Compression of Integrated Gait and Face Recognition

Dr.R.S.Rajesh*, Prof T.Krishnaswamy*, Chettiyar Vani Vivekanand

*Reader, M.S. University, Tirunelveli, Tamilnadu, India, rs_rajesh1@yahoo.co.in

*Professor, Noorul Islam College Of Engg, KK dist, Tamilnadu, India, tkshodece@yahoo.co.uk

0PG Student, Noorul Islam College Of Engg, KK dist, Tamilnadu, India, vani_anjali02@yahoo.co.in

Abstract—Biometrics are methods to automatically recognize a person based on a physiological or behavioral characteristic. Examples of human traits used currently for biometric recognition include fingerprints, speech, face, gait and handwritten signature. For optimal performance, the system must use as many cues as possible and combine them in meaningful ways. In this paper we propose a method for classifier combination for face and gait recognition, and demonstrate both improved performance and better statistical justification for the integration step. We compare MAX, MIN, MEAN, MEDIAN and PRODUCT rules for combining classifier outputs. The results of fusion experiments are demonstrated on the NIST database, which has outdoor gait and face data of 30 subjects. We propose to use optimal wavelet for image compression, given the number of most significant wavelet coefficients to be kept. Simulated Annealing is used to find the optimal wavelet for the given image to be compressed.

I. INTRODUCTION

One of the main goals of computer vision research is to develop methods for recognition of objects and events. A subclass of these problems is the recognition of humans and their activities. Different modalities can be used for identification based on the number of pixels on the individual. If the person is far 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. 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).

Gait recognition is related to the broader problem of human motion modeling, which has very important implications for different areas like surveillance, medical diagnosis, entertainment industry, video communications, etc. Gait recognition is a relatively new area to computer vision researchers. However, significant progress has been made and reasonably good performance on large datasets under controlled circumstances has been achieved. The problems facing this area are poor performance in uncontrolled outdoor situations and the effects of time. The problem in face recognition can be defined as follows. Given an image or a set of images of an individual (known as test images), the problem is to identify the individual from the gallery or decide that he/she is not part of the gallery. The main challenges in face recognition are:

- Varying conditions of illumination between training and test images.
- Different poses of the face in different instances of recording.
- Changes in appearance, make-up and clothing between training and test images.

All of these issues make face recognition an extremely challenging problem. To overcome all these difficulties and to improve the recognition, the face and gait recognition techniques are integrated. It was also shown that the combination of face and gait cues provided slightly better recognition results than either modality alone.

The recognized data are compressed successfully using wavelets. However, for the given image, the choice of the wavelet to use is an important issue. In this paper, we propose to use optimal wavelet for image compression, given the number of most significant wavelet coefficients to be kept. Simulated Annealing is used to find the optimal wavelet for the given image to be compressed. In Simulated Annealing, we need a cost function to minimize. This cost function is defined as the mean square error between the decompressed image and the original image. Wavelets have been successfully used in image compression, to minimize the storage space when used in security applications. It can also be helpful in minimising the bandwidth required for transmission.
II. GAIT RECOGNITION

We present a robust representation for gait recognition that is compact, easy to construct, and affords efficient matching. Instead of a time series based representation comprising of a sequence of raw silhouette frames or of features extracted therein, as has been the practice, we simply align and average the silhouettes over one gait cycle. We then base recognition on the Euclidean distance between these averaged silhouette representations.

A. Averaged Silhouette Representation

The first step is silhouette extraction in each frame based the distance from the background pixel statistics[1]. We compute the background statistics of the RGB values at each image location, \((x, y)\), in terms of the mean \(\mu_B(x, y)\) and the covariances \(\Sigma_B(x, y)\) of the RGB values at each pixel location. Using the distance of a pixel value as the observation, pixels are classified into foreground or background using Expectation Maximization (EM) with a Gaussian mixture model. We found that the process stabilizes within 15 or so iterations. Fig. 1 shows some example silhouettes.

![Example silhouettes](image)

Fig. 1. The first and second rows show samples of the binary silhouettes over one gait cycle, for two subjects, respectively. The third row shows the averaged silhouettes for the subject in the second row; each averaged over a different gait cycle.

The second step is to estimate the gait periodicity, \(N_{gal}\). We simply count the number of foreground pixels in the silhouette in each frame over time, \(N_f(t)\). This number will reach a maximum when the two legs are farthest apart (full stride stance) and drop to a minimum when the legs overlap (heels together stance). To increase the sensitivity, we consider the number of foreground pixels mostly from the legs, which are selected simply by considering only the bottom half of the silhouette [3] and [4]. The third step is average silhouette computation. Given a sequence of silhouettes, \(S = \{S(1), \ldots, S(M)\}\), we partition it into subsequences of gait period length, denoted by \(S_{gal} = \{S(k), \ldots, S(k + N_{gal})\}\). For each subsequence we average the silhouettes to arrive at a set of average silhouettes, \(AS(i), i = 1, \ldots, [M/N_{gal}]\).

\[
AS(i) = \frac{1}{N_{gal}} \sum_{k=i}^{(i+1)N_{gal}-1} S(k)
\]

(1)

Fig. 1 shows examples of the average silhouette representation for a sequence. Note that this representation implicitly captures the shape of the template and, to a lesser extent, the temporal dynamics of gait. The time spent at each stance shows up indirectly as intensity in the average silhouette representation.

B. Similarity Computation

For gait recognition, we need to compute the similarity between a given probe sequence and a stored gallery sequence [2]. Let the average silhouettes from a probe and a gallery be denoted by \(\{AS_p(i)|i = 1, \ldots, N_p\}\) and \(\{AS_g(j)|j = 1, \ldots, N_g\}\), respectively.

\[
Sim(AS_p, AS_g) = \text{Median}(\text{min}_{i=1}^{N_p} \| AS_p(i) - AS_g(j) \|)
\]

(2)

Equation 2 indicates that the similarity is defined as the negative of the median of the Euclidean distance between the averaged silhouettes from the probe and the gallery.

III. FACE RECOGNITION

Face recognition is one of the most important research area in computer vision and pattern recognition and has many real applications such as forensic identification, access control and human computer interface. It is a very difficult problem far from well solved since face object usually has various appearances due to aging; pose variations and other environmental factors.

A new algorithm for automatic face recognition is presented: Reduced Image Eigenfaces based on the Eigenface mode, improvement in the recognition percentage.

Face recognition involves the following stages:

- **Training:** it consists in using some mechanism allowing the system to “learn” the faces that make up the training set. The training type used for the learning will depend to a large extent, on the methodology applied for the recognition.
- **Recognition:** it consists in filling the system, with different images of people, expecting to obtain as result a univocal codification way which allows to identify who the person is, or else to determine that the face is not in the knowledge base.
A. Reduced Image Eigenfaces

This paper proposes a new method consisting in transforming face images into smaller ones in order to allow working directly with the covariance matrix instead of using an approximation of it [5].

The training consists of the following stages:

a) Each image \( I_i \) is divided into blocks \( P \times P \) pixels each, \( P \) being the reduction level. Each of them is averaged and an new image \( I'_i \) of \( D \times D \) is computed, with \( D=N/P \), which is obtained by replacing each block with its average.

b) Each image \( I'_i \) is recognized as a vector \( \Gamma_i \) of size \( D^2 \) whose value is built up as the concatenation of each row of the image, thus composing a matrix of \( D^2 \times M \).

c) The average face \( \Psi \) is obtained according to the formula

\[
\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n
\]

(3)

d) The obtained average face \( \Psi \) is subtracted from each of the images \( \Gamma_i \) thus obtaining a new set of vectors

\[
\varphi_i = \Gamma_i - \Psi
\]

(4)

which compose matrix A= [\( \varphi_1, \varphi_2, \ldots, \varphi_M \)] of \( D^2 \times M \).

e) The covariance matrix is obtained

\[
C = \frac{1}{D^2} AA^T
\]

(5)

Of dimension \( D^2 \times D^2 \).

f) The eigenvectors of \( C \) are obtained, which, ordered from greater to lesser according to their corresponding autovalues, make up matrix \( u \).

g) A pattern of image \( i \) is obtained \( \Omega_i^T = [w_1, w_2, \ldots, w_{D^2}] \) where

\[
W_k = u_k^T ( \Gamma_i - \Psi )
\]

(6)

B. Recognition Process

The recognition process tries to find in the image base the one corresponding to the given face, for which its pattern \( \Omega \) is computed using the minimum distance formula given below [6].

\[
d = \min ( \| \Omega - \Omega_i \|^2 )
\]

(7)

where \( d \) is the distance measure.

IV. INTEGRATED RECOGNITION BY GAIT AND FACE

In this paper we investigate different approaches to classifier combination for face and gait recognition, and demonstrate both improved performance and better statistical justification for the integration step. First, we compare the performance of several common data fusion strategies on our task, and develop statistical interpretations of each[7].

A. Classifier Combination Theoretical Framework

We develop a common theoretical framework for combining classifiers which use distinct pattern representations and show that many existing schemes can be considered as special cases of compound classification where all the pattern representations are used jointly to make a decision [7] and [8]. The two main reasons for combining classifiers are efficiency and accuracy.

Consider a pattern recognition problem where pattern \( Z \) is to be assigned to one of the \( m \) possible classes (\( w_1, \ldots, w_m \)). Let us assume that we have \( R \) classifiers each representing the given pattern by a distinct measurement vector. Denote the measurement vector used by the \( i^{th} \) classifier by \( x_i \). In the measurement space each class \( w_k \) is modeled by the probability density function \( p(x|w_k) \) and its a priori probability of occurrence is denoted \( P(w_k) \). We shall consider the models to be mutually exclusive which means that only one model can be associated with each pattern.

Now, according to the Bayesian theory, given measurements \( x_i, i = 1, \ldots, R \), the pattern, \( Z \), should be assigned to class \( w_j \) provided the a posteriori probability of that interpretation is maximum, i.e.

\[
\text{Assign } Z \rightarrow w_j \text{ if } P( w_j | x_1, \ldots, x_R ) = \max P( w_k | x_1, \ldots, x_R )
\]

(8)

The Bayesian decision rule (8) states that in order to utilize all the available information correctly to reach a decision, it is essential to compute the probabilities of the various hypotheses by considering all the measurements simultaneously. This is, of course, a correct statement of
individual sensing modalities are shown in Table 1. The soft NIST database. The equal error rates obtained using the decisions output by the three verification systems were then evaluated on a claimed identity of an individual: face and gait cues [10].

C. Comparison of Classifier Combination Rules

The effectiveness of various strategies is evaluated on a high-order measurement statistics described in terms of joint probability density functions p( x₁,...,xₖ|wᵢ). The validity of the conditional independence assumption was tested by computing the average within class correlation matrix for the data used in decision making. Since the overall dimensionality of the data exceeds tens of thousands, it is impossible to present a full view of the correlations between the measurements of the respective modalities. However, by adopting a visual representation of the correlation matrix, we will be able to look at the correlations at least in a representative subspace of this highly dimensional feature space. This normalisation process produced average within class correlations taking values in the interval [-1, 1]. For display purposes, we have taken the absolute value of these correlation coefficients. The result of this representation of variable correlations is a matrix with all elements on the diagonal equal to unity (displayed as gray level 255) and the strength of correlation between one and zero mapped onto the gray-level scale 255 to 0. The correlation matrix is shown in Fig. 3.

The correlation matrix exhibits a block diagonal structure, which suggests that the observations generated by each modality are class conditionally dependent. Note that the correlations between features from different modalities are considerably weaker than within modality correlations. This applies in particular to the correlations

---

**Table 1: EQUAL ERROR RATES**

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum</td>
<td>0.7</td>
</tr>
<tr>
<td>Product</td>
<td>1.4</td>
</tr>
<tr>
<td>Maximum</td>
<td>12.2</td>
</tr>
<tr>
<td>Median</td>
<td>1.2</td>
</tr>
<tr>
<td>Minimum</td>
<td>4.5</td>
</tr>
</tbody>
</table>

---

The first experiment is concerned with the problem of personal identity verification. Two different sensing modalities of biometric information are used to check the claimed identity of an individual: face and gait cues [10]. The effectiveness of various strategies is evaluated on a NIST database. The equal error rates obtained using the individual sensing modalities are shown in Table 1. The soft decisions output by the three verification systems were then combined using the various classifier combination strategies discussed in Section IV.B.

**Fig. 3. Correlation of face and gait data.**
We believe that the main reason for the poor performance of Table 2 shows the results of different combining schemes. The results presented in Table 1 show the benefits of classifier combination. It is interesting to note that the sum rule outperformed all the other combination strategies and also the individually best expert.

V. EXPERIMENTS

We have used NIST database, which has outdoor gait and face data of 30 subjects. There are two main degrees of freedom in choosing the combination strategy. First, one has to choose the rule for combining multiple faces. Second, one has to choose the rule for combining the modalities. In our experience, the better performing combination rules – PRODUCT and MEAN – were robust to the changes in temporal fusion of faces. Before starting the experiments with the rules discussed above, we tried a simpler rule, which requires both classifiers to agree on a label for a test example; otherwise, the example is rejected. Since the classifier decisions appear to be uncorrelated (Figure 3), we expect the accuracy to be close to the product of the individual accuracies, which is 45%.

The best performance was achieved by the PRODUCT rule: 89% accuracy. When the quantities combined are estimates of the posterior distribution, and under the independence assumption, this rule can be shown to be equivalent to the likelihood ratio hypothesis test, when the joint likelihood of the combined data is considered under different models, and equal priors are assumed.

<table>
<thead>
<tr>
<th>Combining Rule</th>
<th>Classification Rate %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum Rule</td>
<td>98.05</td>
</tr>
<tr>
<td>Max Rule</td>
<td>78.93</td>
</tr>
<tr>
<td>Min Rule</td>
<td>79.00</td>
</tr>
<tr>
<td>Product Rule</td>
<td>89.00</td>
</tr>
<tr>
<td>Median Rule</td>
<td>98.19</td>
</tr>
</tbody>
</table>

Table 2 shows the results of different combining schemes. We believe that the main reason for the poor performance of the MIN and MAX rules is the high degree of overlap of the distributions of correct and incorrect scores for the classifiers. Both rules rely on order statistics and are likely to suffer from the noise in score assignment more than the more robust MEAN and PRODUCT. Another interesting outcome of our experiments is that the Sum rule as well as the median rule has the best classification results.

VI COMPRESSION

Wavelets have been successfully used in image compression, to minimize the storage space when used in security applications. It can also be helpful in minimising the bandwidth required for transmission. However, for the given image, the choice of the wavelet to use is an important issue [11]. In this paper we use optimal wavelet for image compression, given the number of most significant wavelet coefficients to be kept. Simulated Annealing is used to find the optimal wavelet for the given image to be compressed. In Simulated Annealing, we need a cost function to minimize. This cost function is defined as the mean square error between the decompressed image and the original image.

A. Finding optimal wavelet filter

The choice of the wavelet should make a significant difference in the quality of the compressed image. We use Simulated Annealing to adaptively learn the optimal wavelet filter for the given image to be compressed [12]. The term Simulated Annealing is from the physical process of heating and then slowly cooling down a substance to obtain a crystalline structure. The minimum of the cost function corresponds to this ground state of the substance. In an annealing process, the substance melts at a high temperature and it is disordered. As it cools down, it becomes more ordered and approaches a frozen ground state. The Metropolis step is the fundamental procedure of Simulated Annealing. If the change in energy is negative compared to the previous one, then the change is accepted and the system is updated. If the energy is greater than the previous one, then it is accepted with a probability given by the Boltzman factor. This process is continued for many times until a frozen state is achieved. In Simulated Annealing, we need a cost function to minimize. This cost function is defined as the mean square error between the decompressed image and the original image. The wavelet filter length in our experiments is set to 8.

The following steps obtain the decompressed image:
1. Get the image to be compressed.
2. Perform forward 2D discrete wavelet transform on the image to be compressed.
3. Sort the wavelet coefficients in descending magnitude order and keep the first \( K \) coefficients.
4. Set the rest wavelet coefficients to zero.
5. Conduct inverse 2D wavelet transform so that we obtain the decompressed image.
6. Calculate the PSNR in dB for each compressed image.
7. Compare the results of compression with the D8 wavelet for the different coefficients we keep, in terms of PSNR.
B. Experimental results

We conduct some experiments in Matlab by using the test images. Simulated Annealing is used to find the optimal wavelet filter for the given number of wavelet coefficients a cost function. This is defined as the mean square error between the decompressed image and the original image. The wavelet filter length in our experiments is set to 8. Experimental results in Peak Signal to Noise Ratio (PSNR) are shown in Table 3. The PSNR is defined as

\[
\text{PSNR} = 10 \log_{10} \frac{\sum_{i,j} (B(i,j) - A(i,j))^2}{n^2 \times 256^2}
\]

where \( n^2 \) is the number of pixels in the image, \( B \) is the decompressed image and \( A \) is the original image. We can see that optimal wavelet outperforms D8 for all experiments. In some case, we get nearly 0.6dB improvement over D8 by using optimal wavelet. Table 3 indicates that the choice of the wavelet indeed makes a significant difference in image for compression.

<table>
<thead>
<tr>
<th># of wavelet coefficients</th>
<th>D8 PSNR (dB)</th>
<th>Optimal Wavelet PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>22.77</td>
<td>22.94</td>
</tr>
<tr>
<td>1000</td>
<td>24.82</td>
<td>25.00</td>
</tr>
<tr>
<td>2000</td>
<td>27.42</td>
<td>27.63</td>
</tr>
<tr>
<td>4000</td>
<td>30.88</td>
<td>31.02</td>
</tr>
<tr>
<td>6000</td>
<td>33.46</td>
<td>33.53</td>
</tr>
<tr>
<td>8000</td>
<td>35.55</td>
<td>35.58</td>
</tr>
<tr>
<td>10000</td>
<td>37.34</td>
<td>37.60</td>
</tr>
<tr>
<td>12000</td>
<td>38.94</td>
<td>39.22</td>
</tr>
</tbody>
</table>

TABLE 3
THE PSNR (DB) OF THE ORIGINAL AND THE COMPRESSED IMAGES WITH DIFFERENT NUMBER OF COEFFICIENTS WE KEEP

V. CONCLUSION

We present a robust representation for gait recognition that is compact, easy to construct, and affords efficient matching. A new algorithm for automatic face recognition is presented: Reduced Image Eigenfaces based on the Eigenface mode, improvement in the recognition percentage.

We have developed an approach to combining visual cues for human recognition, as well as for using multiple instances of face classifications, and demonstrated its performance on the example of integrated face and gait recognition. A number of previously proposed combination rules have been empirically compared. While the combination improved the classification accuracy of the system in almost all cases, a classifier that uses product of the posterior probabilities estimated from different modalities obtained the best performance.

Optimal Wavelets were used for compressing the recognized images by stimulated annealing method to save the bandwidth. It was shown that that the choice of the wavelet indeed makes a significant difference in image for compression.

Interesting future work includes extension of this bimodal recognition scheme to additional modalities thus making more complex rules relevant. Exploring the decision-level fusion in a Bayesian context, with regard to the estimated distributions of the correct and incorrect labels, may lead to a more theoretically sound understanding and design of this process.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Gregory Shakhnarovich and Mr. Trevor Darrel, who developed the classifier combinations on which our experiments were based. We would also like to thank the reviewers for their comments.

REFERENCES