IMAGE COMPRESSION USING A CARTOON-TEXTURE DECOMPOSITION TECHNIQUE

Nikola Sprljan, Marta Mrak, Ebroul Izquierdo

Queen Mary, University of London
Mile End Road, E1 4NS, London, UK
{nikola.sprljan, marta.mrak, ebroul.izquierdo}@elec.qmul.ac.uk

ABSTRACT

Based on the assumption that an image is composed of two separated layers - smooth regions and textures, a new image compression technique is established. In our multi-layered compression scenario, a non-linear diffusion is used to smooth out texture and to decompose images into a large scale piecewise smooth component and a residual part consisting of "pure texture". Since the extracted layers have different characteristics, suitable transforms are used to encode each layer individually. Both *PSNR* results and subjective quality are comparable to popular image compression techniques.

1. INTRODUCTION

Conventional wavelet based compression algorithms decompose images into suitable subbands before coding. These methods use the same transform basis to represent large scale piecewise smooth image regions as well as textured areas with high frequency components. Since a single basis cannot approximate these two structurally different components with the same degree of accuracy, conventional wavelet compression does not always lead to optimal results in terms of compression ratio and image quality.

Recently, new strategies have been developed to achieve higher compression ratios using techniques that select adaptively the transform basis in order to exploit structural properties of images in the greater extent [1], [2]. In [2], the wedgelet transform for modelling of an image as a combination of edge cartoon and remaining texture is used, but the overall performance is usually impaired by inefficiency of wedgelet coding. Method from [1] suffers from two main shortcomings: it is hard to determine automatically the bit-allocation for each layer and it generally fails to preserve edges in the cartoon component. Although objective quality measures show good performance, these shortcomings lead to visually annoying artefacts around edges, especially at low bit rates.

In this paper we exploit the fact that image edges represent perceptually relevant image information, in order to design a coding scheme targeting high compression ratio and better subjective image quality. Contrasting single-layer techniques, we propose the image decomposition into independent layers and coding of each layer using the most suitable basis. Two basic image components we consider are the "cartoon-like" layer and the residual part representing "pure texture". This structural image decomposition is achieved by non-linear diffusion. The resulting image components can be efficiently approximated by different transforms and a high degree of compression can be achieved using suitable basis for encoding of each layer individually.

The results of a comprehensive experimental evaluation show that for low bit-rates the proposed technique achieves objectively similar and subjectively superior results as compared with popular image compression algorithms.

2. MULTILAYERED IMAGE CODING

The basic building blocks in a multilayared image coding system are layer decomposition unit, image layer coding unit and rate controller. Let $\boldsymbol{I}$, $\boldsymbol{C}$, $\boldsymbol{T}$, $\boldsymbol{C'}$, $\boldsymbol{T'}$ and $\boldsymbol{I'}$ denote the original image, cartoon, texture, reconstructed cartoon, reconstructed texture and reconstructed image matrices respectively. The transform basis for cartoon and texture are T_C and T_T. A block diagram of the coding system is shown in Fig. 1.

As the decomposition process has to meet compression requirements, i.e. cartoon and texture must be properly separated, a distortion measure is used to evaluate the decomposition process. Since the usual transform basis used in compression are orthogonal or near-orthogonal biorthogonal basis, error estimation can be performed in the transform domain (e.g. $\|\boldsymbol{C} - \boldsymbol{C'}\| \approx \|\hat{\boldsymbol{C}} - \hat{\boldsymbol{C}}'\|$), where $\hat{\boldsymbol{A}} = \mathbf{T}(\boldsymbol{A})$ is the transformed $\boldsymbol{A}$ and $\|A\|$ is the Frobenius norm of matrix $\boldsymbol{A}$ in the notation of L_2 vector norm, or energy of the signal represented with $\boldsymbol{A}$. The relation for

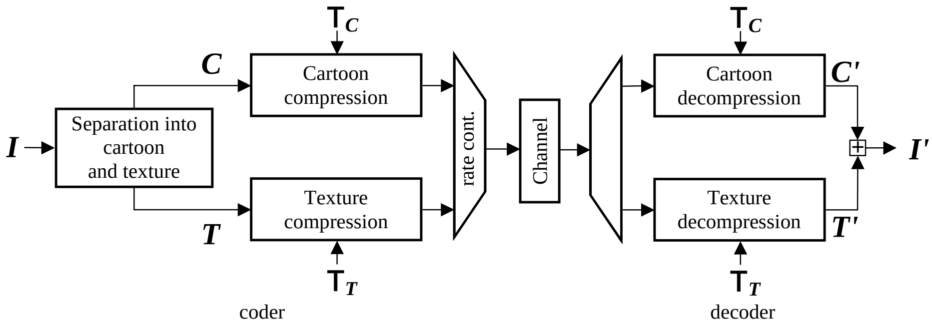


Fig. 1: A cartoon-texture coding system

the distortion D can then be expressed as:

$$\begin{aligned} D &= \|\boldsymbol{I}-\boldsymbol{I}'\|^2 \\ &= \|(\boldsymbol{C}+\boldsymbol{T})-(\boldsymbol{C}'+\boldsymbol{T}')\|^2 \\ &= \|\boldsymbol{C}-\boldsymbol{C}'\|^2+\|\boldsymbol{T}-\boldsymbol{T}'\|^2+2\langle \boldsymbol{C}-\boldsymbol{C}',\boldsymbol{T}-\boldsymbol{T}'\rangle \\ &\approx \left\|\hat{\boldsymbol{E}}_{\boldsymbol{C}}\right\|^2+\left\|\hat{\boldsymbol{E}}_{\boldsymbol{T}}\right\|^2+\Re(\boldsymbol{CT}) \end{aligned}, \qquad (1)$$

where $\boldsymbol{E}$ is the error image and $\Re(\boldsymbol{CT})$ consists of cross-correlation products $\langle \boldsymbol{CT}\rangle$, $\langle \boldsymbol{CT'}\rangle$, $\langle \boldsymbol{C'T}\rangle$ and $\langle \boldsymbol{C'T'}\rangle$ (in the notation of vector dot-product). If $\boldsymbol{C}$ and $\boldsymbol{T}$ are very weakly correlated, then these products tend to cancel each other out and as the result $\Re(\boldsymbol{CT})$ term can be neglected in the distortion estimation. Therefore, the contribution of this term into overall error estimation can be determined by the value of $\langle \boldsymbol{CT}\rangle$, which should be kept as low as possible. Expressed in other words, by targeting the orthogonality of cartoon and texture part the redundancy of the same content in separate compression methods is diminished.

2.1. Structural Layer Decomposition

The objective of the layer decomposition is to decompose the image into a cartoon and a texture layers. The cartoon layer should be textureless and with well preserved edges while the residual texture should consist of highly oscillatory areas. This decomposition can be achieved by non-linear diffusion filtering. The classical Perona-Malik non-linear diffusion process is modelled by the following partial differential equation (PDE):

$$\frac{\partial I}{\partial t} = div\left[c(I,\sigma,k)\nabla I\right], \qquad (2)$$

where I is the original image, and $c(I, \sigma, k)$ is the conductivity function calculated for each image pixel. In this work $c(I, \sigma, k)$ is defined as:

$$c(I,\sigma,k) = e^{-\left(\frac{\nabla(I*G_\sigma)}{k}\right)^2}. \qquad (3)$$

In (3) the parameter k is threshold for the image gradient. It defines the boundary between edge preservation (and enhancement) and diffusion in particular image areas. By convolving the image with the Gaussian kernel G_σ of standard deviation σ, noise with impulse gradient lower than k is smoothed out. Throughout our experiments we noticed that function c, defined by (3), cannot distinguish edge-like textures from real edges. Therefore, we use edge

a) original image

b) gradient map

Fig. 2: Test image and associated gradient map used in the process of layer decomposition

a) original image b) cartoon c) texture

Fig. 3: Result of non-linear diffusion for image “Barbara” (11 iterations)

information obtained prior to diffusion by computing the gradient of the image convolved with Gaussian with σ_k large enough to smooth out false edge information in textured areas (Fig. 2). Edge information adjusts automatically the parameter k to be small for areas with clear edges (bright areas on Fig. 2.b) and to be large for other image parts. The relation between edge information and parameter k was set empirically.

The amount of change in the energy of the texture and cartoon layers at each iteration step is used as a stopping criterion of the iterative filtering process. When diffusion saturates and the "cartoon difference" between two consecutive iterations falls below a certain threshold the iteration is stopped. Additionally, the value of $\langle \boldsymbol{CT} \rangle$ is monitored in order to provide an estimation of the value of $\Re(\boldsymbol{CT})$. By comparing this value with the total energy of the original image a measure of the "goodness" of the cartoon-texture decomposition can be obtained. In most cases this value remains low indicating that our method achieves high degree of decorrelation between texture and cartoon layers.

In Fig. 3, an example of the layer decomposition for the test image "Barbara" is given. Only a part of whole image is shown for 11 iterations.

2.2. Cartoon and Texture Compression

Due to their vanishing moments, i.e. polynomial reproduction properties, some wavelet basis are good candidates to approximate cartoon-like images. For this reason, for the compression of the cartoon layer we use a dyadic wavelet transform, which optimally captures large scale trends in the images, and set partitioning in hierarchical trees (SPIHT) algorithm [3]. Several experiments were conducted with different wavelet basis to find out that the biorthogonal coiflet 22/14 produces the best results. This wavelet is constructed via factorisation of Daubechies polynomials and has 7 and 5 vanishing moments, for analysis and synthesis filters, respectively.

Due to their ability to capture and represent oscillatory patterns, the texture layer is coded either with local cosines (LCT) [4] or fast wavelet packets (FWP) [5]. The performance of these two coders was assessed to find out the best choice. A statistical evaluation of the results revealed that these two methods performed comparably. However, the performance of the LCT strongly depends on several adjustable parameters, e.g., bell type, steepness of the bell, quantisation step, that do not always give the optimal results with the default settings. For that reason, FWP was finally selected.

3. EXPERIMENTAL RESULTS

The proposed approach was compared with SPIHT, JPEG2000 and FWP coding techniques. An objective evaluation in terms of *PSNR* was carried out for different choices of bit allocation between the two layers. It was observed that if the energy of the texture layer after process of separation is less than 5% of the energy of the original image, coding of cartoon and texture layers independently could not achieve good compression gain. For that reason low textured image are not decomposed and the whole image is considered to be cartoon layer. The obtained gain was observable only for the low bit rates, since not completely decorrelated catoon and texture part reduce the overall performance usually not before the magnitude of the correlated features become significant, i.e. when the quantization step is small enough.

Fig. 4 summarises the result of the objective comparison for different compression techniques for test image "Barbara". In this representation b is total bit rate of the compressed image and b_c is the amount of bits assigned to the cartoon layer. The proposed compression algorithm is labelled as C+T (cartoon+texture).

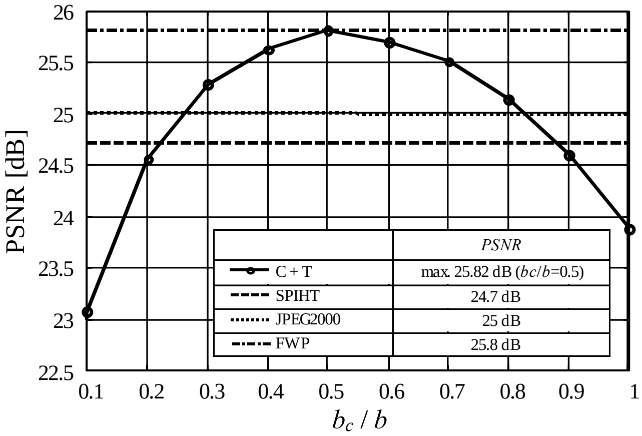
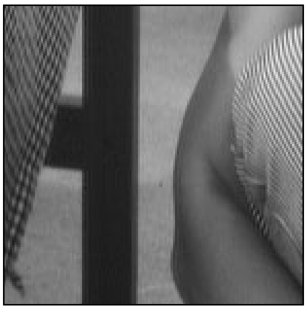


Fig. 4: Comparison of performance at 0.1bpp



original

SPIHT

JPEG 2000

FWP

C+T

Fig. 5: Visual comparison of compressed images at 0.1bpp

Fig. 5 shows the visual appearance of the decoded images using tested methods. The results show that SPIHT and JPEG2000 failed to completely capture texture patterns. While FWP achieves comparable compression ratio for the texture as C+T, it shows the typical artefacts of wavelet packets: highly noticeable ringing around edges and large scale errors. These artefacts are less noticeable in the C+T compression. Comparative results with other techniques and test images are similar. For more comprehensive results, reader is referred to [6].

4. CONCLUSION

Targeting highly efficient still image compression application, multilayered image representation is established. Its basic idea is to separate an image in two layers, namely cartoon and texture, which have different statistical properties. Decorrelated layers are used as inputs to an efficient lossy coding system. For each of image layers, suitable coding mechanism is chosen. The aim of this process is to achieve good subjective image quality on targeting bit-rate.

Experimental results show that the proposed technique obtains objectively similar and subjectively superior results on lower bit-rates compared to well-established image compression algorithms. However, on higher bit-rates, the performance deteriorates due to the redundancy imposed by imperfect cartoon and texture layer separation.

Acknowledgment

We wish to acknowledge support provided by the project European Network of Excellence SCHEMA (Grant Nr. IST-2001-32795).

References:

[1] F.G. Meyer, A.Z. Averbuch and R.R. Coifman, “Multilayered image representation: application to image compression”, *IEEE Trans. Image Proc.*, vol. 11, no. 9, pp. 1072-1080, Sep. 2002.
[2] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk, "Rate-distortion optimized image compression using wedgelets", *Proc. IEEE International Conference on Image Processing (ICIP)*, USA, Sep. 2002.
[3] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partioning in hierarchical trees", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, pp. 243-250, June 1996.
[4] F.G. Meyer, "Image compression with adaptive local cosines: A comparative study", *IEEE Trans. Image Processing*, vol. 41, pp. 616-629, June 2002.
[5] F.G. Meyer, A.Z. Averbuch, and J.-O. Strömberg, "Fast adaptive wavelet packet image compression", *IEEE Trans. Image Processing*, vol. 9, pp. 792-800, May 2000.
[6] N. Sprljan and E. Izquierdo, “New perspectives on image compression using a cartoon-texture decomposition model”, *Proc. 4th EURASIP Conf. on Video/Image Processing and Multimedia Communications*, Croatia, July 2003