Combined approach for face frontal view estimation for video surveillance purposes.

R. Sadykhov[1], D. Lamovsky[2], V. Kharlanov[3], A. Kirienko[3]

1) United Institute of Informatics Problems, Minsk, Belarus
2) Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus,
lamovsky@gmail.com
3) Synesis Vision, Minsk, Belarus, vadimk@synesis.ru, synesis.ru

Abstract: This paper presents the combined approach for face frontal view estimation from video sequences. 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.

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

2. LAYOUT

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).
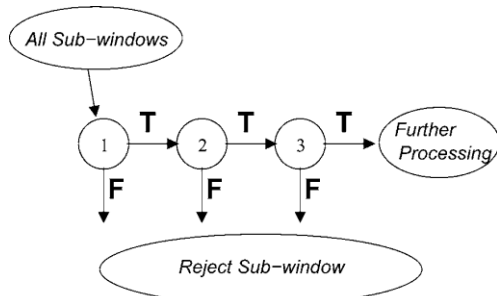
Fig.1 – Cascade of classifiers scheme.

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 $h(x)$ – answer of stage (strong) classifier, n – number of features, $h_i(x)$ – answer of i-th weak classifier, α_i- coefficient that is defined by importance of i-th feature and θ – 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.

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. 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] (2).

$$E_x u + E_y v + E_t = 0, \quad (2)$$

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 (1). 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 level.

We also use optical flow quality estimation procedure to find flow confidence level. This level is used in tracking procedure to select object for connection.

4. 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:

$$H_{OBJ}[i] = \beta \cdot H_{OBJ}[i] + \alpha \cdot H_{IMG}[i], i = \overline{0,255}, \quad (3)$$

where $\beta = 1-\alpha$, α – 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.

5. 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

Fig.2 – Confidence level obtained for different face rotations.

classifier features h_i and threshold θ. 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.

$$c = \sum_{l=1}^{N} \left(\sum_{i=1}^{n_l} \alpha_{l,i} h_{l,i}(x) - \theta_l \right)^2 , \qquad (4)$$

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

7. RESULTS

Figure 2 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.

6. REFERENCES

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

[2] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision,* vol. 57, pp. 137–154, 2004.

[3] R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection," in *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," in *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," in *IEEE Workshop on Intelligent Data Acquisition and Advance Computing Systems: Technology and Applications* Dortmund, Germany, 2007.

[6] C. Cheng and S. Lai, "An Integrated Approach to 3D Face Model Reconstruction from Video," in *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Wahington, USA, 2001, p. p.16.

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

[8] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *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," in *Artificial Intelligence and Soft Computing Conference*, Spain, 2005.

[10] P. Viola and M. Jones, "Robust Real-time Object Detection," in *Second international workshop on statistical and computational theories of vision – modeling, learning, computing, and sampling*, Vancouver, Canada, 2001.

[11] A. Bruhn and J. Weickert, "Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods," *International Journal of Computer Vision,* vol. 6, pp. 211–231, 2005.

[12]   B. K. P. Horn and B. G. Schunk., "Determining optical flow.," *Artificial Intelligence,* vol. 17, pp. 185–203, 1981.