Human Identification using Gait and Face

Rama Chellappa
Department of Electrical and Computer Engineering and
Center for Automation Research
University of Maryland at College Park
College Park, MD 20740

Amit K. Roy-Chowdhury
Department of Electrical Engineering
University of California Riverside
Riverside CA 92521

Amit Kale
Siemens Information Systems Ltd.
Bangalore, India

1. Introduction

Identification of humans from arbitrary view points is an important requirement for different tasks including perceptual interfaces for intelligent environments, covert security and access control etc. Different modalities can be used for identification based on the number of pixels on the individual. If the person is far away from the camera, it is hard to get face information at a high enough resolution for recognition tasks. However when available, it yields a very powerful cue for recognition. A modality which can be detected and measured when the subject is far away from the camera is human gait or the style of walking. For optimal performance, the system must use as many cues as possible and combine them in meaningful ways. Information may be fused in two ways. The data available may be fused and a decision can be made based on the fused data (data fusion) or each signal/feature can be matched separately, using possibly different techniques and the decisions made may be fused (decision fusion).

The gait of a person is best reflected when he/she presents a side view (referred to in this chapter as a canonical view) to the camera. Hence, most gait recognition algorithms rely on the availability of the side view of the subject. For doing face recognition, on the other hand it is desirable to have frontal views of the person's face. The most general solution to perform integrated face and gait recognition from arbitrary views would be to estimate 3-D models for face and gait. While there has been some progress in building 3-D models for faces, the problem of building reliable 3-D models for articulating objects like the human body still remains a hard problem. One way to exploit current recognition algorithms for frontal face and side gait without resorting to 3-D models is to synthesize canonical views, given arbitrary views of the person. In [1], Shakhnarovich

et al. compute an image based visual hull from a set of monocular views which is then used to render virtual canonical views for tracking and recognition. Gait recognition is achieved by matching a set of image features based on moments extracted from the silhouettes of the synthesized probe video to the gallery. The visual hull is also used to render frontal face images. Eigen faces are used for face recognition. They also studied the fusion of face and gait cues for this multi-camera indoor environment. Zhou and Bhanu [2] present an approach to fusion of gait and face which uses side views of non cooperating subjects, making use of PCA and MDA for gait and face features, and combining them at the score level.

2. Overview of Approach

In general the visual-hull approach for performing integrated face and gait recognition requires at least two cameras. In this paper we present experimental results for fusion of face and gait for the *single* camera case. We considered the NIST database which contains outdoor face and gait data for 30 subjects. In the NIST database, subjects walk along an inverted Σ pattern. In [3], we presented a view-invariant gait recognition algorithm for the single camera case along with some experimental evaluations. In this chapter we present the results of our view-invariant gait recognition algorithm in [3] on the NIST database. The algorithm is based on the planar approximation of the person which is valid when the person walks far away from the camera. In [4], an algorithm for probabilistic recognition of human faces from video was proposed and the results were demonstrated on the NIST database. Details of these methods can be found in the respective papers. We give an outline of the fusion strategy here.

We employ decision fusion to combine the results of our

gait recognition algorithm and the face recognition algorithm. We consider two fusion scenarios: hierarchical and holistic. The first involves using the gait recognition algorithm as a filter to pass on a smaller set of candidates to the face recognition algorithm.The second involves combining the similarity scores obtained individually from the face and gait recognition algorithms

To improve efficiency of a multimodal biometric system, one can adopt multistage combination rules whereby subjects may be coarsely classified by a less accurate classifier, passing a smaller set of likely candidates to a more accurate classifier. The results of the gait classifier, for example, can be used to pass a smaller number of candidates to the more accurate face recognition unit. Alternatively, decisions from the different classifiers can be combined directly using simple rules like SUM, PRODUCT etc.. In this case it is first necessary to transform the scores obtained from the different classifiers in order to make them comptible. The transformation should be such that the relative ordering of the scores is not altered. In other words the transformation function should be monotone. Some of the commonly used transformations include linear, logarithmic, exponential and logistic. The purpose of these transformations is, first, to map the scores to the same range of values and second, to change the distribution of the scores. For example, the logarithmic transformation puts strong emphasis on the top ranks, whereas the lower ranked scores which are transformed to very high values, have a quickly decreasing influence. A detailed discussion of score transformation is given in [5] in the context of combining classifiers for face recognition. The face recognition algorithm yields a match score which is a probability while the gait recognition algorithm yields a distance measurement. In order to make the scores comparable before fusing them, we apply an appropriate transformation to the gait scores discussed in the experimental section. Note that the score transformation is necessary only when the scores of the face and gait recognition algorithms are to be directly combined.

3. Experimental Results

We now present the results of the two fusion strategies for face and gait cues. As mentioned before, in order to combine the scores from the face and gait classifiers directly, it is necessary to make them comparable. We used the exponential transformation for converting the scores obtained from the gait recognition viz. given that the match score for a probe X from the gallery gaits is given by $S_{X1}, \cdots, S_{XN}$ we obtain the transformed scores $exp(-S_{X1}), \cdots, exp(-S_{XN})$. Finally we normalize the transformed scores to sum to unity. We also tried logistic and logarithmic score transformation methods. The results obtained using these were comparable to the exponential case. In Figure 1, we show results of our fusion strategy us-
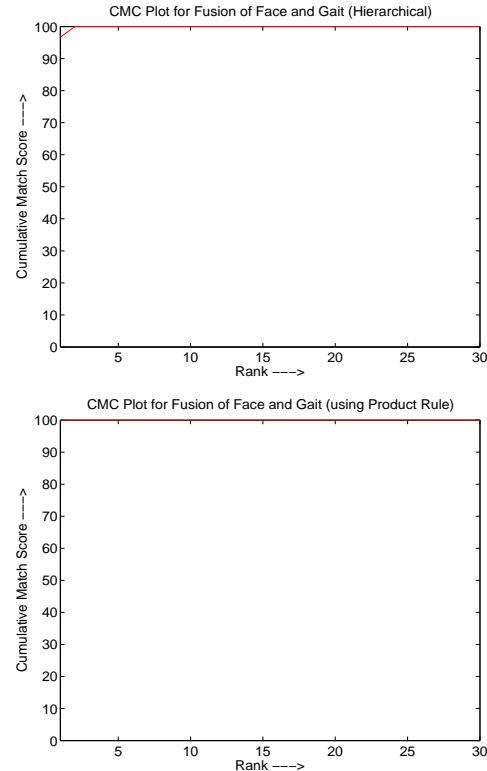

Figure 1. CMC curves for (a) Hierarchical and (b) Holistic Fusion

ing face and gait on the NIST database which consists of 30 people walking along an inverted Σ -shaped walking pattern

References

[1] G. Shakhnarovich and T. Darrell, "On probabilistic combination of face and gait cues for identification," *Proc. of the IEEE International conference on Face and Gesture*, 2002. 1

[2] X. Zhou, B. Bhanu, and J. Han, "Human recognition at a distance in video by integrating face profile and gait," *Proceedings of AVBPA*, pp. 533–543, 2005. 1

[3] A. Kale, A. Roy-Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," *Proceedings of IEEE AVSS*, pp. 143–150, 2003. 1

[4] S. Zhou, V. Krueger, and R. Chellappa, "Probabilistic recognition of human faces from video," *CVIU*, vol. 91, pp. 214–245, 2003. 1

[5] B. Achermann and H. Bunke, "Combination of classifiers on the decision level for face recognition," Tech. Rep., Institut fur Informatik und angewandte, Mathematik,, Universitat Bern, 1996. 2