Algorithms for automation of meeting participant registration and audiovisual recording in intelligent room

Al.L. Ronzhin

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

Abstract

The aim of the research is to develop algorithm for acceleration of image processing at automatic registration of meetings participants based on blurriness estimation and recognition of participants faces as well as audiovisual recording of their talks during meeting. The data captured by the video registration system in the intelligent meeting room are used for calculation variety of person face size in captured image as well as for estimation of face recognition methods. 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%).

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 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.

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, 19]; 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.

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.

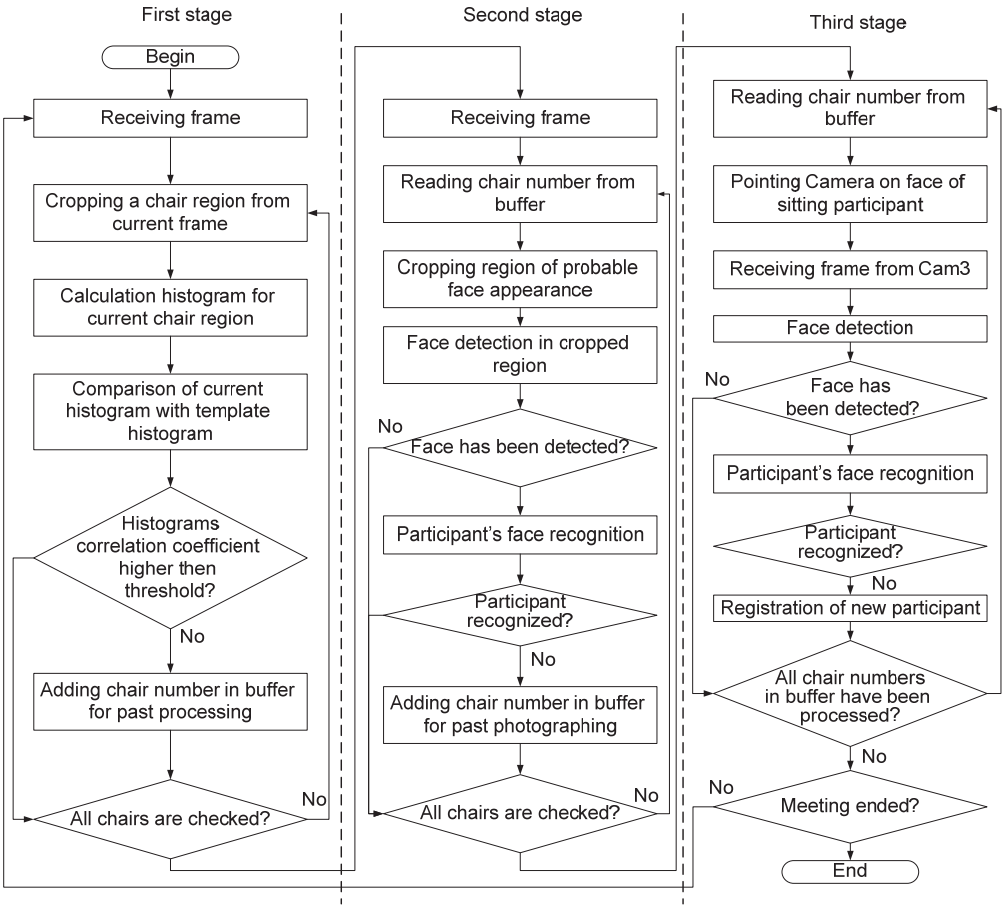

Figure 1: Scheme of the technique for automatic identification and registration of participants in the intelligent meeting room

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. Algorithm of Speaker Recording

The algorithm of camera pointing to the current active speaker in the zone of chairs and following recording of his/her speech should be considered in detail. Sound source localization and object tracking by ceiling camera are implemented here. Both modules work together during the all time of the event in the smart room.

The object detection module carries out the search and following tracking of the participants inside in the room. Also the module marks the occupied chairs, which will be used as hypotheses for speaker position. The scheme of the algorithm of the speaker detection and recording is shown in Figure 2.

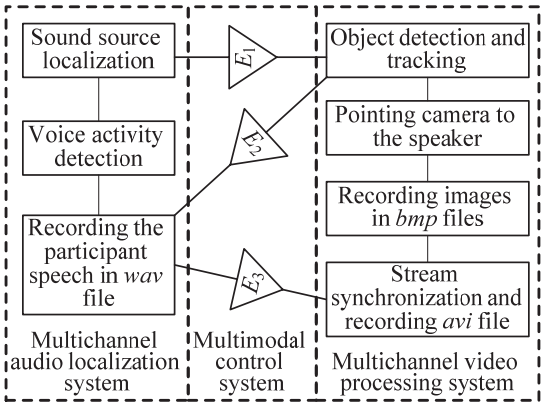

Figure 2: Scheme of the technique for automatic identification and registration of participants in the intelligent meeting room

The appearance of a sound signal in the chair zone launches the voice activity detection process and makes query (the event E_1) to the object detection module in order to check the presence of a participant in the chair, which is closest to the determined coordinates of the sound source. If the chair is marked as occupied then corresponding response is transmitted to the module of speech recording as well as the camera serving this zone is being pointed to the selected chair with the current active participant.

To avoid missing of the speech, the decision about useful sound segment is made every 20 ms. Short remarks and noise with duration less half of second are discarded in order to exclude false speaker detection. The frame rate of the camera, which captures the speaker in the chair zone, achieves thirty frames per second. However, the camera pointing takes up to couple seconds owing to a stabilisation period after mechanical setting of direction angles and zooming of the objective lens. So the recording of images to the *bmp* files is started after the camera pointing is accomplished.

At the same time, in the multichannel audio localization system the *wav* file with participant speech is recorded in case of speech boundaries detection and positive response from the object detection module (the event E_2) about presence of a participant in the selected chair. After recording *wav* file the corresponding notification (the event E_3) with path to the file transmitted to the video processing system. Then the system goes to the sound source detection and localization stage again.

5. Experimental results

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 52,000 for 36 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 1 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 1: Third experiment results

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 LBP 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.

Main attention at estimation of the algorithm of detecting and recording active participant speech was paid on detecting active participants in the zone of chairs. Each tester performed the following scenario: (1) take a sit in the room; (2) wait visual confirmation on a smart board about registration of participant in the chair; (3) pronounce the digit sequence from one to ten; (4) move to another chair.

During the experiments were recorded 36 avi files in a discussion work mode. After manual checking was detected that 89% are files with speaker's speech and 11% false files with noises. Such noises are carrying out in process of tester standing up from a chair, because in such moment chair's mechanical details carry out high noise. Also mistakes in detecting sitting participants influence on appearance of false files. Table 2 shows results of estimation files with speaker's speech.

Table 2: The estimation of algorithm of detecting and recording active participant speech work

L_{b_d}, ms			L_a, ms			N_{f_d}, frames		
Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
80	2440	724	5312	6432	5608	32	104	59

A result of experiments shows, that *avi* file in mean consists of 137 frames, 59 of it are duplicated frames, as well as has length 5 seconds. Calculated mean FPS in video buffer is 24 frames per second, this is due to the fact that rounding of values at calculating a required total amount of additional frames in image packets. The total amount of duplicated frames includes initial delay between audio and video streams. Also such total amount of duplicated frames is carry

out with changing camera FPS as a result of noises in a network devices as well as limited writing speed of storage devices. Analyses of received data shows that *avi* files form by system include all speeches and a small percent of false records.

6. 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%).

7. Acknowledgements

This work is partially supported by the Scholarship of the President of Russian Federation (Project № СП-1805.2013.5).

8. 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*. Springer. 2010.

[3] Yusupov, R.M., Ronzhin, An.L., Prischepa, M., Ronzhin Al.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 / Inside the Smart Home*. Ed. Harper R. London: Springer-Verlag, pp.17-39, 2003.

[5] Lampi, F. *Automatic Lecture Recording. Dissertation*. The University of Mannheim, Germany. 2010.

[6] Calonder, M., Lepetit, V., Fua, P. “BRIEF: Binary Robust Independent Elementary Features”. *In Proceedings of the ECCV’10*, pp. 778-792, 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 '07*, pp. 1-8, 2007.

[8] Ronzhin, A., Budkov, V. “Speaker Turn Detection Based on Multimodal Situation Analysis”. *Springer International Publishing Switzerland. In: M. Zelezny, I. Habernal, A. Ronzhin. (eds.): SPECOM 2013*, LNAI vol. 8113. Springer, Heidelberg, pp. 302–309, 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 the SASR-2008*, Poland, pp. 25-32, 2008.

[13] Ronzhin, Al.L. “An audiovisual system of monitoring of participants in the smart meeting room”, *In Proceedings of the 9th FRUCT*, pp. 127-132, 2011.

[14] Ronzhin, A.L., Budkov, V.Yu, Karpov, A.A. “Multichannel System of Audio-Visual Support of Remote Mobile Participant at E-Meeting”, *In: Balandin S, Dunaytsev R, Koucheryavy Y. (eds.) NEW2AN 2010*. LNCS, vol. 6294, Springer, Heidelberg, pp. 62-71, 2010.

[15] Budkov, V.Yu., Ronzhin, Al.L., Glazkov, S., Ronzhin, An.L. “Event-Driven Content Management System for Smart Meeting Room”, *Springer-Verlag Berlin Heidelberg, S. Balandin et al. (Eds.): NEW2AN/ruSMART 2011*, LNCS vol. 6869, Springer, Heidelberg, pp. 550–560, 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 the CRV'05*, pp. 330-338, 2005.

[19] Castrill´on-Santana, M., D´eniz-Su´arez, O., Guerra-Artal, C., Hern´andez-Tejera, M. “Real-time Detection of Faces in Video Streams”, *In Proceedings of the CRV’05*, 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., Pietikainen, M. “Face Recognition with Local Binary Patterns”, *In Proceedings of the ECCV 2004*, pp. 469–481, 2004.

[23] Kanade, T., Yamada, A. “Multi-subregion based probabilistic approach toward pose-invariant face recognition”, *In Proceedings of IEEE International Symposium on CIRA*, pp. 954–958, 2003.