Dynamic Fusion of Gait and Face Biometrics

K. Bhattacharya, S. Sarkar, D. Goldgof, R. Kasturi
Computer Science and Engineering
University of South Florida, Tampa, FL 33620

Abstract

Automatic analysis of videos using at-a-distance biometric signatures is required for actual operation in typical surveillance scenarios. We propose techniques to automatically determine what at-a-distance biometric can be obtained at which part of a random trajectory. It is easier to extract gait signature from certain portion of the trajectory than others. Face is a good biometric for frontal views. These selected biometric sources at each time instant are then fused using sequential probability ratio test (SPRT) to arrive at a fused verification probability as a function of time. We compare this sequential scheme for fusion with commonly used static strategy based on the sum of the individual. Initial results show significant improvements in verification performance using fusion of biometric classifiers, even on hard datasets.

1. Introduction

Verification performance based on a single biometric is often poor especially in outdoor situations. To overcome problems with a single biometric, biometric classifiers are often fused to boost performance. However, fusing different modalities into a recognition system imposes different requirements for the video data. For example in a typical video surveillance scenario, gait (the manner of walking of a person) biometric [3] is available when the subject is walking fronto-parallely, while face biometric [7] is available when a frontal view of the subject is available. In this paper, we propose a novel approach to automatically select biometric cues which in a video surveillance scenario is to determine whether gait, face or no-biometric is available from each frame of a video sequence of a person walking along a complicated trajectory. To boost performance of a individual classifiers, we combine a gait classifier with a classifier using face as a biometric. We also propose a sequential fusion scheme to boost verification performance.

Biometric fusion strategies employed so far can be classified as being static where individual classifiers are simultaneously available. In [2], authors demonstrate that multi-biometric integration does indeed result in a consistent performance improvement. One can consider inter-modal combination, e.g., combination of face with iris, and intra-modal combination, e.g., combination of outputs of two classifiers on the same modality, or the combination of outputs of two different sensors. For outdoor video surveillance applications one can choose to combine face and gait biometrics. Very recently biometric fusion of gait and face classifiers for video surveillance has been of interest to the research community. In [5], EBGM was used as a face classifier and spatio-temporal correlation was used as a gait classifiers. They showed that even simple fusion schemes can lead to improved recognition performance for outdoor data taken months apart. In [9], a image based visual hull is constructed for canonical virtual views for recognition. For face recognition virtual cameras are placed to capture frontal face appearance and virtual cameras were placed to capture side-view gait recognition. In [4] hierchical and holistic fusion of gait and face classifiers is suggested. The first involves using the gait recognition algorithm as a filter to pass on a smaller set of candidates to the face recognition algorithm and the second involves combining the similarity scores obtained individually using simple combination rules. They, however, do not address the issue of how to select the biometric cues for any random trajectory portion. In this paper we will focus on how to extract such biometric cues like gait and face from arbitrary video sequences and explore sequential fusion schemes for face and gait biometric classifiers for outdoor settings at high distances.

2. Dataset used

The dataset that we used for experimentation is from the NIST sigma trajectory database. The database consists of 29 subjects, 2 views and 4 facial stills per subject. The database has arbitrary trajectory and severe viewpoint changes. It also has severe changes in distances of the subject from the camera, for example, at the farthest points we have only 50 pixels on the subject. Such hard databases are excellent for pushing the technology limits, as is true in our case. Sample frames from this dataset is shown in Figure 2 with corresponding tracks overlaid.
3. Segmentation

Our first step is to estimate the background and to automatically estimate the silhouettes. We estimate the background by simply taking the mean of all frames in the video. Next we compute the difference image for each frame with respect to the estimated background image. The frame differences are then adaptively thresholded with predefined windows and subsequently a morphological closing operation is performed to close small gaps in the subject’s thresholded body parts. Next we perform a connected component analysis over the resulting binary images and the largest connected component is considered as belonging to the subject. The bounding boxes around the subject for each frame is provided by considering the co-ordinates of the left and right extremities for the largest connected component. The silhouettes are then refined by the EM method.

4. Trajectory Analysis

Our task is to find which part of the trajectory has good gait signature and which part has a good face signature, and in which part no biometric signature can be obtained at all. Obviously all portions of the video cannot be used to extract gait and face signatures. For e.g., side views are good for gait and frontal views are good for face. The detailed trajectory in the Sigma database is shown in Figure 2.

To select the biometric cues we approximate the trajectory of the bounding box centres as a piecewise linear model. We use a "split-and-merge" recursive line fitting algorithm for this. The algorithm starts off by estimating a single straight line between the two extremities of the trajectory. If the distance of the farthest point on the trajectory from this straight line is greater than a tolerance limit the initial straight line is divided into two straight lines. The procedure is repeated recursively until the distance of farthest point falls below the tolerance limit. Based on the slope of these fitted straight lines we classify all frames whose bounding box centres are approximated by the line to be proving a particular biometric cue (or no cue).

5. Biometric signatures: Face, Gait

Having knowledge of which biometric is available (or unavailable) in each frame the next step is to extract the biometric signature to be used for verification. We have chosen silhouettes of the moving person as the gait signature for identification. To extract gait signature from a sequence of frames, classified as containing a gait biometric by our trajectory analysis algorithm, the adaptively thresholded images after background subtraction is considered and the portion given by the estimated bounding boxes are cropped and scaled to 128 × 88 size templates using bi-linear interpolation. These silhouette templates are used as features in the gait classifier for which we selected the gait baseline algorithm [8].

For the frames where the identifying biometric classified is a face we still need to locate the exact face portion in the bounding box. For this we select the top one-third of the bounding boxes and use a pretrained Haar face detector [10] to detect the faces. The detected portion is cropped and the cropped rectangular portion of the face image is then preprocessed for aligning the face imagery. In order to do this the exact co-ordinates of the eye in the cropped face rectangle is provided manually. The cropped faces are used as features for the face classifier for which we have chosen CSU’s PCA based classifier [1].

6. Static fusion of biometric classifiers

One obvious strategy for fusion is decision(score) level fusion of biometric classifiers. Before performing a decision level fusion of the scores however, we needed to perform a score normalization. The scores are normalized...
using \( (data - mean)/std\_dev \) and then transformed into range \([0 – 1]\) using min-max normalization. Normalized scores were fused. We have experimented with different combination strategies, i.e., Min, Max, and Mean. If we use the score of the combined classifier as the minimum of normalized scores from a classifier, a conservative approach is followed while taking the maximum of the normalized scores provides a liberal approach.

7. Dynamic fusion of biometric classifiers

We also propose a dynamic fusion scheme using the sequential probability ratio test (SPRT). SPRT was first proposed by Wald [11] and is particularly attractive to apply when the underlying data collection is sequential. SPRT was previously used in verification scenario whether the hypothesis testing is potentially a two-class problem. In [6] SPRT was rephrased into a identification problem using verification as a sub-step and was successfully used for speaker verification. However unlike [6], we don’t train the individual probability distributions from feature vectors, but directly use our normalized scores for individual biometric classifiers as a probabilistic measure. The probability value for frame \( i \) supporting the null hypothesis is the maximum similarity score of all available biometrics in a window starting from this frame. The probability value for frame \( i \) supporting the alternate hypothesis is chosen to be the complement of the mean similarity score of all available biometrics in that frame. Thus for frames 1 to \( t \), if \( P_t \) is the mean score of all biometrics available at time \( t \), the likelihood ratio is defined as:

\[
\lambda_t = \lambda_{t-1} \frac{P_t}{1 - P_t}
\]

As we slide the window the cumulative confidence score builds up. If we consider the number of frames in each video is \( T \) and the number of subjects in the database is \( S \), with \( \lambda_A \) and \( \lambda_B \) as the two constants associated with SPRT, our algorithm is as follows:

for \( t = 0 \) to \( T \) do
    for \( s = 0 \) to \( S \) do
        1. Calculate likelihood ratio.
        2.a If likelihood ratio is greater than \( \lambda_A \), identify as subject is \( s \) and quit.
        2.b If likelihood ratio is less than \( \lambda_B \), remove \( s \) from \( S \) and continue.
        2.c If likelihood ratio is less than \( \lambda_A \) and greater than \( \lambda_B \), no decision is made for subject \( s \) and continue.
    end for
end for

This algorithm returns the identified subject by sequentially accessing the frames until the confidence reaches above a certain threshold. When the likelihood falls below a certain threshold it is decided that the subject \( s \) certainly does not identify the subject to be verified.

8. Results and performance evaluation

In this section we evaluate the performance of our biometric system. For gait the gallery was chosen as the fronto-parallel trajectory segment (track A) from the left camera. Thus gait biometric is matched across views. The face gallery consisted of outdoor facial stills provided with the Sigma database. The morphological structuring required for detecting the bounding boxes was set to isotropic element with size \( 5 \times 5 \). The various steps for the automatic background estimation and bounding box detection for a sample video is shown in Figure 3. As described previously our trajectory analysis system was based on a “split-and-merge” line fitting algorithm, the tolerance limit for which was set to 8 pixels.

We have experimented to see how the confidence in the decision builds up as the subsequence window slides across the trajectory for our dynamic fusion technique using SPRT. The window was chosen to be of length 100. If the window is contained entirely inside a track with a particular biometric only the similarity score for that biometric is used and the confidence score reinforces or drops depending whether the normed score is greater or equal to 0.5. For a no-biometric track the confidence score is provided to be 0.5 and hence as can be seen in Figure 4 the confidence remains constant when the window is entirely inside a track with no biometric cue. When the window overlaps with two tracks which provides different biometric cues the mean score is used. The plots for different subjects are characterised by different line styles. \( \lambda_A \) and \( \lambda_B \) was chosen to be 6 and 0.25 respectively. We summarize the verification performance of individual classifiers and the fused classifiers using different fusion schemes in Figure 5(a). We experimented with all 3 static fusion strategies (min, max and mean) and mean gave the best results. We found, however, SPRT based dynamic fusion scheme clearly outperforms static fusion schemes. As shown in Figure 5(b), in most cases verification was achieved by sampling the first 350 - 450 frames of a probe video sample.

9. Conclusions

We presented a sequential, dynamic scheme for the fusion of gait and face biometrics. Given a sequence of frames of a person walking along an arbitrary track, we decide which biometric source is available and then fuse the information to boost verification performance. As frame data arrives the fusion algorithm incrementally update the running verification score. We compare the performance of this dynamic scheme with the commonly used static fusion scheme.
and showed that it results in a somewhat better performance on the NIST-Sigma dataset of outdoor video sequences.

References