Methods of Automatic Face Angle Recognition for Life Support and Safety Systems

S.G. Nebaba[a], A.A. Zakharova[b]* ,T.V. Sidorenko[c], V.R. Viitman[d]

* Corresponding author: A.A. Zakharova, zaa@tpu.ru

[a]National Research Tomsk Polytechnic University, 30 Av. Lenin, Tomsk, Russia, stepanlfx@tpu.ru
[b]National Research Tomsk Polytechnic University, 30 Av. Lenin, Tomsk, Russia, zaa@tpu.ru
[c]National Research Tomsk Polytechnic University, 84/3 Str. Sovetskaya, Tomsk, Russia, sidorenkot@tpu.ru
[d]National Research Tomsk Polytechnic University, 84/3 Str. Sovetskaya, Tomsk, Russia, vvr@tpu.ru

Abstract

http://dx.doi.org/10.15405/epsbs.2017.01.97

The vision of the surrounding and people that are within eyeshot influences the human well-being and safety. The rationale of system development that allows recognizing faces from difficult perspectives online and informing timely about approaching people is undisputed.
The manuscript describes the methods of automatic detection of equilibrium face points in the bitmap image and methods of forming 3D face model. The optimal search algorithm for equilibrium points has been chosen. The method of forming 3D face model basing on a single bitmap image and building up the face image rotated to the preset angle has been proposed. The algorithm for estimating the angle and algorithm of the face image rotation have been implemented. The manuscript also reviews the existing methods of forming 3D face model. The algorithm for the formation of 3D face model from a single bitmap image and a set of individual 3D models have been proposed as well as the algorithm for forming different face angles with the calculated 3D face model aimed to create biometric vectors cluster. Operation results of the algorithm for face images formation from different angles have been presented.

© 2017 Published by Future Academy www.FutureAcademy.org.uk

Keywords: Computer vision; face recognition; face position estimation; active shape models; 3D face model; deformable face model.

1. Introduction

Methods of object recognition are applied to implement software and hardware-software systems that provide the human well-being and safety. One of the most complicated objects is a human face as

a perfect identification is important to inform about approaching people and to take subsequent decisions.

The task of personal identification based on analysis of the face images is being addressed since the early stages of the computer vision development (Adini, Moses & Ullman, 1997; Bui, Phan & Spitsyn, 2012). Recently the demand for prompt and correct identification of a person in the video stream has been increasing.

To date the task of the person search based on the image has been successfully solved and used in many technical devices. For example, the person search is used by the photo equipment for auto focus select.

The task of the facial recognition (identification) is more complex, and at the moment there is no algorithm that would be close to the facial recognition rate by the man.

In recent years several different approaches to processing, localization and object recognition have been proposed. However, these approaches have insufficient accuracy, reliability and speed in the real context, which is characterized by the presence of noise in the video sequences and variety of shooting conditions. Methods used to solve the problem of face recognition should provide the acceptable recognition accuracy and high speed of video sequences processing.

Identification efficiency can be estimated on the basis of two probabilistic characteristics:

- probability of false positives (false identification of a person – a mistaken identity);
- probability of false negatives (omission of a looked-for person – misrecognition).

These probabilities of the identification systems are interrelated variables in the inverse proportion. In addition, all current systems of personal identification have both types of errors being clearly dependent on such factors as shooting angle of the person to be identified, illumination conditions and face image quality both in the database and those recorded with video surveillance cameras.

The stringent requirements to the shooting angle of the face to be identified are caused by the used algorithms, the best of which do not allow obtaining the resultant satisfactory recognition even at slight turning angle of the facial frontal view towards the axis of the recording camera. Generally, the turning angle is recommended not to exceed 15 degrees.

One of possible solutions of the various rotation angles problem is the evaluation and normalization of the face position in the image. However, there is one more approach to improve recognition efficiency. Recognition method based on measuring the distance between the clusters of biometric vectors requires that each of the faces in the database must be presented with a set of several different images on which the elements of the cluster are computed. This recognition task can be set for one face image (a passport photo, etc.), which makes comparison of biometric vectors the inefficient and unreliable method. However, a variety of face images can be produced with using the rotation of the source on the assumption that the face model is known and corresponds to the image. After such conversion there is a set of face images with different positions which is suitable for formation of biometric vectors cluster.

Due to the above the task of face model calculation basing on its image and generation images of different face angles on the basis of this model becomes relevant.

2. Position estimation methods and methods for automatic placement of equilibrium face points to the image

Monitoring over human face position against the optical axis of the camera is an important issue due to the angle-sensitivity of the recognition algorithms. The best-known method of the angle estimation is POSIT (Pose from Orthography and Scaling with Iterations), offering high performance and rapid convergence (DeMenthon & Davis, 1995).

This algorithm is difficult to be applied into practice since it requires the image with the marked characteristic points of the face, which is generally not done automatically. The existing methods of automatic positioning of the characteristic points are fairly faulty and at small rotation angles it is impossible to evaluate the face angle with the adequate accuracy.

If face characteristic points are properly positioned, the angle estimation algorithm allows obtaining the coefficients to set the angle correction of the face images and calculating the normalized image.

There are several methods for automatic placement of equilibrium face points, all of them are based on Active Shape Models (ASM) developed by Tim Cootes and Chris Taylor in 1995 (Cootes, Taylor, Cooper & Graham, 1995).

This method is widely used to analyze face images, mechanical units and medical images (2D and 3D).

The authors have investigated various modifications of the original ASM algorithm (Cootes, Taylor, Cooper & Graham, 1995; Milborrow, 2014; Xiong, 2013; King, 2007), and have selected the modification which is more stable in case of larger rotation angles.

In general it can be assumed that the accuracy and reliability of points' alignment methods depends largely on the method of constructing the training sampling. Fig. 1 shows the comparison of marking equilibrium face points with large angle of rotation relative to the camera lens for several different ASM versions.

Fig. 1. a) Marking performed with ASM-modification that trained mainly on the frontal faces; b), c) markings made with optimally trained ASM-modifications

Test of POSIT algorithm (8-point model, eye corners + mouth) in combination with the automatic alignment of equilibrium face points algorithm showed that the rotation angle estimation error did not exceed 5 degrees on each axis, which was a good result for those given conditions.

3. Methods of building up 3D face model

There are different algorithms for the task of face position estimation and automatic placement of equilibrium face points. At the same time, the task of angle correction has never been considered as a method of improving the performance of recognition algorithms with bitmap face image. This is primarily due to the problem of determination of the 3D model of the analyzed face.

The issue of obtaining the third missing coordinate can be addressed from different positions:

1. The calculation of the image depth from the illumination specific features. This method does not guarantee the accuracy and uniqueness of interpretation due to the inability to take into account all the possible illumination features in the frame and reflective properties of the surface. However, there are studies on this issue (Zimmermann, 2007).

2. Calculation of the image depth from the character the object motion in the frame sequence. This method is suitable only for analysis of images in the video stream, it has specific demands to the quality of face images, and its accuracy depends on the illumination conditions and properties of the face detector. This method is used in a number of studies, besides its variation with shooting a stationary object is used in 3D-modeling (Agarwal, 2009).

3. Using the previously prepared 3D face model that is aligned with equilibrium points on the image. This method guarantees the accuracy of the depth estimation, but requires 3D model for the each face to be processed, as well as the initial fit of the model and image. Moreover, obtaining of the face model for each person requires special equipment. This method is used mainly in the complex recognition systems with multiple cameras, computing 3D model "on the fly", at the moment of the face detection (Shahriar, 2007).

Due to the above one can conclude that all mentioned methods for building up 3D model have certain shortcomings and are difficult to be implemented. Thus a new method for the face model obtaining needs to be developed.

As a basis the method of using the ready-made 3D face model has been chosen, but to simplify the system the allowance has been made that it is possible to use some unified model as a depth model for the face image, assuming a similar relief of the main features, such as eyes, eyebrows, nose and mouth. This allowance will reduce the accuracy of the model, but avoid its calculation for each person.

4. Method of face position correction

The work (Hassner, Harel, Paz & Enbar, 2015) shows the model that is an array of depth values 181x122 and it is chosen as a unified 3D face model.

The following operations are necessary to solve the problem of face position correction:

- Combining the unified depth model with equilibrium face points obtained by one of the algorithms described above;
- Splitting of the resulting face model into triangles (triangulation). This operation will allow obtaining a set of triangles describing the face area in the image. It can be used to build a normalized face image, but the result will be fairly rough (Fig. 2).
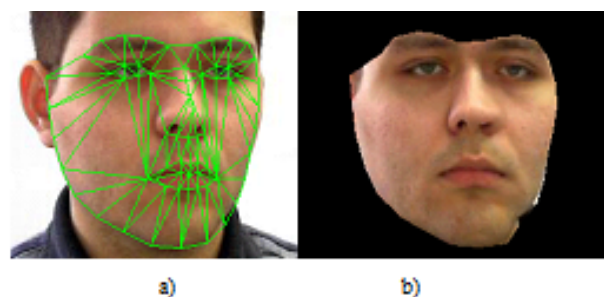
Fig. 2. a) The grid imposed on the original image; b) a normalized image

- Fragmentation of the triangles on the selected criteria that allows reducing the size of the used triangles and approximate used triangular grid to the uniform one. As a subdivision criterion one can take the maximum area of the triangle or the maximum length of the triangle side;

- Each of the triangles is shifted and deformed with rotating operations relative to a given center and affine transformation in accordance with the angle value determined by POSIT algorithm.

As a result of these operations the face image normalized by rotation angles is obtained.

5. Possible solutions of angle correction task

The task of the angle correction may be considered in various ways, and in accordance to the task based on 3D face model it is possible to recover the frontal position of the rotated face in the image as well as to create a set of facial images rotated to different angles relative to the optical axis of the camera (Fig. 3).
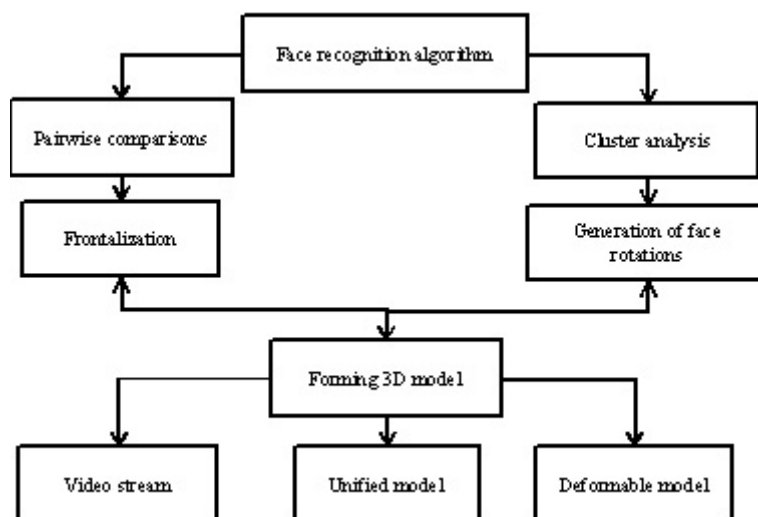
Fig. 3. Possible solutions of angles synchronization task for facial recognition

The majority of face recognition systems for 2D image are based on the principle of clustering of different images of the same face and subsequent comparison by affinity of the biometric vector of the face image obtained from the video stream to one of the clusters (cluster analysis). There is also a variant of face recognition with pairwise comparison of the standard image and images obtained from

video, but the accuracy of this approach is significantly lower, since it is impossible to eliminate completely the influence of light, position and shooting quality, so error probability increases.

Depending on the choice of the face recognition method for the angle correction one can apply the approach of bringing all processing faces to the front view (frontalization) as well as the approach that uses the generation of multiple face images at different angles relative to the optical axis of the camera to form the cluster corresponding to a specific person (generation of face rotations).

The best of these options for face recognition in the video stream is the cluster analysis, as it has such critical advantages as low computational efforts and high recognition accuracy in various shooting conditions.

Regarding the cluster analysis the method of face rotation generation can be used to obtain a set of face images with a single image to build the cluster not degrading the recognition quality.

There are three basic approaches to solve the problem of forming 3D model in the situation where only a set of bitmap images of the human face is available. The first approach is generation of 3D model from the information received through changing the position of the human head in the video ("video stream" in Fig. 3). The second approach uses the unified three-dimensional depth map for the human face with specific equilibrium face points in the image (unified model). The third approach modifies the method of building a unified model with the depth map generated by implicit rules that obtained by the machine learning of the system based on face images, each of which has the corresponding 3D model (a deformable model).

Each of the methods discussed above has advantages and disadvantages as regards parameters as speed, accuracy, dependency on the shooting conditions. Below these properties are considered for each method.

1. Video stream:

+ Individual face model;

- High computational complexity;

- Low accuracy;

- Strong dependence on recording resolution.

2. Unified model:

+ The minimum calculation at the stage of model building;

+ Minimum dependence on recording conditions;

- Moderate accuracy.

3. Deformable model:

+ Moderate computational complexity;

+ Minimum dependence on recording conditions;

+ High accuracy;

- Complex model formation.

Among methods of forming 3D model the method of the video stream is the least suitable for the method of cluster analysis, as in non-ideal conditions of shooting it does not have any advantage, but demands a lot of computational resources. The unified face model is more applicable because of easy

implementation and minimal resource requirements at all stages of the model building. The deformable model has the advantage of the higher model accuracy, but this approach is difficult to implement.

6. Algorithm for building up individual face models

A set of 11 3D face models in the public domain (Paysan, 2009) was taken as a basis for the formation of the deformable face model. Each 3D model contains a cloud of points in three-dimensional space as well as texture information to obtain the original face image in the frontal position and accordingly to find equilibrium points.

If the number of available 3D face models is enough we can take the nearest reference model as required (distance between model equilibrium points and equilibrium points of the analyzed face is a minimal on the average), but the more accurate solution is the interpolation of each point of 3D model by the affinity to the reference models. This method provides a unique model of the analyzed face that is more accurate than any of the reference models.

Due to the above, the following algorithm for the formation of individual facial model has been proposed:

1. Search for the relevance between the found equilibrium points and known 3D models.
2. Normalization of equilibrium points relative to models scale (for example, by the distance between the eye centers) and the center of coordinates (for example, the nose tip).
3. The calculation of distances for each equilibrium point of the face image and the corresponding equilibrium point of each 3D model.
4. Calculation of the depth coordinates for equilibrium points of the face image by interpolating the values of equilibrium points of the models. The distance between the corresponding equilibrium points of the known model and a preset face image indicates the affinity of 3D models to human face in the image.
5. Determination of the depth coordinate for all other points of the desired model with gradual calculation of new points between known equilibrium points until the number of points in the model will reach the number of points in reference models.

All the above operations will allow obtaining the individual 3D face model, which can be used for the algorithm of building up the image of the face rotated by preset angle.

7. Algorithm testing

The obtained 3D model was used to form face images with different angles for testing the approach validity as per following algorithm:

1. Combining the obtained face model with face equilibrium points received through one of ASM method modifications.
2. Splitting the obtained face model into triangles as per equilibrium points (triangulation). This operation will allow obtaining a set of triangles describing the area of the face in the image. Now it can be used for the normalized image building, but the result will be fairly rough everywhere except equilibrium points.

3. Fragmentation of the triangles with the selected criterion aimed to reduce the size of the used triangles and approximate the built irregular triangular mesh to the uniform one. The fragmentation criterion could be the maximum area of the triangle or the maximum length of the side. The criterion as per the maximum length provides more proportional mesh.

4. Each of the triangles is deformed and replaced in accordance with the predetermined rotation angle by rotating operations relative to the given center and affine transformation.

These operations will result in obtaining the face image rotated to a preset angle relative to the optical axis of the camera.

The set of face images rotated by the angle from -20 to +20 degrees relative to the optical axis of the camera was obtained from one front image using this algorithm (Fig. 4).

Fig. 4. a) The initial frontal facial image; b) the face images obtained by rotating along the horizontal axis by 20 degrees left and right

These images used for the formation of the biometrical vectors cluster in the person identification system allow using only one photo of the person photo not compromising the recognition efficiency.

The above methods of the image processing were combined in different ways. The first testing was carried out with the use of the video database. The second testing was carried out on the Caltech Faces database (Paysan, 2009). The obtained feature vectors were compared using the Euclidean metric. EER characteristic was selected to evaluate the effectiveness of feature vector calculation algorithms. EER is an equal FAR and FRR error rate, where FAR is the probability of false detection, that is the probability that the system mistakenly recognizes the identity of the user which is not registered in the system, and the FRR is the probability of goal missing, that is the probability that the system does not recognize the identity of the user registered therein. The lower EER is, the more efficient the algorithm is considered (Kukharev, Kamenskaya, Matveev & Schegoleva, 2013).

Table 1 shows the results of testing recognition system with the database of 12 front face images, database of 15 images of each face, and database of images created using deformable face models.

Table 1. Testing results of face recognition system with various databases

Image database	EER, %
Single face image for one human	8.06867
15 face images for one human	4.81986
15 face images generated by the algorithm	2.17865

Basing the results shown in Table 1 we can conclude that the method of generating face images with different angles does not reduce the efficiency of person recognition and in some cases improves the recognition result.

8. Conclusion

The article analyzes the existing algorithms of auto-search for equilibrium face points on the bitmap image and methods of forming 3D face model. The optimal algorithm for automatic equilibrium points search has been chosen. The method combining unified 3D face model with equilibrium face points has been proposed to be implemented and the algorithm of bringing the face image to specified angle has been designed. The manuscript presents the implementation results of the algorithm estimating the angle and rotation of the face image.

The modified method for forming 3D face model from a single bitmap image and a set of individual 3D models have been proposed as well as the algorithm of the different face positions image forming with the help of the formed 3D face model for building the cluster biometric vectors. The algorithm of 3D model forming and algorithm for generating different positions face images have been described, the results of the image generation have been presented. The images generated by using the described algorithms do not reduce the quality of recognition in a system based on the method of comparing the distances between biometric vectors clusters.

The proposed method of evaluation and correction of the face position can be used in systems of the person recognition by bitmap face image, in particular for compensation of inappropriate shooting angle of photos that form the face database.

Designed face image generation algorithm allows expanding the clusters and thus increasing the recognition accuracy by adding artificially created face images to the existing database. It is offered be applied in recognition systems using nearest neighbor method as a recognition criterion and set of biometric vectors clusters derived from bitmap face images. The authors suggest using the presented algorithms in well-being and safety support systems to recognize faces that allow handicapped people or people with restricted perception, for instance, with vision deficiency to feel at ease and to be informed timely about their social surrounding.

References

Adini, Y., Moses, Y., & Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 721-732.

Agarwal, S. (2011). Building Rome in a Day. *Communications of the ACM*, 54, 105-112.

Bui, T. T. T., Phan N. H., & Spitsyn, V. G. (2012). Face and hand gesture recognition algorithm based on wavelet transforms and principal component analysis. *Proceedings of 7th International Forum on Strategic Technology*, 1-4. doi: 10.1109/IFOST.2012.6357626

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding*, 61, 38-59.

DeMenthon, D. F., & Davis, L. S. (1995). Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, *15*, 123-141.

Hassner, T., Harel, S., Paz, E., & Enbar R. (2015). Effective face frontalization in unconstrained images. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4295-4304.

King, D. E. (2007). Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, *10*, 1755-1758.

Kukharev, G. A., Kamenskaya, E. I., Matveev, Yu. N., & Schegoleva, N. L. (2013). *Metody obrabotki i raspoznavaniya izobrazheniy lits v zadachakh biometrii*. [Face recognition and processing techniques as biometric tasks]. Ed. By M. V. Khitrov, St. Petersburg, Polutekhnika Publ., 242-251.

Milborrow, S. (2014). Active Shape Models with SIFT Descriptors and MARS. *The 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 119-133.

Paysan, P. (2009). A 3D Face Model for Pose and Illumination Invariant Face Recognition. *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 296-301.

Shahriar, H. M. (2007). Inexpensive Construction of a 3D Face Model from Stereo Images. *10th international conference on Computer and information technology (ICCIT)*, 1-6.

Xiong, X. (2013). Supervised Descent Method and its Applications to Face Alignment. *The IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 532-539.

Zimmermann, J. (2007). SilSketch: Automated Sketch-Based Editing of Surface Meshes. *EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling*, 23-30.