Application of the background subtraction method and Viola-Jones algorithm for increasing the speed of face tracking in the personality recognition system[*]

S.G. Nebaba, A.A. Zakharova[**]

Tomsk Polytechnic University, Tomsk, Russia

Existing method of automatic tracking of moving objects in the video stream in real time is described in the paper in relation to the task of person's recognition from his face image. A method of face tracking in the video stream based on a combination of the background subtraction method and the Viola-Jones algorithm for the face area detection in the frame and reducing requirements for computing resources of person's recognition systems is proposed. Results of testing the proposed face tracking algorithm in the video stream are shown.

Keywords: computer vision; face detection; facial recognition; tracking; Viola-Jones algorithm; background subtraction; real-time video stream processing

Introduction. The works on solving the problem of identifying a person on the basis of face images analysis are conducted from the earliest stages of the development of computer vision. Recently, the demand for quick and correct recognition of a person's identity in the real time video stream is growing in various areas of activity, especially in the areas related to security and people search.

The problem of finding the face area on a raster image was solved one of the first and its solution has long been used in a variety of technical devices, for example, in photo and video equipment for automatic focus.

The problem of automated person identification (recognition) with his face image is much more complicated than simple face detection, and until now there is no algorithm that would be able to recognize person's identity with face image in real conditions as effectively as a human. Every year new approaches to localization, processing and recognition of objects appear, but often these approaches do not have sufficient accuracy, speed and reliability in a real environment, which is characterized by the presence of noise in video sequences, various shooting conditions, such as illumination and the angle in which faces are recorded [1,2]. The methods used to solve the problem of facial recognition should improve the accuracy and reliability of recognition, without having a significant effect on the processing speed of video sequences.

For the use of resource-intensive and effective methods of facial recognition in the real time video stream, it is important to minimize the computational load of the auxiliary algorithms used in

---

[*] The research is done within the frame of the government task №. 2.1642.2017/4.6 for the project «Cognitive methods for visualization and analysis of multidimensional data in the modeling of nonlinear dynamic systems».
[**] E-mail: stepan-1fx@mail.ru, zaa@tpu.ru.

all recognition systems. The most resource-intensive of them is the Viola-Jones algorithm for the face area detection in the frame.

Therefore, the problem of creating a face tracking algorithm in a sequence of frames, which allows to reduce the computational cost of the face detector for real-time recognition systems, is actual.

Tracking of objects in the video stream. Typically, the same person is present in the field of view of the surveillance camera for some time interval, so its face image can be found in a sequence of frames. This provides an opportunity to perform tracking, in other words, to determine the location of a moving object on the video.

Often the first step to solving this problem is the selection of the foreground. For example, the paper [3] considers methods of a background subtraction. There are many problems that have to deal with when using such methods: a changing background, moving shadows, changing illumination, and camera noise.

Most of described problems are not relevant in the task of human face tracking, since the search for an object is carried out using the well-studied and repeatedly verified Viola-Jones algorithm.

The algorithm of P. Viola and M. Jones was published in 2001 [4]. It does not specialize in selecting only faces and can be used to find other classes of objects, which requires the use of the appropriate classifier. Nevertheless, it was most widely used to solve the task of detecting faces on a raster image.

In the process of using the Viola-Jones algorithm for the task, a set of rectangular areas corresponding to the detected faces will be found in each frame.

Since tracking is used solely for the purpose of reducing the amount of computation by combining sequences of face images in a video stream (or for the purpose of obtaining a set of several images of one face), then it should not be required in absolute accuracy and applicability to a variety of different objects. In this case, it is enough to use an algorithm that operates with information about the position of the face images in the frame, which will make possible to correlate most of these images to two adjacent frames.

Tracking method based on the known face position in the frame. For the development of the tracking algorithm, the simplest method, which operates only with data about detected faces position in the frame, is taken as a basis.

The video stream can be represented as a sequence of frames

$$V_s = [f_1, f_2 \dots f_N],$$

where N is the total number of frames, f_i is the current frame. Any face found in the frame is described as a set

$$Fc_{ij} = \{If_{ij}, Rf_{ij}\},$$

where If_{ij} is the face image, Rf_{ij} is its position in the frame, i is the frame number, j is the face number in the frame. Wherein Rf can be represented as a set of two points

$$Rf = \{p_1=(x_1; y_1), p_2=(x_2; y_2)\},$$

where p_1 is the upper left point of the rectangular area, p_2 is the lower right point, x and y are their coordinates.

Then the face track k in a video stream is a sequence of faces Fc_{ik}, corresponding to one person:

$$Tr_k = [Fc_{1k}, Fc_{2k} \ldots Fc_{Nk}].$$

The main task of tracking is comparison of target object positions in the sequence of frames. Thus, tracking systems usually use a model that describes how the image of the target object can be changed with all sorts of its different motions [5]. In the case of tracking the face position, it is assumed that the objects position in the next frame is changed slightly. In this regard, the following tracking method is proposed: the found face image belongs to the track Tr_k, if the point-center of the rectangle Rf of face Fc_{ij}, $p_c = (\dfrac{x_1 + x_2}{2}; \dfrac{y_1 + y_2}{2})$, belongs to the area of the rectangle Rf of face $Fc_{i-1,j}$, found on the previous frame and belonging to the track Tr_k.

This method is very simple from a computational point of view, but it is quite effective for solving the problem of combining face images in sequences on video while refining the resulting estimation of belong the face image to a person in recognition systems.

This method can be improved in two ways: in the direction of increasing the quality of tracking, that is, eliminating gaps in motion due to failure of the detector, and in the direction of increasing the speed of computation.

The speed of tracking plays a greater role in considered problem, associated with increasing the efficiency and speed of video sequences processing by facial recognition systems. It makes sense to perform a preliminary analysis of video frames to find areas potentially containing faces and to reduce computational load. In this regard, the fast methods of background subtraction and motion detection, consisting of the simplest operations of calculating the difference between two frames of the video stream and being easy from the computational point of view, are the most promising [6].

Tracking algorithm based on a combination of the background subtraction method and the Viola-Jones face detector. In the course of this work, an algorithm for accelerating the basic face tracking using the face position in the frame is proposed, it consists a following steps:

1. Calculation of the difference between the current frame and the previous one to determine the existence of motion in the video stream:

$$f_M = f_i - f_{i-1},$$

where f_i is the current frame of the video stream, f_M is the matrix, each element of which determines the difference of the corresponding pixels f_i and f_{i-1} of video stream frames.

2. Calculation of the difference between the current frame and the first one, which determines the existence of non-background objects in the video stream:

$$f_T = f_i - f_0,$$

where f_i is the current frame of the video stream, f_T is the matrix, each element of which determines the difference of the corresponding pixels f_i and f_0 of video stream frames.

3. Calculation of binary images f_T and f_M with a threshold that depends on the quality of the video.

4. Evaluation of white pixels distribution of a binary image for f_T and f_M using a sliding window with a side equal to half the minimum window side of the face detector. Those areas of the image where the sum of the white pixels exceeds the selected threshold are marked as suitable for further analysis. Resulting areas for f_T and f_M images are combined, which helps to include moving objects and objects that are in the frame without moving in the general tracking area.

5. Using of the face detector in the selected image area allows to find people faces among the moving objects.

Fig. 1 shows a binary image with a selected area of motion in the video stream frame.
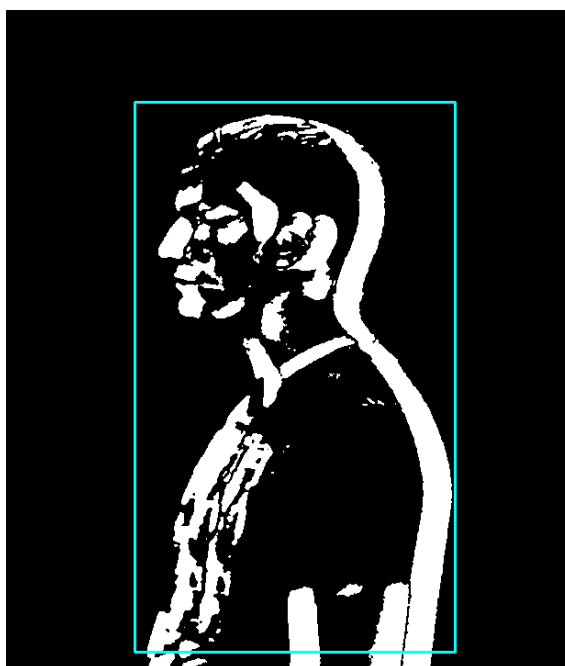

Fig. 1. Area of motion on a binary image f_M

In general case, the selected area with objects will be smaller than the entire frame; it leads to a decrease in the number of operations with the Viola-Jones face detector. In the ideal case, when there are no motion and objects in the frame, the detector does not perform its calculations at all; it frees additional resources to process other video stream frames in the I/O buffer.

The proposed algorithm can work effectively only in limited conditions, in the presence of a static background, that is, with video streams from stationary CCTV cameras. In the case of high scene dynamics, this way of processing acceleration will not be effective and may lead to a slow down search of faces in the image. Minor background changes can be compensated for by regular

replacement of the basic frame at the moments when the total difference of neighboring frames is generally less than some critical threshold [7].

Testing. For initial comparative testing, two variants of tracking were taken: tracking based only on Viola-Jones method and tracking by the proposed algorithm using background and motion subtraction in the frame.

Testing was performed on the following computer configuration: Intel Core i7-3770 processor 3.4 GHz, 16 GB RAM, Windows 7 64-bit.

The video was received on the IP CCTV camera RVi-IPC21WDN with the lens RVi- 1240AIR, as well as using a webcam Microsoft «LifeCam Studio For Business» 5WH-00002.

The resulting video database consists of 263 video sequences in 1280*720 (HD) format with duration from 5 to 20 seconds. The video recording involved 12 people (figure 2), who perform various movements and rotate relative to the optical axis of the camera at angles up to 90 degrees. In total 21,735 face images were found in the frames using the Viola-Jones method in the implementation of the openCV library.
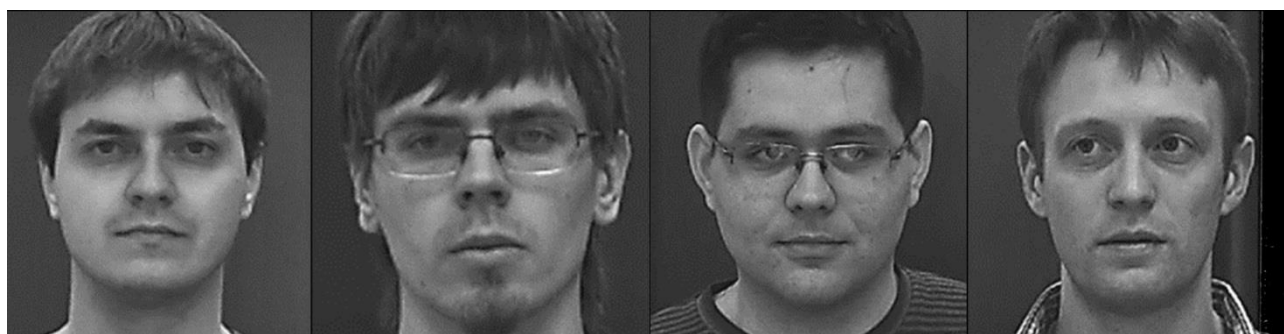

Fig. 2. Example of frontal images of persons from video recordings

The implementation of the Viola-Jones algorithm is taken from the open source computer vision library openCV [8], the minimum detectable face size is 100*100 pixels.

Testing of the algorithm using only the Viola-Jones method on our own video database yielded the following results: 1846 tracks was obtained from the initially found 21,735 face images, the average processing time of one frame was 16.75 ms.

The algorithm, which uses a preliminary analysis of background and motion, allowed to reduce the average processing time of one frame to 2.36 ms, mainly due to the fact that frames, which do not differ from the background and moving objects, eliminates before frames are processed by the face detector. Meanwhile, the average processing time of a single frame containing moving objects has not decreased significantly - on average to 12.14 ms. The total number of tracks also became less - 1834 - due to the fact that some false positive responses of the face detector were excluded from the subsequent processing as a result of the reduction of the analysis area. The average length of each track is 11 frames.

The algorithm for accelerating the search for faces is applicable for situations in which a stationary survey with a simple static background is conducted, but in the case of a dynamic background, this algorithm may lead to a drop in system performance by the time of preliminary processing with background subtraction – 2-8 ms.

The processing time of each frame depends on the parameters of the detector of face area on the image, in the test examples it is about 10-15 ms, and does not require additional calculations after the face detection in the frame. The proposed algorithm for accelerated tracking of objects on a static background does not prevent to apply other methods to improve the accuracy of objects tracking on video, including those that provide higher tracking accuracy with a considerable reduction in speed [9]. Thus, the speed of the proposed tracking algorithm for video with a static background is 27% higher compared to tracking based on the selection of face area and not using preliminary background and motion analysis.

Conclusion. In this article a method of face tracking in the video stream, based on using a joint application of the background subtraction method and the Viola-Jones algorithm to detect the face area in the frame, is proposed. An algorithm, based on the proposed tracking method, is developed, and the results of its work are presented.

The proposed tracking algorithm allows to reduce the requirements for computing resources necessary for the functioning of the face recognition system, as well as to reduce the probability of false detector response with the correct threshold setting of background and motion selection, that is confirmed by the results obtained on the test video.

This algorithm can be used in systems of face recognition in real time video stream provided that the video stream is received from the fixed point camera, which excludes background changes that can neutralize the effect of accelerating the frames processing.

References

1. Bronstein A., Bronstein M., Kimmel R. Expression-Invariant 3D Face Recognition // Proceedings of International Conference on Audio- and Video-Based Person Authentication. - 2003. - pp. 62-70.

2. Bui T.T.T., Phan N.H., Spitsyn V.G., Bolotova Yu.A. Development of algorithms for face and character recognition based on wavelet transforms, PCA and neural networks // Proceedings of International Siberian Conference on Control and Communications (SIBCON) - 2015. – pp. 1-6.

3. Vrazhnov D.A., Shapovalov A.V., Nikolayev V.V. O kachestve raboty algoritmov slezheniya za ob"yektami na video // Komp'yuternyye issledovaniya i modelirovaniye. – 2012. - T. 4, № 2, c. 303–313.

4. Viola P., Jones M.J. Rapid Object Detection using a Boosted Cascade of Simple Features // Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001). - 2001. - vol. 1. - pp. 511-518.

5. Gaganov V.A., Konushin A.S. Segmentatsiya dvizhushchikhsya ob"yektov v videopotoke // Komp'yuternaya grafika i mul'timedia. - 2004. - №2(3). Rezhim dostupa: http://cgm.computergraphics.ru /content/view/67, svobodnyy.

6. Aggarwal A., Biswas S., Singh S., Sural S., Majumdar A.K. Object tracking using background subtraction and motion estimation in MPEG videos // Asian Conference on Computer Vision. – Springer Berlin Heidelberg, 2006. – pp. 121-130.

7. Zamalieva D., Yilmaz A. Background subtraction for the moving camera: A geometric approach // Computer Vision and Image Understanding. – 2014. – T. 127. – pp 73-85.

8. OpenCV: Face Detection using Haar Cascades [Elektronnyy resurs]: [ofits. sayt] - Rezhim dostupa: http://docs.opencv.org/3.1.0/d7/d8b/tutorial_py_face_detection.html, svobodnyy. - Zagl. s ekrana (data obrashcheniya: 20.05.2017).

9. Lychkov I.I., Alfimtsev A.N., Devyatkov V.V. Otslezhivaniye dvizhushchikhsya ob"yektov dlya monitoringa transportnogo potoka // Informatsionnyye tekhnologii i sistemy: materialy 34-y konf. molodykh uchenykh i spetsialistov IPPI RAI, 2-7 okt. – 2011. – pp. 31.

Authors:

Zakharova Alena Aleksandrovna, Professor, Doctor of Technical Sciences, School of Computer Science & Robotics of National Research Tomsk Polytechnic University (30, Lenin Avenue, Tomsk, Russia)

Nebaba Stepan Gennad'yevich, engineer, Candidate of Technical Sciences, School of Computer Science & Robotics of the National Research Tomsk Polytechnic University (30, Lenin Avenue, Tomsk, Russia)

УДК 519.6

Применение метода вычитания фона и алгоритма Виолы-Джонса для увеличения скорости трекинга лиц в системе распознавания личности[*]

С.Г. Небаба, А.А. Захарова^{**}

Национальный исследовательский Томский политехнический университет, Томск, Российская Федерация

В статье рассмотрен существующий метод автоматического сопровождения движущихся объектов в видеопотоке в режиме реального времени применительно к задаче распознавания личности человека по изображению его лица. Предложен метод сопровождения лиц в видеопотоке, основанный на комбинации метода вычитания фона и алгоритма Виолы-Джонса для обнаружения области лица в кадре и снижающий требования к вычислительным ресурсам систем распознавания личности. Представлены результаты тестирования работы предложенного алгоритма сопровождения лиц в видеопотоке.

Ключевые слова: компьютерное зрение; обнаружение лиц; распознавание лиц, трекинг; алгоритм Виолы-Джонса; вычитание фона; обработка видеопотока в режиме реального времени

Авторы:

Захарова Алена Александровна, профессор, доктор технических наук, Инженерная школа информационных технологий и робототехники Национального исследовательского Томского политехнического университета (Россия, Томск, проспект Ленина, 30)

Небаба Степан Геннадьевич, инженер, кандидат технических наук, Инженерная школа информационных технологий и робототехники Национального исследовательского Томского политехнического университета (Россия, Томск, проспект Ленина, 30)

---

^{**} E-mail: stepan-1fx@mail.ru, zaa@tpu.ru.