Compression For Distributed Face Recognition

Navid Serrano[1], Antonio Ortega[1], Shuang Wu[2], Kazunori Okada[3], and Christoph von der Malsburg[4]

[1]Department of Electrical Engineering-Systems
[2]Department of Physics and Astronomy
[3]Department of Computer Science
University of Southern California
{navidser,aortega,shuangwu,malsburg}@usc.edu

[3]Siemens Corporate Research, Inc.
kazunori.okada@scr.siemens.com

Abstract

We investigate the application of a face recognition system in a distributed environment. Images of faces are captured by clients remotely and transmitted to a server for recognition or authentication using a central database. In many distributed scenarios, bandwidth may be limited and transmission of image data may not be feasible. We assume the client does not have processing limitations and can extract and transmit compressed features. In this paper we explore the impact of feature compression on face recognition performance. Specifically, we consider the Bochum/USC face recognition system and propose an embedded coding scheme for Gabor-based wavelet features extracted from optimally selected landmarks on the face. Our results show that the impact on recognition rates—even at the highest compression rates—is minimal.

1. Introduction

Distributed systems are fast becoming an important area of research. The paradigm of centralized computing is changing as mobile devices and sensors with enhanced processing power and multimedia capabilities proliferate. The emergence of distributed systems adds flexibility to existing problems and simultaneously poses new challenges. Applications such as user authentication can now be carried out using a multitude of devices in a wide variety of locales but must account for bandwidth limitations.

The bandwidth problem can be mitigated if the client shoulders the processing burden entirely. However, there are scenarios where this is neither possible nor desirable. In authentication systems, for example, it is often necessary to store the database centrally in order to add users dynamically and ensure security. In addition, a central database is easier to maintain. Hence, communication is inevitable and some form of compression is required. In some cases, it may be advantageous for the client to extract and transmit the features instead of the image data. To further preserve bandwidth, the features themselves could be compressed prior to transmission.

Compression for distributed systems is a growing area of research. Applications such as distributed speech recognition [1], distributed image classification [2], and distributed sensor networks [3] have been studied. Yet, compression for distributed face recognition remains unexplored. We are interested, in particular, in a scenario where the client has sufficient processing power but limited bandwidth. Under such an assumption, the distributed system can have a far wider coverage area and clients are guaranteed access from a larger number of locales. Furthermore, we assume the server maintains the database centrally for enhanced security and storage capacity.

Based on the above assumptions, the Bochum/USC face recognition system [4] is a good study case. This system can easily be adapted to a distributed environment because it is based on general principles rather than statistical learning. In fact, classification is based on a simple nearest neighbor distance metric. This implies that faces can be easily added to the database without having to dynamically retrain the classifier. The fundamental part of the algorithm involves an elastic bunch graph technique used to locate specific landmarks on the face from which Gabor wavelet features are extracted.

Prior research has shown that in bandwidth limited cases—where compression is inevitable—the decrease in classification accuracy is less when compressing feature vectors as opposed to the signal itself [1]. Therefore, we attempt to exploit the underlying structure of the Gabor wavelet features as a first step. Although the wavelet features are extracted from selected landmarks on the face, they demonstrate some tendencies common to natural images, including energy concentration in the low-frequency sub-bands. Given these observations, we propose a modified bit-plane coding technique similar to the embedded zerotree wavelet coder proposed by Shapiro [5]. The compression scheme also has the advantage of being fully embedded, meaning that finer renditions of the features are transmitted progressively. We show that the embedded coding scheme can achieve a compression ratio of roughly 6 to 1 while decreasing the overall face recognition rate on average by only 1%. Finally, we show that face recognition performance is considerably better using embedded feature coding compared to image compression using the state of the art JPEG2000 [5] technology.

The paper is organized as follows. We first provide an overview of the Bochum/USC face recognition system, then discuss feature vector compression, including our proposed embedded coding technique, and conclude with an evaluation of the impact of compression on face recognition.

2. The Bochum/USC Face Recognition System

As discussed earlier, the performance of the Bochum/USC face recognition system is based largely on the efficacy of a bunch graph matching technique [6] used to optimally locate specific landmarks on the face. Gabor wavelet features are then extracted from these landmarks and classified using a simple similarity measure. In terms of feature compression, it is noteworthy that the algorithm does not rely heavily on the generalization capability of a classifier engine. Otherwise, distortion resulting from compression could displace vectors in feature space and consequently affect recognition performance. Finally, it should also be noted that the Bochum/USC face recognition system achieved the best performance among competing algorithms in a FERET test administered between September 1996 and March 1997 [7].

Each elastic graph has 48 nodes placed at specific landmarks on the face. The face recognition system addresses pose variations but for our present purposes we only consider frontal views of the face, as shown in Figure 1.

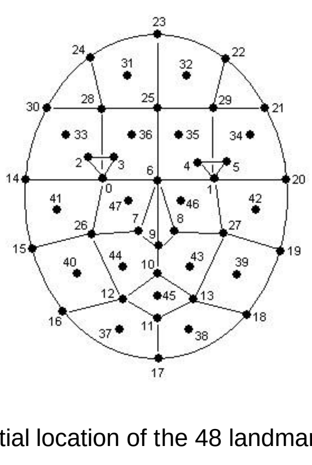

Figure 1. Spatial location of the 48 landmarks on the face.

At each node location, a feature vector, or jet, is extracted using a Gabor-based wavelet expansion:

$$\psi_{\mathbf{k}}(\mathbf{x}) = \frac{\mathbf{k}^2}{\sigma^2}\exp\left[-\frac{\mathbf{k}^2\mathbf{x}^2}{2\sigma^2}\right]\left[\exp(i\mathbf{k}\mathbf{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right], \qquad (1)$$

where $\mathbf{x}$ are the image coordinates, $\mathbf{k}\in\Re^{LD}$ is the wave vector, $l=1,\ldots,L$ is the frequency index, $d=1,\ldots,D$ is the direction index, and $\phi_d=\pi d/D$ is the kernel direction. Let $J_{\mathbf{k}}(\mathbf{x})$ be the Gabor jet extracted at each landmark by convolving the image $I(\mathbf{x})$ with the Gabor wavelet of Eq. (1):

$$J_{\mathbf{k}}(\mathbf{x}) = \int I(\mathbf{x})\psi_{\mathbf{k}}(\mathbf{x}-\mathbf{x}')d\mathbf{x}' . \qquad (2)$$

In the Bochum/USC face recognition system, the Gabor wavelet is sampled over L=5 levels in D=8 directions:

$$J_j = a_j e^{i\phi_j} , \qquad (3)$$

where a_j and ϕ_j, $j=1,\ldots,LD$, are the wavelet coefficient magnitude and phase respectively. A model graph is the collection of Gabor jets (i.e. feature vectors) at each of the N=48 landmarks. Therefore, the Gabor jet at each landmark will have 80 elements (2LD) organized in the following manner:

$$\begin{bmatrix} J(\mathbf{x}^1) \\ \vdots \\ J(\mathbf{x}^N) \end{bmatrix} = \begin{bmatrix} a_1^1 & \phi_1^1 & \cdots & a_{LD}^1 & \phi_{LD}^1 \\ \vdots & \vdots & & \vdots & \vdots \\ a_1^N & \phi_1^N & \cdots & a_{LD}^N & \phi_{LD}^N \end{bmatrix} \qquad (4)$$

The similarity between any two jets J and J' can be computed by:

$$S(J,J') = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}} \qquad (5)$$

The similarity function defined by Eq. (5) is phase insensitive but other similarity measures could be used. The similarity between two faces can be obtained by computing the node-wise similarity between corresponding jets on each face, according to Eq. (5).

3. Feature Vector Compression

One of the challenges in devising a compression scheme for feature vectors is to identify redundancy, irrelevance, and structure. Assumptions can be made about the features used here. In terms of irrelevance, it is reasonable to assume that certain landmarks impact recognition more than others. In such a case, one option would be to apply a coarser quantization to less important landmarks. Another option is to simply eliminate less important landmarks—equivalent to a bit-rate of 0. In fact, the application of principal component analysis to represent the Gabor wavelet features using fewer parameters was studied in [8].

In terms of redundancy, the symmetry of the facial landmarks could be exploited. For instance, it is reasonable to assume that the Gabor jets extracted from the landmarks on the left eye will be similar to the Gabor jets extracted from the landmarks on the right eye. A cursory study of the images in a database of 800 images showed that the features extracted from the left eye were strongly correlated with features extracted from the right eye. On the other hand, the study also showed that most of the other symmetric landmarks—particularly those near the edge of the face—showed little correlation.

Another option is to exploit the underlying structure of the wavelet coefficients. It has been shown that wavelet coefficients extracted from natural images have a strong energy concentration in the low-frequency sub-bands [9]. This trend can be used favorably for compression, as demonstrated by the use of wavelet coding in the new JPEG2000 image compression standard [10]. However, the Gabor wavelet coefficients are only extracted from selected landmarks. Still, our evaluation of the Gabor jets showed that the wavelet coefficient energy was concentrated in the low-frequency sub-bands and furthermore, the energy of coefficients in high-frequency sub-bands tended to be correlated with the energy in low-frequency sub-bands.

3.1. Embedded Coder

The embedded zerotree wavelet (EZW) algorithm was introduced by Shapiro [5]. The EZW algorithm is embedded in that it codes bit-planes in order of importance from the LSB to the MSB plane. The quantization is applied by eliminating wavelet coefficients below a certain energy threshold. The bit-plane coding is accomplished by assigning one of four labels to the wavelet coefficients. If a coefficient is above a threshold T (half the value of the bit-plane), it is assigned a *significant positive* or *significant negative* label, depending on its sign. If the wavelet coefficient is below the threshold and all of its descendants (i.e. the coefficients in higher frequency sub-bands) are also below the threshold, then it is assigned a *zerotree root* label. Otherwise, it is assigned an *isolated zero* label. This means that only four labels are needed to code the coefficients at each bit-plane. Furthermore, if a coefficient is a zerotree root, it does not have to be transmitted at higher bit-planes. The reconstruction value for significant coefficients at each bit-plane is simply $3T/2$

EZW also includes a refinement stage where the reconstructed value of significant coefficients is refined. The refinement is simply $\pm T/4$ and only requires one bit for transmission. Thus, at each bit-plane two bit-streams are transmitted: the significant coefficient bit-stream s and the refinement bit-stream r. Assuming the wavelet coefficients have low-frequency energy compaction, this technique can yield significant savings. EZW is an embedded coder in that the bit-stream is generated and transmitted in order of importance.

In this case, the Gabor wavelet coefficients are self-contained in feature vectors, which makes the typical EZW memory considerations less important. It should be noted that other embedded coders have been developed since EZW and could be applied here but we choose EZW because of its simplicity.

3.2. Modified Embedded Coder for Gabor Jets

We modify the EZW principle to apply it to the Gabor jets extracted using the Bochum/USC face recognition system. The coefficients in the Gabor Jets are infinite precision. Therefore, in order to apply bit-plane coding, they were scaled to 8-bit finite precision. The scaling can be done by normalizing relative to the maximum component of each Gabor jet and then sending this value as side information. Alternatively, the scaling can be done using a global scale factor for each landmark. There is no reason why the coefficients could not be scaled to higher bit representations. The only change would be the number of bit-planes to code.

The first modification to the EZW algorithm is the elimination of the significant negative label since the Gabor jet coefficients are all positive. Another modification is the establishment of parent-child relationships and a scanning order to determine zerotree roots. In natural images, a parent is a wavelet coefficient at any sub-band. Children are coefficients in higher frequency subbands corresponding to the same spatial location in the original image. The scanning is done in zig-zag fashion, according to the traditional QMF-pyramid representation. In the case of the Gabor jets, each of the eight directions is coded separately. We establish parent-child relationships in each direction. The scanning order is from the low-frequency coefficients (l=5) toward the highest frequency coefficients (l=1). Thus, the entry in the bit-stream at bit-plane i for a given wavelet coefficient $a_{l,d}$ is assigned one of three values:

$$s_i = \begin{cases} 1, & a_{l,d} \geq T \\ 2, & a_{l+j,d} < T, \forall j = 0,\ldots,L \\ 3, & a_{l,d} < T \end{cases}, \tag{6}$$

Where, l=1,…,L is the decomposition level, d=1,…,8 is the direction, and the symbols 1, 2, and 3 represent the significant positive, zerotree root, and isolated zero coefficients respectively. The threshold T at each bit-plane i=0,…,7 is equal to 2^i. Example Gabor wavelet coefficients extracted from landmark 19 are shown in Figure 2 (averaged over all images in the database of faces). The decreasing energy trend from low frequency to high frequency coefficients, where high pixel intensities correspond to high wavelet coefficient energy values, can also be seen in Figure 2.

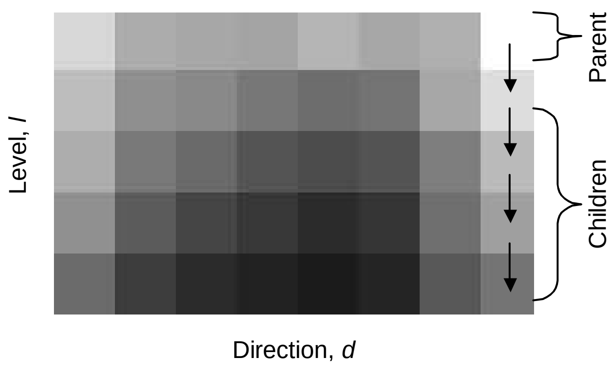

Figure 2. Parent-child dependency and scanning order.

4. Evaluation

We evaluated the proposed compression scheme on a database of 800 faces (frontal view). For each facial image there are 48 Gabor jets. Each jet is coded separately and weighted equally. The embedded coder can be applied to both the magnitude and phase elements but we only consider the magnitude here since the similarity measure of Eq. 5 is phase insensitive. For each bit-plane, the coefficients are assigned one of three labels, zerotrees are established and the entropy of the bit-stream is computed. Since the bit-stream is embedded, the entropy of the bit-stream at any bit-plane b is equal to the sum of entropies of itself and previous bit-planes:

$$H = \sum_{i=0}^{b} \left[\frac{S_i \sum_{j=0}^{2} P(s_i = j) \log P(s_i = j) + R_i \sum_{j=0}^{1} P(r_i = j) \log P(r_i = j)}{S_i + R_i} \right] \quad (7)$$

Here, s_i and r_i are the significant and refinement bit-streams at bit-plane i, respectively, and S_i and R_i are the lengths of s_i and r_i. For an entire model graph, the entropy of each Gabor jet was averaged. The compression ratio CR for the embedded coder is simply:

$$CR = \frac{\sum_{k=0}^{255} P(J_{ij} = k) \log P(J_{ij} = k)}{H} \quad (8)$$

Where, $P(J_{ij}=k)$ is the probability that the ith Gabor jet element of the jth landmark is equal to the 8-bit value k. We use the mean squared error MSE to evaluate distortion of each Gabor jet:

$$MSE = \sqrt{\frac{1}{LDN} \sum_{i=1}^{LD} \sum_{j=1}^{N} \left(J_{ij} - \hat{J}_{ij} \right)^2} , \quad (9)$$

where $\hat{J}_{ij}$ is the ith element of the compressed Gabor jet corresponding to landmark j, LD is the number of wavelet coefficients, and N is the number of landmarks. The MSE is computed over the entire jet, including magnitude and phase reconstruction values.

4.1. Distortion Impact on Classification

As discussed earlier, standard rate-distortion performance measures are not sufficient to gauge the impact of the compression on classification. We thus devised the following experiment using a database of 800 images containing two frontal views of 400 different individuals. One half of the database was left uncompressed and treated as the central server database. The other 400 images were treated as client faces. The model graphs corresponding to the client faces were compressed using the embedded coding technique described above. The similarity was then computed between each compressed model graph and the model graphs of each face in the central database using Eq. 5. The recognition rate was obtained using:

$$P(Match \leq X) = \frac{\text{No. of graph models matched within } X \text{ tries}}{\text{No. of images in the database}} \quad (10)$$

For evaluation purposes we compared the rate-distortion performance of the embedded quantizer to a standard scalar quantizer. We also compared the performance of the feature compression using embedded coding to the performance obtained by compressing the images using JPEG2000 prior to feature extraction.

5. Results

Figure 3 shows the rate distortion performance of the embedded and scalar quantizers averaged over all images in the database.

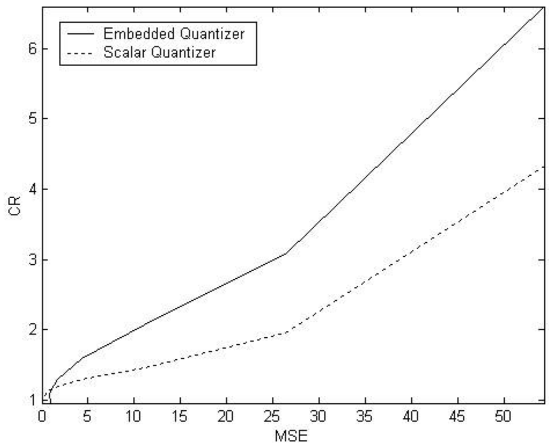

Figure 3. Rate-distortion performance.

Clearly, the embedded coder achieves better compression at equal distortion. This is encouraging in that it shows that the embedded coder is taking advantage of the Gabor wavelet structure. The embedded coder achieved a maximum compression ratio of roughly 6 to 1 compared to 4 to 1 for scalar quantization. Still, these results do not provide any indication of how the incurred distortion affects the face recognition performance. Furthermore, it is unclear what the compression ratio means relative to the original images. The entropy calculated using Eq. (7) measures bits per landmark coefficient. In order to compare the rate performance of the embedded coder to image coding performance, the entropy H was scaled to bits per pixel (bpp):

$$H_{bpp} = H\frac{LDN}{I_1 I_2} \tag{11}$$

Where, LD is the number of wavelet coefficients, N is the number of landmarks, and I_1 and I_2 are the original image dimensions (128 x 128 pixels in this case). Given this scaling, the embedded coder can be compared to the performance of a standard image compression algorithm. We have chosen to use JPEG2000 for this evaluation, as it is the state of the art technology of its kind.

The face recognition rates were evaluated for the three lowest bit-rates in (bits per pixel) using Eq. (10). The recognition rates using the embedded coder are shown in Table 1 compared to the recognition rates using only uncompressed features. The bit-rates shown in Table 1 were averaged over all images in the database.

Table 1. Recognition rates using the embedded coder

$P(Match \leq X)$	Uncompressed	0.13 bpp	0.27 bpp	0.41 bpp
X=1	96.1%	96.1%	96.1%	96.1%
X=2	96.7%	96.1%	96.4%	96.7%
X=5	97.0%	96.1%	96.7%	97.0%

The recognition rates obtained for model graphs extracted from JPEG2000 compressed images is shown in Table 2. The images were compressed at the target bit-rates shown in Table 2, which are very close to those obtained using the embedded coder.

Table 2. Recognition rates using JPEG2000

$P(Match \leq X)$	Uncompressed	0.15 bpp	0.27 bpp	0.41 bpp
X=1	96.1%	40.8%	92.5%	94.3%
X=2	96.7%	45.7%	93.7%	95.5%
X=5	97.0%	53.8%	94.3%	95.8%

As can be seen from Table 1, the embedded coding for feature compression impacts recognition rates minimally. Even at the lowest bit-rate (6 to 1 compression ratio) there is an average decrease in classification performance of only 1%. At the higher bit-rates, the recognition rates are equivalent. The impact of the result is most notable when compared to the performance using JPEG2000 image compression on the images prior to feature extraction. At the lowest bit-rate, the recognition rates are very poor. The classification rates increase significantly for higher bit-rates but are still below the embedded coder rates. The dramatic impact on recognition when compressing images at low bit-rates can be seen in Figure 4, which shows as example an uncompressed face image drawn from the database, together with the compressed images at the target bit-rates shown in Table 2. Clearly, the face in the image compressed at 0.15 bpp is unidentifiable.

The fact that the compressed feature recognition rates on average drop by only 1% compared to the uncompressed features, suggests that the structure of the features is being adequately preserved and large gains in compression can be made without a significant impact in classification performance. Finally, in comparing the results of Tables 1 and 2, the conclusion can be drawn that compressing feature vectors as opposed to images is preferable in distributed classification applications.

Figure 4. Original (a), 0.40 (b), 0.30 (c), and 0.15 bpp (d).

6. Conclusions and Future Work

We addressed compression for distributed face recognition by investigating the impact of feature compression on overall face recognition rates. Given that the Bochum/USC face recognition system employs Gabor wavelet features, we propose using a modified embedded coding scheme. Our evaluation showed that the embedded coder achieves compression rates as high as 6 to 1 with minimal impact on recognition rates (a 1% decrease on average). Furthermore, our evaluation showed that the classification performance is significantly better when compressing feature vectors compared to the classification performance obtained from features extracted from compressed images.

Given these promising results, we believe it will be worthwhile to study variable coding rates for the facial landmarks. The variable rates could be determined using existing knowledge obtained from previous studies of the Bochum/USC face recognition system with parametric linear subspaces [8]. In addition, other coding schemes could be explored such as arithmetic coding.

7. References

[1] N. Srinivasamurthy, A. Ortega and S. Narayanan, ``Towards Optimal Encoding for Classification with Applications to Distributed Speech Recognition," *Eurospeech 2003*, Geneva, Switzerland September 2003..

[2] H. Xie and A. Ortega, "Entropy- and Complexity-constrained Classified Quantizer Design for Distributed Image Classification," *IEEE International workshop on Multimedia Signal Processing,* Virgin Islands, December 2002.

[3] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed Compression in a Dense Microsensor Network," *IEEE Signal Processing Magazine*, March 2002.

[4] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. von der Malsburg, "The Bochum/USC Face Regognition System and How it Fared in the FERET Phase III Test," *Face Recognition: From Theory to Applications*, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman Soulie, T.S. Huang (Eds.). Springer-Verlag, pp.186-205, 1998.

[5] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Trans. Signal Processing,* vol. 41, December 1993.

[6] L Wiskott, J.-M. Fellous, N. Krueger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 19, 1997.

[7] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 22, 2000.

[8] K. Okada and C. von der Malsburg, "Pose-Invariant Face Recognition with Parametric Linear Subspaces," *Fifth International Conference on Automatic Face and Gesture Recognition*, Washington DC, May 2002.

[9] S. G. Mallat, "A Theory of Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, July 1989.

[10] D. S. Taubman and M. W. Marcellin, "JPEG2000: Image Compression Fundamentals, Standards and Practice," Kluwer Academic Publishers, 2002.