fig.4. Divide the test image into three sub-images by the color features of candidate logo

This algorithm begins to traverse the color lumps in sub-image of the first primary color to find whether it intersects with other color lumps in the sub-images of the second and third primary color. Thus it can find the candidate search space which contains the same colors as the target logo，as shown in Fig.5.

Two color lumps intersect is strictly defined as whether their bounding rectangles intersect each other, as defined in Eq. (1).

$$a = \max(x1, a1), b = \min(y1, b1)$$
$$c = \min(x2, a2), d = \max(y2, b2) \quad (1)$$

If a<c and b>d, then the two bounding rectangles are considered to be intersected, as shown in Fig. 6.

Fig.5. Obtain the candidate region by the positional relationship among the color lumps

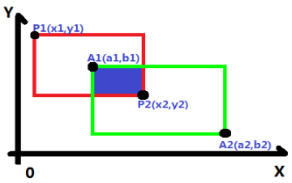

Fig.6. The blue area is the intersection of two bounding rectangle

B. *Sample database construction using affine transformation and geometric correction*

The candidate logo in the real-world images can be under any viewing angle. Thus, to construct a robust SVM Classifier, a large dataset covering most visual angle is needed. However, obtaining such a large dataset is usually very costly. To solve this problem, one novel dataset construction scheme using affine transformation is developed, in which affine transformations is adopted to generate approximation of logos in various angles from a standard logo. We construct a dataset of 10,000 samples of 200 viewing angles from 50 different standard logos, making it possible to quickly construct a new dataset for detecting a new type of logo. Fig.7 shows some examples of the data set.

Fig.7. Using affine transformation to approximate standard logo under various visual angles

Compared with many possible choices for what classifier to be used, for example classification trees and neural networks, the support vector machine (SVM) approach is considered a good selection because of its high generalization performance without the need to add a priori knowledge [14]. Thus, SVM is adopted as the classifier for the task of TV logo recognition in this work.

A large dataset covering 200 visual angles from 50 different standard logos is constructed for training a robust SVM Classifier. However, a large dataset will inevitably aggravate the aliasing between the positive and negative samples, which will lower the accuracy of SVM classifier. By weakening the aliasing between the positive and negative samples in data set, the SVM classifier's decision hyperplane is simplified, as illustrated in Fig.8, to improve the accuracy of the SVM classifier.

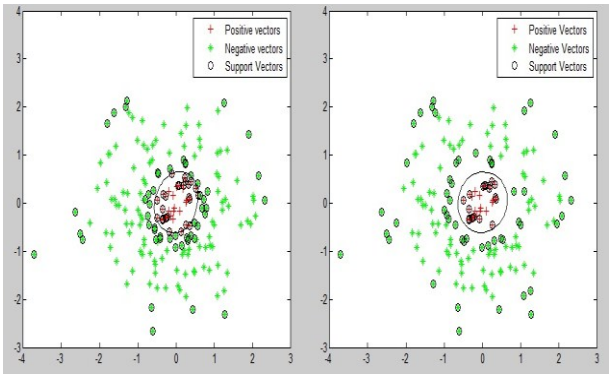

Fig.8.By weakening the aliasing between the positive and negative samples in dataset, the SVM classifier's decision hyperplane is simplified

In order to narrow down the needed sample visual angles, a geometric correction based on minimum rectangle bounding is employed to correct the logo in the candidate region.

The key to recover an image through affine transformation is to find its transformation matrix or its origin bounding rectangle, which is very difficult. We use one minimum bounding rectangle to approach the origin bounding rectangle. The results are shown in Fig.9 and Fig. 10.

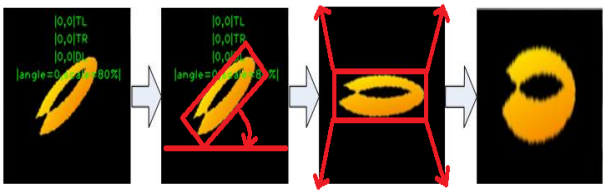

Fig.9. Correct the candidate region using the minimum bounding rectangle.

As illustrated in Fig.9, this algorithm begins with finding the minimum bounding rectangle of the logo (see the logo in second picture of Fig.9). Then, it adjusts the rectangular area to the horizontal direction (see the logo in second and third picture of Fig.9). Finally, the rectangular region is normalized to a fixed size (see the logo in forth picture of Fig.9).

Logo in the candidate region

Angle corrected

Size corrected

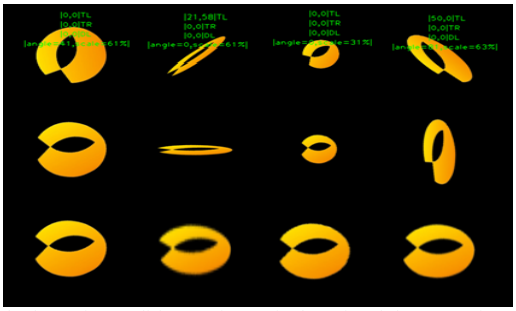

Fig.10. Correct the logos in candidate regions of other visual degrees using the minimum bounding rectangle.

As illustrated in Fig.10, through affine transformation, the logos in the candidate region which are under various visual angles have been corrected to approximate the standard logo.

C. *SVM based Recognition Using HOG*

Based on the fact that TV logos mostly have a simple

shape, the SIFT descriptors [15] and shape contexts [16] can only extract a few features, leading to low accuracy. Dalal and Triggs proposed the Histogram of Oriented Gradients (HOG) feature for pedestrian detection [17]. According to the existing works, HOG algorithm can effectively extracts the contour features, which makes it a good choice for the recognition task of TV logo. Thus, HOG features are selected to recognize the candidate logos finally using SVM classifier.

In each candidate region, pixels are first grouped into smaller spatial units called "cells". For each cell, a histogram feature on gradients orientations is extracted. The magnitude of the gradient is used as the weight for voting into the histogram. Multiple cells form larger spatial units called "blocks". The descriptor of each block is the concatenation of all cell features. A number of strategies are crucial for the final performance: when computing the HOG for each cell, a Gaussian weighting window is applied to each block; each pixel's gradient votes into the histogram using tri-linear interpolation in both spatial and orientation dimensions; the block feature is normalized for invariance to illumination. Inside each detection window, densely sampled and overlapping blocks produce redundant descriptors, which is important for better performance. The descriptor of a window is again the concatenation of all block features. Finally, a RBF Support Vector Machine is used to classify individual candidate region.

III. EXPERIMENTS

The experiment is divided into three main phases for testing purposes: segmentation by low-cost color processing, geometric correction of candidate region by affine transformation, and recognition using HOG+SVM.

A. Segmentation

We first test the validity of the segmentation by low-cost color processing. We consider the baseline method (Method A) as using a slide-window algorithm to get candidate region, while another baseline method (Method B, proposed method) uses low-cost color processing. Both of them use a SVM Classifier based on HOG features of the same dataset. The dataset is constructed by 200 affine transformations from each of 50 sample TV logos, which make it to achieve 100,00 samples in total. The 200 affine transformations try to cover all 360 degrees viewing angles. In total 200 images of 5 types of TV logos, have been tested and reported. Some images are taken from TV, and some of them have logo in the real world image. The results are shown in table Ⅰ and table II. The index A1—A5 indicate five different TV logos respectively, as shown in Fig. 11.

B. Geometric correction

We also test the validity of the geometric correction by affine transformation. Both Method B and D use the candidate region selected by the low-cost color processing. System B use SVM Classifier based on HOG features. While Method D, which is the proposed method, uses geometric correction to adjust the candidate region, and it also uses SVM Classifier based on HOG features. The results are shown in table II and tableⅣ.

C. Recognition by HOG+SVM

We test the validity of the recognition performance of the proposed method using Method C and Method D. Method C uses SIFT algorithm; While Method D, proposed by this paper, uses SVM Classifier based on HOG features. The results are shown in table III and tableⅣ.

Fig.11. Five different TV logos

Table Ⅰ. Method A: slide-window algorithms and recognition by HOG+SVM

	A1	A2	A3	A4	A5	TP	FP	FN	A(%)
A1	24	0	0	8	7	24	15	1	60.0
A2	0	27	0	0	0	27	0	13	67.5
A3	0	0	23	0	0	23	0	17	57.5
A4	7	0	0	25	6	25	13	2	62.5
A5	6	0	0	5	27	27	11	2	67.5
sum						126	39	35	63.0
Time	230ms								

Table II. Method B: low-cost color processing and recognition by HOG+SVM

	A1	A2	A3	A4	A5	TP	FP	FN	A(%)
A1	32	0	0	0	5	32	5	3	80.0
A2	0	29	0	0	0	29	0	11	72.5
A3	0	0	33	0	0	33	0	7	82.5
A4	0	0	0	31	0	31	0	9	77.5
A5	4	0	0	0	28	28	4	8	70.0
sum						151	9	40	75.5
Time	55ms								

Table III. Method C: low-cost color processing and recognition by SIFT algorithm

	A1	A2	A3	A4	A5	TP	FP	FN	A(%)
A1	23	0	0	0	0	23	0	17	57.5
A2	0	36	0	0	0	36	0	4	90.0
A3	0	0	25	0	0	25	0	15	62.5
A4	0	0	0	28	0	28	0	12	70.0
A5	0	0	0	0	32	32	0	8	80.0
sum						144	0	56	72.0
Time	120ms								

Table IV. Method D: low-cost color processing using affine transformation to correct the candidate region and recognition by HOG+SVM

	A1	A2	A3	A4	A5	TP	FP	FN	A(%)
A1	37	0	0	0	2	37	2	1	93.5
A2	0	35	0	0	0	35	0	5	87.5
A3	0	0	40	0	0	40	0	0	100
A4	0	0	0	38	0	38	0	2	95.0
A5	3	0	0	0	36	36	3	1	90.0
sum						186	5	9	93.5
Time	45ms								

D. RESULT

By comparing the results shown in table Ⅰ and table II, it can be found that the detection method using low-cost color processing is more than 5 times faster than that using slide-window algorithm. The reason that slide-window based algorithm is of relatively low efficiency is that, it need to examine about 150 candidate regions to traverse an image, while the low-cost color processing method examines only 20 regions. Moreover, the Recognition rate of Method B is also higher. During the experiment, due to the segmentation using low-cost color processing, a lot of background patterns with similar shape but different colors are excluded, which may be difficult to exclude in the slide-window algorithm in Method A. As is shown in table Ⅰ, due to the similar shape between logo A1 and A4, a few errors is made in Method A. However, Method B distinguishes the two logos perfectly by the obvious difference in colors.

In the second experiment, comparing the data in table II and tableⅣ, the recognition rate of Method D is higher than that of Method B. The ultimate recognition rate reaches as high as 93.5%. The result argues that through affine transformation, the candidate logo under various visual angles has been corrected into a small range. Thus it can weaken the aliasing between the positive and negative samples in dataset, and it also simplify the SVM classifier's decision hyperplane, which can improve the accuracy of the SVM classifier.

In the third experiment, comparing the data in table III and table Ⅳ, the recognition rate of Method D is higher than that of Method C. The ultimate recognition rate reaches as high as 93.5%. In table III, the recognition rate of A1, A3, and A4 is relatively low, and that of A2 reaches a high rate of 90%. That is because, SIFT algorithm extract only a few features from logos with simple shape like A1, A3, and A4, While A2 and A4 reach a relatively high recognition rate due to their complex shape. Based on the results of the experiment and the fact that TV logos mostly have a simple shape, one can draw the conclusion that HOG is more suitable to extract the logo features than SIFT.

IV. CONCLUSIONS

A novel real-time TV logo detection method has been introduced in this work. TV logos both in the TV sequences and in the real-world videos/images can be efficiently detected and recognized with the proposed algorithm. Affine transformation has been used to build the robust dataset, which covers every needed viewing angle, from a few sample logos. Thus, this system makes it possible that every user can construct their own database easily to meet their need. With the low-cost color processing, a lot of background patterns with similar shape but different colors are excluded, leaving only a few candidate regions to be recognized, which greatly improves the speed of the algorithm. By apply the geometric correction to the candidate regions, the needed viewing angles in the data set is greatly reduced from 360 degrees to less than 45 degrees. Thus it weakens the aliasing between the positive and negative samples in dataset, and it can also simplify the SVM classifier's decision hyperplane, which can improve the accuracy of the SVM classifier from 75.5% to 93.5%.

Promising results have been achieved by applying the presented system to a set of broadcast videos. In the future, we plan to test other features and classifiers to detect animated and more transparent logos.

REFERENCES

[1]. N. Guil, J.Gonza, Logotype detection to support semantic-based video annotation, Signal Processing: Image Communication 22 (2007) 669-676

[2]. A. Ekin, R. Braspenning, Spatial detection of TV channel logos as outliers from the content, Proceedings of Visual Communications and Image Processing (VCIP 2006), vol. 6077, pp. 60770X-1—8, 2006.

[3]. Meisinger, K., Troeger, T., Zeller, M., Kaup, A., Automatic TV logo removal using statistical based logo detection and frequency selective in-painting. Proc. ESPC'05(2005).

[4]. A. Albiol, M.J. Fulla, A. Albiol, L. Torres, Detection of TV commercials, Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004), vol. 3,2004, pp. 541–544.

[5]. Y. Wei-Qi, J. Wang, M.S. Kankanhalli, Automatic video logo detection and removal, Multimedia Syst. 10 (5) (2005) , 379–391.

[6]. S. Duffner, C. Garcı´a, A neural scheme for robust detection of transparent logos in TV programs, Lecture Notes in Computer Science—II, vol. 4132, pp.14–23,Springer, Berlin, 2006.

[7]. P. Duygulu, J. Pan, D.A. Forsyth, Towards auto-documentary: tracking the evolution of news stories, ACM Multimedia 2004—Proceedings of the 12th ACM International Conference on Multimedia, pp. 820–827, 2004.

[8]. B. Gu¨ nsel, A. Ferman, A.M. Tekalp, Temporal video segmentation using unsupervised clustering and semantic object tracking, J. Electron. Imaging 7 (3) (1998) 592–604.

[9]. D. Hall, F. Pe´ lisson, O. Riff, J. Crowley, Brand identification using gaussian derivative histograms, Mach. Vision Appl. 16 (1) (2004) 41–46.

[10]. Haritaoglu I, Harwood D, Davis L S, eta1. Real time Surveillance of People and Their Activities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1): 809830, 2000.

[11]. J. R. Smith, "Color for image retrieval," in Image Databases. John Wiley& Sons, Inc., 2002, ch. 11, pp. 285–311.

[12]. S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in Proceedings of IEEE International Conference on Image Processing, Sep. 2002, pp.589–592.

[13]. Z.-K. Huang and D.-H. Liu, "Segmentation of color image using EM algorithm in HSV color space," in Proceedings of IEEE International Conference on Information Acquisition, Jul. 2007, pp. 316–319.

[14]. Olivier Chapelle, Patrick Haffner, and Vladimir N. Vapnik, "Support Vector Machines for Histogram-Based Image Classification", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 10, NO. 5, SEPTEMBER 1999

[15]. M. Brown and D. Lowe. Recognising panoramas. In IEEE Int. Conf. on Computer Vision, pages 1218–1225, 2003. 1

[16]. Belongie, S., Malik, J., &Puzicha, J. (2002). Shape matching and object recognition using shape contexts. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(4), 509-522.

[17]. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR05, pages I: 886–893.