Proceeding of the IEEE
International Conference on Robotics and Biomimetics (ROBIO)
Shenzhen, China, December 2013

A Robot-in-the-Loop Face Detection and Recognition System with Coordinative Control Platform

Xiaohan Li, Yu Cheng, Song Chen, and Tao Zhang, *Senior Member, IEEE*

***Abstract*— This paper presents a robot-in-the-loop face detection and recognition system. This system is constructed based on a robot coordinative control platform. By means of the coordinative control platform, robot can actively capture high quality face images and move according to the image acquisition request, so that the performance of both the face detection and recognition can be remarkably improved. To improve the speed of face detection, a hybrid method consisted of Adaboost algorithm and Skin Color Model for face detection is proposed. In addition, the Embedded Hidden Markov Model (EHMM) is employed to recognize the detected faces. We demonstrate the effectiveness of the proposed approaches on several datasets. The superior property of the proposed system is also shown on a real robot system in the uncontrolled indoor conditions.**

I. INTRODUCTION

The aim of face detection and recognition is to judge the existence of faces in the image, getting the amount and positions of faces and confirming the identities of the faces in different scenes [1]. In the recent years, with the development of pattern recognition and artificial intelligence, face detection and recognition has become an important research subject promoted by the demand of airport security, video surveillance, et al. Meanwhile, the researchers in the field of robot have concerned on effective interaction between users and robots by vision. In the meanwhile, robot acquiring information by vision is one of the most important sensing methods [2]. Thus, developing face detection and recognition algorithms for robot system has good value and future prospects.

There are many literatures proposing various algorithms of face detection and recognition. Generally, the performance of face detection or recognition task has been largely effected by the quality of acquired face images [3], [4], for both training and testing. Many systems proposed and developed robust strategies to handle passively acquired face images in uncontrolled environment. [5] describes a face detection framework that is capable of processing images rapidly while achieving high detection rates in uncontrolled environment. [6] prosed a novel face detection approach based on a convolutional neural architecture, designed to robustly detect highly variable face patterns, in complex real world. [7] presents a fast image clustering algorithm with complex background, which can be used in an identification biometrics or a face classification system for robot vision.

Xiaohan Li and Tao Zhang are both with the Department of Automation, Tsinghua University, Beijing 100084, China, (email: kay1531@163.com, taozhang@mails.tsinghua.edu.cn)

Yu Cheng is with IBM TJ Watston Research Center, Yorktown Heights, New York 10598, and Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL 60208, USA, (email: yucheng@us.ibm.com)

Song Chen is with Beijing Institute of Spacecraft System Engineering, Beijing 100094, China, (email: chensong@hq.cast)

Only a little attention has been paid to actively controlling the image quality during the acquisition processing, which has many vision applications with non-cooperative avatars, like surveillance, robot vision of single camera are limited. [8]–[10] consider the resolutions as an measurement of the image quality. When capturing face images, the system exploit cameras zoom to FOI (field of interest) to capture high resolution face images. [11] tries to get the face image with good position with single camera. Frontal pose or profile pose is expected in their system to achieve high recognition rate in the experiments. These methods exploit single rules to measure the quality of face images and design active images acquisition systems.

In this paper, we proposed a robot-in-the-loop face detection and recognition system which can interact with human and actively acquire the face image. This system is constructed based on a robot coordinative control platform. To evaluate the quality of the candidate images, we propose a novel method which consider both pose variation and resolution measure. Then the best candidate image is selected in each sampling round and the robot can be moved according to the position of the target object. The quality of the image for face detection and recognition can be remarkably improved and the accuracy of the face detection and recognition can get high. In addition, a hybrid method consisted of Adaboost algorithm and color information for face detection is proposed, which improves the processing speed of face detection by integrating two rapid algorithms. The Embedded Hidden Markov Model (EHMM) is employed to recognize the detected faces. The main contribution of our paper is:

- We propose strategies for the system to select the face images by considering pose variation and resolution, and guide the robot to move.
- We proposed efficient face detection and recognition approaches in uncontrolled environment.
- The experimental results on several face image datasets demonstrate the effectiveness of our proposed method. We also perform the experiments with a real robot system and the results outperform the other baselines.

In this following of this paper, Section II introduces the proposed robot-in-the-loop real-time face detection and recognition system in detail, including coordinative control platform, face images acquisition, robot behavior control as

978-1-4799-2744-9/13/$31.00 ©2013 IEEE

well as face detection and recognition algorithms. In Section III the effectiveness of the proposed system is evaluated by experimental results on several benchmark datasets and on a real robot system in the uncontrolled indoor conditions. Finally, conclusions and future work are drawn in Section IV.

II. Robot-in-the-Loop Real-Time Face Detection and Recognition System

A. Knowledge Model-Based Robot Coordinative Control Platform

1) Knowledge Model-Based Coordinative Control Platform: The Knowledge Model-Based Intelligent Coordinative Control Platform (K-ICCP) provides a central module, which acts as blackboard, knowledge processing brain, memory, and do the judgment, task planning and execution. It also provides software tools necessary for integration of various existing modules over a TCP/IP network. K-ICCP is a frame-based knowledge engineering environment. The features of this network platform are ”platform-independent” as existing robots and software modules often rely on different platforms or operation systems, ”network-aware” as the modules must interact on a network, supporting ”software agent” and being ”user friendly”. K-ICCP is targeted to be the platform on which a group of co-operative robots (or their agents) operate on top of frame knowledge. K-ICCP consists of the following software components:

- GUI Interface: A user-friendly graphical interface to the internal knowledge manager and the inference engines. It provides the users direct access to the frame-based knowledge.
- Knowledge Database and Knowledge Manager: This is the K-ICCP core module that maintains the frame systems as Java class hierarchy, and performs knowledge conversion to/from XML format.
- Inference Engines: Verify and process information from external modules that may result in instantiation or destruction of frame instances in the knowledge manager, and execution of predefined actions.
- JavaScript Interpreter: Interprets JavaScript code which is used for defining conditions and procedural slots in a frame. It also provides access to a rich set of standard Java class libraries that can be used for customizing K-ICCP to a specific application.
- Basic Class for Software Agent: Provide basic functionality for developing software agents that reside on networked robots.
- Network Gateway: This is a daemon program allowing networked software agents to access knowledge stored in K-ICCP. All K-ICCP network traffics are processed here.

Fig. 1 illustrates the K-ICCP knowledge editor showing the frames hierarchy for human-robot system.

In K-ICCP, a human-robot system is described by XML format according to its knowledge representation. XML is a markup language for documents containing structured information. With text-based XML format, frame hierarchy can be serialized and stored in a local file. It can be also transmitted over the network to a remote K-ICCP. In addition, the frame system can be illustrated in the K-ICCP Graphic User Interface. Corresponding to XML file, there is an interpreter to translate XML specification into relative commands. With the XML format, the knowledge model of human-robot system can be defined in K-ICCP and the coordinative control of robots can be implemented as the following explanation.

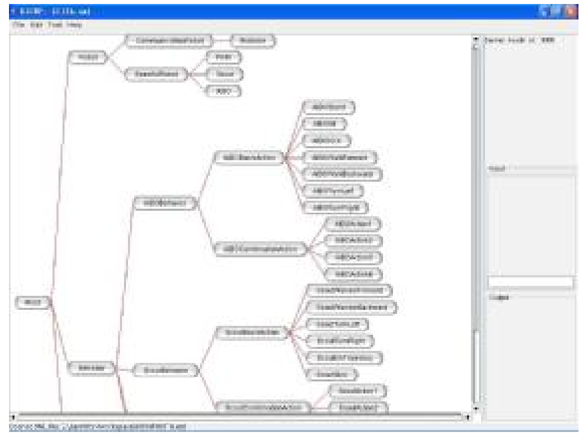

Fig. 1. K-ICCP knowledge editor showing the frames hierarchy for human-robot system.

In order to implement robot behaviors for face detection and recognition by means of K-ICCP, the results of face detection of recognition are essential to trigger the various behaviors of multiple robots. In addition, human-robot interaction by use of speech as well as sentence parsing is also adopted to assist the process of face detection and recognition. In the behavior frames, many independent programs for performing various functions of robots are adopted by the specific slot of ”on-Instantiate”. If these behavior frames are activated, these functions will be called and robots will conduct their relative actions. The detail explanation of robot behaviors for face detection and recognition will be given in the following.

B. Robot Behavior Control with Active Image Acquisition

We consider the image acquisition as well as robot control strategy in our single camera system. When only one candidate object exists in the targeting area, the situation is simple and the camera can serve for it. While the number of object is more than one, active selection is needed among all the candidates and the strategy needs to consider which target should be captured each time. In this paper, we do not consider multi-agent robot system [12].

Previous method exploit either pose-based or resolution-based method to select the target objects, which only use single measurement. In this paper, we propose a robust method by combining both pose and resolution factors. Before each capturing time, candidate objects position and moving direction at the sampling time are estimated based on the method in [11]. Pose-based and resolution-based measure are calculated, with the hypothesis that face orientation is the same as the direction as candidate object moves in. Then the robot system can move according to the position of the target objects. We call this method as joint-pose-resolution (Joint-Pose-Res) and the measured value with both as resolution and pose as rpEval in this paper. The detail of the main procedure is given as below:

1) Input: given area A_1, all the N candidate images $\mathrm{obj_1}, \cdots, \mathrm{obj_N}$ in A_1, the camera in area A_2 and its velocity v_{Cam}.
   Output: for each sampled time t, the selection of target object $\mathrm{obj_{Tar}}$ from $\mathrm{obj_1}, \cdots, \mathrm{obj_N}$.
2) Estimating the possible position range R_{Cam} of the camera in time t:
$$A^t_{\mathrm{Cam}} = \{Area | A \leq Pos^{t-1}_{\mathrm{Cam}} + v_{\mathrm{Cam}} \cdot dt\} \quad (1)$$
   where Pos^{t-1}_{Cam} is last position of the camera. Then We sample K possible positions from R_{Cam} with $P(k)^t_{\mathrm{Cam}} = Random(R_{\mathrm{Cam}})$.
3) Estimating the position Pos^t_i and orientation Ori^t_i for the i candidate object $\mathrm{obj_i}$ at time t.
$$\begin{aligned} Pos^t_i &= Pos^{t-1}_i + v_{\mathrm{Cam}} \cdot dt \\ Ori^t_i &= [0, 0, atan2(Ori^t_i \cdot v^y_{\mathrm{Cam}}, Ori^t_i \cdot v^x_{\mathrm{Cam}})]^T \end{aligned} \quad (2)$$
4) Calculating the position and resolution measured value $pEval(k)^t_i$, $rEval(k)^t_i$ between the candidate i image and the possible sampled positions $P(k)^t_{\mathrm{Cam}}$ of the camera:
$$\begin{aligned} pEval(k)^t_i &= InterSection(-Ori^t_i, Pos^t_i - P(k)^t_{\mathrm{Cam}}) \\ rEval(k)^t_i &= \frac{\mathrm{fWidth^t_i}}{\mathrm{fLenth_{Cam}}} \cdot Dis(Pos^t_i, P(k)^t_{\mathrm{Cam}}) \end{aligned} \quad (3)$$
   where $\mathrm{fWidth^t_i}$ is the face width of the i object image and $\mathrm{fLenth_{Cam}}$ is the focal length of the camera. $Dis()$ is the distance function.
5) After we get the $pEval(k)_i$ and $rEval(k)_i$ for all the k, the combined evaluation $rpEval_i$ can be calculated by:
   - For $i = 0$ to k, if $rEval(k)_i > rEval_{th}$, $pSet_i = pEval(k)_i \bigcup pSet_i$.
   - $rpEval_i = Mean(pSet_i)$, where $Mean()$ is the mean distance function for all the points for the set $pSet_i$;
   - $rpGain_i = rpEval^t_i - rpEval^{t-1}_i$
   - Choose the largest $rpGain_i$ and the target object $\mathrm{obj_{Tar}}$ in sample time t is $\mathrm{obj_i}$.
6) Update the training image pools. Robot moves to the position Pos^t_i and the orientation Ori^t_i of object $\mathrm{obj_i}$.

C. Face Detection

In the proposed system, to efficiently detect human faces, we introduced a method integrating the Adaboost algorithm and skin color model method.

Adaboost algorithm was first proposed by Freund Y. and Schapire R.E. in 1996, which was an iterative and very important machine learning method [13]. Based on this method, Viola, P. and Jones, M. proposed an improved machine learning approach, which could process the images extremely efficiently and achieve high detection rates, for visual objects detection [5], [14].

Adaboost detector employs the extended set of Haar-like features which is proposed by Lienhart et al. [15], [16] to make classification. Each feature is consisted of 2 to 3 rectangles to detect edge features, center-surround features and line features. Fig. 2 illustrates the set of features. In order to compute the features rapidly, Viola P. introduced the integral image representation for images. The integral image at location (x, y) means the sum of the pixels at the upper left side of (x, y), as
$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x, y) \quad (4)$$
where $ii(x, y)$ is the integral image and $i(x, y)$ is the original image.

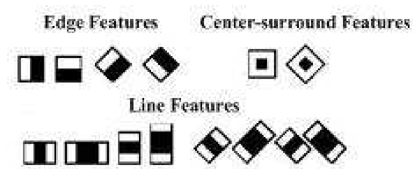

Fig. 2. Haar-like features for Adaboost algorithm.

With the feature set and a training set of positive and negative images, Adaboost algorithm is used to train strong classifiers and a cascade of classifiers is constructed to achieve increased detection performance while reducing computation time. The main procedure is as below:

1) Given example images $(x_1, y_1), \cdots, (x_n, y_n)$, and $y_i = 1, 0$ for negative and positive example respectively.
2) Initialize weights
$$w_{1,i} = \begin{cases} \frac{1}{2m}, & m \text{ is the number of negative examples} \\ \frac{1}{2n}, & n \text{ is the number of positive examples} \end{cases} \quad (5)$$
   for $y_i = 1, 0$.
3) For $t = 1, \cdots, T$:
   - Normalize the weights $w_{t,i} \leftarrow w_{t,i} / \sum^n_{j=1} w_{t,i}$, so that w_t is a probability distribution.
   - For each feature j trains a classifier h_j which is restricted to using a single feature. The error is $\varsigma_j = \sum_i w_i |h_j(x_i) - y_i|$.
   - Choose the classifier with the lowest error.
   - Update the weights: $w_{t+1,i} = w_{t,i} \beta^{1-e_i}_t$, where $e_i = 0$ if example x_i is classified correctly, otherwise $e_i = 1, \beta_t = \varsigma_t / (1 - \varsigma_t)$.
4) The final strong classifier is:
$$h(x) = \begin{cases} 1, & \sum^T_{t=1} \alpha_t h_t(x) \geq (\sum^T_{t=1} \alpha_t)/2 \\ o, & \text{otherwise} \end{cases} \quad (6)$$

where $\alpha_t = log(1/\beta_t)$:

5) Constructed the cascade of strong classifiers.

On the other hand, skin color model has been used and proven to be an effective feature in face detection [17], [18]. Several color spaces have been utilized to label pixels as skin including RGB, normalized RGB, HSV, YCrCb, YIQ, CIE XYZ and CIE LUV. YCrCb model is influenced little by illumination variations and the most simplest and effective model. Considering both the detection rates and speed, we exploit the YCrCb model and classify a pixel simply to have skin tone if its values fall within a range, which is learned from the training dataset. In Fig. 3 shows the distribution according to the statistics of 200 images in our dataset.

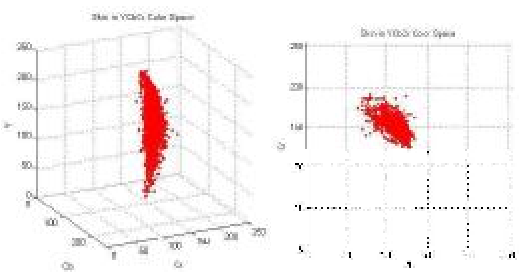

Fig. 3. Distribution of skin color in YCrCb model.

The skin color model is used to verify the regions detected by Adaboost algorithm. If the ratio of skin color pixels and total pixels is larger than the threshold, which is set as 0.355 got by experiments, the regions are considered as faces finally.

D. Face Recognition

In the subsection, Embedded Hidden Markov Model (EHMM) [19] is employed to recognize the detected faces. We know Hidden Markov Model is a set of statistical models used to characterize the statistical properties of a signal, which consists of five elements: (1) N, the number of states in the model; (2) M, the number of different observation symbols; (3) A, the state transition probability matrix; (4) B, the observation symbol probability matrix; (5)Π, the initial state distribution. With the factors, a HMM is defined as

$$\lambda = (A, B, \Pi) \tag{7}$$

Each of facial regions (hair, forehead, eyes, nose, mouth or chin) is assigned to a state in a left to right 1-demension continuous HMM. Then the observation vectors consist of a set of 2D DCT coefficients that are extracted from each region. After the HMM $\lambda = (A, B, \Pi)$ is being initialized, the training data is uniformly segmented from top to bottom in $N = 5$ or 6 states and the observation vectors associated with each state are used to obtain initial estimates of the observation probability matrix B. The initial values for A and Π are set given the left to right structure of the face model.

EHMM is similar to HMM, but consists of s set of super states. Super states divide face into different regions which are hair, forehead, eyes, nose and mouth, from top to bottom. Each super state also consists of some states named embedded states. The faces are divided into some blocks both vertically and horizontally. 2D DCT features are extracted from the blocks as the observation vectors. EHMM utilizes more information of faces than HMM and the recognition performance of EHMM is good. Fig. 4 describes the face structure of EHMM and Fig. 5 is the program flow of EHMM.

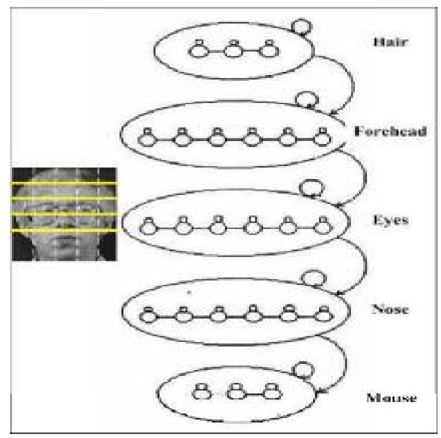

Fig. 4. Face structure of EHMM.

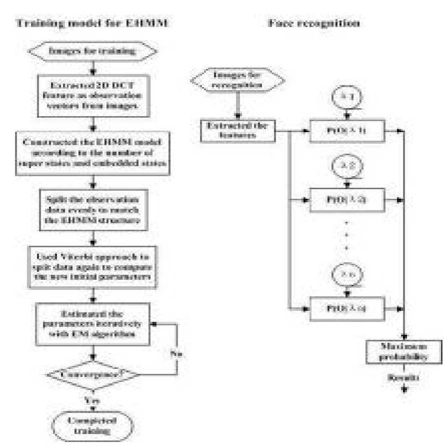

Fig. 5. Flowchart of EHMM.

III. Results

In this section we present the performance of our proposed algorithm on several real-world datasets. We first describe the results of face detection and recognition on some benchmark datasets, and then present our experimental results with a real robot system.

A. Face Detection and Recognition Evaluation

GTAV face database [20] is used to test the effectiveness of the proposed face detection algorithm in this paper. GTAV database includes a total of 44 persons with 34 pictures per person which corresponding to different pose views $(0°, \pm30°, \pm45°, \pm60°, \pm90°)$ under three different illuminations (environment or natural light, strong light source from an angle of $45°$, and finally an almost frontal mid-strong light source). Furthermore, at least 20 more additional frontal view pictures are included with different occlusions and facial expression variations. The resolution of the images is 320×240 with BMP format. In the experiment, 40 persons with 15 images per person with different illuminations and facial expressions are selected for the experiments. Half of the faces images are used to train the cascade of classifiers. We divided the test dataset into groups. Part of face detection results are shown in Fig. 6 and Table I summarizes the detection rate and speed. Our method can achieve a good performance efficiently.

Fig. 6. Part of face detection results on GTAV.

TABLE I
FACE DETECTION RESULTS ON GTAV DATASET

	# of Images	# of Persons	Detection rate	Speed
1	115	19	85.8%	0.22s/image
2	165	25	79.6%	0.31s/image

We use the ORL face database [21] to evaluate the recognition algorithm. The ORL database consists of 40 persons with 10 pictures per person. Half of the data (fiver pictures per person) are used to train the EHMM while the others are for test. We run the experiments with different EHMM embedded states. The recognition results are shown in table II with the recognition speed. The EHMM can achieve excellent recognition results under certain defined situation.

B. Robot System

A real robot system has been used to evaluate the proposed approach in real-time performance. The adopted robot system is shown in Fig. 7. Then the face detection and recognition results are obtained under the system condition.

TABLE II
FACE RECOGNITION RESULTS ON ORL DATASET

	Super states number	Embedded states number	Recognition rate	Recognition speed
1	5	3,6,6,6,3	100%	4.1s/image
2	5	1,1,1,1,1	92%	1.3s/image

Face database is established by capturing the face images from 15 persons in the laboratory, with 13 pictures per person. Different images acquisition methods are performed in this processing, performed by the robot system with matched control strategies. Given these datasets, half of images captured are used in the training stage for detection and recognition. The test set including all the possible face images to be captured in the scene, where each target random cross the area differently.

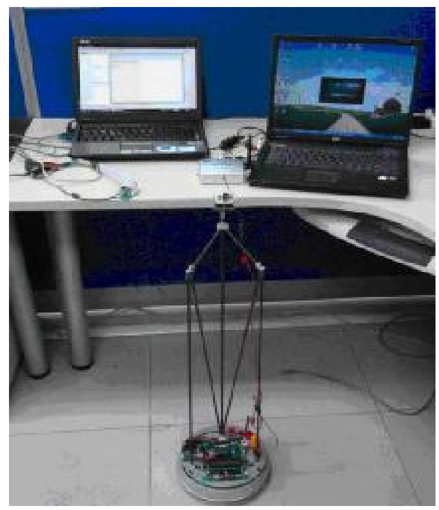

Fig. 7. The robot system in the experiment.

Different results of different method (including Joint-Pose-Res proposed in this paper) are compared. The baseline methods are: (1) Resolution: capture the images with high resolutions [8]–[10]; (2) Single Pose: acquiring images with poses close to frontal or profile pose [11]; (3) Random: assigns a random image to a camera each time.

TABLE III
PERFORMANCE COMPARISON WITH DIFFERENT IMAGES ACQUISITION METHODS

Image Acquisition Method	rpEval	Detection rate	Recognition rate
Resolution	3.60	47.4%	22.1%
Single Pose	2.86	42.1%	18.4%
Random	3.10	37.5%	25.3%
Joint-Pose-Res	4.19	62.1%	39.8%

To evaluation the performances with different method, we use rpEval, Detection rate and Recognition rate as the evaluation matrix. As table III illustrates, the performance of Joint-Pose-Res outperforms the other methods on the real-time system. The detection rate can be more than 60% and recognition rate can achieve about 40%. The resolution and single pose have similar performances, while they achieve lower recognition rate than the random sampling. It is shown that they both are not very robust methods. With more deliberative recognition methods, the rate is expected to be even higher. The strategy proposed in this paper is ubiquitous for various recognition methods and keep the original recognition power.

IV. CONCLUSIONS

The paper introduces a robot-in-the-loop face detection and recognition system with coordinative control platform. This system is developed for robots to have the capability of interaction with people in the uncontrolled indoor environments. The target images are selected by the proposed joint-pose-resolution method and robot can move according to the position of the target images. In addition, Adaboost algorithm and skin color model are used for efficient human face detection in the complex background rapidly and efficiently. EHMM algorithm is employed to recognize the detected faces with high recognition rate. The simulation and experimental work by use of an actual robot verified the effectiveness of the proposed system.

There are several direction for the future work. First, we would like to extend current method to multiple cameras system and develop algorithms for robot control. Second, we will improve the current EHMM to increase the recognition speed. In addition, active learning approaches which is designed for stream robot vision data [22] could be incorporated.

V. ACKNOWLEDGEMENT

This work is supported by the National Basic Research Program of China (973 Program, No.2010CB731800). We would like to thank Hao Sun, Xuedong Chen and Heyi Li for their useful comments and insightful suggestions.

REFERENCES

[1] E. Garcia, M. Jimenez, P. De Santos, and M. Armada, "The evolution of robotics research," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 90–103, Mar. 2007.

[2] S. Helmer, D. Meger, P.-E. Forssén, S. McCann, T. Southey, M. Baumann, K. Lai, B. Dow, J. J. Little, and D. G. Lowe, "Curious george: The ubc semantic robot vision system," AAAI Technical Report Series, Tech. Rep. AAAI-WS-08-XX, October 2007.

[3] N. Dadashi, A. Stedmon, and T. Pridmore, "Automatic components of integrated cctv surveillance systems: Functionality, accuracy and confidence," in *Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, ser. AVSS '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 376–381.

[4] T. Shan, S. Chen, C. Sanderson, and B. C. Lovell, "Towards robust face recognition for intelligent-cctv based surveillance using one gallery image," in *Proceedings of the 2007 IEEE Conference on Advanced Video and Signal Based Surveillance*, ser. AVSS '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 470–475.

[5] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[6] C. Garcia and M. Delakis, "Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 26, no. 11, pp. 1408–1423, Nov. 2004.

[7] Y. Cheng, T. Zhang, and S. Chen, "Fast person-specific image retrieval using a simple and efficient clustering method," in *Proceedings of the 2009 international conference on Robotics and biomimetics*, ser. ROBIO'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1973–1977.

[8] M. Tistarelli and E. Grosso, "Active vision-based face authentication," *Image and Vision Computing*, vol. 18, no. 4, pp. 299–314, 2000.

[9] A. Bakhtari and B. Benhabib, "An active vision system for multitarget surveillance in dynamic environments," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, no. 1, pp. 190–198, 2007.

[10] A. Bakhtari, P. Adragna, and B. Benhabib, "Autonomous surveillance of a dynamic, multi-subject environment: an implementation for automated facial recognition," in *Measurement Systems for Homeland Security, Contraband Detection and Personal Safety Workshop, 2005.(IMS 2005) Proceedings of the 2005 IEEE International Workshop on*. IEEE, 2005, pp. 21–26.

[11] M. Hanheide, S. Wrede, C. Lang, and G. Sagerer, "Who am i talking with? a face memory for social robots," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. IEEE, 2008, pp. 3660–3665.

[12] Y. Cheng, T. Zhang, and J. Wang, "Multi-agent system model for urban traffic simulation and optimizing based on random walk," in *Advances in Neural Network Research and Applications*, ser. Lecture Notes in Electrical Engineering, Z. Zeng and J. Wang, Eds., vol. 67. Springer Berlin Heidelberg, 2010, pp. 703–711.

[13] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Thirteenth International Conference on Machine Learning*. San Francisco: Morgan Kaufmann, 1996, pp. 148–156.

[14] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *ICIP (1)*, 2002, pp. 900–903.

[15] R. Lienhart, E. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *In DAGM 25th Pattern Recognition Symposium*, 2003, pp. 297–304.

[16] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[17] J.-C. Terrillon and S. Akamatsu, "Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images," in *Proc. of the12th Conf. on Vision Interface (VI '99*, 2000, pp. 180–187.

[18] A. V. Nefian, M. H. Hayes, and III, "Hidden markov models for face recognition," in *Proc. International Conf. on Acoustics, Speech and Signal Processing*, 2005, pp. 2721–2724.

[19] H. Othman and T. Aboulnasr, "A separable low complexity 2d hmm with application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1229–1238, Oct. 2003.

[20] E. D. A Albiol, L Torres, "Fully automatic face recognition system using a combined audio-visual approach," *Vision, Image and Signal Processing, IEE Proceedings*, vol. 3, no. 152, pp. 318–326, 06 2005.

[21] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Machine Learning (CVPR'07)*, 2007.

[22] Y. Cheng, K. Zhang, Y. Xie, A. Agrawal, and A. Choudhary, "On active learning in hierarchical classification," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, ser. CIKM '12. New York, NY, USA: ACM, 2012, pp. 2467–2470. [Online]. Available: http://doi.acm.org/10.1145/2396761.2398668