COMBINING ADAPTIVE PRINCIPAL COMPONENT ANALYSIS AND A HYBRID CODEC APPROACH FOR FACE CODING IN VIDEO SEQUENCES*

Roger Piqué
rpique@ieee.org

Luis Torres
luis@gps.tsc.upc.es

Technical University of Catalonia, Barcelona, Spain

ABSTRACT

This paper presents a new proposal for a video coding scheme intended for human faces in video sequences using an eigenspace approach. The scheme greatly improves previous results [1] by proposing a new adaptive eigenspace technique in combination with the new upcoming standard AVC [2], [3]. The basics of the approach have been already presented in [4]. The description of the face is also included in the bit stream using the MPEG-7 descriptors [5] thus providing search and browsing functionalities along with coding efficiency. The total bit stream presents better or same coding efficiency than AVC / H.264 and offers a very competitive alternative to B-predictive frames. Results are provided in the paper for bit-rates around 2.5 Kbits/s.

1. INTRODUCTION

Image and video coding are one of the most important topics in image processing and digital 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 and other functionalities such as scalable bitstreams, content description, error resilience, error concealment, etc. [1].

There is a need to provide novel compression schemes to code faces present in video sequences. Although the emerging standard H.264 / MPEG-4 part 10 [2] along with other model-based proposed schemes [6], [7] achieve high compression ratios for this particular application, we still believe that further compression is needed in video transmission through channels with very limited capacity such those in mobile or internet applications [8].

Having in mind these applications, we present a novel scheme to encode faces in video sequences based on an adaptive eigenspace approach. The eigenface concept for still image coding has been already presented in a face recognition framework in [9], and further explored in [10] and [1]. These works have proven the validity of the eigenspace approach for image and video coding but it has been also clear that more work is needed to improve the overall scheme.

It is in this context that the main contributions of this paper are a new way to adapt the eigenspace to take into account the different poses, expressions and lighting conditions of the faces and the combined use of the upcoming standard AVC to encode the face updates. As an added value to the encoded images, the MPEG-7 descriptors of the image faces have been found and added to the bit stream to allow search and browsing functionalities.

The paper is divided as follows: Section 2 presents the video coding scheme based on Adaptive Principal Component Analysis (APCA). Section 3 provides results and comparisons against the AVC hybrid coder and Section 4 draws some conclusions.

2. FACE CODER SCHEME

The video coding algorithm (Figure 1) proposed in this section is designed specifically for human faces. Thus, the face image is extracted, and the rest of the image is discarded. In the encoding stage, the extracted face image is coded combining an adaptive eigenspace approach with the AVC upcoming standard.

The first image in the sequence is coded INTRA using AVC. Once decoded, this image is used to set up the first eigenface, and to establish the motion reference image for the first update. The MPEG-7

* This work has been partially supported by the European Project MASCOT IST/FET (2000-26467) and by the grant TIC2001-0996 of the Spanish Government

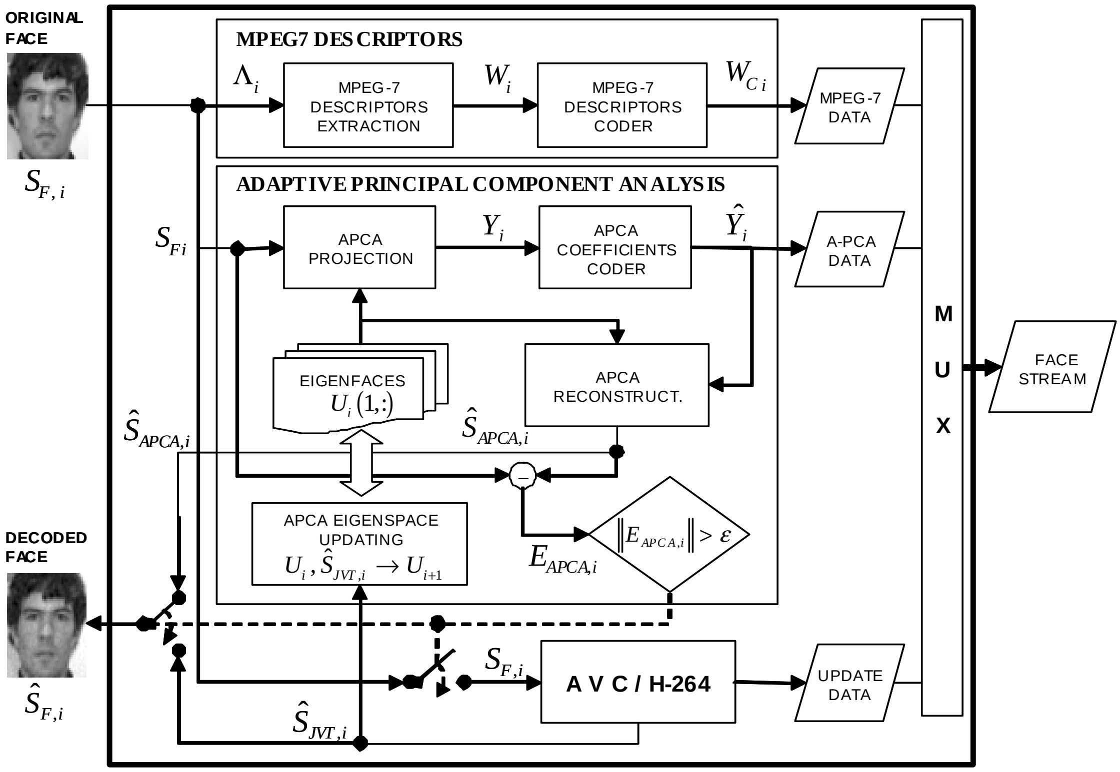

Figure 1 Face coder scheme

descriptors are also extracted from this image. No other images of the sequence are coded INTRA.

The second frame and the following ones are projected to the adaptive eigenspace to obtain the APCA coefficients. If the image is reconstructed with enough quality, it is coded using only these coefficients (APCA frame). On the contrary, if the face image cannot be correctly represented by the current eigenspace the image is coded through AVC using the last update image as motion reference (Update frame). In addition, this update image is used to adapt the APCA eigenspace simultaneously at both the coder and the decoder side.

Step by step, the coding algorithm works as follows:

1) APCA eigenspace projection.

$$y_i = U_i^T \ s_{F,i} \tag{1.1}$$

2) APCA coefficients coding. The coefficients are predicted from the previous frame ones. The difference is uniformly quantized and entropy coded using an UVLC (Universal Variable Length Code)
3) The image is reconstructed using the decoded APCA coefficients(1.2), and the reconstruction error is evaluated(1.3).

$$\hat{s}_{APCA,i} = U_i \ \hat{y}_i \tag{1.2}$$

$$e_{APCA,i} = s_{F,i} - \hat{s}_{APCA,i} \tag{1.3}$$

a) If MSE < ε → frame type = APCA:
   i) Only APCA coefficients are sent
b) Else if MSE > ε → frame type = Update:
   i) Do not send APCA coefficients
   ii) Encode the original image as an update frame using AVC. We use the P picture type with the last update as reference image.
   iii) Adapt APCA eigenspace using the SVD update technique described in [12] and [13]. The maximum number of eigenfaces is limited to N, so those with the smallest singular value are discarded.

In summary, images are coded in 3 different types: (INTRA, APCA and UPDATE). And five parameters, which are fixed constant for the whole sequence, are used to control the face coder:

– Q_I→ INTRA picture quality (AVC [3] coder parameter)
– Δ → APCA coefficients quantization step
– ε → Update threshold
– Q_U→ Update picture quality (INTER P picture quality parameter of AVC [3] coder)
– N → Maximum APCA eigenspace size

In addition, the face coder has been integrated in the AVC reference implementation JM2.1 [11], reusing its NAL structure and bitstream syntax.

In order to obtain a good performance is important that the face extraction step correctly aligns all the faces of the original sequence. There are several algorithms that can be used. We have used the same technique that is required in the extraction of the MPEG-7 face recognition descriptors [5] due to the following reasons. Firstly, it only requires the location of two points, the center of each eye in the original image. Secondly, with only these two points correctly located we can achieve a high coding efficiency. And finally, it allows us to easily find the MPEG-7 descriptors from the same extracted face image.

3. RESULTS

In order to show the potentiality of this new coding scheme, some results are presented and compared to AVC (JM 2.1 implementation [11]). Two test sequences have been used with an image size of 56x46 pixels (for MPEG-7 compatibility) and 25 frames/sec. Moreover, in order to make a better evaluation of the new coding algorithm, the first 30 frames of the coded sequence and the MPEG-7 descriptors are discarded from the BR and PSNR computation.

Figure 2 shows that for the Miss America sequence, the proposed face coder, obtains an efficiency comparable to AVC2B (IBBPBBPBBP…) and better results than AVC 4B (IBBBBPBBBBP…) or AVC 0B (IPPPPPPP…) for a PSNR in the 29.5-31.5 dB range.

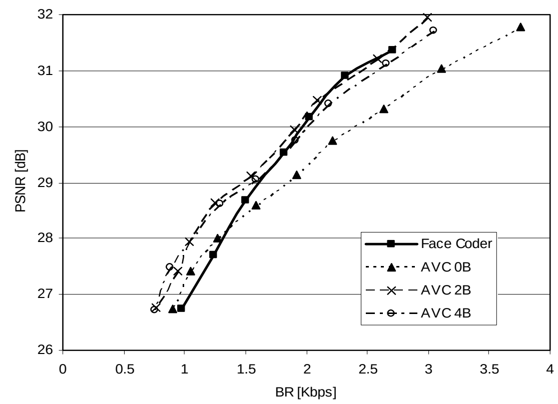

Figure 2. PSNR of coded sequence vs. BR, "Miss_America" sequence.

A thorough analysis of each frame reveals that the APCA coding algorithm achieves the highest performance when the eyes are located precisely and

Esquema	Avg. BR [Kbps]	Avg. PSNR [db]	% APCA o B Frames
FACE CODER	2.323	30.90	79.83%
AVC 0B	3.108	31.04	0 %
AVC 2B	2.579	31.22	66.7%
AVC 4B	2.652	31.13	80%

Table 1 Average results for "Miss_America" at a PNSR ≈ 31 dB.

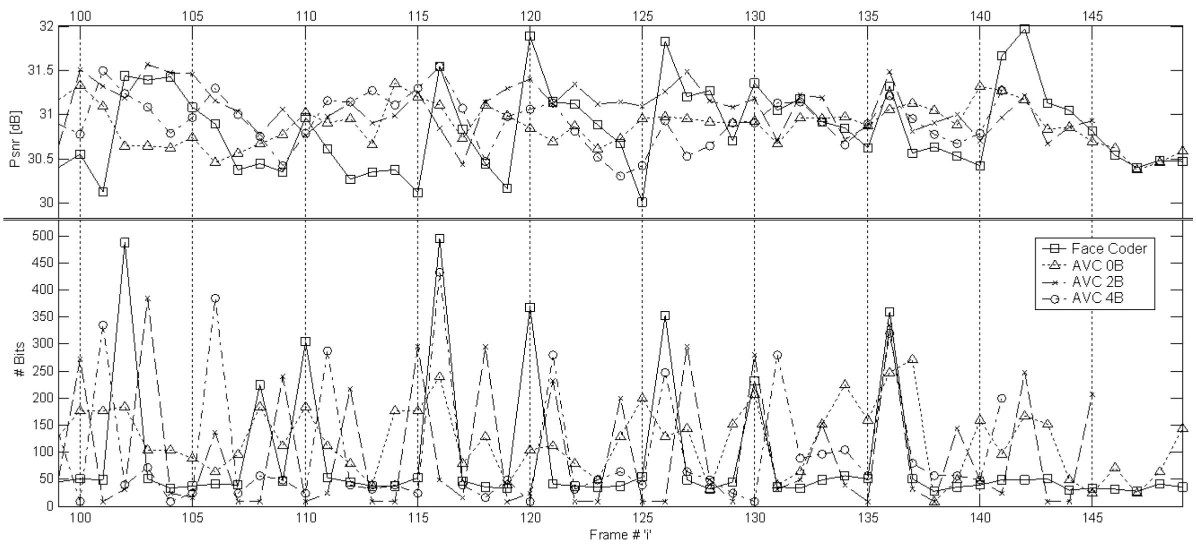

Figure 3. PSNR & BITS/FRAME evolution of the last 50 frames of "Miss_America" sequence of Table 1

the face image to encode has a previously coded one with a similar expression. In such situations, the image will be coded with good quality using only APCA coefficients. These APCA frames achieve the major bit-rate savings in our proposed algorithm.

In Figure 3, we compare the performance of APCA frames versus B type frames, which offer the major bit-rate savings in a classical hybrid scheme. In AVC the efficiency of B type frames relies on the frame structure, which is difficult to adapt. Thus, the GOP is usually prefixed using a compromise. For example, if we use a frame structure with many B-type frames like AVC 4B, it can be said that a better efficiency will be obtained, but the reference pictures are more spaced out, so depending on the motion activity the coding efficiency may decay. This is the reason why in Figure 2 AVC 2B obtains a better efficiency using fewer B frames. Moreover, the number of consecutive B's should be limited because the coding order is not equal to the presentation order. Therefore, it may break the synchronization between lips and speech, or introduce an uncomfortable delay in a talk.

On the other hand, the face coding algorithm dynamically adapts its frame structure. We use APCA frames, whenever the face is suitable to be coded with enough quality in this way. In our case, the number of consecutive APCA frames is not limited because the coding order is the same as the presentation order. Therefore, the updates are selected only when they are necessary, as the quality goes below a threshold. As a result, an efficient way to update the eigenspace is provided that takes into account new views of the face such as different poses, expressions and lighting conditions.

	Original	Face Cod.	AVC 0B	AVC 2B	AVC 4B	
PSNR		30.73	30.96	30.87	31.10	dB
#BITS		34	80	8	48	
Fr. type		APCA	P	B	B	

Figure 4. Visual results at frame #68 of Table 1

	Original	Face Cod.	AVC 0B	AVC 2B	AVC 4B	
PSNR		31.67	31.28	30.97	31.28	dB
#BITS		49	96	24	200	
Frame type		APCA	P	B	P	

Figure 5. Visual results at frame #141 of Table 1

Additionally, visual results give a quality according to the PSNR statistics. Figures 4 and 5 provide the reconstructed images for frames #68 and #141 of "Miss_America" sequence.

Finally, Figure 6 shows the results with the "Carphone" sequence. This sequence has a worse eye location accuracy, so very few frames (14.2%) are coded as APCA frames. In such cases, our encoding scheme sends a lot of consecutive update frames (…PPPP…), and offers the same behavior as AVC 0B (IPPPP…). Therefore, in the worst case, our scheme will obtain a similar efficiency to that of AVC 0B.

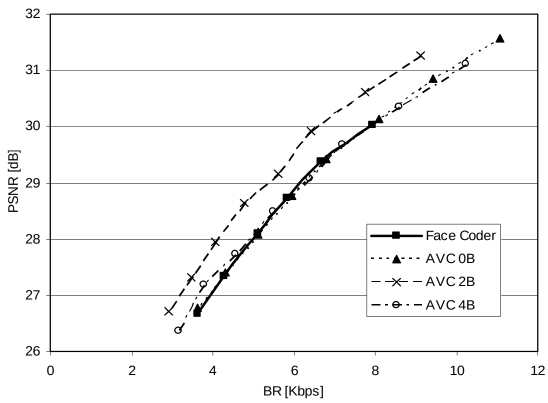

Figure 6. PSNR of coded sequence vs. BR, "Carphone" sequence.

4. CONCLUSIONS

A new approach for encoding face images in video sequences has been presented with very promising results. Three main advantages can be considered:

Firstly, when the face is located and extracted correctly, better results than AVC are obtained. On the contrary, if the face is incorrectly located, the results are similar to AVC 0B (IPPPP…).

Secondly, APCA frames, those coded using only APCA coefficients, offer great bit-rate savings without the limitations of B frames. Moreover, APCA frames do not require motion compensation and have an execution time ten times faster, on average, than motion compensated frames used in hybrid video coding systems. However, a more complete analysis considering all possible optimizations should be done to draw a more definite conclusion about computational complexity.

Additionally, the embedded MPEG-7 face recognition description identifies the face and allows searching and browsing functionalities.

On the other hand, the main disadvantages are the following:

Firstly, the amount of memory necessary to store the eigenspace is bigger than that of the motion reference buffer of a standalone hybrid scheme. However, this inconvenient is less important for small sized images.

Secondly, the face coder efficiency gain over AVC depends on the performance of the face tracking and extraction block.

Finally, our scheme can also be improved in several ways, namely: encoding APCA coefficients using an arithmetic coder (like CABAC used in JVT), and improving the update decision using a Rate-Distortion framework. Work is in progress to include these approaches.

5. REFERENCES

[1] L. Torres, D. Prado, "A proposal for high compression of faces in video sequences using adaptive eigenspaces", *IEEE International Conference on Image Processing*, Rochester, USA, September 22-25, 2002.

[2] H. Schwarz, T. Wiegand, "The Emerging JVT/H.26L Video Coding Standard", *Proc. of IBC 2002*, Amsterdam, NL, September 2002.

[3] T. Wiegand, *Editor's Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Draft 7*, October 2002. http://bs.hhi.de/~wiegand/jvt.html

[4] R. Piqué, L. Torres "Efficient face coding in video sequences combining adaptive principal component analysis and a hybrid codec approach", *International Conference on Acoustics, Speech and Signal Processing*, Hong-Kong, China, April 6-10, 2003.

[5] ISO/IEC 15938-3: 2002, Information technology -- Multimedia content description interface -- Part 3: Visual.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] JVT software page, JM/TML Software Coordination, http://bs.hhi.de/~suehring/tml/

[12] S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler and H. Zhang, "An eigenspace update algorithm for image analysis", *Graphical Models And Image Processing*, Vol. 59, No. 5, September 1997, pp. 321–332.

[13] M. Gu and S. C. Eisenstat, *A Stable and Fast Algorithm for Updating the Singular Value Decomposition*, Research Report YALEU/DCS/RR-966, Yale University, New Haven, CT, 1994.