ISSN 2464-4617 (print) WSCG 2016 - 24th Conference on Computer Graphics, Visualization and Computer Vision 2016
ISSN 2464-4625 (CD-ROM)

Implementation of Face Recognition Methods as a First Step for Human Behaviour Analysis in Intelligent Room

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

Irina Vatamaniuk
SPIIRAS
39, 14th line
199178, St. Petersburg, Russia
vatamaniuk@iias.spb.su

Miloš Železný
University of West Bohemia
Univerzitní 8
306 14, Pilseň, Czech Republic
zelezny@kky.zcu.cz

ABSTRACT

This paper discuss an intelligent space concepts, goals and developments. The review presents an analysis of seven intelligent meeting rooms, equipment and developed services. One of the main goals of intelligent space is the development of proactive services. Realization of such types of services requires the analysis of current situation in a room as well as behaviour of the audience. Implementation of methods for facial features analysis are the first step for this goal destination. This paper presents an algorithm for facial image normalization and recognition as well as a developed system for automatic registration and identification of meeting participants in the intelligent room. The developed algorithm is intended for processing of low resolution facial images with approximate size of 30x30 pixels. The experiments shows distribution of participants recognition rate by variety of their face sizes. In addition, implementation of blur estimation method at preliminary stage of facial image processing shows an increase of the whole system speed as well as a decrease in face recognition error rate.

Keywords

Computer vision, face recognition, image normalization, low resolution image, image quality estimation, intelligent room.

1 INTRODUCTION

The development of technologies for the automated processing of audiovisual and other types of data is required at a constant growth of volumes of recorded data. The significant amount of these data is related to photo and video recording. The automatic image processing usually includes preliminary processing (elimination of images of unsatisfactory quality, light normalization, noise removal, etc.), feature extraction, object segmentation and pattern recognition. Often it is necessary to select some images of high quality corresponding to the requirements of a particular technical task from an image database. For an operator such work is not difficult, however, if the amount of data is large, especially at the stage of preliminary processing, it is advisable to automate the process of image quality evaluating and elimination of the frames with no reliable data. In addition, implementation of blur estimation methods at preliminary stage allows increasing accuracy of pattern recognition methods.

This paper is organized as follows. Section 2 presents a concept and goals of an intelligent space. Section 3 discusses facial image normalization methods. Section 4 describes the developed algorithm for facial image normalization and recognition. Section 5 presents the experiments, conditions and results.

2 INTELLIGENT SPACE

The research work of recent years in the development of prototypes of intelligent spaces (in most cases, intelligent meeting rooms) for scientific projects has been analyzed. Let us consider the functionality of these prototypes of intelligent meeting rooms. By functionality we understand their equipment and methods for the implementation of information support services and the automation of the support of activities. The sensory equipment of the intelligent meeting room can be divided into two types: primary equipment for the recording of events in the entire space of the conference room and supporting equipment for a more detailed analysis of specific areas of interest identified by the main equipment.The list of sensory and multimedia equipment of rooms and the methods used for processing audiovisual signals are given in [Ronzhin15a]. Most intelligent meeting rooms are equipped with audio and video recording equipment as well as with advanced I/O multimedia devices. An analysis of the data [Ronzhin15a] shows that the services for recording, broadcasting, and logging of activities, as well as the creation of test audiovisual databases, are implemented in almost all considered intelligent meeting rooms. However, services that are important for the automation of the course of activities, such as automatic participant recording, current speaker detection, integration of participant devices with the intelligent meeting room's system, and automatic generation of multimedia report, were implemented only in half of these intelligent meeting rooms.

3 METHODS FOR IMAGE NORMALIZATION

The paper [Xie08] presents a framework for solving the illumination problem in face recognition tasks. Features extracted from face images may be divided into two categories small- and large-scale features also may be called as small and large intrinsic details respectively. The first aim in [Xie08] was decomposition of face image into two features images by implementation of logarithm total variation method (LTV) with interpoint second-order cone program (SCCP) algorithm for approximate estimation of LTV model [Chen06]. The second aim was processing of decomposed face image elements for smoothing of small-scale and illumination normalization of large-scale features. For illumination normalization in work [Xie08] methods non-point light quotient image relighting (NPL-QI) and truncating discrete cosine transform coefficients in logarithm domain (LOG-DCT) were separately employed and estimated. All processed features are used for reconstruction of normalized face image. It shows [Xie11] that implementation of NPL-QI allows one to extend the illumination estimation from single light source to any type illumination conditions and simulate face images under arbitrary that conditions. The LOG-DCT is not so likely for image normalization because it loss some low-frequency information during image restoration.

The paper [Liu08] presents a method of illumination transition of an image (ITI) for illumination estimation and image re-rendering. This method is based on the calculation of face similarity factor, which is considered: the more similarity comparison of an image with more weighted pre-referenced personal face image allows generating a person-specific reference face $T_{x,z}$:

$$T_{x,z} = \frac{\Sigma_i k_i A_i z + \varepsilon}{\Sigma_i k_i A_i x + \varepsilon}$$

where $k = (k_1, k_2, \ldots, k_N^T$ is the weight coefficient, A_i is an images array of subject i, x and z are current and expected illumination parameters respectively, ε is a small constant ($\varepsilon = 1$), which is used in [Liu08] in case of being divided by zero. The experimental results [Liu08] show that current method may be used for illumination normalization of facial images as preprocessing in face recognition methods.

In continuation of works [Xie08, Xie11, Liu08], the paper [Xie14] presents the implementation of empirical mode decomposition (EMD) [Huang98] for illumination preprocessing of facial images. This method can adaptively decompose a complex signal into intrinsic mode functions (IMFs), which are relevant to intrinsic physical significances. A face image may be represented as Lambertian reflectance assumption [Xie08]:

$$I(x,y) = R(x,y)L(x,y),$$

where $I(x,y)$ is the observed intensity, $R(x,y)$ is the reflectance (albedo), and $L(x,y)$ is the shading or illumination component. The right side of this formula is multiplicative and the EMD decomposes signal into IMFs in an additive form. So, factorization in the logarithm domain of EMD (LEMD) was used in [Liu08] for conversion of multiplicative model into additive:

$$f = \log I = \log R + \log L.$$

Because the EMD can decompose a signal into a set of IMFs with different frequencies, the R and L components may be estimated as:

$$R = \exp\left(\sum_{k=1}^{K_0} d_k\right), L = \exp\left(\sum_{(} k = K_0)^K d_k + r\right),$$

where d_k are the IMFs with different frequencies k, r is residue, and K_0 is a settable parameter, in [Xie14] $K_0 = 2$.

The complete algorithm for illumination normalization in facial images, presented in [Xie14], is:

1. Computation of the logarithm of image;
2. Perform the 1-D EMD of gathered logarithm (experiment result in [Xie14] shows that 1-D EMD method have better recognition performances);
3. Detection of shadow regions. In this stage each IMFs have been analyzed. There is shadow in a IMFs if $D \cdot e^{m_{IMF}^2} > \theta_1$ is satisfied the binarization operator with threshold $\theta_2 = m_{IMF}/2$ is applied to analyzed IMF or residue, where D and m_{IMF} are variance and mean pixel values recently as well as the threshold $\theta_1 = 0.1$ [Xie14], and then each connected black area is marked as a shadow region;
4. Grayscale adjustment. Gray level Substitution of each detected region by average gray level (with the use of Gaussian weighting) of its neighboring region.
5. Restoration of the image using processed IMFs and residue and conversion of it to original image space.

For experiments only frontal-face images from same databases as in work [Xie11] were carried out.

4 ALGORITHM FOR FACIAL IMAGE NORMALIZATION AND RECOGNITION

The proposed algorithm of participant identification based on group photo processing is a part of an intelligent meeting room software, which is described in detail in [Ronzhin15a]. First of all, what is the group photo? The classic or porters photo is an image with simple background and several peoples, who stay or sit one by one on the foreground and look at the camera point. But the group photo, which was captured during a meeting, has more complexity, including: 1) complex background; 2) occlusions of participants by each other; 3) illumination and shadow variety; 4) blurred participant's faces (by their own movements at the shooting moment); 5) small size of participant's face region on an image.

In this stage of research the algorithm (Fig. 1) will decide first three and fifth problem. The image quality estimation methods would not be implemented in the algorithm, because state of the art methods cannot make a decision about blurriness of an image with a low resolution (approximately 30x30 pixels), details of such problem are presented in [Ronzhin15b]. The high resolution image (1280x1024 pixels), which is a group photo of participants sitting in chairs, is processed for finding their faces by face detection procedure [Viola03]. Then each face region is processed for estimation of facial reference points such as eyes, nose and mouth. This information is used for images dividing into two categories – frontal-face and other (with angle more than 15 degree from camera point) by the head orientation estimation procedure. For frontal-face images the corrected region was extracted as rectangle and for other images as ellipsoid. These procedures are used for decision of complex background and occlusion problems by reduction of background patches in a facial region.

After extraction of the final participant facial region, information about distance to the participant and illumination in the room was analyzed. Such information was gathered from the audiovisual monitoring system of the intelligent meeting room [Ronzhin15a], description of the room can be found in [Yusupov11].

For image decomposition in the proposed algorithm, the LTV model and LEMD method based on the result in [Xie11, Xie14] will be used separately. In addition, after decomposition by LTV model, the threshold-average filtering will be used for smoothing the small-scale features. As presented in [Xie14], the LEMD decompose facial image into several IMFs, and the illumination-insensitive facial representation is stored in first IMFs. So, in this case, for large-scale features normalization will be IMFs from second to residue. In LTV model application for this aim NPL-QI and LOG-DCT algorithms would be separately used for normalization of shadow and illumination as well as for reconstruction of normalized image. The final step of the algorithm is the identification of the participant by recognition of his face, which is based on Local Binary Patterns method (LBP) [Ahonen06].

5 EXPERIMENTS

For the experimental evaluation of the automatic participants registration system during the events in the intelligent meeting room, participants photos were accumulated only at the first algorithm of the system. As a result, the number of accumulated photos was more than 55000 for 36 participants. The training database contains 20 photos for each participant. For estimation of the influence of participants face size change on recognition rate, it was decided to divide them into several groups. Figure (Fig. 2a) shows distribution of participants by variety difference of their face sizes in ranges from 0 to 10, from 10 to 20, and so on. Most of participants have a difference between minimum and maximum face sizes in range from 30 to 40 pixels. (Fig. 2b) shows a distribution of recognition rate for three methods for the groups of participants. From (Fig. 2b) it is obvious that with increasing of the participant's face size variety difference, the recognition accuracy gradually decreases. This is due to the fact that at the normalization of images to a uniform size, a distortion in certain facial features like eyes, nose, mouth may occur.

6 CONCLUSION

Personalized user service and satisfaction of user needs in an unobtrusive and almost invisible form by analyzing their behaviour and determining the current situation is the main idea in the concept of an ambient intelligent space. The awareness of the conference room of the spatial position of participants, their current activities, their role in the activity, and their preferences provides intelligent control over multimedia and other equipment in the room. In order to determine the current behaviour of participants and analyze the situation, modern interconnected methods and a software for audiovisual processing of data that automatically control audio and video recording equipment are required. Considering the experimental conditions (different distances from the camera to the participant, lighting, movement of participants while taking pictures) that influence the quality and quantity of the facial features extracted from the image, which are directly influenced by the accuracy of recognition and occurrence of false positives, we can conclude that the best results are shown by the method LBP 79,5%.

7 ACKNOWLEDGMENTS

This research is financially supported by the Ministry of Education and Science of the Russian Federation, agreement no. 14.616.21.0056 (reference

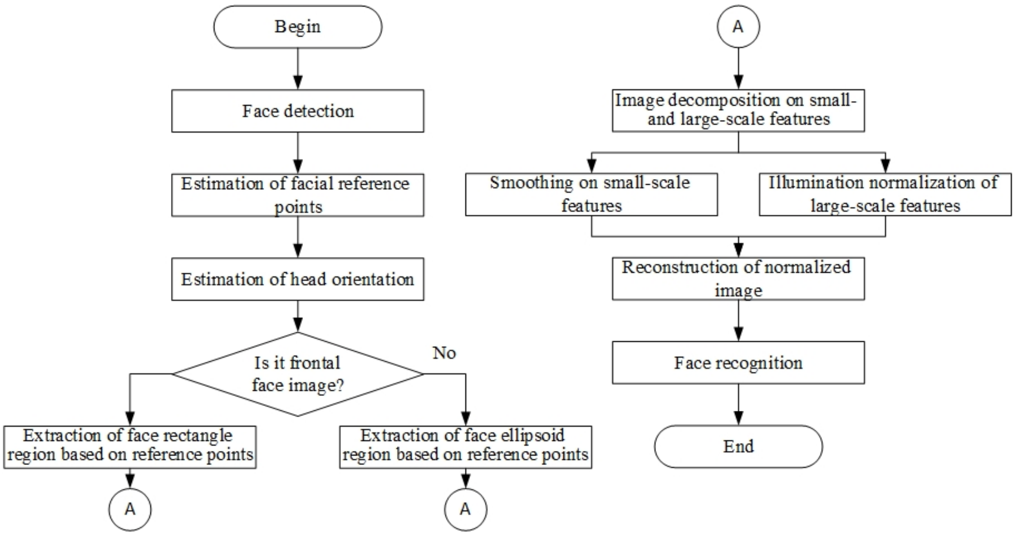
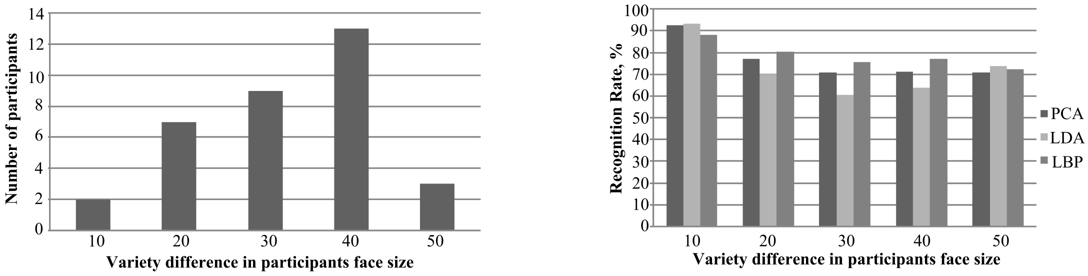


Figure 1: Scheme of the participant identification algorithm



a b

Figure 2: Distribution of participants recognition rate by variety of their face sizes

RFMEFI61615X0056), as well as by the Czech Ministry of Education, Youth and Sports, project no. LO1506.

8 REFERENCES

[Ahonen06] Ahonen, T., Hadid, A., Pietikainen, M. Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006. Vol. 28. No. 12. pp. 2037–2041.

[Chen06] Chen, T. et al. Total variation models for variable lighting face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006. Vol. 28. No. 9. pp. 1519–1524.

[Huang98] Huang, N. E. et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society, 1998. Vol. 454. No. 1971. pp. 903–995.

[Liu08] Liu J. et al. Illumination transition image: parameter-based illumination estimation and re-rendering. In Proceedings of the 19th ICPR. 2008. pp. 1–4.

[Xie08] Xie, X. et al. Face illumination normalization on large and small scale features. IEEE Conference on Computer Vision and Pattern Recognition CVPR. 2008. pp. 1–8.

[Ronzhin15a] Ronzhin, A., Karpov, A. A software system for the audiovisual monitoring of an intelligent meeting room in support of scientific and education activities. Pattern Recognition and Image Analysis. 2015. Vol. 25. No. 2. pp. 237–254.

[Ronzhin15b] Ronzhin, An., Vatamanyuk, I., Ronzhin, Al., and Zelezny, M. Mathematical Methods to Estimate Image Blur and Recognize Faces in the System of Automatic Conference Participant Registration. Automation and Remote Control, 2015, Vol. 76, No. 11, pp. 1623–1632.

[Viola03] Viola, P., Jones, M., Snow, D. Detecting pedestrians using patterns of motion and appearance. In Proc. of IEEE ICCV. 2003. pp. 734–741.

[Xie11] Xie, X. et al. Normalization of face illumination based on large-and small-scale features. IEEE Transactions on Image Processing. 2011. Vol. 20. No. 7. pp. 1807–1821.

[Xie14] Xie, X. Illumination preprocessing for face images based on empirical mode decomposition. Signal Processing. 2014. Vol. 103. pp. 250–257.

[Yusupov11] Yusupov, R. M. et al. Models and hardware-software solutions for automatic control of intelligent hall. Automation and Remote Control. 2011. Vol. 72. No. 7. pp. 1389–1397.