computing@computingonline.net
www.computingonline.net

ISSN 1727-6209
International Journal of Computing

COMBINED APPROACH FOR FACE FRONTAL VIEW ESTIMATION FOR REAL TIME FACES DETECTION IN MULTICAMERA SYSTEM

Denis V. Lamovsky [1], Rauf Kh. Sadykhov [2], Vadim A. Kharlanov [3], Alexandr S. Kirienko [3]

[1] Belarusian State University of Informatics and Radioelectronics, 220013 P.Brovki str., Minsk, Belarus, lamovsky@gmail.com, http://bsuir.by
[2] United Institute of Informatics Problems, 220012 Surganova str. 6, Minsk, Belarus, http://uiip.bas-net.by/structure/l_is/
[3] Synesis Vision, 220043, Nezavisimosti av. 95, room 316, Minsk, Belarus, http://synesisvision.com

Abstract: *This paper presents the combined approach for face frontal view estimation from video sequences in a multiview camera setup. This task is important for person identification by face image in video surveillance systems. Face tracking algorithm was developed based on optical flow and cascade face detector. We also found way to estimate quality of face detection. This quality is used as base for best frontal view estimation.*

Keywords: *face detection, tracking, and frontal view estimation.*

1. INTRODUCTION

The problem of face detection is important in the area of computer vision. It has a lot of applications in both static and dynamic images analysis. The best advances in the area of face detection during recent years are connected with two methods. These are neural networks based approach of Rowley [1] and weak classifier cascade based approach of Viola and Jones [2]. Neural networks are more robust in the case of partially occluded faces and faces with strong shadows. They also can be applied for rotated face detection. From other side weak cascade classifier based face detection is computationally effective and can by applied for real time face detection on video. Many methods were presented to increase robustness of Viola and Jones method. Lienhard [3] extended feature set for weak classifiers and fast computational scheme for new features estimation. That's resulted in reduction of false positive rate on 10%. Huang [4] constructed tree structure instead of cascade. This allowed handling the detection of faces in different pose with a higher speed. Sachenko and Paliy [5] combined neural network approach with cascade weak classifiers. They've used components of Rowley's NN as classifiers at first several stages of cascade. As a result – significant reduction of false patterns that are able to pass first stages.

Analyzing the face in dynamic (so called face tracking) gives you an extra features and abilities that extend the range of applications where face detection can be employed. The main advantage of face tracking is the ability to use not only spatial resolution but temporal resolution as well.

For instance, face tracking can be a base of face 3D modeling system. Only one camera can be used to retrieve the sequence of facial images of the same human. While moving in front of the camera the same face will be captured with the different perspective and the combination of them can allow you to reconstruct 3D face model [6]. The model in its turn can be used for face recognition mission.

Another important opportunity of face tracking is an analysis of face regions. Face tracking system output can be used for such actual applications as lip reading (as an alternative or addition to speech recognition) [7], gaze tracking (attention detection in digital signage analytics) [8], face expression analysis (as an input interface system) [9] etc.

As you may see the only limit for face tracking system applications is our imagination. However besides lots of advantages there are many difficulties.

One of such difficulties that were the main obstacle for face tracking methods implementation and developing are high computation demands of face detection algorithms. Due to this a real time face tracking with appropriate resolution was non-trivial task in recent past. Nowadays this problem is not so urgent as computer systems performance continue growing.

Another important problem of face tracking is a quality of multi-faces tracking. The primary issue here is the potentiality of faces overlapping. Since face tracking is the core component of different analytics systems (as it was described above) and ordinary visual scenes usually contain more than one human face it is extremely important to develop an accurate and resolute algorithm of real time multiple faces tracking.

One of application of face detection and tracking is people identification in video. Accurate frontal facial view is very important in this case. Dynamic data usage gives an additional ability to estimate best facial view. This means that system has to analyze facial patterns at each frame during tracking and determines which one is the best for recognition.

Another way to increase face detection and recognition system is to use multicamera approach. Several cameras are used in this case for best face view estimation.

In this paper we present system for face detection and tracking in multicamera setup for face frontal view estimation.

2. FACE DETECTION

Face detection module of presented system is based on cascade of classifiers [2]. We describe it briefly in order to show theoretical basis for face quality estimation.

Classical approaches for appearance based face detection use redundant search. Image is scanned by fixed window on different scales. Face/noface classifier is applied for pattern in each window position. It leads to huge amount of patterns that have to be checked. Viola and Jones proposed cascade of classifiers (figure 1).

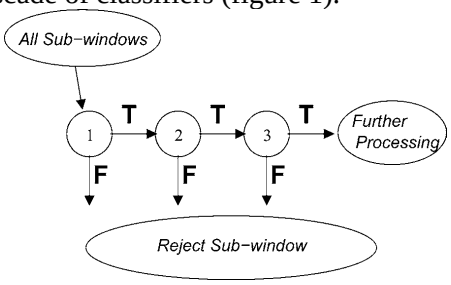

Fig. 1 – Cascade of classifiers scheme

Each stage in cascade can provide relatively high false positive rate. But it has to operate extremely fast. Cascade of such classifiers allows quickly eliminate most non face patterns and provides low positive rate at the same time. Time of classification of one window depends on pattern and face similarity. Size of cascade provide low positive rate.

Stage classifier consists of set so called weak classifiers. Each of them is based on two-dimensional Haar-like function (feature) (see figure 1). Haar-like features are used because of their computational simplicity. The set of features for each classifier stage is obtained using boosting algorithm during learning stage.

Stage classifier provides boolean answer that depends on sum of feature values:

$$h(x) = \begin{cases} 1 & \sum_{i=1}^{n} \alpha_i h_i(x) \geq \theta \\ 0 & otherwise \end{cases} \quad (1)$$

where $\alpha_i(x)$ – number of features, $h_i(x)$ – answer of i-th weak classifier, $б_i$- coefficient that is defined by importance of i-th feature and u – strong classifier threshold.

Additional information can be found in original work [10]. Method of face confidence level estimation based on equation (1) will be described below.

2. STEREO-FACE MATCHING

We use sparse stereo matching algorithm for stereo-faces obtaining. Stereo-face here is a set of face images obtained from different views. 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. Fig 2 shows stereo pair example with detected faces and corresponding epilines.

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.

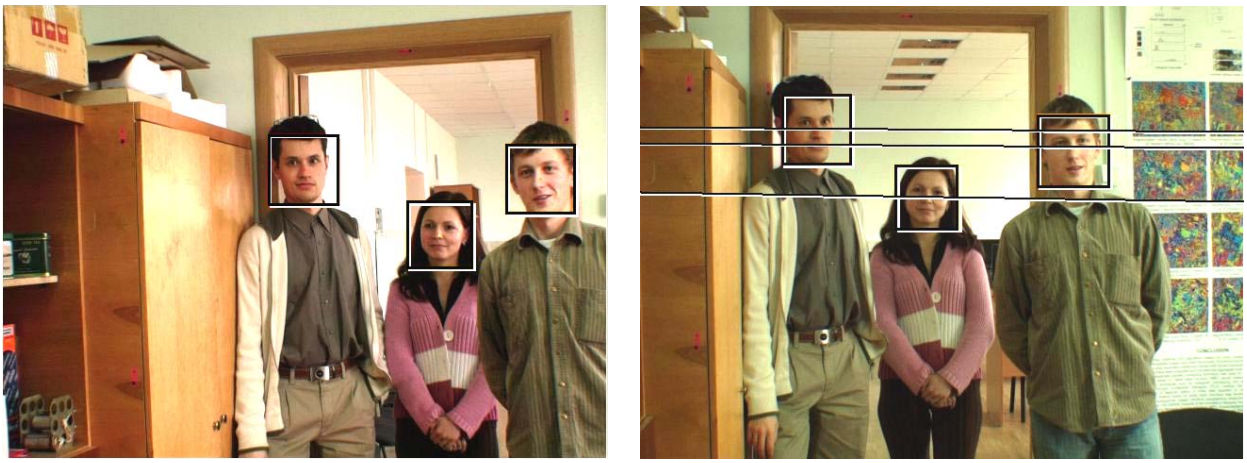

Fig. 2 – Stereo frame with detected faces and corresponding epipolar lines

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 center 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.

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 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 K3 then is estimated as:

$$K3 = 1 - histERR\,, \qquad (1)$$

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.

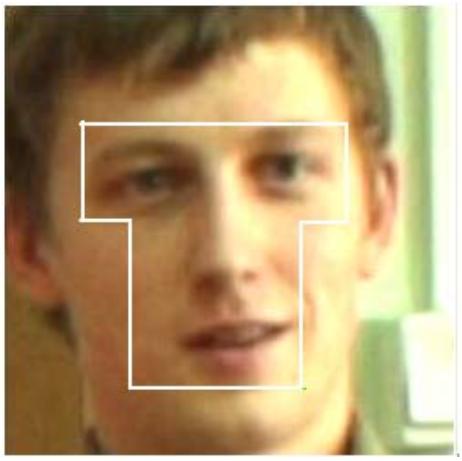

Fig. 3 – Histogram area

3. FACE TRACKING

The main task of face tracking system is to link static face images to dynamic objects. Each object contains the history of face moving, including preceding trajectory and a set of dynamic (or temporal) features. The accuracy of linking is very important as each false link has a dramatic effect to the whole tracking system performance. However, to

provide a confident tracking, the system should implement a number of essential techniques. The most important are:

- Precise moving object prediction algorithm. Solving this issue will allow you neatly track single faces.
- High-quality matching algorithm. It is necessary to solve situations when you need to track face groups – several face objects that concentrated in the local region of the scene and generally have short speed vector.
- Face overlapping resistance. Face objects during their movement can partially or fully overlap each other leading to additional tracking difficulties.

Our system uses intends to solve difficulties described above using original combined algorithm. The basis of our tracking system is a face position prediction using local optical flow. Face histogram matching is used to resolve complex situations with the face groups. Face histogram matching is also a part of overlapped faces tracking algorithm. Another important part of it is a face position prediction based on history of movement.

4. FACE POSITION PREDICTION BASED ON OPTICAL FLOW

To predict the position of the face at current frame we calculate global optical flow in face area. The predicted position is used in two ways. It is consider like connection point if traced face was detected at current frame, and is used as new face position otherwise. Second situation is possible if the observed person turn head away from camera of partially hide face. Optical flow allows tracking face at such frames until it will be detected again.

Optical flow is a two dimensional field that represents directions and velocities in each point. Global optical flow shows movement vector for the whole zone of interest. We use multi-scale differential approach to estimate visual movement of tracked face. It based on so called optic flow constraint [11]:

$$E_x u + E_y v + E_t = 0 \ , \qquad (3)$$

where E_x, E_y, E_t – local gradients of frame point (one channel representation), (u,v) – optical flow vector.

We consider that optical flow is constant in all points of the face area. It leads to linear system of equations (3). Calculation of gradients is performed using classical scheme from [12]. We consider only points there at least one gradient is not zero.

Multi-scale image pyramid is built for accurate estimation of movement vector. Each level of pyramid is two times less than previous. Biggest movement is presenter at top level. We use coarse to fine strategy to estimate precise motion vector. Vector of displacement from current level is used to correct initial area position at next level. Finite vector of movement is obtained as sum

of vector from each level multiplied by 2^l. There l is number of the level.

We also use optical flow quality estimation procedure to find flow confidence level. This level shows how accurate vector of movement was calculated.

5. FACE POSITION PREDICTION BASED ON HISTORY OF MOVEMENT

Optical flow can be calculated with significant error. It can happen if big movement is presented or face image is blurry. In this case algorithm relies on other face position prediction procedure.

As face object is a temporal entity, we can use the history of its trajectory to predict it current position. For this purposes we employ such natural parameters as trajectory points (central points of face images) and corresponding timestamps. Let's say we have a trajectory of n frames. In this way the average speed for x and y axis can be estimated as a first derivative of the path function:

$$\begin{aligned} \overline{V_X} &= \frac{1}{n} \cdot \sum_{i=0}^{n-1} \frac{x_{i+1} - x_i}{t_{i+1} - t_i} \\ \overline{V_Y} &= \frac{1}{n} \cdot \sum_{i=0}^{n-1} \frac{y_{i+1} - y_i}{t_{i+1} - t_i} \end{aligned} \ , \qquad (4)$$

where $\overline{V_X}$ and $\overline{V_Y}$ – average speeds in x and y directions respectively, (x_i, y_i) – position of the face in time t_i.

Using the average speed value, predicted coordinates can be estimated as follows:

$$\begin{aligned} x_{i+1} &= x_i + \overline{V_X} \cdot (t_{i+1} - t_i) \\ y_{i+1} &= y_i + \overline{V_Y} \cdot (t_{i+1} - t_i) \end{aligned} \ , \qquad (5)$$

6. HISTOGRAM MATCHING

To resolve complex situations when the optical flow has insufficient values for unambiguously linking face objects to currently detected faces we used the idea of histogram matching algorithm used in [13]. Corresponding algorithm was adapted for using face histogram as a face object dynamic feature. The word "dynamic" means that each time new face image is linked to face object, object's histogram feature is updated using the following formula:

Fig. 4 – Confidence level obtained for different face rotations

$$H_{OBJ}[i] = (1-\alpha)\cdot H_{OBJ}[i] + \alpha \cdot H_{IMG}[i], \qquad (6)$$

where *б* – is the fractional coefficient in range [0;1] that describes influence of newly linked face image histogram to present histogram statistics, H_{OBJ} – histogram of face object, H_{IMG} – histogram of face image and i – index of color.

7. BEST FRONTAL VIEW ESTIMATION

We estimate frontal view by using face quality obtained from classifier. Equation (1) and figure 1 show that classifier cascade gives only binary response. Originally no face quality measure is provided. We calculate the level of similarity between detected pattern and face appearance using difference between sum of classifier features hi and threshold *u*. Threshold define the border between face and nonface. Distance between stage classifier sum and threshold show how much pattern looks like average face. Average face is defined at training step and depends on faces appearance in training dataset. We use cascade that was built to detect frontal faces. It leads to high similarity level for strong frontal faces and near low – for rotated faces. We calculate the sum of square similarity measures at all levels (7) and consider it as confidence level for whole cascade.

$$c = \sum_{l=1}^{N}\left(\sum_{i=1}^{n_l}\alpha_{l,i}h_{l,i}(x) - \theta_l\right)^2, \qquad (7)$$

where N – number of cascade stages. Other variables came from (1).

8. RESULTS

The multiview face matching 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 1. 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	

Figure 4 shows several examples of confidence level estimation using presented approach for different face appearance. Absolute value of measure depends on concrete cascade. We also can't estimate confidence level maximum. Described measure is used only for comparison of several detections. In the case of face tracking we use it to find best face during face observation. In the case of multi-camera face analysis described measure can be used for best view estimation.

Application of confidence level measurement algorithm is presented at figure 5. Best views from two cameras are highlighted automatically. It is used

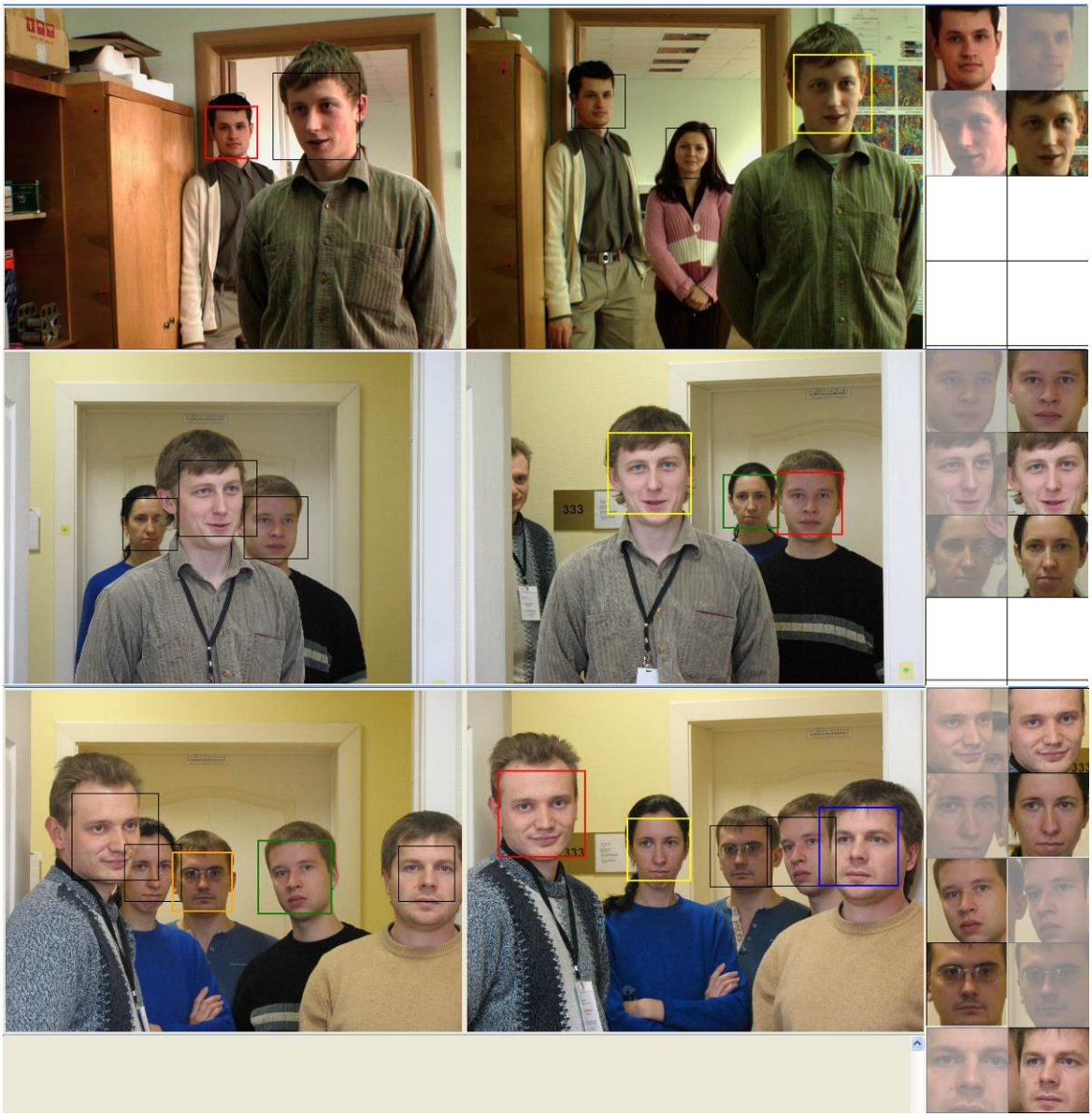

Fig. 5 – Best face view estimation examples

for incising of the accuracy of the recognition algorithm.

9. REFERENCES

[1] H. A. Rowley, S. Baluja, and T. Kanade, Neural network-based face detection, *IEEE Trans. Pattern Analogous. Machine Intelligence*, (1998). pp. 23-38.

[2] P. Viola and M. J. Jones, Robust real-time face detection, *International Journal of Computer Vision*, (57) (2004).pp. 137-154.

[3] R. Lienhart and J. Maydt, An extended set of Haar-like features for rapid object detection, *IEEE International Conference on Image Processing*, (2002). pp. 900-903.

[4] C. Huang, H. Ai, Y. Li, and S. Lao, Vector boosting for rotation invariant multi-view face detection, *IEEE International Conference on Computer Vision*, (2005). pp. 446-453.

[5] A. Sachenko, I. Paliy, Y. Kurylyak, V. Kapura, R. Sadykhov, and D. Lamovsky, Face detection algorithms for video-surveillance system, *IEEE Workshop on Intelligent Data Acquisition and Advance Computing Systems: Technology and Applications Dortmund*, Dortmund, Germany, (6-8 September 2007). – pp. 594-598.

[6] C. Cheng and S. Lai, An integrated approach to 3D face model reconstruction from video, *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Washington, USA, (2001), p. 16.

[7] T. Chen, Audiovisual speech processing, *IEEE Signal Processing Magazine*, (18) (2001). pp. 9-21.

[8] Y. Zhai and M. Shah, Visual attention detection in video sequences using spatiotemporal cues, *14th annual ACM international conference on Multimedia*, Santa Barbara, USA, (2006).

pp. 815-824.
[9] F. Ren, N. Nagano, D. B. Bracewell, S. Kuroiwa, T. Tanioka, Z. Zhang, and C. Zong, Facial feature based expression recognition for an effective interface, *Artificial Intelligence and Soft Computing Conference*, Spain, (2005).
[10] P. Viola and M. Jones, Robust real-time object detection, *Second international workshop on statistical and computational theories of vision – modeling, learning, computing, and sampling*, Vancouver, Canada, (2001).
[11] A. Bruhn, J. Weickert, and C. Schnörr, Lucas/Kanade Meets Horn/Schunck: Combining local and global optic flow methods, *International Journal of Computer Vision*, (6) (2005). pp. 211-231.
[12] B. K. P. Horn and B. G. Schunk, Determining optical flow, *Artificial Intelligence*, (17) (1981). pp. 185-203.

Rauf Kh. Sadykhov, *in 1967 graduated from Azerbaijan Polytechnic Institute (Baku) on the specialty "Mathematical and Computing the Instruments and Devices". After graduation from the Institute he attended the postgraduate course at the Institute of Engineering Cybernetics in Minsk. In 1991 he defended his thesis for a scientific degree of a doctor of engineering science in the field of computing science and in 1992 has obtained a professor's scientific rank.*

Since 1995 R.Kh. Sadykhov is a head of Computer System Department in Belarusian State University of Informatics and Radioelectronics and simultaneously he is a head of System Identification laboratory, Institute of Engineering cybernetics of the Belarusian Academy of Sciences. R.Kh. Sadykhov has published more than 350 scientific works, including books, patents, papers and reports at the International Conferences, Symposiums and Workshop.

The area of the scientific investigations includes: digital signal and image processing, recognition of handwritten symbol and signature identification, remote-sensing object recognition, computer vision system for the control and recognition, intellectual neural systems, multi-agent systems, parallel architectures for digital signal and image processing.

Prof. R.Sadykhov is the vice-chairman of Belarusian Association of Pattern Recognition (IAPR) and Belarus SIG of International Neural Network Society (INNS), the member of IEE (United Kingdom).

Denis V. Lamovsky, *in 2004 graduated from Byelorussian State University of Informatics and Radioelectronics (Minsk) on the specialty "Electronic computing machines, systems and networks". After graduation from the University he attended the master's course at the same institution and received master's degree. In 2009 Denis Lamovsky defended his thesis and received a PhD degree.*

The area of the scientific investigations includes: digital image and video processing, face and iris recognition, computer vision system for the control and recognition, parallel architectures for digital signal and image processing.

Vadim A. Kharlanov, *in 2004 graduated from Byelorussian State University of Informatics and Radioelectronics (Minsk) on the specialty "Electronic computing machines, systems and networks". After graduation from the University he attended the master's course at the same institution and received master's degree in 2005.*

Alexander S. Kirienko, *in 2009 graduated from Byelorussian State University of Informatics and Radioelectronics (Minsk) on the specialty "Electronic computing machines, systems and netrworks". After graduation from the University he attended the master's course at the same institution.*

The area of the scientific investigations includes: digital image and video processing, face recognition, computer vision system for the control and recognition, parallel architectures for image processing, 3D scene analysis.