Automatic Fiducial Points Detection for Facial Expressions Using Scale Invariant Feature

Tie Yun [#1], Ling Guan[#2]

[#] *Ryerson Multimedia Research Lab*

Ryerson University, Toronto

[1] ytie@ee.ryerson.ca

[2] lguan@ee.ryerson.ca

***Abstract*—Detecting fiducial points successfully in facial images or video sequences can play an important role in numerous facial image interpretation tasks such as face detection and identification, facial expression recognition, emotion recognition, and face image database management. In this paper we propose an automatic and robust method of facial fiducial point's detection for facial expressions analysis in video sequences using scale invariant feature based Adaboost classifiers. Face region is first located using the face detector with local normalization and optimal adaptive correlation technique. Candidate points are then selected over the face region using local scale-space extrema detection. The scale invariant feature for each candidate point is extracted for further examination. We choose 26 fiducial points on the face region from training samples to build the fiducial point detectors with Adaboost classifiers. All the candidate points in the test samples are examined through these detectors. Finally, all the 26 facial fiducial points are located on each frame of the test samples. Cohn-Kanade database and Mind Reading DVD are used for experiment. The results show that our method achieves a good performance of 90.69% average recognition rate.**

I. INTRODUCTION

Facial expression contains much more powerful, natural and immediate information for human-to-human communication and social life, allowing people to understand each other beyond the verbal domain [1]. Automatically analyzing facial expression in real time without human intervention will help intelligent computer system to understand people's behavior well. It can be used in many potential applications in areas such as human-computer intelligent interaction (HCII), emotion analysis and recognition, computer vision, images understanding and synthetic face animation [2].

Many difficulties are involved in facial expression recognition techniques due to head pose, clutter, variations in lighting conditions, the variation across the human population and to the context-dependent variation even for the same individual [3]. Facial expression recognition attempts to find the most appropriate way to represent the facial expressions. It is the focal element from arising variability of facial expressions, and is an unwanted source for face recognition, where the uniqueness of a face is the central recognition criterion.

MMSP'09, October 5-7, 2009, Rio de Janeiro, Brazil.
978-1-4244-4464-9/09/$25.00 ©2009 IEEE.

Different methods have been explored to perform facial expression analysis in the past, which can be roughly categorized into two groups, the geometric feature-based methods and appearance-based methods [4]. The geometric facial feature-based methods present the shape, texture and/or location information of prominent components such as the mouth, eyes, nose, eyebrow, and chin, which can cover the variation in the appearance of the facial expression. The appearance-based methods, on the other hand, using image filters such as Gabor wavelets, generate the facial feature for either the whole-face or specific regions in a face image.

Fiducial points are a set of facial salient points [3], usually located on the corners of the eyes, corners of the eyebrows, corners and outer mid points of the lips, corners of the nostrils, tip of the nose, and the tip of the chin. Automatically detecting fiducial points can extract the prominent characteristics of facial expressions with the distances between points and the relative sizes of the facial components and form the feature vector. Additionally, choosing the feature points should represent the most important characteristics on the face and be extracted easily. In other words, the number of feature points should represent enough information and not be too many.

A. Related Works

Most of the existing approaches for facial expressions recognition are generally attempted to recognize Action Units (AUs) of the Facial Action Coding System (FACS) [5] and are based on deliberate and exaggerated facial displays. FACS is used to describe facial actions of facial muscles for detecting subtle changes of facial expressions.

Tian et al. developed an Automatic Face Analysis (AFA) system [6] to analyze facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features (deepening of facial furrows) in a nearly frontal-view face image sequence. This system recognizes fine-grained changes in facial expression into AUs and has achieved average recognition rates of 96.4 %. Petar and Katsaggelos [7] presented an automatic hidden Markov models (HMM) facial expression recognition system to utilized facial animation parameters (FAPs). The proposed multistream HMM facial expression system, which utilizes stream reliability weights, achieves relative reduction of the facial expression recognition error of 44% compared to the

single-stream HMM system. Yeasin et al. [8] presented a spatio–temporal approach in recognizing six universal facial expressions from visual data and using them to compute levels of interest. This approach achieved an average recognition rate of 90.9%. Sung and Kim [9] proposed a pose-robust face tracking and facial expression recognition method using a view-based 2D + 3D active appearance model (AAM) that extended the 2D + 3D AAM to the view-based approach, where one independent face model was used for a specific view and an appropriate face model was selected for the input face image. Three 2D + 3D AAMs were compared for the facial expression recognition performances and the proposed view-based 2D + 3D AAM demonstrated the best outperformance.

Due to high computing complexity, fiducial points are not well studied in the past. Cohn *et al.* in [10] presented an optical-flow based approach, which automatically tracked the selected facial features with a hierarchical algorithm for estimating the optical flow. Maghami *et al.* [11] selected facial feature points from the first frame to the last using a maximum cross-correlation algorithm followed by Kalman filter. The extracted feature vector was given to different classifiers to classify the face expressions within six basic emotions. The results showed that Bayes optimal classifier can reach the average correct classification rate of 93.72 % by this method. Lai et al. [12] proposed to use the integral optical density (IOD) to detect the fiducial points for near frontal face images and showed that the proposed algorithm was insensitive to the facial expressions, small rotation, different types of glasses and hairstyle. Ersi and Kiana [13] presented a feature based hybrid method to analyze local facial features located by a meta-version of the specification algorithm in the context of Local Feature Analysis (LFA) technique. Fiducial points were determined based on genetic algorithm and the output points were decorrelated. Valstar and Pantic [14] built an automatic facial expression reorganization system from face video. 20 facial fiducial points were detected by a localization method using individual feature GentleBoost templates. Then, a particle filter scheme was exploited to track the facial points. The AUs displayed in the input video and their temporal segments were recognized finally by Support Vector Machines (SVM). They got 90.2% average recognition rate.

B. Proposed System

In this paper we propose an automatic and fast method of 26 fiducial points' localization for facial expressions in video sequences using scale invariant feature based Adaboost classifiers. These 26 points are selected upon AUs with FACS and the descriptions of 26 fiducial points are shown in Table I.

TABLE I

FACIAL FIDUCIAL POINTS DESCRIPTION

Points No.	**Fiducial Points Description**
1	Top of the head
2	Tip of the chin
3	Left of the head
4	Right of the head
5	Left eye inner corner
6	Top of the left eye
7	Left eye outer corner
8	Bottom of the left eye
9	Right eye inner corner
10	Top of the right eye
11	Right eye outer corner
12	Bottom of the right eye
13	Left eyebrow inner corner
14	Top of the left eyebrow
15	Left eyebrow outer corner
16	Right eyebrow inner corner
17	Top of the right eyebrow
18	Right eyebrow outer corner
19	Top of the nose
20	Left nose corner
21	The medial point between left and right nostril centers
22	Right nose corner
23	Left corner of the mouth
24	Top of the upper lip
25	Right corner of the mouth
26	Bottom of the lower lip

Our method includes three major steps. The first step is to locate face region in the input images. Candidate points are then chosen and extracted to form the feature vectors using scale invariant feature in the second step. The third step is to localize fiducial points with Adaboost classifiers. The block diagram of this method is shown in Fig. 1.

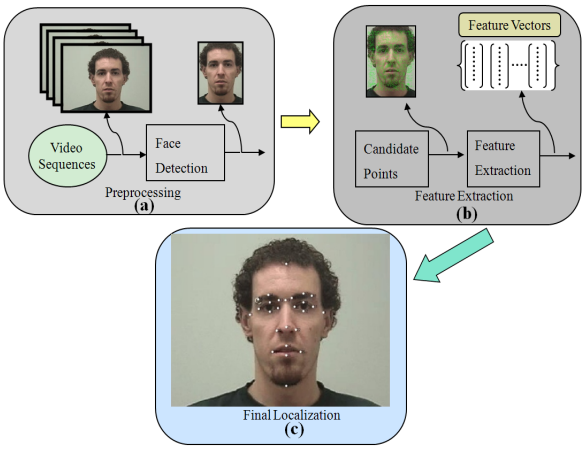


Fig. 1 Block diagram of the proposed method, (a) Face region detection, (b) Features Extraction, (c) Fiducial Points Localization

The remainder of this paper is organized as follows: Section 2 presents face region detection method. The feature extraction and final fiducial point's localization are discussed in Section 3 and 4 respectively. Section 5 presents the experimental setup and results. Finally, conclusions are drawn in Section 6.

II. FACE DETECTION

Face detection is important to the automatic fiducial detecting and tracking system and affects the overall system

performances. In this paper, we use the face detection method proposed in [15], which incorporates local normalization with optimal adaptive correlation (OAC) technique into a conventional face detector to alleviate illumination variation problem. The input video sequence is first regularized by local normalization to facilitate accurate and robust feature extraction and face detection, including gamma intensity correction, difference of Gaussian, local histogram matching and local normal distribution. The system's resistance can be evaluated to the most common classes of natural illumination variations.

Face candidate regions are roughly located by OAC and kernel canonical correlation analysis (KCCA). For a given normalized image, we can estimate the face candidates in the segmented image using the OAC value

$$C^* = \arg\max(a_i / b_i) \tag{1}$$

where a_i and b_i are the projections of OAC detector and the entire image OAC transform onto kernel space repetitively. The segmentation image mask *M(x,y)* for the original image *s(x,y)* is then generated from the correlation image *c(x,y)* as

$$\mathrm{M(x,y)} = \begin{cases} 0 & \mathrm{Th}_{c(x,y)} < C^* \\ 1 & \mathrm{Th}_{c(x,y)} \geq C^* \end{cases} \tag{2}$$

where $\mathrm{Th}_{c(x,y)}$ is the thresholds parameter for pixels corresponding to the face candidates. Compared with common automatic face detection algorithm, this process does not need to use the pyramid of downscaled copies of the input image and thus speeds up the processing.

The Gabor wavelets filters are applied for local feature extraction after candidates' selection and the face region is finally detected through a cascade classifier consisting of detectors with Adaboost algorithm.

III. Feature Extractiom

A. Candidates Selection

Most of the automatic feature point detection algorithms are implemented in such a way that every pixel in the input image is examined through feature detectors one by one to construct the feature vectors. Then the classifiers are applied on the feature vectors to perform the feature point detection. Practically, the number of feature points or equivalently the number of times the classification will be processed is typically in the tens of thousand depending on the image size and demagnification factor.

We propose to use scale space extrema method mentioned in [16] to efficiently detect the locations of the interest candidate points in the face region from last step. Compared with common automatic feature point detection algorithm, this method does not need to classify every pixel of the input image and thus speeds up the processing.

The scale space extrema can be located using the Gaussian convolution kernel function convolved with the input image. The description function L of input image in different scale space is expressed as:

$$L(x, y, \sigma) = G(x, y, \sigma) * s(x, y) \tag{3}$$

$L(x, y, \sigma)$ is the spatial scale images, where $s(x, y)$ indicts input image of face region, and $G(x, y, \sigma)$ is the Gaussian convolution kernel function that:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp[-(x^2 + y^2) / 2\sigma^2] \tag{4}$$

σ is the scale factor. The image zoom with σ and the smoothness of the image $s(x, y)$ would be varied with the changing of σ. Consequently, a series of scale image could be obtained. The scale space extrema are computed by the difference of Gaussian (DoG) function of the input image, which calculates the difference of two nearby scales separated by a constant multiplicative factor k

$$\begin{aligned} & D(x, y, \sigma) \\ & = [G(x, y, k\sigma) - G(x, y, \sigma)] * s(x, y) \\ & = L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{5}$$

where $G(x, y, \sigma)$ is the DoG function of the input image. Each pixel in the DoG images is compared to its eight neighbors on the same scale and each of its nine neighbors one scale up and down. The pixels with the local maximal or minimal values are chosen as interest candidates, including the adjacent scale, the position and scale of the local extreme point. These points are generally the feature points of the image, located on contour, corner and edges.

B. Feature Vectors Generation

After the position and scale σ of the interest candidates are determined from the input image, a gradient orientation histogram is calculated for the direction of each interest point in its neighborhood. The gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ are computed using pixel differences as:

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \tag{6}$$

$$\theta(x, y) = \arctan[\frac{L(x+1, y) - L(x-1, y)}{L(x, y+1) - L(x, y-1)}] \tag{7}$$

where L is the scale of feature from (3). By selecting a neighborhood F by the center of each interest point and calculating the directions of points in F, we can obtain the direction distribution and the statistical histogram. The gradient magnitude orientation is divided into 36 portions so as to be convenient in obtaining the direction distribution. The direction of the interest candidate point is the maximal component of the 36 phases in the statistical histogram.

C. Fiducial Point Detector

To detect the fiducial points from the interest candidate points, a set of fiducial point detectors with the feature description for the gradient orientation histogram of the input images are constructed.

At the center of each fiducial point, a 16×16 pixel neighborhood window F is selected and divided into 16 subregions by 4×4. Using equations (6) and (7), the directions and amplitudes for all pixels in the subregions are obtained, and then accumulated into orientation histograms summarizing the contents over 4×4 subregions. These are eight direction distributions in the ranges of (0, π /4, π /2,3 π

/4, π ,5 π /4,3 π /2,7 π /4,2 π) with the length corresponding to the sum of the gradient magnitudes near that direction within the region. The amplitude and Gaussian function are also applied on the eight direction distributions to create the direction statistical histogram of subfields. The feature description of each fiducial point is obtained by connecting the direction descriptions of all subfields. The total of the direction descriptions is 16, so the length of a fiducial point detector is 128=16×8, and should be normalized in order to ensure the illumination invariance.

IV. ADABOOST CLASSIFICATION

Boost algorithm has been proposed to reduce the redundancies of the high dimensional feature space and computational cost. The Adaboost algorithm by Viola and Jones [17] for face detection is a typically successful example as it has a very low false positive rate and can detect faces in real time. It can be trained for different levels of computational complexity, speed and detection rate which are suitable for specific applications. The performances of RealAdaboost [18], GentleAdaboost [19] and ModestAdaboost [20] for fiducail point detectors are compared in our work based on video sequences using GML AdaBoost Matlab Toolbox.

The boosting methods are generated on data classes and have a significant overlap. RealAdaboost is the generalization of a basic Adaboost algorithm and treated as a fundamental boosting algorithm. GentleAdaboost is a more robust and stable version of RealAdaboost. It is identified that GentleAdaboost performs slightly better than RealAdaboost on regular data, and considerably better on noisy data. The reason is that RealAdaboost over-emphasizes the atypical examples, which eventually results in inferior rules. GentleBoost gives less emphasis to misclassified examples since the increase in the weight of the example is quadratic in the negative margin, rather than exponential [14]. It is also much more resistant to outliers. ModestAdaboost is a regularized tradeoff of Adaboost, mostly aimed for better generalization capability and resistance for certain specific sets of training data.

V. EXPERIMENT AND RESULTS

A. Database

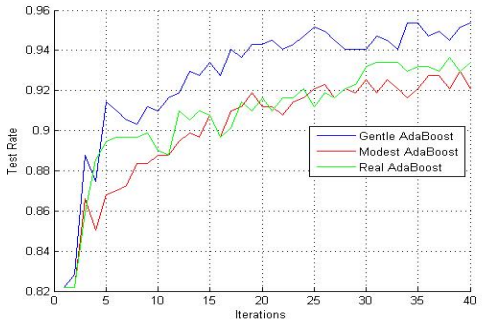
In this section, we evaluate the performance of the proposed method for face fiducial point detection using two video datasets, Cohn-Kanade database and Mind Reading DVD. Cohn-Kanade database consists of approximately 2000 image sequences in nearly frontal view from over 200 subjects, who are 18 to 50 years old; 69 % female and 31 % male; and 81 % Caucasian, 13 % African, and 6 % other groups. Each video pictures a single facial expression and ends at the apex of that expression while the first frame of every video sequence shows a neutral face. Image sequences from neutral to target display are digitized into 640×480 pixel arrays with either 8-bit gray-scale or 24-bit color values.

The Mind Reading DVD [21] is an interactive computer-based resource for face emotional expressions, developed by Cohen and his psychologist team. It consists of 2472 faces, 2472 voices and 2472 stories. Each video pictures the frontal face with a single facial expression of one actor (30 actors in total) of varying age ranges and ethnic origins. All the videos are recorded at 30 frames per second, last between 5 to 8 seconds, and the resolution is 320×240.

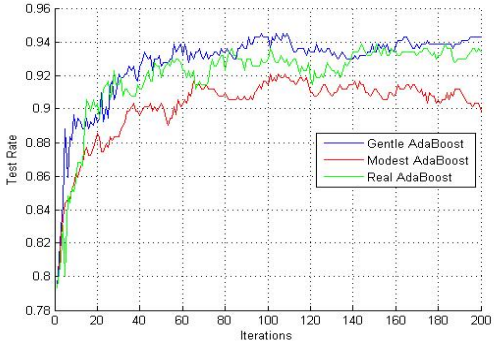
360 image sequences of 100 subjects from Cohn-Kanade database and 120 image sequences of 20 subjects from Mind Reading DVD are selected for our work, which constitute a total of 480 image sequences of 120 subjects.

B. Samples Training

We divide all the 480 image sequences into training and testing subsets containing 240 sequences each. For the training of the 26 face fiducial point detectors, the representative sets of positive and negative samples are selected from face region. We use 10 frames from each training sequence and manually label each fiducial point on the face region. We get 10×240 positive samples for each detector. We also choose another 5 arbitrary points in the same frame and get 5×10×240 negative samples, and in total have 14400 samples for each detector. As discussed in 3.3, the feature length of each one of the detectors is 128, so we have a 128×6 size feature vectors from one frame and a 768×2400 feature matrix for each training detector. The representation of features is highly redundant and computing the complete set is computationally expensive. Thus, we apply the Adaboost algorithms for the dimensionality reduction and detector classification.



a)



b)

Fig. 2 Test rates from Adaboost algorithms, a) 40 boosting iterations, b) 200 boosting iterations

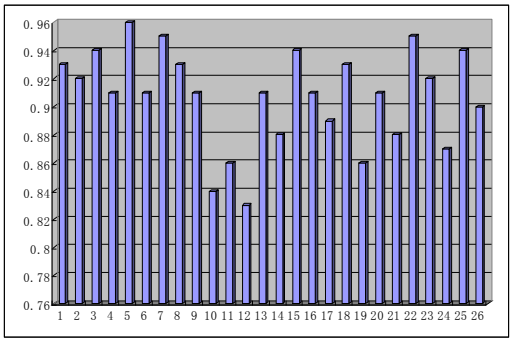
RealAdaboost, GentleAdaboost and ModestAdaboost are compared for error checks with 40 and 200 boosting iterations, shown in Fig. 2. GentleAdaboost returns the best detection rates from the results, and is selected as the classification algorithm for our system.

C. Testing and Results

Experiment results are presented in this section with the aforementioned databases. 240 image sequences are used for testing. We compare the automatically located fiducial points with the manually located points to evaluate the performance of the method. As the location of each fiducial point is at the center of a 16×16 pixel neighborhood window, and the feature vector for point detectors are extracted from this region, we consider that automatically detected points displaced 5 pixels distance from relevant true facial points are as successful detection points in the final results. The overall detection rates for each point are shown in Table II. And the proposed method achieves 90.69% average detection rate for the fiducial points' detection.

TABLE II

FINAL 26 FIDUCIAL POINTS DETECTION RATE



We illustrate some representative cases in Fig. 3. The proposed method is applied on each frame of the input video sequences, and the 26 fiducial points are automatically detected. We can also see that the detection is actually succeeded even under varying illumination condition.

Fig. 3 Sample sequences from the test videos for facial point detection

D. Comparison With State-Of-The-Art

A comparison of the detection rates achieved by feature point based methods for facial expressions recognition or face recognition is depicted at Table III. It shows that the proposed method has achieved the second best detection rate among the recent state-of-the-art methods. The best method has been the one proposed in [11], where a total 93.72% recognition rate has been achieved using a method based on Cross Correlation analysis. The main drawback of the method in [11] is that it is only tested in perfect manually aligned image sequences and no experiments in fully automatic conditions have been presented.

The proposed method also can reduce the computation time by using the scale invariant features extraction contrary to any other methods that examine every pixel one by one**.** Moreover, it can be used for both image and video based facial expression recognition.

TABLE III

COMPARISON OF FIDUCIAL POINT DECTION METHODS

Ref.	Initialization	Sequences	Features	DR
Our Prop.	Automatically	480	Scale Invariant Feature	90.69%
[6]	Automatically	300	Gabor	90.2%
[10]	Manually	504	Optical Flow	87.3%
[11]	Manually	13	Cross Correlation	93.72%
[12]	Automatically	743	Color and Edge	86%
[13]	Automatically	400	Gabor wavelet	87%

VI. Conclusion And Future works

In this paper, we proposed an automatic fiducial point detection method in video sequences based on Adaboost classifiers. Scale invariant features are extracted for candidate points generating and examined by the detectors. The experiment results show that our proposed method can get the fiducial points accurately with reduced processing time. The main contributions of this work are: a) constructing a set of 26 fiducial point detectors with scale invariant feature, b) decreasing computing time by candidates' localization, c) locating feature points automatically on a single frame and making it possible to eliminate the manual initiation step from tracking algorithm.

To further improve this work, we propose to do the followings: improve the accuracy and make the detected location closer to the true point; the processing time for each frame is much higher than general videos display time, and we will investigate a proper tracking method to reduce the total computation time; dynamic information will be generated for the purpose of face and human facial expression recognition.

REFERENCES

[1] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, T. J. Sejnowski, "Classifying Facial Actions", *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Vol. 21, No. 10, October 1999

[2] M. Pantic and M. S. Bartlett, *Machine Analysis of Facial Expressions*, *Face Recognition*, I-Tech Education and Publishing, 2007

[3] Z. Zeng, M. Pantic, G. I. Roisman, T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Vol. 31, No. 1, January 2009

[4] Y. Tian, T. Kanade, J. F. Cohn, *Facial expression Analysis*, Handbook of Face Recognition, Springer, 2005

[5] P. Ekman, W.V. Friesen, J.C. Hager, *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*, San Francisco: Consulting Psychologist, 2002

[6] Y. Tian, T. Kanade, J. F. Cohn, "Recognizing Action Units for Facial Expression Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Vol. 23, No. 2, February 2001

[7] P. S. Aleksic and A. K. Katsaggelos, "Automatic Facial Expression Recognition Using Facial Animation Parameters and Multistream HMMs", *IEEE Transactions on Information Forensics and Security*, Vol. 1, No. 1, March 2006

[8] M. Yeasin, B. Bullot, R. Sharma, "Recognition of Facial Expressions and Measurement of Levels of Interest From Video" *IEEE Transactions on Multimedia*, Vol. 8, No. 3, June 2006

[9] J. Sung and D. Kim, "Pose-Robust Facial Expression Recognition Using View-Based 2D + 3D AAM", *IEEE Transactions on System, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS,* Vol. 38, No. 4, July 2008

[10] J.F. Cohn, A.J Zlochower, J.J. Lien, T. Kanade, "Feature point tracking by optical flow discriminates subtle differences in facial expression", *Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998

[11] M. Maghami, R. A. Zoroofi, B. N. Araabi, M. Shiva and E. Vahedi, "Kalman Filter Tracking for Facial Expression Recognition using Noticeable Feature Selection", *International Conference on Intelligent and Advanced Systems,* 2007

[12] J. H. Lai, P.C. Yuen, W. S. Chen, S. Lao, M. Kawade, "Robust facial feature point detection under nonlinear illuminations", *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems,* July 2001

[13] E. F. Ersi and K. Hajebi, "Face recognition by fiducial point analysis", *IEEE CCECE*, Vol.2, May 2003

[14] M. Valstar and M. Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis", *Computer Vision and Pattern Recognition Workshop*, June 2006

[15] T. Yun and L. Guan, "Automatic face detection in video sequences using local normalization and optimal adaptive correlation techniques", *Pattern Recognition*, 10.1016/j.patcog.2008.11.026 (10 pages), December 2008

[16] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, Vol. 60, No. 2, 2004

[17] Viola and Jones, "Robust Real Time Object Detection", *Proceedings 2nd International Workshop on Statistical and Computational Theories of Vision*, 2001

[18] B. Wu, H. Ai, C. Huang, S. Lao, "Fast rotation invariant multi-view face detection based on RealAdaboost", *Proceedings IEEE International Conference on Automatic Face and Gesture Recognition*, May 2004

[19] J. Friedman, T. Hastie, R. Tibshirani, "Additive logistic regression: A statistical view of boosting", *The Annals of Statistics*, April 2000

[20] A. Vezhnevets, V. Vezhnevets, *Modest AdaBoost - Teaching AdaBoost to Generalize Better*, Graphicon-2005, Novosibirsk Akademgorodok, 2005

[21] *Mind Reading: The Interactive Guide to Emotions*, London, Jessica Kingsley, 2004