Face identification in videos from mobile cameras

Meiru Mu, Luuk Spreeuwers, Raymond Veldhuis
SCS group, University of Twente
Enschede, The Netherlands
{M.Mu,L.J.Spreeuwers,R.N.J.Veldhuis}@utwente.nl

Abstract

It is still challenging to recognize faces reliably in videos from mobile camera, although mature automatic face recognition technology for still images has been available for quite some time. Suppose we want to be alerted when suspects appear in the recording of a police Body-Cam, even a good face matcher on still images would give many false alarms due to the uncontrolled conditions. This paper presents an approach to identify faces in videos from mobile cameras. A commercial face matcher FaceVACS is used to process the face recognition frame by frame. On a video of certain length, in order to suppress the false alarms, we propose to count the recognized identities and set thresholds to the counts, as well as to the matching scores for still-image face recognition. In this way, the facial information of a single subject over time is exploited without implementing face tracking, which is complicated and more difficult for low-quality unconstrained videos. For experiments, videos are recorded by two type of mobile cameras, which provide different video qualities. The results demonstrate the efficiency of our proposed approach.

1. Introduction

Face recognition in the context of visual surveillance application has been a topic of growing interest in computer vision. The object of our work is to develop a suitable approach to face recognition from mobile police cameras, which is expected to warn the holder of the camera whether someone on a blacklist is spotted by the camera. Currently, the mobile cameras on policemen or police cars are primarily intended for the recording of events. The application we are considering to supplement in this work can be addressed as recognizing targeted subjects on a list from a sequence of uncontrolled video frames by face recognition. Figure 1 illustrates the considered system framework. Before face identification, face extraction and feature extraction (or part of them) can be executed on mobile camera, or the complete video recording needs to be forwarded to a central server to process. It depends on the camera processing power and the available bandwidth of network connection. An alert is expected to be transferred to the camera when some target subject is identified. The extracted faces of one subject from a long-term video frames can be represented a set of vectors. The faces of targeted subjects on watch list, saved as mug-shots, can be respectively represented as one vector. Accordingly, we address our considered video face recognition issue as a set-to-one similarity measurement problem, or query set and still image matching issue as some other papers present [1]. The faces of a same subject from the long-term observations vary in scale, pose, illumination, and expressions etc. In addition, because of the moving camera, the frame images are of poor quality due to the compression and of heavy motion blur. In this case, we highlight the identity similarities, i.e. determine whether the face set from video and some face from mug-shots are of the same subject, while ignoring the image similarities.

In the context of face recognition from a set of images obtained over time, how to integrate information from multiple observations has been well studied [2]. The possible schemes for integrating information include selecting the best observation from each set or simply averaging all observations prior to classification, and constructing some statistical models from multiple observations to compare the existing face models. However, it is challenging to first...
Figure 2. Illustration of the proposed idea for assisting identification determination on video, assuming that the gallery set consists of \{A - Z\} 26 mugshots and multiple faces are extracted from each frame and rank-1 identified for faces. By counting the resulting identities and ordering the counts, subject A is suspected in the video.

achieve those face observations of the same subject to integrate information, given each video frame includes several faces of multiple subjects. Generally, face detection and tracking are needed for that [3]. But, for the common surveillance videos from mobile camera as we consider in this paper, face tracking is extremely challenging due to the poor image quality and unconstrained subjects behavior.

In this paper, we present an approach to face recognition on surveillance videos from mobile camera. Each frame includes multiple subjects. We highlight the measurement way of identity similarity between the extracted faces from a sequence of video and the stored mug-shots. The face tracking is not involved, which is usually complicated in this considered scenario. The face extraction and identification algorithms are carried out by a commercial face matcher FaceVacs [4], since the face recognition scheme based on still images is out of our discussion. Section 2 presents the proposed approach. Experimental results are given in Section 3. Section 4 is about some conclusion and discussion.

2. Identity determination in video

So far there have been mature automated face recognition technologies available in still images. However, it is challenging to recognize faces in online videos. In case of identifying suspects to assist law enforcement, the unconstrained videos usually provide images of low-quality. If we carry out the single-image face matching frame by frame, the false matching rate could be significantly high. Suppose we set an alarm when the query face is matched with someone in a background set of mugshots, probably we will receive many false alarms. But for a video of certain frames, a simple fact is that a single face may appear in many frames. And if we frequently receive alarms of a same subject, then chances are higher for this subject to genuinely appear in the video. Figure 2 illustrates this idea by assuming that the gallery set consists of \{A - Z\} 26 mugshots and five subjects simultaneously appear on one frame. Faces are firstly extracted frame by frame and then face matchings on still images are carried out. By analyzing the rank-1 identification results, one can guess that chances are the most higher that subject A appearances in the video.

Intuitively, it is not a trustable alarm by only counting the identification results from all the extracted faces in a video. The matching score, which indicates the similarity of two faces, should play an important role for identity determination in the same way as that for classical face recognition task on still images. Figure 3 illustrates how we make a final identity determination by setting thresholds on the identity counts and matching scores. The involved processing steps can be listed as following:

1. Given a video of certain frames, faces are extracted from frame by frame. The rank-1 face identification is subsequently carried out by some mature face matching technology. It provides each face a maximum matching score and its corresponding identity. It should be noted that there are some non-faces resulting from face extraction due to the unavoidable false face detections. Some extracted faces are of badly poor quality. The commercial face matcher FaceVacs which we use for experiments filters them out before feature extraction by implementing eyes detection. As it shows in Figure 3, the identification results of some faces are denoted by #.

2. For each single face, threshold \( T_1 \) is set to reduce the false identification rate. And those low-quality faces and non-faces, which don’t get identification results from the face matcher, are removed from the list. Then the number of each recognized identity are counted, and their corresponding matching scores are group into one score by maximum.

3. Threshold \( T_2 \) is set to the grouped maximal score, and threshold \( T_3 \) is set to the percent of identity counts. Note that thresholds \( T_1 \) and \( T_2 \) are depended on the quality of single images and of course the face matching algorithm itself. While the threshold \( T_3 \) is more depended on the video content. They are adjustable as the input videos vary and the applied face recognition algorithm differs.

3. Experimental results

Video recording data and face matcher: The videos for our experiments are recorded by two types of Zepcam, which are respectively ChestCam and ShoulderCam. The Zepcam, developed and manufactured in Netherlands, consists of a recorder unit, a camera, and a wireless remote control, which has been used in many European police forces as well as security organizations [5]. We suppose that the policeman starts a recording when he finds someone who is
Figure 3. Illustration of our proposed approach. Input: video data. Output: determined identity by supposing subjects \{A, B, C, D\} are on the watch list. Face extraction and identification, which are carried out by a commercial face matcher 'FaceVacs', provide a list of identities and corresponding scores from all of extracted faces. For the final identity determination, the list are processed by thresholds.

Thresholds: For finding out the suitable thresholds \{T_1, T_2, T_3\} as introduced in Section 2, the distributions of genuine and impostor scores, which are from grouping the list by maximum, are analyzed. Figure 4 plots these score distributions by taking a video from ChestCam as example. As it shows, the genuine scores are primarily ranged in \([0.2, 0.6]\), and the impostor scores are mainly in \([0.2, 0.4]\). Accordingly, \(T_1 = 0.2\) and \(T_1 = 0.4\) are determined. For determining \(T_3\), the percentages of recognized identity counts are calculated. The right axis, as Figure 5 shows, plots the percentages we get. Based on those plots of count frequencies, \(T_3 = 0.1\), i.e. \(T_3 = 10\%\) is determined. It should be noted that there are more videos collected in our experiments to investigate the suitable thresholds. \(\{T_1 = 0.2, T_2 = 0.4, T_3 = 0.1\}\) are feasible for a moderate video condition. They are adjustable as the video quality varies.

Identity determination results: Based on the three thresholds we can easily filter out most of the false recognition resulting from still image-to-image matching. Figure 5 displays the identity determination results by thresholds, in which the red line shows the identities to stimulate alerts making. For example, we are alarmed that there are three suspects in the video ChestCam_Group1 and one suspect in the video ShoulderCam_0276. Assuming without the count percentage threshold \(T_3\) given, the video ChestCam_Group1 and ShoulderCam_0276 will both make four alarms, as shown by the blue makers above the blue dash line.

4. Conclusion and discussion

In this paper, we present an approach to automatically determine the subject’s identity over a surveillance video from mobile camera, under the assistance of face recognition technology. By setting a threshold to the counts frequency of recognized faces as well as two thresholds to the matching scores, the false alarms are suppressed. The
alert of final determined identity is more reliable. The results demonstrate the simplicity and reliability of the proposed approach. Besides, since there is no face tracking involved, the approach can be efficiently implemented in real-time applications. In order to further decrease the false alarms for our considered application scenario, the optimal face matcher may not be FaceVACS. And for enhancing the reliability of determined identity by our proposed approach, some guidance on the video recording for users would help. These issues could be our future work. A demonstration system working on it, including mobile camera recording, wireless connection, face recognition scheme would help looking for a better algorithm solution for more realistic application.

References


Figure 4. Plots of genuine and impostor scores by taking a video from ChestCam as example. Three subjects {195,200,203} in gallery set appear in the video, and one mistakenly recognized identity 61 is randomly chosen to display the distribution of impostor scores.

Figure 5. Results display by taking two videos as examples, which are respectively from ChestCam and ShoulderCam.