Application of Face Recognition Methods for Process Automation in Intelligent Meeting Room

Alexander L. Ronzhin

SPIIRAS, 39, 14th line, St. Petersburg, 199178, Russia
ronzhinal@iias.spb.su

Abstract. This paper describes the automatic registration technique based on face recognition of meeting participants, which has been implemented in the intelligent meeting room. This technique provides unobtrusive recognition and picture making of participant faces. Application of the developed technique makes it possible to reduce the work of secretaries and videographers; it also allows participants to concentrate on the discussed issues at the expense of automated control of sensory equipment. For the experimental evaluation of the developed methods and the technique about 52,000 photographs for 36 participants were accumulated from a high resolution camera. During the experiments three face recognition methods LBP, PCA and LDA were compared. The experimental results showed that method LBP has the highest recognition accuracy 79,3%, but the PCA method has the lowest percentage of the false positives 1,3%, which is important aspect in the participants identification.

Keywords: Multichannel signal processing, intelligent meeting room, computer vision, face recognition, processes automation.

1 Introduction

Application of intelligent information technologies in business and education, including at carrying out distributed events and for automation of the speaker's talk recording at the meeting, is important issue due to the increasing mobility of people and necessity to control the quality of decisions [1, 2]. Nowadays, the evolvement of a scientific paradigm of the intellectual space has shaped several models of intelligent environment that may serve users in a confined space: intelligent room, house, lecture hall, meeting room [3, 4, 5]. Development of the tools for capture and processing of audiovisual signals, which are capable to contactless evaluate the current situation in the room, is one of the main fields of research in this area.

When designing intelligent rooms for meetings, lectures, scientific and educational activities the following methods of audio and video signals processing are now most widely used: 1) detection and tracking of participants based on video monitoring [6]; 2) estimation of head orientation and face recognition [7]; 3) sound source localization [8, 9]; 4) speech recognition [10]; 5) speaker diarization [11]; 6) speech synthesis [12]. Application of these methods and their combination makes it possible to

S. Balandin et al. (Eds.): NEW2AN/ruSMART 2014, LNCS 8638, pp. 156–163, 2014.

develop tools for automatic recording of the speakers' talks, organizing of television broadcasts, journaling and archiving of audiovisual recordings after the event.

Let us to consider SPIIRAS intelligent meeting room. For its design ergonomic aspects of multimedia, audio-visual recording equipment location were taken into account to provide coverage and service of the greatest possible number of participants. Functionality of the intelligent meeting room includes its equipment as well as methods needed to the implementation of information support services and automation of events. At implementation of the participants system monitoring in the intelligent room, which based on distributed audiovisual signals processing were used as the existing methods of digital data processing (image segmentation, calculation and comparisons of the histograms, etc.), and developed own proprietary methods, such as method for meeting participants registration, the method for audiovisual recording of their performances. Detailed description of the equipment and audiovisual data processing methods used in the intelligent room is presented in [3, 8, 13, 14, 15].

This paper is organized as follows. The second section discusses methods of biometric identification based on face recognition. The third section describes the specifics of the developed technique and methods of automatic registration of meeting participants based on face recognition. The fourth section presents the experiments, conditions and results of the evaluation of face recognition methods.

2 Biometric Identification Methods Based on Face Recognition

Biometric identification is the automatic identification or user verification technology on the basis of physical characteristics and personal traits. Biometric characteristics and traits are divided into two categories: behavioral and physical. Behavioral characteristics include events such as signature and typing rhythm. Biometric systems for physical characteristics used to identify eyes, fingers, hands, voice and face.

Face recognition system is a computer application for the automatic identification or verification of the digital image part, with video frame read from the video source [16, 17]. Face recognition system allows user to be identified just by walking past the surveillance camera. People often recognize each other by unique facial features which is the most successful form of human observation. On this phenomenon the biometric face recognition technology is based. Currently, the interest in face recognition technology has been excited by the availability and low cost of video equipment as well as their unobtrusive presence. Such systems are already available on the market and are claimed to accurately, efficiently and rapidly recognize the human face. But there are some restrictions: 1) recognition errors occur due to the external conditions, such as ambient light and change of the position of the face relative to the camera, etc., and 2) the lack of accuracy appears during the processing of frames with a gestures variety in them.

In most cases, face recognition algorithm may be divided into two stages: 1) determining the position of the user face in the picture with a simple or complex background [18]; 2) recognizing face to identify a user. At both stages feature extraction procedure is performed, which converts pixels of the face region on the image

into vector. Moreover, at these stages, a block for forms recognition is used. This block performs a search for and comparison of characteristics in the pre-arranged database to determine the best resemblance with the received face image of the user.

Further we consider in more detail the processes performed at each stage of image processing. Detecting the presence of a face on the image is a simple task for humans, but not for the computer [19]. Computer needs to determine the image pixels belonging to a face region and to the background. On the classic passport photo where the background image is clean, i.e. uniform and solid, it is not difficult to identify the face region. But when the background consists of several layers, where other objects are present, this problem becomes quite complex. Typically, methods for the face region detection are based on the face key point's determination such as eyes, or on the analysis of the color space of the image, as well as other face features, are used.

When the face region is determined (separated from the background image), normalization procedure should be performed. This means that the image must be standardized in terms of size, orientation, brightness and other parameters relative to the images in the database. In order to perform the normalization, the face key points must be accurately identified. The use of these key points in the normalization algorithm allows modification of images based on statistical conclusions or approximation. Recognition can be done properly only if the original image and the images in the database have the same parameters such as orientation, rotation, scale, size, etc. Only after the normalization features extraction and face recognition can be performed. At features extraction mathematical representation is generated. This representation is called biometric template or biometric reference [20], it is stored in the database and is the basis for any recognition.

The most common face recognition technologies such as PCA [20], LDA [21] and LBP [22] are applied with different success. In addition, in the Kanade and Yamada paper [23] authors describe the advantages of rigid components application, where the weigh coefficients for each of face fragments are predefined during processing of set of prepared photos. These coefficients depend on face position on the image. Methods based on holistic and rigid components of face representation are similar to the applied classification mechanism, because both of them compare image set with region "points" of feature space. In the approaches based on the use of rigid components several points are used, where each point is located in a distinct and to a considerable degree independent feature space. Until now, the advantage of application of free components to automatic face recognition with the presents of pose mismatches hasn't been fully investigated.

3 The Developed Technique for Automatic Registration of Meeting Participants

At the development of technique for automatic registration of meeting participants three cameras were implemented. So, this technique can be divided into three stages, as it is shown on Figure 2.

At the beginning of the first stage, frame from ceiling panoramic camera is received. Then in a cycle on total number of chairs located in the room, the procedure for cropping of a chair region from video stream is performed. Each chair region has predetermined size and position. After that, a histogram of color distribution on a received frame region is composed. Next, the created histogram is compared with prearranged template histogram of current chair region for calculation of the correlation coefficient. All numbers of occupied chairs, which correlation coefficient was more than threshold value, were added to list of chairs for processing at the next stage.

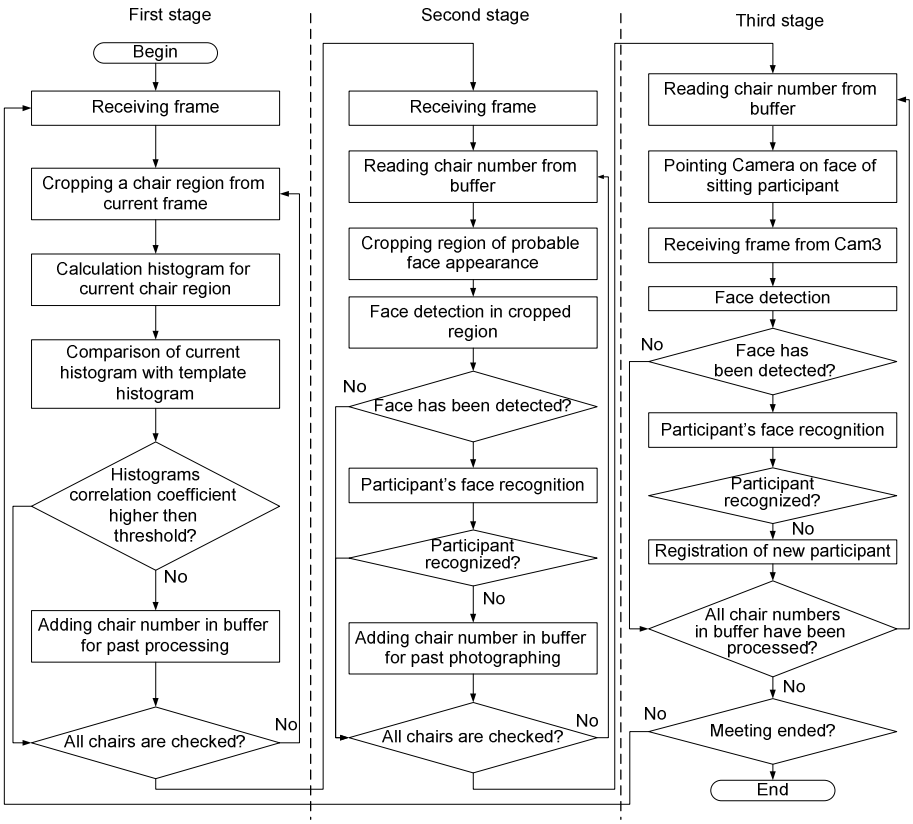

Fig. 1. Scheme of the technique for automatic identification and registration of participants in the intelligent meeting room

At the end of processing for all chairs, in the second stage a frame from the high resolution camera is received and processed by procedure for searching participant faces in areas of possible appearance of face corresponding to occupied chairs. Further, all zones in which faces have been found are processed using the face recognition method. Then a list of chair numbers is formed with unidentified participants, as well as a set of control commands, which are used in the for camera pointing on faces of the participants.

At the third stage, PTZ camera is pointing in close-up to the face of each previously unidentified participant. After checking the presence faces in the frame will be carried participant re-identification. If face hasn't been founded then number of chair with sitting participant passes into the end of the queue of unidentified participants. In case of absence a participant in the database new participant registration process is launched.

4 Experimental Results

First, we consider the dependence between the participant face size in the frame and the increasing of the distance between him/her and the stationary high resolution camera, as shown in Table 1. Distance from the camera to the near chair is 3m, and to the farthest – 9.5 m.

Table 1. Maximum participant's face size for chairs of the room

	Columns of chairs					
Chair rows	66	64	65	59	55	43
	56	54	50	43	44	38
	50	43	40	36	36	36
	38	36	36	34	33	30

Data presented in Table 1 shows that when the distance from the camera to chairs changes then the maximum detected participant face size gradually decreases, and at the farthest chair it becomes more than two times smaller.

For the experimental evaluation of a method of automatic identification of participants during the events in the intelligent room the accumulation of participant photos was produced only at the second stage of the method. As a result, the number of accumulated photos was more than 40,000 for 30 participants. In the database for training face recognition models, photos of each participant were added in order to have a difference between participant's head orientation and the direction of view from the Cam4 on every photo. As a result the created database contains 20 photos for each participant.

At the preliminary stage of experiments have been decided to determine the threshold for the three face recognition methods: 1) recognition of monolithic face representation with PCA method (PCA) [20]; 2) Linear Discriminant Analysis (LDA) [21]; 3) local binary patterns (LBP) method [22]. During this experiment threshold was calculated for each participant added to the recognition model, a maximum value of a correct recognition hypothesis for the LBPH method ranged from 60 to 100, for the PCA method from 1656 to 3576, for the LDA method from 281 to 858. As a consequence, for the further experiments were selected minimum threshold value - 60, 1656 and 281, respectively, for these methods. Table 2 presents average values of face recognition accuracy, as well as first (False Alarm (FA) rate) and second (Miss rate (MR)) errors type for each selected method.

Table 2. First experiment results of face recognition for selected methods

Method	FA, %	MR, %	Accuracy, %
LBP	10,5	10,1	79,3
PCA	0,8	25	74,1
LDA	7,6	15,3	77,1

For the training of recognition model during the second experiment some photos from the set of photos for 3 participants were replaced in such a way, that new photos of the equidistant from the camera chairs were added. Average values of the second experiment results with same parameters as in the first experiment are shown in Table 3.The presented results in Table 3 that the model trained on the photos of equidistant from the camera chairs improves the recognition accuracy and reduce the number of false positives for methods LBP and PCA. However, for the LDA results decreased significantly.

Table 3. Second experiment results of face recognition for selected methods

Method	FA, %	MR, %	Accuracy, %
LBP	10,3	10,3	79,4
PCA	0,8	24,5	74,7
LDA	9,8	15,6	74,6

In the third experiment photos for 13 participants were added to trained model in such a way as in the second experiment. The total amount of photos for each participant was equal to 20. In addition, training data base was increased to 36 participants as well as test data base includes more than 52 thousands photos. Average results values of the third experiment are shown in Table 4.

Table 4. Third experiment results of face recognition for selected methods

Method	FA, %	MR, %	Accuracy, %
LBP	12	8,5	79,5
PCA	1,3	23,5	75,2
LDA	19,2	7,8	73

The high value of false positives and miss rate errors due to the fact that the photos were stored in the course of actual of events, without a prepared scenario and focusing participants on a single object. Hereupon at the time of photographing participants can move freely, according to their face in the photos could be blurred or partially hidden. The experimental results showed that for systems of meetings process automation for face recognition PCA method should be used, this conclusion is based on the fact that this method as it is inferior in accuracy to LBPH method by 3-5%, but it has the lowest value of false alarm errors, which is an important aspect in the identification of meetings participants.

5 Conclusion

Information-control services provision based on human behavior and situation analysis is the main idea of intelligent space concept. An example of such space is intelligent meeting room, which is equipped by network of program modules, activation devices, multimedia facilities, and audiovisual sensors. Application of biometric identification technology based on face recognition methods provides automation of registration processes of meeting participants, thus reducing the work of secretaries and videographers; it also allows participants to concentrate on the discussing issues at the expense of automation control of sensory equipment.

During the research the technique for automatic registration of meeting participants was developed. It provides unobtrusive recognition and picture making of participants faces. At this stage of research for the experimental evaluation of the technique 52 thousands of photos for 36 participants were used. During experiments three face recognition methods LBP, PCA and LDA were compared. The results shows that LBP method has highest recognition accuracy (79,5%) as well as PCA method has the lowest false alarm rate (1,3%).

Acknowledgments. This work is partially supported by the Scholarship of the President of Russian Federation (Project № СП-1805.2013.5) and by the Committee on Science and Higher Education of the Administration of St. Petersburg.

References

1. Fillinger, A., Hamchi, I., Degré, S., Diduch, L., Rose, T., Fiscus, J., Stanford, V.: Middleware and Metrology for the Pervasive Future. IEEE Pervasive Computing Mobile and Ubiquitous Systems 8(3), 74–83 (2009)
2. Nakashima, H., Aghajan, H.K., Augusto, J.C.: Handbook of Ambient Intelligence and Smart Environments, 1294 p. Springer (2010)
3. Yusupov, R.M., Ronzhin, A.L., Prischepa, M., Ronzhin, A.L.: Models and Hardware-Software Solutions for Automatic Control of Intelligent Hall. Automation and Remote Control 72(7), 1389–1397 (2011)
4. Aldrich, F.: Smart Homes: Past, Present and Future. In: Harper, R. (ed.) Inside the Smart Home, pp. 17–39. Springer, London (2003)
5. Lampi, F.: Automatic Lecture Recording, 229 p. The University of Mannheim, Germany (2010)
6. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary Robust Independent Elementary Features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
7. Ekenel, H.K., Fischer, M., Jin, Q., Stiefelhagen, R.: Multi-modal Person Identification in a Smart Environment. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8 (2007)
8. Ronzhin, A., Budkov, V.: Speaker Turn Detection Based on Multimodal Situation Analysis. In: Železný, M., Habernal, I., Ronzhin, A. (eds.) SPECOM 2013. LNCS (LLNAI), vol. 8113, pp. 302–309. Springer, Heidelberg (2013)

9. Zhang, C., Yin, P., Rui, Y., Cutler, R., Viola, P., Sun, X., Pinto, N., Zhang, Z.: Boosting-Based Multimodal Speaker Detection for Distributed Meeting Videos. IEEE Transactions on Multimedia 10(8), 1541–1552 (2008)
10. Karpov, A., Markov, K., Kipyatkova, I., Vazhenina, D., Ronzhin, A.: Large vocabulary Russian speech recognition using syntactico-statistical language modeling. Speech Communication 56, 213–228 (2014)
11. Imseng, D., Friedland, G.: Tuning-Robust Initialization Methods for Speaker Diarization. IEEE Transactions on Audio, Speech, and Language Processing 18(8), 2028–2037 (2010)
12. Lobanov, B., Tsirulnik, L., Ronzhin, A., Karpov, A.: A Model of Personalized Audio-Visual TTS-synthesis for Russian. In: Proceedings of International Conference on Speech Analysis, Synthesis and Recognition, Applications in Systems for Homeland Security, SASR-2008, Poland, pp. 25–32 (2008)
13. Ronzhin, A. L.: An audiovisual system of monitoring of participants in the smart meeting room. In: Proceedings of the 9th Conference of Open Innovations Framework Program FRUCT – Russia, Petrozavodsk, pp. 127–132 (2011)
14. Ronzhin, A.L., Budkov, V.Y., Karpov, A.A.: Multichannel System of Audio-Visual Support of Remote Mobile Participant at E-Meeting. In: Balandin, S., Dunaytsev, R., Koucheryavy, Y. (eds.) ruSMART/NEW2AN 2010. LNCS, vol. 6294, pp. 62–71. Springer, Heidelberg (2010)
15. Budkov, V.Y., Ronzhin, A.L., Glazkov, S.V., Ronzhin, A.L.: Event-Driven Content Management System for Smart Meeting Room. In: Balandin, S., Koucheryavy, Y., Hu, H. (eds.) NEW2AN 2011 and ruSMART 2011. LNCS, vol. 6869, pp. 550–560. Springer, Heidelberg (2011)
16. Lin, S.-H.: An Introduction to Face Recognition Technology. Information special issue on Multimedia Informing Technologies Part-2, 3(1), 1–7 (2000)
17. Rajeshwari, J., Karibasappa, K.: Face Recognition in Video Streams on Homogeneous Distributed Systems. International Journal of Advanced Computer and Mathematical Sciences 4(1), 143–147 (2013)
18. Gorodnichy, M.D.: Video-Based Framework for Face Recognition in Video. In: Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV 2005), pp. 330–338 (2005)
19. Castrillon-Santana, M., Deniz-Suarez, O., Guerra-Artal, C., Hernandez-Tejera, M.: Real-time Detection of Faces in Video Streams. In: Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV 2005), pp. 298–305 (2005)
20. Georgescu, D.: A Real-Time Face Recognition System Using Eigenfaces. Journal of Mobile, Embedded and Distributed Systems 3(4), 193–204 (2011)
21. Belhumeur, P.N., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 19(7), 711–720 (1997)
22. Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
23. Kanade, T., Yamada, A.: Multi-subregion based probabilistic approach toward pose-invariant face recognition. In: Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp. 954–958 (2003)