Algorithm for Real Time Faces Detection in 3D Space.

Rauf Kh. Sadykhov
Belarusian State University of Informatics and Radioelectronics,
220125 Brovki str. 6 Minsk, Belarus
Email: rsadykhov@bsuir.by

Denis V. Lamovsky
Belarusian State University of Informatics and Radioelectronics,
220125 Brovki str. 6 Minsk, Belarus
Email: lamovsky@bsuir.by

Abstract—**This paper presents the algorithm of stereo-faces detection in video sequences. A Stereo-face is a face of a man presented by set of images obtained from different points of view. Such data can be used for faces structure estimation. Our algorithm is based on computationally effective method of face detection in mono image. Information about face positions is then combined using sparse stereo matching algorithm. Sparse means that stereo correspondence is estimated not for all scene points but only for points of interest. This allows obtaining the low computation cost of algorithm. We use few criteria to estimate correspondence. These are: epipolar constraint, size correspondence, 3D region of interest constraint and histogram correspondence.**

Object distance estimation method that does not use projective transformations of stereo planes is also considered.

I. Introduction

3D FACE modeling is widely used nowadays in data processing systems. These models are applied in such fields as man-computer interface, multimedia applications, biometric identification in surveillance and access control systems. Three-dimensional face recognition became more popular in recent time [1], [2], [3]. Its methods promise to achieve better accuracy in comparison with the more traditional 2D face recognition methods. It is rapidly developed area in computer vision.

Special equipment is usually used to obtain 3D models. It can consist of several video cameras, structured light source, special laser scanners and other. When accuracy of the model is not critical 3D models can be obtained from static images or frames of video sequence [4]. Building such coarse models sometimes requires user assistance [5]. Mentioned approaches allow estimating three-dimensional face structure with appropriate quality but have disadvantages. They can't be used when it is necessary to obtain such models in real time, in uncontrolled conditions and when user assistance is unavailable.

Person identification systems which use 3D face models as biometric feature require high speed of the model estimation and minimum requirements to the object position. Existing systems can not provide such requirements. A lot of cameras of surveillance systems are used in present time. Such equipment can provide obtaining frames with high resolution. They can be used for 3D face estimation and can be done in several ways. The first approach is based on mono view image processing, the second are multi view (stereo) images processing. In the first case frontal face position is strongly required for acceptable model accuracy. Such models can not accurately represent real parameters of the face. A more accurate face model can be obtained using images from different points of view. Stereo reconstruction methods can be used in this case for additional information estimation. Stereo base face detection methods [6] require additional efforts for depth estimation. It also usually need that cameras have parallel optical axes. In other case it will be necessary to perform transformation of the images to a standard parallel stereo view [7].

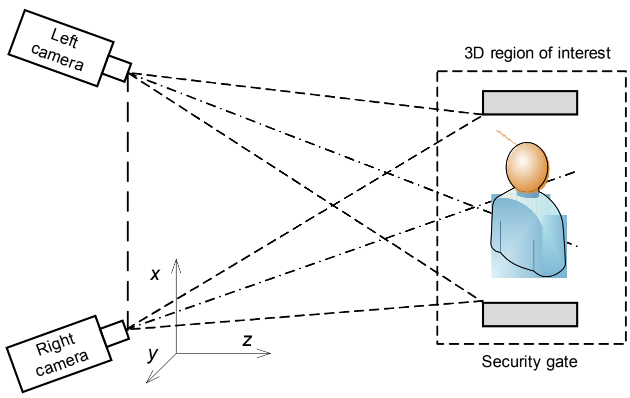

Fig. 1. 3D face processing in security applications

Object detection is usually one of the stages of image processing. We speak about face detection in the case of person identification by portrait. In this paper we present the algorithm of face detection and matching in frames of stereo images sequence. Our goal is to detect and localize stereo-faces of people moving in the field of view of several video cameras (see fig. 1). Stereo-face here is a set of face images obtained from different views. Main requirements are the next: number of faces from 1 to 5; face size not less than 100*100 pixels; and frame size is 1024*768 pixels.

II. Camera Calibration

Camera calibration in stereo vision classically means computation of relative positions of the camera optical centers. It is not necessary for current task and it is acceptable to compute fundamental matrix that represents epipolar geometry of

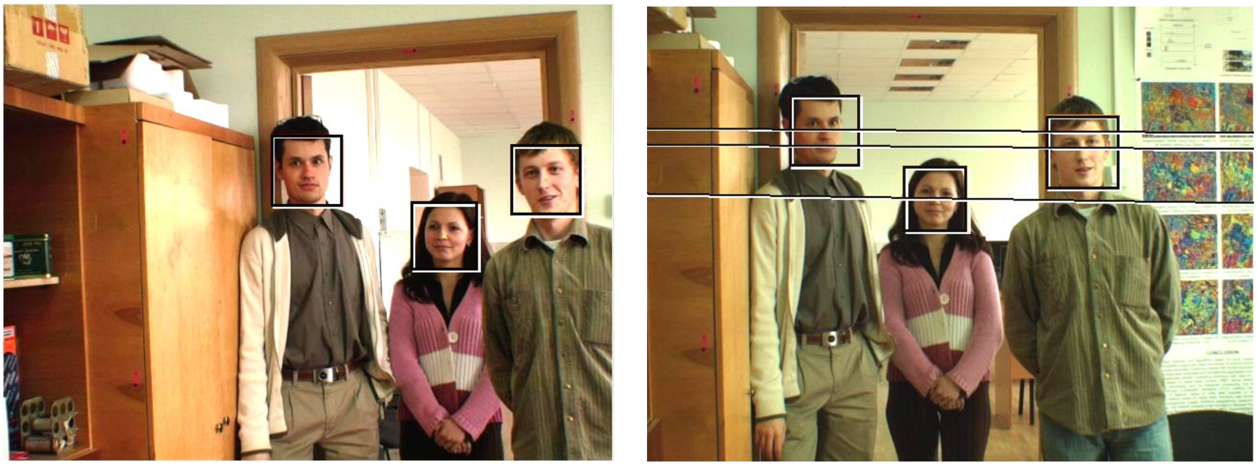

Fig. 2. Stereo frame with detected faces and corresponding epipolar lines.

camera views. The fundamental matrix allows calculating so called epipolar line in one frame that corresponds to given point in another frame. Epipolar geometry greatly reduces search space for points correspondence estimation. We used RANSAC method [8] that is based on eight points algorithm [9] for fundamental matrix estimation.

User has to mark at least eight pairs of correspondent points for calibration. Obtained matrix can be applied for epipolar line estimation by using next relation:

$$l = Fx, \tag{1}$$

where x – given point, F – fundamental matrix and l – sought value.

III. Stereo-Faces Detection

We use sparse stereo matching algorithm for stereo-faces obtaining. It works with small amount of scene points that leads to fast processing speed. In classical scheme of stereo reconstruction correspondence is estimated for each point of stereo frames. Our algorithm processes only points of interests. These points correspond to human faces found at each view. We use cascade of weak Haar classifiers [10] for face detection in each frame. Fig 1 shows stereo pair example with detected faces and corresponding epilines.

A. Matching Algorithm

Applied approach for face corresponding estimation is based on geometric characteristics analysis, spatial face position and visual correlation. Faces found in each view are characterized by their geometric characteristics and pattern. The algorithm calculates correspondence degree for objects from two frames based on these characteristics. The goal of this procedure is selection of most probable pairs and rejection of faces without pair.

Pair correspondence is defined by multiplication of the set of coefficients. Each coefficient corresponds to some feature correlation:

- K1 – represents the faces positions accuracy in compliance with epipolar geometry;
- K2 – represents face sizes correspondence;
- K3 – represents histogram correspondence of the face areas.

All calculation are performed in the coordinate system of one frame from stereo pair. Algorithm handles only pairs that comply with epipolar constraint. In other words, pair candidates for the face from one view have to lie on correspondent epiline at another view. In fig.2 each man's face in first image has two candidate faces in second image, but woman's face has only one. This constraint allows greatly reducing the amount of false pairs. Then size and position coefficients are calculated.

Coefficient K1 is defined as ratio between distance from face area centre to correspondent epipolar line and minimal face size in pair. It possesses the value 1 when the center of the face coincides with the line and is less than 1 otherwise. Coefficient K2 is defined as ratio between smaller and bigger area sizes. It also is less or equal to 1.

It is necessary to estimate face position in the 3D space for coefficient K3 calculation.

B. Histogram matching.

Coefficients K1 and K2 allow rejecting big amount of false pairs. They are not enough because are based on geometrical features that can be obtained with essential error. That is why histogram matching based coefficient K3 is

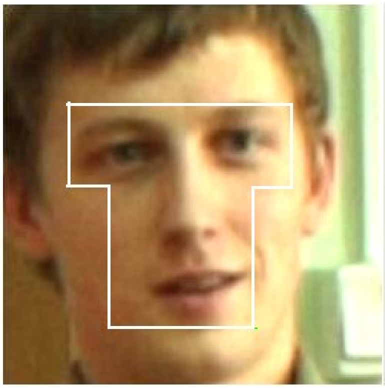

Fig. 3. Histogram area

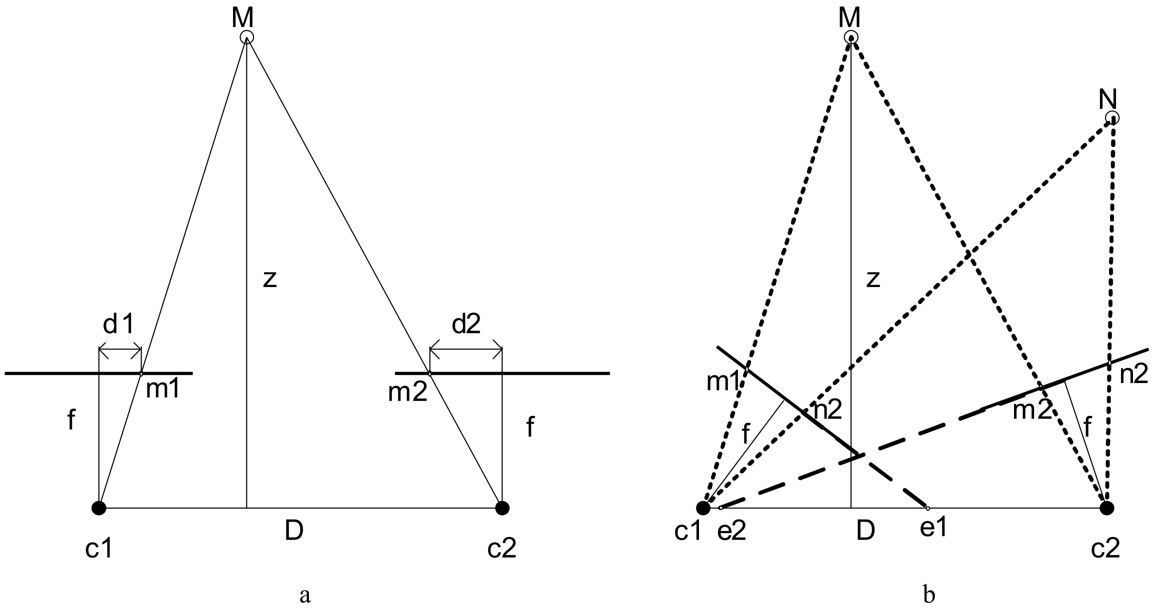

Fig. 4. Stereo geometry a) simple, b) with arbitrary cameras positions.

computed for the rest small amount of possible pairs. Histogram matching is used because of its relative computational simplicity in comparison with other metrics.

Histogram is built not for all face area to avoid influence of background, hair, clothes pieces that can be not presented at both views. Only central part of face area is used where eyes, nose and mouth are apparently situated (see fig. 3). Median filter and area normalization are used to bring histogram to common form.

Sum of Absolute Differences (SAD) and Sum of Squared Differences were used for histogram matching. The SAD applied to three-cannel histograms shows the best result. Error is calculated as square of Euclidian distance between each RGB component's errors. Coefficient K4 then is estimated as:

$$K3 = 1 - histERR, \tag{2}$$

where *histERR* is histogram correspondence error.

Hence maximization of the product *K1*K2*K3* gives us the most probable pairs. Pairs formed by faces without right correspondence are ignored by threshold. Such pairs can appear because only one face view can be visible for stereo cameras. The threshold is applied to compound coefficient (the product). Its value was estimated empirically and is equal to 0,3.

IV. Face 3D Position Estimation.

It is known from stereomerty that distance to object can be estimated by using triangulation. Fig 3a. shows simple stereo system geometry. The indications are: c1 and c2 – camera optical centers, M – point of the scene, m1 and m2 – scene point projections, *f*–focal length of optical system and *D*–distance between optical centers (so called base line). Unknown distance *z* can be found by using next equation:

$$\frac{z}{D} = \frac{f}{d1 + d2}, \tag{3}$$

where sum *d1+d2* is disparity of projections' positions.

Real stereo systems rarely have such ideal configuration. Fig.4b represents the geometry of a stereo system with arbitrary cameras position and orientation. Image planes in such system are not parallel to base line. It is necessary to perform projective transformation of each plane in order to use eq. 3 for distance estimation. It leads to additional computational cost.

Optical centers' projections can not be used as reference points for disparity computation in this case. It is seen in fig. 4b that disparity for space points M and N relative to image centers are equal. But these points lie at different distance from base line. It is necessary to choose reference points for disparity computation that allow estimating distance independently from camera configuration. We use epipolar centers as such points. These centers are points where all epipolar lines of respective view come together. The centers are indicated as e1 and e2 at fig. 4b. These points can be calculated by using next equations:

$$\begin{aligned} e_1 F &= 0, \\ e_2 F^T &= 0, \end{aligned} \tag{4}$$

where *F* is fundamental matrix.

Disparity computed relative to epicenters can not be used for real distance calculation but allows estimating its value. Disparity evaluations can have large absolute values. Especially in the case when image plane has very small angle relative to base line. Such values are not informative. We consider the crossing point of cameras' optical centers as zero point (Z=0) to avoid large disparity values. Base disparity *d*0 is calculated for this point. Distance to the object is estimated as difference between the point and base disparities. For example absolute disparities *dm* and *dn* for points M and N at fig. 2b can be estimated as:

$$\begin{aligned} dm &= dist(m1,e1)+dist(m2,e2), \\ dn &= dist(n1,e1)+dist(n2,e2), \end{aligned} \tag{4}$$

where *dist* is distance function between two points. Distance (or depth) evaluation can be computed as $dm-d0$ and $dn-d0$ respectively.

Hence three-dimensional face position accuracy in compliance with 3D region of interest can be computed as:

$$K_{ROI} = 1 - \frac{|d_{face} - d_0|}{\Delta d}, \tag{5}$$

where Δd is disparity scattering determined by 3D region of interest.

V. Position estimation discussion

It seems that the coefficient K_{ROI} can to be used as additional weight for pair correspondence estimation. It means that it can be added to the complex coefficient ($K1*K2*K3*K_{ROI}$). Figure 5 illustrates this idea.

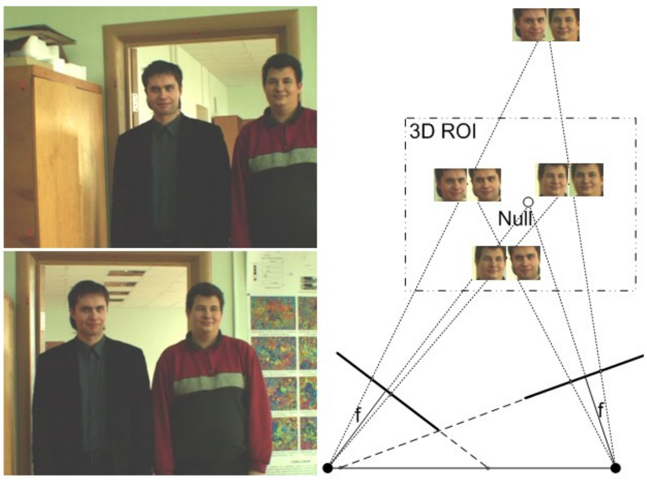

Fig. 5. Peoples stay close to zero point.

Men's stay quite close to zero point inside 3D ROI. False face pairs are located behind and in front of true pair's positions. It means that difference between true face disparity and zero point disparity is less than the same value in false face case. False face means that it is formed by false pair. Coefficient K_{ROI} (5) will be mach less for such pairs than for correct pairs. It allows greatly increasing of the right correspondence rate.

It was good idea, but its application meat problems. Some cases exist then this condition does not help. Figure 6 represent the example then people stay close (at the same distance) to each other but relatively far from zero point.

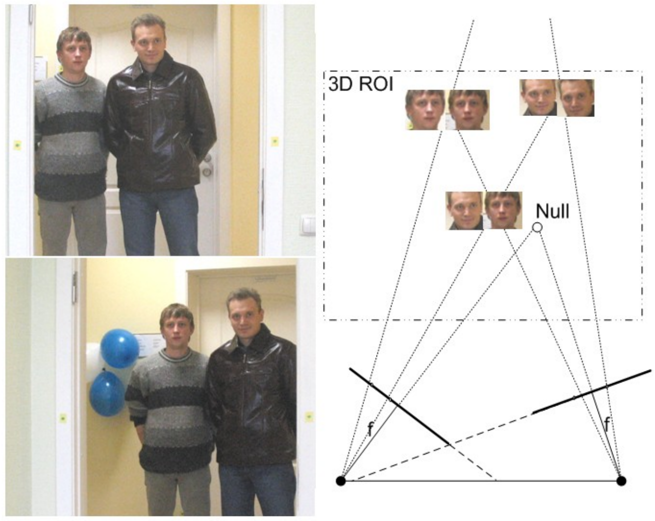

Fig. 6. Peoples stay far behind zero point.

Here false pair has greater weight because it lies closer to zero point. Right and face pairs positions are unpredictable. It is impossible to say even that they always positioned at the same cornets of the quadrilateral (fig. 7).

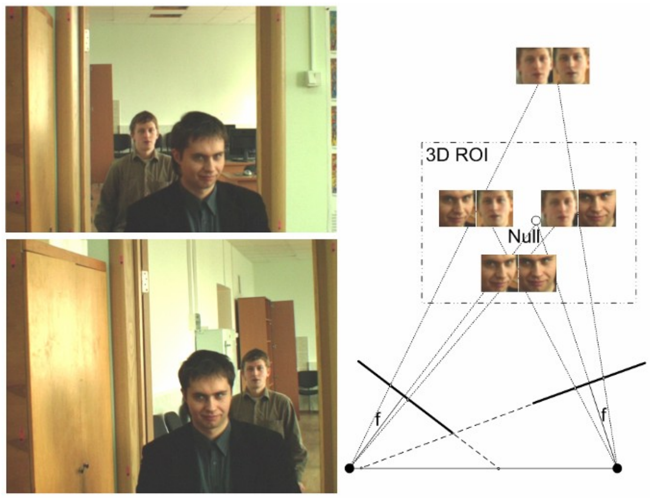

Fig. 7. Peoples stay at the different sides from zero point.

The obstacle problem also leads to unpredictable values of the coefficients for correct and false pairs. We considered different cases of occlusion then a face can close one ore more other faces on different views. We skip its detailed description.

As the result we can say that coefficient K_{ROI} can't be used for correspondence estimation as it was described. But it can be applied for face ranging. It can be usable to process faces with predictable parameters. For example face identification method can be configured to operate with appropriate size faces. This size must correspond to visible face size then corresponding man stay at the zero point distance.

Fig. 8. Missed views estimation based on epipolar geometry.

VI. Missed Views Estimation

Described algorithm can be used in a system this two and more cameras. The situation then face was not found at all vies is possible. It may be causes by to big angle of the head rotation, occlusions, mistakes and so on. If face was detected at least at two frames its position can be computed at other views based on epipolar geometry.

Position of the missed face view is defined by crossing two ore more epipolar lines which correspond to known views. For example, face was detected at first and second frames (Fig. 8) and they were matched by presented algorithm. It is necessary to compute epipolar line at third frame that correspond to the face at first frame and also epipolar line that correspond to the face at second frame. Face must be positioned on each epipolar line. That's why the crossing is the result (Fig 8. right image).

VII. Results

The algorithm has been tested with images obtained by stereo device that includes two high resolution cameras. Stereo frames are color images with resolution 1024*768. It was considered that maximum face size is 100*100 pixels. Test set includes 200 stereo pairs and Table 1 represents test results.

The matching accuracy came to 93.5%. Time costs for processing is about 50 ms. per stereo frame. Main part of this processing time employs face detection module. Hence presented algorithm allows face views matching with high processing speed.

Table I.
Algorithm Test Results

Parameter	Numerical value	Percentage
Number of faces in left view	336	-
Number of faces in right view	372	-
Detected faces (left)	326	97%
Detected faces (right)	366	98%
Total amount of correct pairs	306	-
Correctly matched pairs	286	**93,5%**
False rejections by threshold	4	1,3%
False matched pairs	2	0,65%
Faces without pair (left)	30	13%
Faces without pair (right)	61	

VIII. Future Work

Described algorithm was developed to be used in biometric identification applications which operate in real time. Several ways to use the results of this algorithm for this purpose exist.

First of all stereo face can be used for best view estimation. 2D face recognition algorithms mostly based on frontal face processing. Information about relative positions of eyes, nose and other feature point can be used for appropriate image selection. Needed information can be obtained using methods of stereo analysis.

From other hand good frontal view can be not represented at frames. It is very probable situation especially if only two cameras are used. In this case stereo face can be used to develop frontal face view.

The most preferable case of stereo face usage is 3D model estimation (Fig. 9). Such techniques are still very computationally expensive and require a lot of efforts to be launched in real time.

IX. Conclusion

Developed algorithm allows obtaining 3D position of stereo-faces with high computational efficiency. It is based on sparse stereo matching of objects of interest. Fast speed processing was also reached by using distance estimation procedure without projective transformation. It can be used in real time surveillance systems as base for 3D face identification.

References

[1] S. Gupta, M. K. Markey, and A. C. Bovik, "Advances and challenges in 3D and 2D+3D human face recognition," in *Pattern recognition in biology*, M. S. Corrigan, Ed., Nova Science Punlisher Inc., 2007, pp. 63-103.

[2] L. Akarun, B. Gokberk, and A. A. Salah, "3D face recognition for biometric applications," in *Proc. 13th European Signal Processing Conference,* Antalya, Turkey, 2005

[3] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal of Computer Vision*, vol. 64, no. 1, 2005, pp. 5-30.

[4] V. Blanz, and T. Vetter, "A Morphable model for the synthesis of 3D faces," in *Proc. 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187–194.

[5] Z. Liu, Z. Zhang, C. Jacobs, and M Cohen, "Rapid modeling of animated faces from video" unpublished.

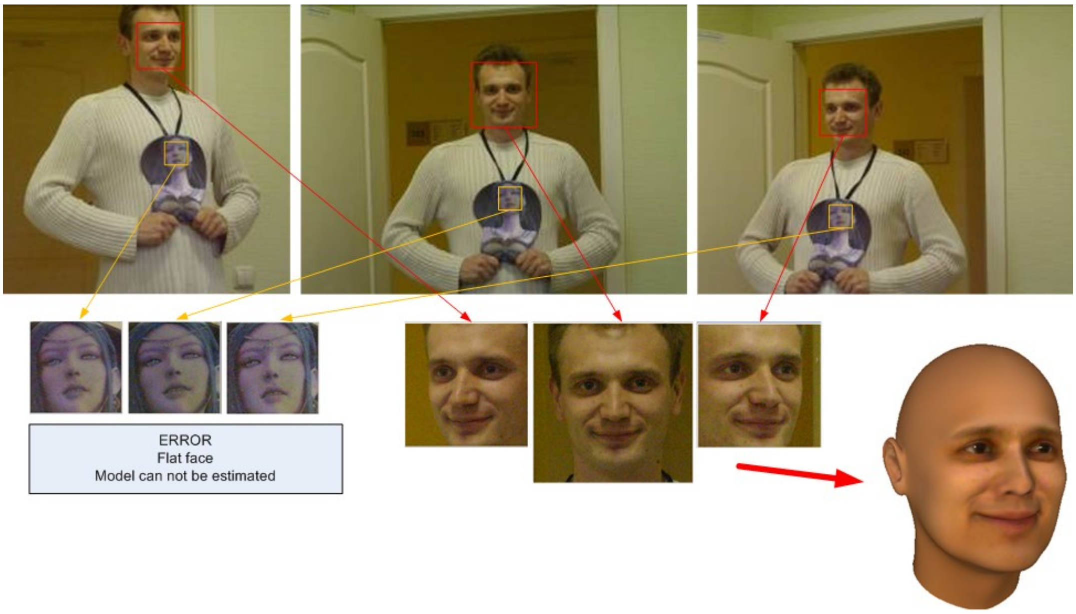

Fig. 9. 3D face model estimation using stereo faces.

[6] C. Park, and J. Paik, "Face recognition using optimized 3D information from stereo images," *Image Analysis and Recognition*, 2005, pp. 1048-1056.

[7] R. Niese, A. Al-Hamadi, and B. Michaelis, "A novel method for 3D face detection and normalization," *Journal of Multimedia, Academy Publisher* vol. 2, no. 5, 2007, pp. 19-26.

[8] A. J. Lacey, N. Pinitkarn, and N. A. Thacker, "An evaluation of the performance of RANSAC algorithms for stereo camera calibration," in *Proc. 11th British Machine Vision Conference*, Bristol, 2000.

[9] I. H. Richard, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Analysis Machine. Intelligent,* vol. 19, no. 6, 1997, pp. 580-593.

[10] P. Viola, and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, 2004, pp. 137-154.