HIGH COMPRESSION OF FACES IN VIDEO SEQUENCES FOR MULTIMEDIA APPLICATIONS [*]

Luis Torres   *Daniel Prado*

Technical University of Catalonia, Barcelona, Spain
{luis, aldanip}@gps.tsc.upc.es

ABSTRACT

This paper presents a proposal for a novel video coding scheme intended to encode human faces in video sequences at very high compression using a recognition and reconstruction approach. The scheme is based on the well-known eigenspace concepts used in face recognition systems, which have been modified to cope with the video compression application. An adaptive mechanism is presented to update the eigenspace which improves the overall approach. Results are presented which look promising.

1. INTRODUCTION

Image and video coding are one of the most important topics in multimedia processing and communications. During the last thirty years we have witnessed a tremendous explosion in research and applications in the visual communications field. However, and in spite of all this effort, there are some applications that still demand higher compression ratios than those provided by state of the art technologies.

In particular, and due to its high applicability, there is a need to provide novel compression schemes to encode faces present in video sequences. Although the new standards H.263+ [1] and the synthetic part of MPEG-4 [2] along with other model-based proposed schemes [3] achieve high compression ratios for this particular application, we still believe that further compression is needed, among others, for mobile and video streaming environments.

It is in this context that we present a novel scheme to encode faces in video sequences based on an eigenspace approach. The eigenface concept for still image coding has been already presented in a face recognition framework in [4]. However, to the best of our knowledge, our approach is original and adapts the eigenspace to the video sequence to take into account the different poses, expressions and lighting conditions of the faces.

Section 2 presents an introduction to the topic of very high compression and the basic eigenspace concepts on which our scheme is based. Section 3 presents a fixed eigenspace approach while Section 4 presents the basics of the adaptive scheme. Section 5 presents some results and finally Section 6 draws some conclusions.

2. IMAGE CODING THROUGH RECOGNITION

2.1 Introduction

Many proposals have been made in the last years for image and video coding. In particular H.263+ is mainly intended for low to high data rate robust compression and is based on a block-based redundancy removal scheme [1]. In addition, MPEG-4 combines frame-based and segmentation-based approaches along with model-based video coding in the facial animation part of the standard which allows efficient coding as well as content access and manipulation [2]. It can be said that H.263+ and MPEG-4 represent the state of the art in video coding [6].

Our proposal relies on fourth generation video coding techniques based on recognition and reconstruction [5]. Recognition and reconstruction approaches rely on the understanding of the content. In particular, if it is know that an image contains a face, a house, and a car, recognition techniques to identify the content can be developed as a previous step to coding. Once the content is recognized, content-based coding techniques can be applied to encode each specific object. MPEG-4 provides a partial answer to this approach by using specific techniques to encode faces and to animate them [2].

[*] This work has been supported partially by grant CICYT TIC2001-0996 of the Spanish Government and by the NATO Science Fellowship Programme of Spain.

2.2 Face coding using a Principal Component Analysis approach

Let us simplify the visual content by assuming that we are interested in the coding of faces in a video sequence. Let us also assume that automatic tools to detect a face in a video sequence are available. Then, some experiments show that a face can be well represented by very few coefficients found through the projection of the face on an eigenspace previously defined [4]. The image face can be well reconstructed (decoded), up to a certain quality, by coding only very few coefficients.

Face coding can be done using a fixed eigenspace, what decreases the quality of the reconstructed image, or by adaptively changing the eigenspace according to the changes experienced by the face to be encoded as appears in the video sequence. Prior to coding, the corresponding face is detected and separated from the background using some segmentation techniques [7].

Our coding technique is based on a face recognition approach, which has been modified to cope with the coding application [8]. It assumes that a set of training images for *each* person contained in the video sequence is previously known. Once these training images have been found (usually coming from an image database or from a video sequence), a Principal Component Analysis (PCA) is performed for each individual using the corresponding training set of each person. This means that a PCA decomposition for every face image to be coded is obtained. The PCA is done previously to the encoding process.

After the PCA, the face to be coded is projected and reconstructed using each set of different eigenvectors (called eigenfaces) obtained in the PCA stage. If the reconstruction error using a specific set of eigenfaces is below a threshold, then the face is said to match the training image which generated this set of eigenfaces. In this case the recognized face is coded by quantizing only the most important coefficients used in the reconstruction. The size of the coded image has to be previously normalized for PCA purposes and then denormalized at the decoder.

3. FIXED EIGENSPACE APPROACH

In order to check the validity of the eigenspace approach for image coding, results using a fixed eigenspace will be first presented. These results will be useful to point out the main drawbacks of the eigensapce approach and to fully understand the adaptive eigenspace proposed in the next section. Figure 1 shows faces of the original sequence to be coded and the corresponding coded images. The original sequence is 8 bits 68x105 pixels at 12.5 frames/s which implies 714 kbits/s. The number of training images has been set to 30 and 16 coefficients per image have been used to encode each frame. The encoded sequence has an average bit-rate of 0.62 kbits/s and an average PSNR of 29.6 dB. Other similar sequences present a very similar performance. A dynamic coefficient selection scheme could have been designed but our purpose here is just to check the validity of the eigen approach. A first order optimized DPCM encoding scheme has been designed to encode the projection coefficients.

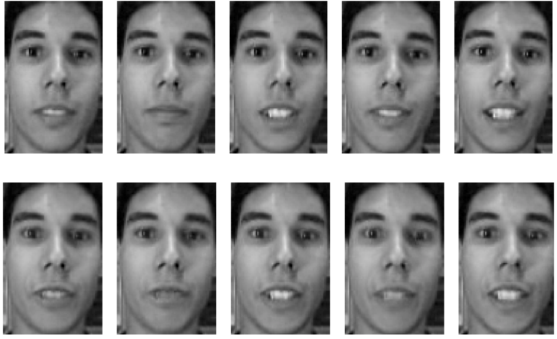

Figure 1. Above: original images at 714 kbits/s; below: coded images at of 0.62 kbits/s.

All the images have been coded intraframe and no motion compensation has been used. This approach is mainly a still image coder (except for the DPCM encoding of the coefficients) although we are interested in using video sequences to check the adaptive eigenspace presented in the next section. When viewed in video mode, the coded sequence presents a very acceptable visual quality.

In order to check how the quality of the coded image varies over the sequence, Figure 2 shows the PSNR of the coded sequence over 500 frames.

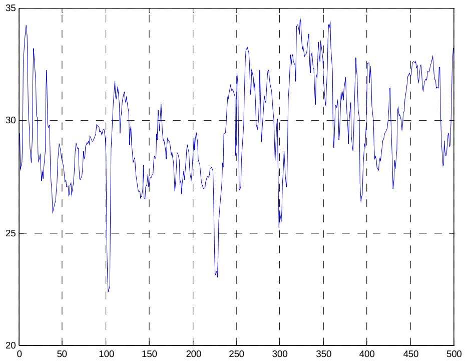

Figure 2. PSNR of coded sequence.

An important decrease in quality can be appreciated in some frames of the sequence. This corresponds to a change of expression in the face image. Any important change in the expression of the face that is not included in the training images, will lead the system to poor quality reconstruction. Next section presents an adaptive eigenspace approach to cope with these situations.

4. ADAPTIVE EIGENSPACE APPROACH

4.1 Introduction

In this section we propose an eigen coding approach that adapts itself to the face content appearing in the video sequence. Notice that a related technique has been presented in [9] although the approach is significantly different.

The initial encoding scheme follows that of the fixed eigenspace explained in Section 3. In that scheme, any important change in the expression of the face will lead at a poor performance of the scheme. In order to overcome this problem, we have designed a fall back mode system, which consists of a quality of reconstruction evaluation block followed by an upgrade mechanism of the coder and decoder eigenspace databases.

If the error obtained from the reconstruction process is low, only the coefficients of the reconstruction are sent to the decoder (this corresponds to the fixed eigenspace approach). If an important change in the expression of the face leads to a high error result, that face will be coded using one of the available *off-the-shelf* coding techniques like JPEG. In addition, an upgrade mechanism has been implemented, and the corresponding eigenspace is recalculated. In this way the faces of that person contained in the following frames, will be better represented using the upgraded eigenspace. This provides a basic approach for updating the eigenspace. Next section provides details of the upgrade mechanism.

4.2 Eigenspace database updating system

The updating PCA system proposed here works as follows: let us assume the eigenfaces of a person have been obtained by calculating the PCA of N training images of this person. The encoder starts coding the faces of the video sequence by transmitting the coefficients of the projection of the original faces over the eigenspace. The system continuosly evaluates the mean square error of the reconstruction. When the error increases over a specified limit, the eigenspace has to be updated at both the coder and the decoder.

A first updating system approach can be designed by substituting the *oldest* face in the group by the face which is being now coded. However, this system has an obvious problem. We are wasting the face that is updated in the substitution process. It would be more useful to store the information of any existing frames, as the images of a ‘talking-head’ sequence present short-term and long-term correlation. To illustrate this, let us assume we have a 10 face training database, formed mainly with the first frames of a sequence where the person has a *happy* expression. During the encoding process, if the person changes its expression to a *sad* one, the system will update the face database. But at the same time, if the oldest face is substituted, part of the database will be destroyed and the system will not be able to reconstruct good *happy* faces anymore using only projection coefficients.

There are two solutions to this problem. The size of the training images set can be increased without eliminating any face each time an updated-frame has to be sent (unless a maximum number of training images is fixed). However, as the system increases the size of the training images, the eigenspace generation becomes more and more time-consuming, and the system will not be able to perform the necessary operations in a reasonable time. The other solution consists of a multiple low-dimension eigenspace database. The system starts using a single eigenspace as before. When the eigenspace has to be updated, the training set is duplicated, and the substitution process is performed over one of the resulting sets. A new PCA is then calculated over the new set and the system obtains two eigenspaces. The following frames are reconstructed with the two existing eigenspaces and the one with the minimum reconstruction error is selected. A maximum number of eigenspaces per person is fixed, and when it is reached, the updates are made by direct substitution of the oldest face of the selected eigenspace. This approach is less expensive, computationally speaking. Firstly, because every time an update must be done, the first system has to calculate only an eigenspace of dimension 5, whereas the dimension is 50 in the second one. Secondly, because the PCA is a much more computationally intensive process than the projection and reconstruction one. To calculate 10 reconstructions per frame is not as time consuming as calculating the eigenspace.

5. RESULTS

Preliminary results will be shown here. Figure 3 presents a PSNR comparison of coded sequences with and without eigenspace adaptation for the same sequence above. An update occurs whenever the PSNR of the reconstruction error falls below 24 dB. Any other threshold may be used but we have found that this provides good visual results for this application. For clarity purposes, only frames 1-100 are shown. The corresponding bit-rate is 4.8 Kbits/s. Notice that after the update, the PSNR of the decoded frames increases with respect to the static case, at the expenses of obtaining a higher bit-rate. This increase in bit-rate corresponds to the JPEG images used in the update process. Some other image coding scheme different than JPEG may be used which would improve the coding efficiency.

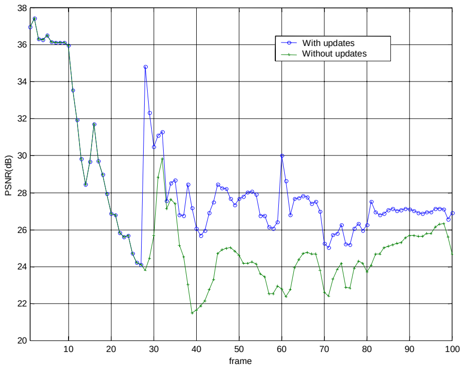

Figure 3. PSNR of coded sequences with and without updates

As another example, Figure 4 shows some results for a sequence with moderate expression changes (see the mouth zone).

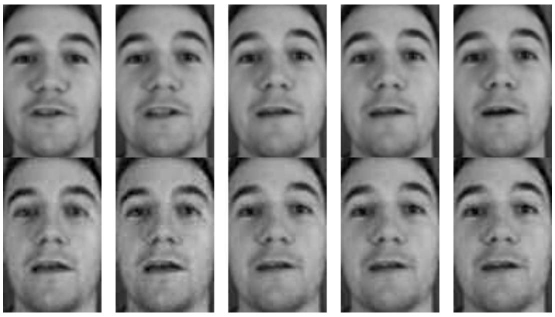

Figure 4. Above: original image; below: coded image at 3.93 kbits/s.

The first two frames have been coded using only the projection coefficients while the third image has been coded using JPEG. The fourth and fifth images (coded only with their reconstruction coefficients) present a better visual quality than the first ones due to the update process. The total bit-rate is 3.93 kbits/s. Notice that the visual quality of the coded image can be improved by designing reconstruction error coding schemes suited to these kind images. This would also provide scalable schemes.

6. CONCLUSIONS

An eigenspace approach for encoding moving faces has been presented. Very acceptable results below 1 kbit/s have been obtained for moderate changes of expression. An adaptive eigenspace scheme has been also designed to cope with more active sequences which provides bit-rates at around 4 Kbits/s.

7. REFERENCES

[1] G. Côté, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates", *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 8, no. 7, November 1998.

[2] ISO/IEC ISO/IEC 14496-2: 1999: "Information technology – Coding of audio visual objects – Part 2: Visual", December 1999.

[3] P. Eisert, B. Girod, "Analyzing facial expressions for virtual videoconferencing", *IEEE Computer Graphics and Applications*, Vol. 18, No. 5, pp. 70 - 78, 1998.

[4] B. Moghaddam, A. Pentland, "Probabilistic visual learning for object representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 696-710, July 1997.

[5] H. Harashima, K.Aizawa, and T. Saito, "Model-based analysis synthesis coding of video-telephone images – conception and basic study of intelligent image coding", *Transactions IEICE*, vol. E72, no. 5, pp. 452-458, 1989.

[6] L. Torres, E. Delp, "New trends in image and video compression", *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Tampere, Finland, September 5-8, 2000.

[7] F. Marqués, V. Vilaplana, and A. Buxes, "Human face segmentation and tracking using connected operators and partition projection", *IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[8] L. Torres, J. Vilà, "*Automatic face recognition for video indexing applications*", Invited paper, Pattern Recognition. Vol 35/3, pp 615-625, December 2001.

[9] W E. Vieux, K. Schwerdt and J.L. Crowley, "Face-tracking and Coding for Video-Compression", First International Conference on Computer Vision Systems, Las Palmas, Spain, January, 1999.