CODING OF ARBITRARILY SHAPED VIDEO OBJECTS USING B-SPLINE WAVELETS

Mervi Sepponen
Machine Vision and Media Processing Group
Infotech Oulu, Dept. of EE
University of Oulu
FIN-90570 Oulu, FINLAND
e-mail: mese@ee.oulu.fi

Visa Koivunen
Signal Processing Laboratory
Tampere University of Technology
FIN-33101 Tampere, FINLAND
e-mail: visa@cs.tut.fi

ABSTRACT

This paper addresses the problem of encoding and decoding image patches in a model-based image coding scheme. Any arbitrarily shaped video object can be represented using a triangular mesh. A method for coding triangular image blocks using B-spline wavelet filter banks is developed here. Triangles are considered degenerate rectangular blocks hence the same coding can be used for rectangular blocks as well. Triangles at the boundary of a video object are allowed to have curved boundaries in order to have more accurate shape representation using fewer elements. In the decoding end, patches are warped using affine motion estimates. Examples are given using real image data.

1. INTRODUCTION

A growing interest on content-based video coding methods, which are based on realistic structural and motion models, has emerged [1], [10]. This type of techniques provide good quality reconstructed image data at low bit-rates. Existing video compression standards such as H.261, MPEG-1 and MPEG-2 do not perform well at very low bit-rates. In content-based video coding, shape and texture of individual arbitrarily shaped video objects have to be captured. Shape is commonly represented with a contour, and texture is coded using block-based DCT. However, methods using only rectangular blocks are not well-suited for coding arbitrary shapes [3], [9].

In this paper the problem of coding of arbitrarily shaped video objects is addressed. As an example, coding of face images which are typical in videoconferencing and multimedia, is studied. A 2-D mesh representation for the face is obtained by using detected feature points and then constructing a mesh using constrained Delaunay triangulation. The coding of triangular mesh elements is done by using a B-spline wavelet filter bank [2] with a dyadic tree structure. Such filter bank requires a rectangular arrangement of points as an input. Therefore, triangles are considered as degenerate rectangles and data within triangles are resampled using a method employing barycentric coordinates. Finally, a rectangular arrangement of data is obtained. Consequently, similar filter bank for both triangular and rectangular image patches may be used. Using B-splines as scaling functions allows exploitation of their affine invariance property in image warping. Moreover, local deformations can be performed by changing wavelet coefficients [4]. This coding approach may be extended to 3-D.

This paper is organized as follows. In section 2, B-spline wavelets and their properties are studied. In section 3, a method for coding triangular image patches is introduced. Examples of patch coding and image sequence coding are provided in section 4. Finally section 5 concludes the paper.

2. B-SPLINE WAVELETS AND FILTER BANKS

First we define the normalized B-spline basis function $N_{i,k}$ of order k, by the following Cox-deBoor recursion formulas

$$N_{i,1}(t) = \begin{cases} 1 & \text{, if } x_i \le t < x_{i+1} \\ 0 & \text{, otherwise} \end{cases} \tag{2.1}$$

and

$$N_{i,k}(t) = \frac{(t - x_i)N_{i,k-1}(t)}{x_{i+k-1} - x_i} + \frac{(x_{i+k} - t)N_{i+1,k-1}(t)}{x_{i+k} - x_{i+1}}, \tag{2.2}$$

where the values of x_i are elements of a knot vector. The endpoint-interpolating B-splines of order k with open end condition result when the first and last k knots are set to zero and one, respectively. This forces the spline to pass exactly through the endpoints.

When using the endpoint-interpolating B-splines of order k as scaling functions, the length of the data required for the filter bank must be $2^j + k - 1$, where j denotes the level of filtering. Filtering equations are

$$\Phi^{j-1}(x) = \Phi^j(x)P^j$$
$$\Psi^{j-1}(x) = \Phi^j(x)Q^j,$$

where $\Phi^j(x)$ is a matrix of scaling functions for level j and $\Psi^j(x)$ is a matrix of wavelets for level j. In the case of B-splines the matrix P^j can be generated using the formulas (2.1) and (2.2). Resulting matrix P^j is sparse because the scaling functions are locally supported. The matrix Q^j is determined by requiring each column to have a minimal number of consecutive non-zeros thus imposing a banded structure [4].

Let a signal be expressed in terms of the B-spline scaling function basis written as a column matrix C^j. A lower resolution version of C^j is created by using dyadic analysis filter bank expressed as matrix equations $C^{j-1} = A^j C^j$ and $D^{j-1} = B^j C^j$. The reconstruction is done using dyadic synthesis filter bank i.e. original coefficients C^j can be reconstructed from lower resolution coefficients C^{j-1} and detail coefficients D^{j-1} by

$$C^j = P^j C^{j-1} + Q^j D^{j-1}.$$

Analysis and synthesis scheme is depicted in Figure 1 where processing of coefficients is done in the middle.

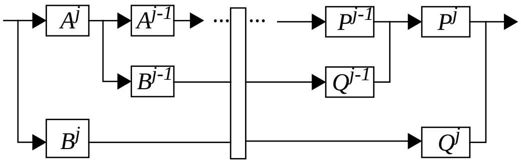

Figure 1: The matrices A^j and B^j are the analysis filters and P^j and Q^j are synthesis filters.

Computing of low-resolution and detail parts from the original signal may be done by solving the sparse linear system

$$\left[P^j \middle| Q^j\right] \left[\frac{C^{j-1}}{D^{j-1}}\right] = C^j$$

exploiting the banded structures of matrices P^j and Q^j. As a result the filter bank does not necessarily involve forming of the analysis filters A^j and B^j.

This scheme can be generalized into two dimensions performing decomposition and reconstruction along rows and columns of the tensor product surface, treating each row or column as a one-dimensional curve. Using the endpoint-interpolating B-splines as scaling functions results in semi-orthogonal wavelet basis i.e. the wavelets are orthogonal to the scaling functions at the same level but not to each other, except for the piecewise-constant case. Perfect reconstruction is not possible when using semiorthogonal wavelet basis, again except the piecewise-constant basis with orthogonality.

B-splines possess affine invariance property which is useful in warping image patches based on affine motion parameters, i.e. coefficients C^j may be multiplied by an affine transformation matrix and the result transforms accordingly. Also local deformations may be performed at different scales in case affine warping is not sufficient [4]. This is important for example in coding facial images where facial expressions have to be captured.

3. CODING OF TRIANGULAR PATCHES

Commonly wavelet coding and filter banks are applied to rectangular image blocks. Rectangular blocks, however, are not suitable for coding arbitrarily shaped video objects. A triangular mesh can model objects of any shape and topology accurately. One way to perform multiresolution analysis using triangular mesh is based on subdivision surfaces [6]. In order to code triangular mesh elements using the B-spline wavelet filter bank, we consider a triangle as a degenerate rectangle. Consequently, both triangular and rectangular image patches may be processed using similar filter bank. Standard rectangular blocks and the same filter bank could be used if image analysis used for creating and tracking the triangular mesh fails. The data inside the triangle are resampled in order to form a rectangular arrangement of pixels [5]. Desired m-by-n rectangular patch is obtained by using barycentric coordinates for triangular patch as follows. Let a, b and c be the vertices of a triangle shown in Figure 2. Any point p in the triangle can be written as barycentric combination of vertices a, b and c i.e. $p = p(u,v,w) = ua + vb + wc$ where the coefficients (u,v,w) are barycentric coordinates of point p such that $u + v + w = 1$.

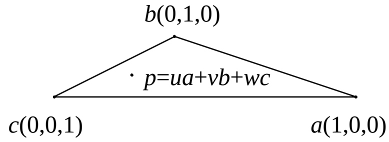

Figure 2: Barycentric coordinates of vertices.

In practise resampling is started along the longest edge of the triangle and is implemented as follows.

1. For the first row of n samples, let one of the barycentric coordinates, e.g., $v = 0$. Uniform sampling is performed by changing the values of two other coordinates u and w. The first sample has coordinates (1,0,0). For the second sample $u = 1 - \delta$ and $w = \delta$ where $\delta = 1/(n - 1)$. The last sample has coordinates (0,0,1).
2. For i th row, $v = (i - 1)\rho$ with increment $\rho = 1/(m - 1)$. Uniform sampling is performed using $\delta = (1 - \rho\,(i - 1))/(n - 1)$ and requiring that $u + v + w = 1$.
3. The last row has n samples with coordinates (0,1,0).

The actual sample value is obtained as a weighted mean of 4 closest pixels to point p obtained by barycentric coordinates. The weights are inversely proportional to the distance from point p.

The resulting rectangular arrangement of points is then coded using the B-spline wavelet filter bank. B-spline wavelet basis allows for performing the analysis-synthesis even if sample spacing is not uniform [7],[8]. Sampled values near each vertex are very similar. In the encoding stage they cause small wavelet coefficients, which are quantized and small coefficients are set to zero. Hence, no extra data have to be transmitted. The coefficients are highly concentrated and they are compressed using entropy coding. In the decoding end, the coefficients are decompressed and a rectangular patch is reconstructed using P and Q filters.

After reconstruction the rectangular patch may be warped according to affine motion model. The motion of the vertices of a triangle uniquely determines the transformation. Because of the affine invariance property of B-splines, one may just multiply the coefficients by the transformation matrix and interpolate. Warping may also be done by using barycentric coordinates for the triangular patch values.

Face images, for example, have curved boundaries. An accurate approximation of 2-D shape of such an object requires plenty of straight edges. The number of edges can be decreased or accuracy improved by using curved edges. Cubic Bezier curves are employed here [5]. The bounding Bezier curve is determined by using Hermite interpolation. In order to compute the boundary, we need to know the vertices of the triangles on the boundary and the direction vector at each vertex. The direction vector is obtained using edges of neighbouring triangles as follows. Let p_0, p_1, p_2 and p_3 denote four adjacent vertices on the boundary and m_0 and m_1 two tangent vectors at vertices p_1 and p_2, respectively. The tangent vector m_0 is determined by

$$m_0 = s\frac{|s|}{|s|+|r|} + r\frac{|r|}{|s|+|r|} ,$$

where $s = p_1 - p_0$, $r = p_2 - p_1$ and $|\,.\,|$ denotes the length of the vector. Tangent vector m_1 is obtained using three vertices p_1, p_2 and p_3 in a similar way. The cubic curve that interpolates to p_1, p_2, m_0 and m_1 is obtained by Hermite interpolation:

$$p(t) = p_1 H_0(t) + m_0 H_1(t) + m_1 H_2(t) + p_2 H_3(t) ,$$

where $H_0 = (1 - t)^3 + 3t\,(1 - t)^2$, $H_1 = t\,(1 - t)^2$, $H_2 = - t^2(1 - t)$ and $H_3 = 3t^2(1 - t) + t^3$.

In Figure 3 there is the first frame of Miss America image sequence with constrained Delaunay triangulation using cubic Bezier curves as boundaries.

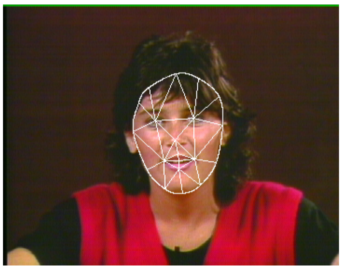

Figure 3: Contour of the face is represented using Bezier curves.

4. EXAMPLES

In this section the performance of the proposed coding method is investigated. Experiments are implemented using Matlab and patches used are from the first frame of Miss America image sequence. A rectangular 33 x 33 block shown in Figure 4(a) was encoded using the B-spline wavelet filter bank with second order scaling functions. The detail coefficients are highly concentrated about zero, thus most of them can be neglected. In this case a scalar quantization was used to restrict the values of the coefficients to integer values. Quantization was employed at level 3 and coefficients less than 6.0 were neglected. The result after the quantization and decoding is shown in Figure 4(b). The Peak Signal to Noise Ratio (PSNR) of the decoded image was 36.27 dB and 74.93 per cent of the wavelet coefficients was zero.

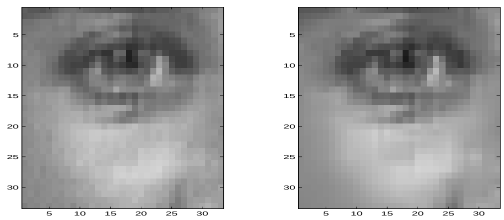

Figure 4: (a) Original and (b) decoded rectangular patch.

The coding of the triangular patch is considered next. Second order basis functions were used in the filter bank as scaling functions, scalar quantization was applied at level 3 and coefficients less than 6.0 were neglected. The resampling is performed starting from the longest edge of the triangle and processing towards the vertex where two other edges meet. The resampled data set included 65 x 65 pixels. The original triangular patch is presented in Figure 5(a) and image after resampling, coding and decoding is shown in Figure 5(b). Figure 6 shows the histogram of the image data before and after the transformation. The PSNR of the decoded image was 37.50 dB and 89.61 per cent of the wavelet coefficients was zero.

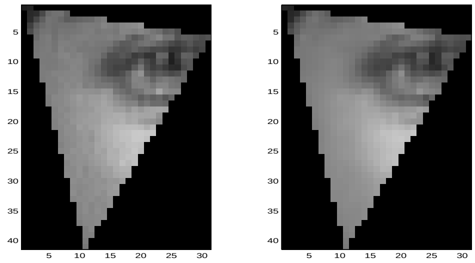

Figure 5: (a) Original and (b) decoded triangular patch.

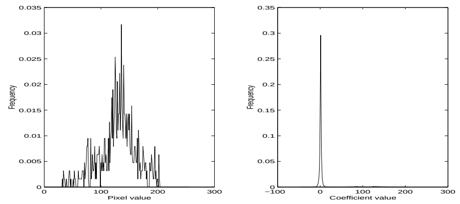

Figure 6: (a) Histogram before and (b) after the transform.

As an example, triangular patch shown in Figure 5(a) is warped by rotating it -40 degrees and scaling horizontally by value 1.2 and vertically by value 1.5. Warping was done by multiplying coefficients by the affine transformation matrix and then interpolating. The result is shown in Figure 7. Affine warping works reliably and only very few bits needs to be transmitted as long model failure does not occur. Therefore, the size of the triangles should be such that the data within the triangle undergoes the same transformation. In any case, in order to achieve good performance high quality motion estimates are required.

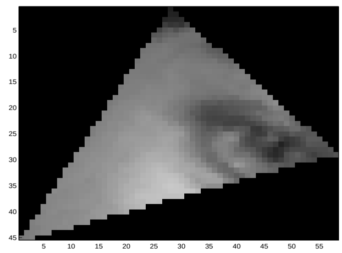

Figure 7: Warping of a triangular patch by rotating and scaling.

Coding of a set of consecutive frames from Miss America sequence is considered next. Faces are first segmented from the background. The mesh over the face region is obtained using a collection of detected feature points and performing Delaunay triangulation on the point set. A total of 82 triangles were used in coding. Image data within triangles is coded as explained earlier. The first frame from original sequence is presented in Figure 8(a) and decoded frame in Figure 8(b). The PSNR of the decoded face image was 36.76

dB and number of non-zero coefficients was 6059 and the total number of pixels was 8017.

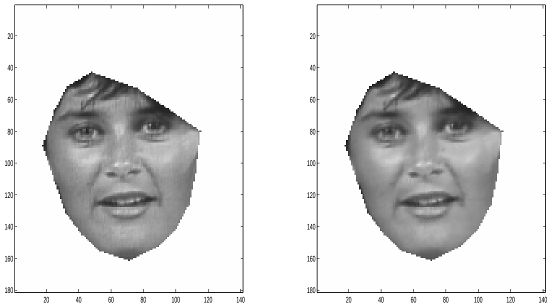

Figure 8: (a) Original frame (1st) and (b) reconstructed frame.

Next coding of the fourth frame is studied. Mouth and eyes are coded separately from the fourth frame because model failure occurs there. In practise model failure means that the new image block cannot be obtained through affine warp of the corresponding block in the preceding frame. Other areas are warped using the data from preceding frame and estimated affine motion parameters. The original frame is presented in Figure 9(a) and reconstructed frame in Figure 9(b). The PSNR of the warped image was 29.59 dB and number of non-zero coefficients was 1257 and the total number of pixels was 7136.

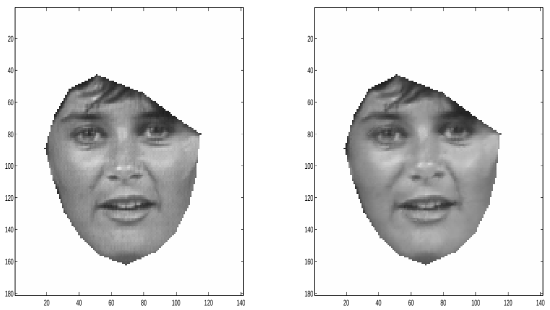

Figure 9: (a) Original frame (4th) and (b) warped 1st frame.

Same coding is applied also to the seventh frame of the image sequence. Mouth and eyes are coded separately from the seventh frame. The original frame is presented in Figure 10(a) and reconstructed frame in Figure 10(b). The PSNR of the warped image was 28.73 dB and number of non-zero coefficients was 747 and the total number of pixels was 7163.

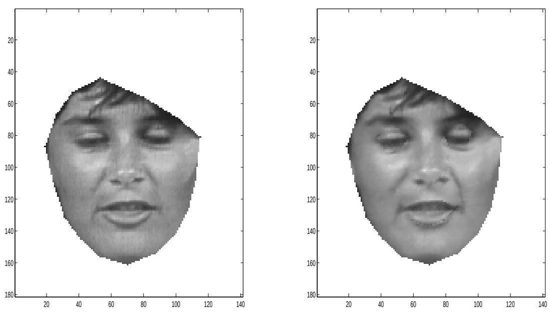

Figure 10: (a) Original frame (7th) and (b) warped 1st frame.

5. CONCLUSION

Triangles can represent objects of arbitrary shape and topology. In this paper, a coding method for triangular (and rectangular) image patches based on B-spline wavelet filter banks and affine warps is introduced. A resampling process is developed in order to obtain a rectangular arrangement of points from triangles. This is required to be able to apply a dyadic filter bank to the data. Shape (contour) of the video object is represented using Bezier curves computed from the mesh elements at the boundary. Filter banks yield high quality reconstruction in simulation and the obtained detail coefficients are highly concentrated. Affine motion model and B-spline basis allows for warping mesh elements accordingly as well as applying local deformations with the filter bank. The performance of the proposed method for coding image sequences depends on the correctness of estimated motion parameters and on the rate of occurring model failures. If no model failure occurs only very few bits/frame need to be transmitted.

In the proposed method, there is a fundamental trade-off between achieving a high compression ratio and avoiding model failures. Small triangles are more likely to obey affine motion model and consequently warping works better whereas model failures occur more frequently for larger triangles. It is easier to achieve a high compression ratio by coding large triangles with the filter bank. Using small triangles requires that large number of vertices is tracked and transmitted. Moreover, small triangles allow only very few levels of analysis filtering.

6. REFERENCES

[1] Aizawa, K., Huang, T.S., “Model-Based Image Coding: Advanced Video Coding Techniques for Very Low Bit-Rate Applications”, *Proceedings of the IEEE*, Vol. 83, No. 2, pp. 259-271, 1995.

[2] Chui, C.K., *An Introduction to Wavelets*, Academic Press, 266 p., 1992.

[3] Egger, O., “Region Representation Using Nonlinear Techniques with Applications to Image and Video Coding”, *Ph.D. thesis, EPFL*, Lausanne, 1997.

[4] Finkelstein, A., Salesin, D.H., “Multiresolution Curves”, *Proceedings of SIGGRAPH ´94*, pp. 261-268, ACM, 1994.

[5] Koivunen, V., Vezien, J-M., “Machine Vision Tools for CAGD”, *Int. Journal of Pattern Recognition and Artificial Intelligence*, Vol. 10, No. 2, pp. 165-182, 1996.

[6] Lounsbery, M., “Multiresolution analysis of surfaces of arbitrary topology”, *Ph.D. thesis, University of Washington*, 1994.

[7] Lyche, T., Mørken, K., “Spline-wavelets of minimal support”, *Numerical Methods of Approximation Theory*, Vol. 9, (Braess, D., Schumaker, L.L., Eds), pp. 177-194, Birkhäuser, Basel, 1992.

[8] Quak, E., Weyrich, N., “Decomposition and Reconstruction Algorithms for Spline Wavelets on a Bounded Interval”, *Applied and Computational Harmonic Analysis*, Vol. 1, No. 3, pp. 217-231, June 1994.

[9] Sikora T., “Low-Complexity Shape-Adaptive DCT for Coding Arbitrarily Shaped Image Segments”, *Signal Processing: Image Communication*, Vol. 7, pp. 381-395, 1995.

[10] Sikora, T., “MPEG Digital Video-Coding Standards”, *IEEE Signal Processing Magazine*, Vol. 14, No. 5, pp. 82-100, September 1997.