REGION BASED CODING SCHEME WITH SCALABILITY FEATURES

Olivier Egger, Frank Bossen, and Touradj Ebrahimi
Signal Processing Laboratory
Swiss Federal Institute of Technology at Lausanne
CH–1015 Lausanne, Switzerland
Email: egger@lts.de.epfl.ch

ABSTRACT

In order to satisfy the needs of new applications in a multimedia environment the problem of object–oriented coding has to be addressed. In this paper two main approaches are presented to tackle this problem. First, an algorithm for shape coding is presented. It is based on a chain coding algorithm where powerful modeling techniques are used to increase the compression ratio. Second, an algorithm for interior coding is described. It is based on an arbitrarily–shaped subband transform followed by a generalized embedded zerotree wavelet algorithm. It is shown in the paper that it achieves good compression results and has additional properties such as supporting arbitrarily–shaped regions, being computationally efficient, keeping the same dimensionality in the transformed domain, allowing perfect reconstruction and an intrinsic rate control mechanism. The presented results show that the two algorithms build an efficient basis to design object–oriented video coding schemes.

1 INTRODUCTION

New applications in a multimedia environment arise the need of additional functionalities and requirements for video coding schemes. Most of them are closely related to object–oriented manipulations. Object–oriented functionalities require a prior segmentation of the scene into regions of interest. In general, these regions are of arbitrary shapes. There are two important issues involved in a region–based scheme. The first one is the representation and coding of the region shapes. The second is the represention model and coding of the interior of the so–called objects (regions). This work presents an algorithm addressing both of these issues.

The problem of representation of region shapes has been investigated in the past for regions with arbitrary and particular shapes [1, 2]. The easiest and also the most obvious way is to represent each region using its contour. More complex solutions use a tree structure for the segmented data [3].

In this paper we concentrate on the problem of region representation using the contour information. This problem becomes less obvious when representing all regions in a segmented image, and when the goal is to reduce the redundancies as much as possible. In fact, in this case the contours representing a segmented image will be coded twice, if regions are coded independently. It is possible of course to remove one of the two contours. However, in this case the contour pixels themselves should be marked so that the decoder can assign them to the correct region. In this paper we use a technique, which will code only one contour in between regions. This allows a unique way to assign the correct region labels to the pixels on the contours without a need for additional labeling. Quite an extensive literature has already treated the problem of contour coding [3–8]. The most popular technique consists in coding the position of the pixels in the contour relative to their neighbors, which is referred to as *chain coding*. In the most classical chain coding, the first pixel in a contour is coded in its absolute coordinates. All the remaining pixels are then coded by just indicating their relative position compared to the last encoded neighboring pixel. Variants of this technique based on the same principle can be used in order to improve the efficiency of the coding [1, 6–8].

The other main issue of object–oriented techniques is the texture coding. An efficient representation of the region interior should provide a good decorrelation of the data in order to obtain a high energy compaction. In the case of compressing rectangular pictures this operation is performed using appropriate block transforms such as DCT or subband/wavelet transforms. Various approaches have been proposed to generalize the block–based techniques to arbitrary shaped regions. These techniques are either computationally very expensive [9, 10] and/or do not perform a full decorrelation of neighbouring pixels in a given region [11]. The technique described in this paper is based on a subband transform applied to arbitrarily–shaped regions.

2 SHAPE CODING

One of the major issues in every object–oriented coding approach is that of shape coding. When considering the coding of one single region in a segmented image, it is logical to assume that its contour is closed. However, when coding all the regions, if one wants to avoid the problem of coding contours twice in the segmented image, the closeness of contours should be relaxed. In this case, an additional code for the end of contour is needed. But if all the already coded information is available at the decoder, no additional code for the end of contour is required, because once again, it is possible to discover the end of a contour when it hits another one.

We adopted a variant of chain coding which takes into account the special characteristics of the typical contours obtained by classical segmentation techniques. These characteristics are: 1) The contours are open; 2) They are 4-connected; 3) They typically contain long runs in the same direction.

Thus, this shape coding process is subdivided into three steps: first the shape mask is transformed into a contour representation, then the contours are chain-coded, and finally the stream produced by the chain coder is entropy encoded. The next paragraphs present each of the these three steps.

The shape of an object O is given by a binary mask M^O, where

$$M^O_{i,j} = \begin{cases} 1 & \text{if the pixel (i,j) belongs} \\ & \text{to the object } O \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The contour representation C^O of the shape of O is defined as

$$C^O_{i,j} = \begin{cases} 0 & \text{if } M^O_{i,j}, M^O_{i-1,j}, M^O_{i,j-1}, \\ & \text{and } M^O_{i-1,j-1} \text{ have the} \\ & \text{same value} \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

A *contour pixel* is defined as a pixel (i,j) such that $C^O_{i,j} = 1$.

Although the transformation from M^O to C^O is straightforward, this is not true of the inverse which requires a filling algorithm. Moreover, the inverse transform does not always exist. The original mask can only be recovered if there are no i and j such that (i,j), $(i,j+1)$, $(i+1,j)$, and $(i+1,j+1)$ are contour pixels. To guarantee the existence of the inverse, the mask can be either morphologically filtered, or magnified by a factor of two [12]. The first solution has been preferred in the frame of this work, because the segmentation mask of the test sequence is quite noisy.

Probably the most popular contour coding technique consists in coding the position of the contour pixels relative to their neighbors referred to as chain coding. In the proposed approach we use a classical 4 connected contours in order to obtain a stream of direction symbols representing the shape of contours to be coded. The latter is further compressed by means of an appropriate entropy coding as described below.

The PPM (prediction by partial matching) [13] compression scheme is used to efficiently compress the stream generated by the chain coder. PPM is a finite-context statistical modeling technique that blends together several fixed-order context models to predict the next symbol in the stream. Prediction probabilities are computed from frequency counts, which are updated as symbols are processed. The symbol that actually occurs is encoded relative to its predicted distribution using arithmetic coding.

An improvement specific to chain coding symbols is further applied. Since backward moves (for example a *south* move following a *north* one) are impossible, a zero probability is assigned to the direction opposite to the previous one (the *south* move has zero probability in the previous example).

3 INTERIOR CODING

Another major issue in an object–oriented coding scheme is the problem of the interior coding. The proposed approach can be seen as a generalization of the embedded zerotree wavelet (EZW) algorithm [14,15] from rectangular images to arbitrarily shaped regions. Special care has been dedicated to retain all the advantages of the EZW algorithm for arbitrarily shaped regions without any compromise. As for the original EZW algorithm the proposed technique is based on three basic blocks, namely the transformation, the zerotree prediction and the successive approximation quantization (SAQ).

The transformation block (generalized to abritrarily shaped regions) performs the decorrelation of the data into a multiresolution structure using the wavelet transform. The subband transform is performed in a separable way as in conventional decomposition schemes. The zerotree prediction in turn allows a further improvement of the energy compaction by taking into account the remaining intra–band correlation. Moreover, the zerotree wavelet algorithm has been generalized to wavelet packets by means of an appropriate pruning of the parent–children relationship tree. This procedure is illustrated in Figure 1. Finally, the SAQ provides an embedded bit stream and allows an exact rate control. The embedded bit stream together with the multiresolution structure allows additional features such as progressive transmission, retrieval and data browsing.

4 RESULTS OF THE PROPOSED TECHNIQUES

The segmentation mask of the person in the foreground of the *Sean* sequence is encoded. The spatial resolution is 176×144 pixels (QCIF), and the temporal one 10 Hz. The statistical model is either reset (i.e. all frequency counts are reset to zero) after every frame, never reset, or reset after every 10 frames. The order of the model is 4, and 7 if the model is never reset.

Results are shown in Figure 2. They do not take into account the overhead needed to encode the number

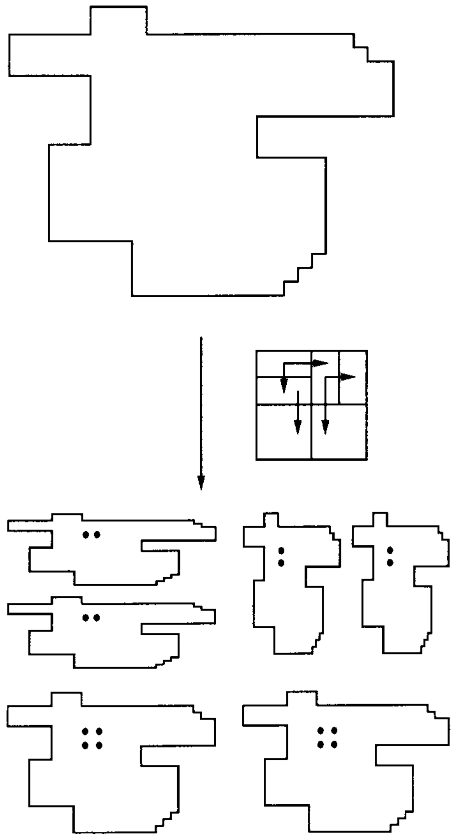


Figure 1: *Illustration of the parent–children relationship of a general decomposition into six subbands.*

of chains, and their starting positions (~ 20 bits per frame).

Figure 2 shows that never resetting the statistical model yields the best results (280.46 bits per frame on average). However this method has a significant drawback. To decode the n-th frame, all the previous frames need to be decoded first. As resetting the model after every frame doesn't yield very good performance (401.57 bits per frame on average), it is best to reset the model every once in a while. Resetting every 10 frames seems reasonable (309.54 bits per frame on average). The effect of resetting the model is clearly visible in Figure 2 (the cost of the first frame after a reset is much higher).

Results to illustrate the interior coding have been performed by coding and decoding an actual bit stream. In this way, the correctness of the algorithm is assured. In Figure 3 it is shown that it is possible to control the rate of each object in the scene. The picture in the middle has been coded at a rate of 0.35 bits/pel. The PSNR of the reconstructed picture is 28.91 dB, of the person only 28.56 dB and of the background 29.07 dB. The shape has been coded by means of the algorithm presented in Section 2. For the given frame, a cost of 400 bits has been achieved. In the example proposed in Figure 3c, 0.023 bits/pel were allocated to the background and 0.316 bits/pel for the person in the foreground,

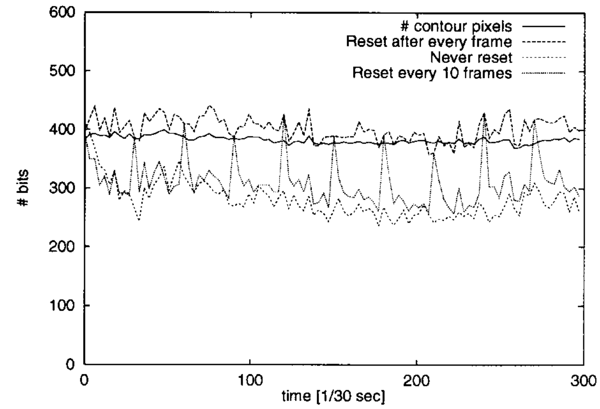


Figure 2: Results

giving a total rate of approximately 0.35 bits/pel. The PSNR of the reconstructed objects are 32.515 dB for the person in the foreground, and 22.41 dB for the background.

5 CONCLUSIONS

In this paper the problem of object–oriented coding is discussed. Two algorithms are presented to approach the problem. The first one deals with the coding of the shapes. It is based on three steps. 1) The tranformation of the shape mask into a contour representation. 2) Chain coding of the contours. 3) Entropy encoding of the produced symbol stream.

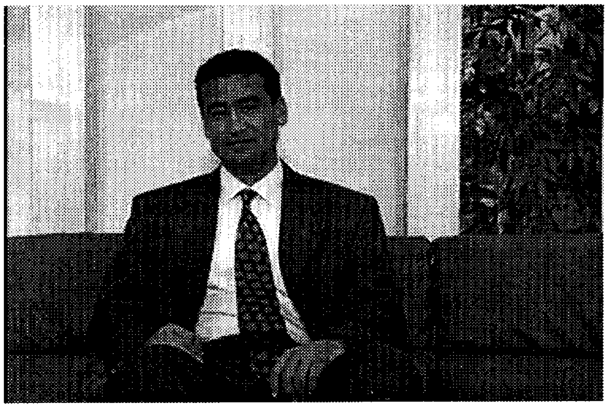
The second one deals with the coding of the region interior. It is based on a subband transform followed by an adaptation of the embedded zerotree wavelet algorithm. The proposed algorithm can be characterized by the following properties: 1) It can be applied to arbitrarily shaped regions. 2) The whole process is computationally efficient and does not require any iterative procedure. 3) The representation of the transformed data keeps the same dimensionality and preserves the shape information of the original data. 4) The transformation allows a perfect reconstruction with a multiresolution representation. 5) The SAQ allows an exact rate control together with a progressive bitstream.

Results of the proposed techniques show that they are well suited to be included in a general object–oriented video coding scheme.

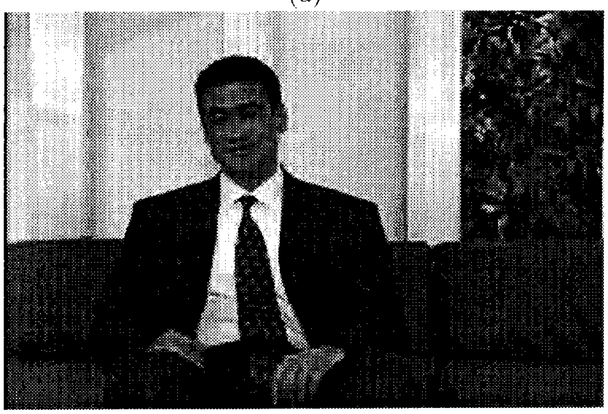
References

[1] D. H. Ballard and C. M. Brown. *Computer Vision.* Prentice Hall, Inc., New Jersey, 1982.

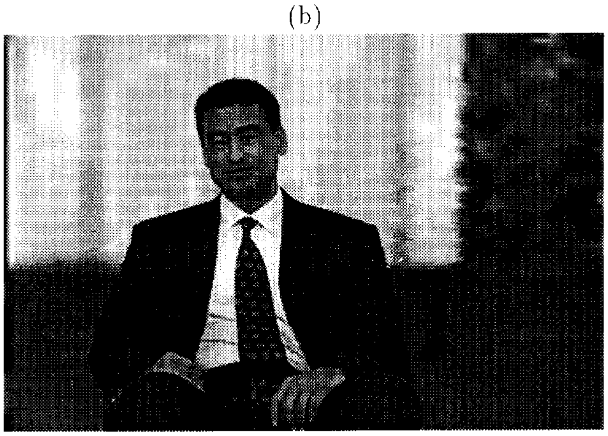
[2] J. Foley, A. Van Dam, S. Feiner, and J. Hughes, editors. *Computer Graphics: Principles and Practice.* Addison -Wesley Publishing Company, 1987.



(a)



(b)



(c)

Figure 3: *Illustration of object editability on the picture "scan". Total cost of each coded picture is 0.35 bits/pel. (a) Original. (b) Total picture coded with the EZW algorithm. (c) Separate coding of the objects, higher rate to the main object.*

[3] H. Samet. "Region representation: Quadtrees from boundary codes". *Commun. of the ACM*, Vol. 23, No. 3, pp. 163–170, March 1980.

[4] H. Freeman. "On the encoding of arbitrary geometric configurations". *IRE Trans. on Elec. Comp.*, Vol. EC-10, pp. 260–268, June 1961.

[5] C.R. Dyer, A. Rosenfeld, and H. Samet. "Region representation: Boundary codes from quadtrees". *Commun. of the ACM*, Vol. 23, No. 3, pp. 171–179, March 1980.

[6] M. Eden and M. Kocher. "On the performance of a contour coding algorithm in the context of image coding Part I: contour segment coding". *Signal Processing*, Vol. vol. 8, no. 10, pp. 381-386, 1985.

[7] C. Lu and J.G. Dunham. "Highly efficient coding schemes for contour lines based on chain code representation". *IEEE Trans. on Communications*, Vol. 39, No. 10, pp. 1511-1514, October 1991.

[8] T. Kaneko and M. Okudaira. "Encoding of arbitrary curves based on chain code representation". *IEEE Trans. on Communications*, Vol. 33, pp. 697–707, July 1985.

[9] M. Gilge, T. Engelhardt, and R. Mehlan. "Coding of Arbitrary Shaped Image Segments Based on a Generalized Orthogonal Transform". *Image communication*, Vol. 1, No. 2, pp. 153–180, October 1989.

[10] H. H. Chen, M. R. Civanlar, and B. G. Haskell. "A Block Transform Coder for Arbitrarily Shaped Image Segments". In *Proceedings of the International Conference on Image Processing ICIP*, Vol. I, pp. 85–89, Austin, USA, November 1994.

[11] T. Sikora and B. Makai. "Shape–adaptive DCT for Generic Coding of Video". *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 5, No. 1, February 1995.

[12] Frank Bossen and Touradj Ebrahimi. "Region shape coding". Technical Report M0318, ISO/IEC JTC1/SC29/WG11 MPEG, November 1995.

[13] John G. Cleary and Ian H. Witten. "Data compression using adaptive coding and partial string matching". *IEEE Transactions on Communications*, Vol. COM-32, No. 4, April 1984.

[14] J. M. Shapiro. "Embedded Image Coding Using Zerotrees of Wavelet Coefficients". *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, pp. 3445–3462, December 1993.

[15] O. Egger, A. Nicoulin, and W. Li. "Embedded Zerotree Based Image Coding Using Linear and Morphological Filter Banks". In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP*, Detroit, USA, May 1995.