ON THE TRADE-OFF BETWEEN CONTOUR-ADAPTIVE TEXTURE CODING AND LOSSY SHAPE CODING

Karsten Schröder

Lehrstuhl für Nachrichtentechnik University of Dortmund 44221 Dortmund, Germany sr@nt.e-technik.uni-dortmund.de

ABSTRACT

Shape-adaptive texture coding is often being perceived as a mean to achieve increased coding efficiency, e.g. due to a better representation of correlation properties along the boundary between two objects which should lead to a higher transform gain. However, if data rates demand a lossy encoding of contours, an improvement in transform gain becomes questionable.

This paper investigates the potential coding gain which may result from shape-adaptive texture coding using DCT basis functions under the constraint of both lossless and lossy shape coding. For this purpose, a 1D texture model is derived with reflects synthetical as well as "natural" borders between objects.

1. INTRODUCTION

A discussion of benefits and theoretical limits of objectbased coding techniques accompanies the ongoing standardisation of MPEG-4: Clearly, with objects being extracted from a video sequence, coded individually in a layered manner and composed at the receiver's side, an enhanced accessibility of video material is achieved, providing so-called *functionalities* [ISO94].

The layered composition of video sequences as described above requires an encoding of object contours. Besides enabling certain functionalities, a potentially higher prediction gain is expected since motion discontinuities at the object's contour are likely and thus easy to employ. Furthermore, an improvement in transform coding may result from adjusting the transform domain to highly correlated object textures. Obviously, the accuracy of the encoded shape is of significant importance for the eventual coding gain. This has to be taken into account especially for object-based coding at very low bit rates, when lossy shape coding becomes inevitable.

In the following, the interdependencies between lossy shape coding and methods for contour-adaptive texture cod-

ing are investigated. Since most region-based transform coding techniques currently considered for standardisation are derived from the DCT (SADCT [Siko95], block padding [ISO96]), the use of DCT basis functions will be assumed throughout this paper.

2. OBJECT BORDERS: TEXTURE MODEL AND ADAPTED TRANSFORM CODING

The following investigations will be carried out using a onedimensional model of a horizontal image line at the transition between two adjacent objects. Both the left and the right object textures shall be represented by first order autoregressive processes [Jain89] with correlation coefficients ρ_L and ρ_R , and variances σ_L^2 and σ_R^2 , respectively, forming a multivariate Gaussian distribution with zero-mean random variables. Moreover, correlation coefficients between random variables belonging to different object textures are set to a constant value of r_u .

Both model textures are combined at edge location κ with κ pixels belonging to the left object, and $N - \kappa$ belonging to the right object. As depicted in Fig. 1, the resulting texture is finally being low-pass filtered with matrix Φ , containing shifted duplicates of the symmetrical impulse response with length 2V + 1. This filtering is intended as a simple model of the optical filtering inherently involved during image acquisition.

A number of 20000 image vectors were arbitrarily selected from sequence Akiyo and were used to estimate the values given in Table 1 in order to establish a set of representative model parameters.

The texture model introduced above is used to investigate the transform gain achievable by an contour-adaptive transform based on separate DCTs. In order to take lossy encoding of the object's contour into account, a first transform is carried out over a domain of $\tilde{\kappa} = \kappa + \varepsilon$ pixels and the second transform over the remaining $N - \tilde{\kappa}$ pixel, where ε denotes a potential contour approximation error introduced



Figure 1: Model of the texture in the transition area of two adjacent objects and subsequent transform

| texture of Akiyo (general) | $\sigma_{a} = 58,40$ | $ \rho_a = 0.9632 $ |
|------------------------------------|--------------------------------|------------------------------|
| texture of Akiyo (left side) | $\widetilde{\sigma}_a = 74,32$ | $\widetilde{ ho}_a = 0,7276$ |
| background texture (general) | $\sigma_h = 31,67$ | $ \rho_h = 0,9916 $ |
| background texture (left of akiyo) | $\widetilde{\sigma}_h = 17,97$ | $\widetilde{ ho}_h = 0,7012$ |

Table 1: Model parameters of object textures extracted from MPEG-4 sequence Akiyo

by lossy shape coding.

This procedure generates an orthogonal transform of

$$\widehat{\mathbf{X}} = \widetilde{\mathbf{D}} \cdot \mathbf{X}$$
 with $\widetilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_{\widetilde{\kappa}} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{N-\widetilde{\kappa}} \end{bmatrix}$, (1)

where $\mathbf{D}_{\tilde{\kappa}}$ and $\mathbf{D}_{N-\tilde{\kappa}}$ are representing transform matrices of a $\tilde{\kappa} - DCT$ and $(N - \tilde{\kappa}) - DCT$, respectively (refer to Fig. 2). The coefficients δ_{ij} of the $\xi \times \xi$ matrix corresponding to a $\xi - DCT$ are defined as [Jain89]

$$\delta_{ij}(\xi) = \alpha(i) \cdot \cos \frac{(2j+1) i\pi}{2\xi}$$
(2)

with

$$\alpha(i) = \begin{cases} \sqrt{\frac{1}{\xi}} & i = 0 \\ \sqrt{\frac{2}{\xi}} & else \end{cases}$$
(3)

 $i, j \in \{0, \ldots, \xi - 1\}.$

As a reference, a single DCT over N pixels will be used during evaluation. In contrast to this regular N-DCT, a contour-adaptive transform with matrix $\tilde{\mathbf{D}}$ will be referred to as dual DCT (Eq. (1)).

3. COMPUTATION AND RESULTS

Both contour adaptive DCT and regular N-DCT are evaluated by means of the commonly used transform gain G_{TC} [Jaya84][Siko95] (Eq. (4)) that measures the distortion improvement of transform coding over PCM at equal data rates. However, the coding overhead introduced by shape coding shall be neglected in this comparison, since shape coding can be perceived as an inevitable step in order to achieve object based functionalities.

The variances $\hat{\sigma}^2$ of random vector $\hat{\mathbf{X}}$ can be extracted from the covariance matrix $\hat{\mathbf{C}} = \mathbf{D} \mathbf{C} \mathbf{D}^{\mathsf{T}}$, with \mathbf{C} denoting the covariance matrix of \mathbf{X} before transformation, and \mathbf{D} corresponds either with the *N*-DCT (N = 8) or with the dual DCT.

$$G_{TC} = 10 \log \frac{\frac{1}{N} \sum_{i=0}^{N-1} \widehat{\sigma}_i^2}{\left(\prod_{i=0}^{N-1} \widehat{\sigma}_i^2\right)^{\frac{1}{N}}}$$
(4)

The following diagrams show comparisons between the introduced model and corresponding simulations which are based on image vectors taken directly from sequence Akiyo.

Synthetical edges (Fig. 3, left) can be understood as a model for artificially combined textures common to computer graphics. A large number of image vectors have been selected from arbitrarily regions of both objects (akiyo and background) to yield simulation results, whereas $\sigma_L = \sigma_a$, $\sigma_R = \sigma_h$, $\rho_L = \rho_a$, and $\rho_R = \rho_h$ have been chosen for the model-based computation. With $r_u = 0$ and $\Phi = E$ both textures are ideally de-correlated from each other. As expected, the dual DCT outperforms the *N*-DCT for $1 \leq N < 7$ in simulation as well as in computation.

In contrast to synthetical edges, the model for "natural" edges involves parameters $\sigma_L = \tilde{\sigma}_a$, $\sigma_R = \tilde{\sigma}_h$, $\rho_L = \tilde{\rho}_a$, and $\rho_R = \tilde{\rho}_h$, respectively. Moreover, r_u has been set to 0.5 and filtering with a low-pass filter (impulse response



Figure 2: Contour-based DCT transform of image vector X



Figure 3: Transform gain over pixels in edge domain as a function of edge position κ along synthetical edges (left) and filtered edges (right)

 $\{0.25; 0.5; 0.25\}$) is carried out via matrix Φ . Computational results for such filtered edges are depicted in Fig. 3 (right), together with transform gains determined in simulations. The gain of the contour-based transform is reduced substantially by non-zero correlations between pixels from different object textures. However, an average improvement of 1.5 dB is still perceivable.

The results above hold only for lossless contour coding. Since the transform kernel of the dual DCT does no longer match the actual correlation profile if the contour approximation error ε differs from zero, the potential coding gain is diminished.

More general results can be derived by feeding the model with a continously increasing ratio of the variances of both textures and averaging the resulting transform gain over all possible edge position κ , provided that κ is uniformly distributed.

For the case of lossless shape coding, contour-based texture coding leads to an improvement in the averaged transform gain especially if the underlying video material has been composed synthetically. However, it should be noted that the significant increases in transform gain depicted in Fig. 4 and Fig. 5 hold only for pixels along the object's border. In the case of lossy shape coding, the potential transform gain is reduced significantly such that the dual DCT may even perform worse compared against a regular N-DCT. Furthermore, if there is a substantial difference in texture variances of both objects, an error towards the object with greater variance is extremely critical, whereas enlarging the same object leads to a slighter deterioration or — in certain situations — still to an improvement over the N-DCT.

4. CONCLUSION

In this paper, a one-dimensional model for the transition of textures along object borders has been established. Based on this model, the transform gain of two separate DCT transforms with transform domains adapted to the corresponding objects was evaluated and compared against a regular DCT with constant length N = 8. While significant improvements were perceivable for lossless shape coding, it was shown that for lossy shape coding improvements are reduced or even inverted into a deterioration. It could be further demonstrated that an enlargement of the object with higher variance is preferable to its minification, if texture variances differ significantly from each other.



Figure 4: Averaged transform gain over pixels in edge domain as a function of the contour approximation error ε with respect to synthetical edges



Figure 5: Averaged transform gain over pixels in edge domain as a function of the contour approximation error ε with respect to filtered edges

5. REFERENCES

- [ISO94] MPEG-4: Package Proposal Description. ISO/IEC JTC1/SC29/WG11 N0821, Nov. 1994
- [ISO96] MPEG-4: Video Verification Model Version 4.0, ISO/IEC JTC1/SC29/ WG11 N1380, Oct. 1996
- [Jain89] A. K. Jain: Fundamentals of Digital Image Processing. Prentice-Hall, New Jersey, 1989

- [Jaya84] N. S. Jayant, P. Noll: *Digital Coding of Wave*forms. Prentice-Hall, New Jersey, 1984
- [Schr97] K. Schröder: Objektbeschreibende Bewegtbildcodierung mit gittergestützter Prädiktion.
 Ph. D. Dissertation, University of Dortmund.
 Fortschritt-Berichte VDI No. 10/128. VDI-Verlag, Düsseldorf, 1997 (in German)
- [Siko95] T. Sikora, B. Makai: Shape-Adaptive DCT for Generic Coding of Video. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, Feb. 1995