Improved Segmentation and Extrapolation for Block-Based Shape-Adaptive Image Coding

Ying Zhao and David Malah
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
y_zhao@hotmail.com   malah@ee.technion.ac.il

Abstract

In this work, we address the issues of image segmentation and boundary block padding involved in shape-adaptive image coding. Image segmentation helps to exploit human visual system (HVS) characteristics for bit-rate reduction in coding an image. In the context of block-based shape-adaptive coding, segmentation allows the use of more effective boundary block coding techniques than conventional methods. Segmentation algorithms considered are based on mathematical morphology tools. Following a brief discussion of the drawbacks of two reference segmentation algorithms, an improved *edge detection, local-activity classification* segmentation algorithm is proposed. Simulation results indicate that the proposed algorithm enjoys the advantages of subjectively accurate contour location, simple image partition and lower computational load. In order to code boundary blocks efficiently, an optimal block padding approach, which minimizes the l_1 norm of the corresponding transform coefficients, is proposed. The proposed scheme draws on a basis pursuit problem, which uses linear programming for its solution. It is shown that the proposed scheme provides better coding results (in terms of bit-rate reduction) than various other block-based shape-adaptive coding techniques.

1 Introduction

Block-based 2D-DCT is a conventional transform coding technique used for image compression (e.g., JPEG [1]). Artifacts known as *blockiness* and the *blurring of edges* arise at lower bit rates. One reason is that HVS puts special emphasis on edges and, therefore, instead of coding all the blocks in the same manner, blocks with an edge in it should be treated differently. Therefore, various segmentation-based shape-adaptive image coding techniques were proposed, which exploit the HVS properties to overcome the problems stated above. Generally, such a system consists of the following building blocks:

- Image segmentation – partitions the original image into arbitrarily shaped homogeneous segments, such that each segment has a specific local characteristic.
- Contour coding – codes the segment shape (contour) information.
- Texture coding – represents and codes the segment content.

There are two main approaches for shape-adaptive texture coding – region-based [2, 3] and block-based [4, 5]. The focus of this work is on image segmentation and block-based shape-adaptive coding [6]. We examined two Mathematical-Morphology-based segmentation algorithms [7, 8]. Our experiments have shown that these algorithms suffer from over-segmentation and false contours. Therefore, we propose a different algorithm which is better suited to the coding task. Two shape-adaptive coding techniques, proposed in [4], namely, shape-adaptive DCT (SADCT) and low-pass extrapolation (LPE) padding, are also investigated. These schemes, although simple, do not provide sufficient bit reduction. Therefore, a novel boundary-block coding scheme, which is optimal (in minimal l_1 sense) is proposed.

The paper is organized as follows: In section 2, the proposed segmentation algorithm, which is based on *edge detection*, and *local-activity classification* is presented. Various block-based shape-adaptive coding techniques are discussed in section 3, including SADCT, LPE padding and the proposed novel scheme - *minimal l_1 norm* block padding. Section 4 demonstrates the coding results obtained by a JPEG baseline system in comparison with the various shape-adaptive coding systems discussed. Conclusions and suggestions for further studies are given in section 5.

2 Image Segmentation Algorithm

The purpose of image segmentation is to partition the input image into segments as the HVS does. Therefore, a segmentation algorithm should be able to detect adjacent regions with different texture characteristics and define boundaries to separate them. An algorithm based on edge-detection and local-activity classification is proposed in this work [6]. It consists of the following stages:

- **Simplification Stage:** Morphological opening-closing by reconstruction (MOCR) filters [7] have a good ability in image simplification. That is, they remove less important regions without causing severe distortions to the region boundaries. But, if small size filters are used, they do not result in sufficient simplification, while large size filters are computational costly. To cope with these drawbacks of morphological filters, we proposed to combine a local-activity classification technique and morphological filters. Three operations are executed in the simplification stage of the proposed scheme. The original image is first filtered by MOCR filters to remove noise and less important detail. Then, a pixel-wise local activity measure is calculated by applying a modified Prewit operator [9] on the simplified image. This results in a new four-level image, which we call the *characteristic image*. Each pixel in the characteristic image is classified into one of four categories: flat, low structured, high structured and edge. The corresponding gray levels assigned to these classes are 255, 180, 100 and 0, respectively. Moreover, this characteristic image is further filtered by MOCR filters with a smaller size. The resulting simplified characteristic image, which also has four levels can be used for marker extraction. Empirical results indicate that the size of the morphological filters can be chosen to be 5×5 in the initial step and 3×3 in the following step.

- **Marker Extraction Stage:** Within the simplified characteristic image, a certain number of regions having large size (flat zones) are produced. The boundaries of these regions are well defined, so that we can use a marker extraction technique to detect the inner part of the regions.

- **Region Growing Stage:** A region growing algorithm is used to assign undecided pixels to adjacent regions. The region growing procedure is performed by appending undecided pixels to a specific neighboring marker, if the absolute difference between the pixel gray-level and the mean gray-level of the marker in the original image is less than a specified threshold value. When all the pixels complying with this threshold have been appended, the threshold value is increased and the procedure is repeated until all pixels of the image are assigned.

- **Region Refinement Stage:** Adjacent segments with a perceived difference along their boundary are easily identified after the region growing. However, in the simplification stage, some important detail could have been eliminated and merged into a big segment, because of their comparatively small size or low contrast. Therefore, an optional region refinement stage is introduced. The purpose of region refinement is to retrieve this missing detail from big segments. At present visual inspection is used to decide the necessity of this stage. The experiments indicate that in most applications, the biggest segment generated by the previous region growing stage, denoted as background segment, is the integration of the poorly segmented regions. The refinement stage consists of the following steps:

  1. Histogram equalization [10]: An operation applied to the original image to increase the contrast between regions with poor contrast.
  2. Simplification: The equalized image is simplified, as in the previous simplification step. Flat zones, which indicate the inner parts of the low contrast regions, are generated.
  3. Marker Extraction (within poorly segmented regions): The flat zones generated by the simplification step stated above, can be identified by the marker extraction technique. Since only the poorly segmented regions need to be refined, the marker extraction is performed under the constraint of the previous segmentation output (the point "A" shown in the block diagram).
  4. Region Growing: Region growing is redone to redefine the contours of the regions.

- **Region Merging & Contour Simplification Stage:** Because over-segmentation typically exists in the segmentation output, i.e., a perceptually single segment may be split into several fragments, small segments exist in the segmentation output. Thus, region merging techniques are performed as follows: To merge split segments, the average contrast of pairs of adjacent regions is calculated. If the value is lower than a threshold, the boundary which separates this pair of regions is removed and the two regions are merged into one region. This merging procedure is carried out by successively eliminating the boundary which is between regions with the lowest average contrast.

  The segmentation algorithm does not have a tool to control the complexity of the contours generated. Therefore, simplification of the segments contours is usually necessary. A majority filter [11] is used for this task. By changing the size of the majority filter used, the smoothness of the generated contour can be controlled.

  To eliminate small regions, which are generated by contour simplification or the segmentation algorithm itself, the contrast values between a small region and all its adjacent regions are calculated, and the region is merged to its lowest contrast neighbor.

2.1 Demonstration of Segmentation Results

Several additional 512×512 images segmented by the proposed algorithm are demonstrated in this section. It is seen in the pictures shown in Fig. 1 and Fig. 2, that adjacent regions are separated if steep edges exist between them, while regions with similar characteristic are clustered into one segment. The advantages of the resulting partitions are: First, neighboring regions with similar characteristic are grouped into one region, so that the following shape-adaptive coding can fully exploit the specific characteristic of the segment to achieve better compression. Second, the contours generated by this algorithm are relatively simple, therefore, reducing the contour-coding cost. The examples in these figures show that the proposed segmentation algorithm matches a wide range of images, besides the advantages of a well defined segmentation output and a relatively lower computational load as compared to the algorithms of [7, 8].

3 Coding of Region Content

As we mentioned above, the main issue in block-based shape-adaptive coding is the coding of boundary blocks. Currently used methods are SADCT, proposed by T.Sikora et. al. in [5], and LPE padding which is suggested for MPEG-4 [4]. These methods even though simple, are not efficient in the sense of bit rate reduction.

Motivated by the data extrapolation idea, we were looking for an optimal block data extrapolation technique. As shown in the next subsection, we were able to state a data extrapolation problem which is equivalent to optimal basis selection (from an over-complete bases set), called basis pursuit in [12]. This technique minimizes the l_1 norm of the corresponding transform coefficients and has the property that it results in a number of representation coefficients equal to the signal dimension.

3.1 Optimal Extrapolation

The standard 2D-IDCT of a $N \times N$ matrix A is given by: $S_{N\times N} = C^T_{N\times N} A_{N\times N} C_{N\times N}$. It is well known that it can be calculated by a corresponding 1D-IDCT: $\underline{s} = [\mathcal{C}_{N^2\times N^2}]\underline{a}$. This result is obtained by mapping matrices A and S into vectors $\underline{a}$ and $\underline{s}$, respectively, (by row ordering) and replacing the cosine transform matrix $C_{N\times N}$ by $\mathcal{C}_{N^2\times N^2} = C^T_{N\times N} \otimes C^T_{N\times N}$, where $\otimes$ denotes a Kronecker product. Moreover, the data can be arranged such that $\underline{s} = [\underline{\beta}^T_{m\times 1} \quad \underline{\theta}^T_{(N-m)\times 1}]^T$, where $\underline{\beta}_{m\times 1}$ denotes the part of known data, and $\underline{\theta}_{(N-m)\times 1}$ denotes the remaining unknown data in $\underline{s}$ (to be extrapolated). Accordingly, the resulting inverse transform matrix, denoted as Φ, can be decomposed as:

$$\left[\begin{array}{c} \Phi' \\ \Phi'' \end{array}\right]_{N\times N} \underline{a} = \left[\begin{array}{c} \underline{\beta} \\ \underline{\theta} \end{array}\right]_{N\times 1} \tag{1}$$

Since the only known data is $\underline{\beta}$, the solution for $\underline{a}$ is non-unique. By considering a minimal l_1 norm solution, the following constrained minimization problem is obtained:

$$\min \|\underline{a}\|_1 \;\text{ subject to }\; \Phi' \underline{a} = \underline{\beta} \tag{2}$$

An equivalence to the basis pursuit problem posed in [12] is obtained if we consider the N columns of Φ' in (1), (2) to be an over-complete set of basis vectors representing $\underline{\beta}$ (which is of dimension $m < N$). As shown in [12], the solution to (2) can be efficiently obtained by linear programming (LP), and has the property that the number of non-zero components in $\underline{a}$ is equal to m – the dimension of $\underline{\beta}$.

3.2 Linear Programming Solution

Standard linear programming solves a constrained optimization problem defined in terms of a variable $\underline{x} \in R^N$ by

$$\min M^T \underline{x} \;\text{ subject to }\; A\underline{x} = \underline{b}, \;\text{ and to }\; \underline{x} \geq \mathbf{0} \tag{3}$$

where $M^T \underline{x}$ is the objective function, $A\underline{x} = \underline{b}$ is the collection of equality constraints, and $\underline{x} \geq \mathbf{0}$ is a non-negativity constraint.

Now, if we define a $N \times 1$ vector $\underline{a} = \underline{u} - \underline{v}$, $\underline{u}, \underline{v} \geq \mathbf{0}$, then, $\min \|\underline{a}\|_1 = \min \underline{\mathbf{1}}^T \underline{u} + \underline{\mathbf{1}}^T \underline{v}$. The equivalence relies on the fact that at an optimal solution, only one of the two variables u_i or v_i can be nonzero. Hence, $\Phi' \underline{a} = \Phi'(\underline{u} - \underline{v}) = [\Phi' \quad -\Phi'][\underline{u}^T \quad \underline{v}^T]^T$. The problem stated in (2) is then finally reformulated into a standard linear programming problem defined in R^{2N}:

$$\min \underline{\mathbf{1}}^T \left[\begin{array}{c} \underline{u} \\ \underline{v} \end{array}\right] \;\text{ subject to }\; [\Phi' \;\; -\Phi'] \left[\begin{array}{c} \underline{u} \\ \underline{v} \end{array}\right] = \underline{\beta}, \;\; \underline{u}, \underline{v} \geq \mathbf{0} \tag{4}$$

A generic property of the LP solution is that the number of non-zero entries in the solution $\underline{a}$ is exactly the same as the number of known data elements, and the signal energy concentrates in quite few nonzero coefficients. This property benefits the following coding process of the transform coefficients. However, the positions of the non-zero coefficients do not have a predictable pattern.

Figure 1: Segmentation Results of the Proposed Algorithm for the Images ”Lena” and ”Peppers”

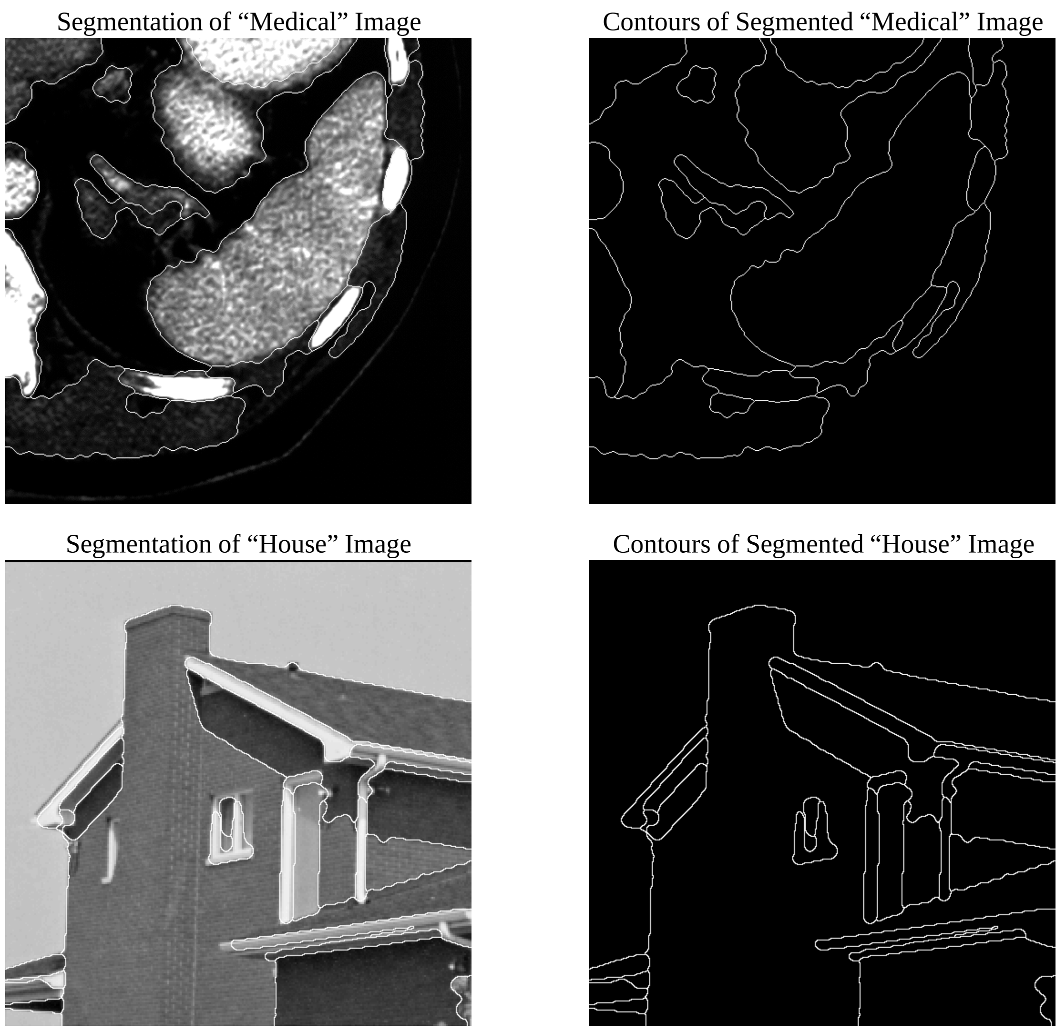

Figure 2: Segmentation Results of the Proposed Algorithm for the Images ”Medical” and ”House”

4 Coding Results

4.1 Block-Based Shape-Adaptive Coding Systems

Following the classical 8×8 block image partitioning, sequential blocks fall into either inner block or boundary block categories. Inner blocks can simply be coded by a JPEG codec [1] (but we can allow rougher quantization). Boundary blocks are coded by the methods we specify below.

We investigated three alternative transform coding methods. The coding methods of SADCT and LPE are similar to those proposed in [5, 4]. To code the transform coefficients calculated by the proposed minimal l_1 norm padding method, we apply a uniform quantizer, and the quantized coefficients are zig-zag scanned. Since the positions of the high energy coefficients are not predictable, it was expected (and was indeed supported by experiments) that the default VLC table in JPEG is not matched to the coefficients statistics. Therefore, arithmetic coding was used instead. Based on the quantized coefficients, a sequence, a_i, of the amplitudes of the nonzero coefficients, and a sequence, z_i, of the number of zeros before each nonzero coefficients, are generated. Arithmetic coding is then applied to encode these sequences. On the decoding side, an inverse procedure is conducted to obtain the extrapolated data block. Since the shape information is transmitted to the decoder, the extrapolated data can then be discarded.

4.2 Simulation Results

This section demonstrates block-based image coding results of two 512×512 still images. We apply the *modified chain coding* method of [13] to code the contours, so that a comprehensive comparison between different shape-adaptive coding methods can be made. Simulation results indicate an advantage (in bit-rate reduction) of the proposed minimal l_1 norm extrapolation method over SADCT and LPE padding. The comparison is summarized in Tables 1 and 2. It is seen that the reconstructed images produced by JPEG and minimal l_1 norm have almost the same PSNR and about the same perceptual quality, but lower rate for minimal l_1 norm. The reason is that a lower reconstruction error is obtained by coding minimal l_1 norm extrapolated boundary blocks, while rougher quantization is used for inner blocks, as mentioned earlier, which compensate each other in terms of PSNR.

In general, block-based shape-adaptive coding techniques follow the conventional rectangular-grid image partitioning (while using also region boundary information). The advantage thus is that they are quite compatible with current widely used image coding systems. On the other hand, since these coding methods can not exploit the global characteristic of a segment, the potential compression capability of segment-based coding is not fully exploited. In particular, in still image coding, a boundary block belongs to more than one segment, so that its coding must be done more than once, which increases the bit-rate. The simulation results indicate that coding with SADCT and LPE do not provide better compression than JPEG. Yet coding with minimal l_1 norm padding outperforms JPEG in some cases. Further studies are needed to identify the kind of images which are better matched to minimal l_1 norm padding. Besides, the complexity of the minimal l_1 norm extrapolation algorithm is still too high for general use. As discussed in the next section, the advantage of the minimal l_1 norm extrapolation technique is expected to be much more pronounced in video coding.

Method	Texture bpp	Contours bpp	Total bpp	PSNR
JPEG	/	/	0.4154	36.134
	/	/	0.2917	33.793
Min. l_1	0.3692	0.0240	0.3932	35.973
	0.2660	0.0240	0.2900	33.794
LPE Padding	0.4609	0.0240	0.4849	36.171
	0.3222	0.0240	0.3462	33.962
SADCT	0.4923	0.0240	0.5163	35.630
	0.3435	0.0240	0.3675	33.831

Table 1: Comparison of Coding Results of the Image "House" by Block-Based Techniques

Method	Texture bpp	Contours bpp	Total bpp	PSNR
JPEG	/	/	0.7320	33.448
	/	/	0.4488	31.879
Min. l_1	0.6172	0.0574	0.6746	33.324
	0.3835	0.0574	0.4409	31.682
LPE Padding	0.7282	0.0574	0.7856	33.093
	0.4463	0.0574	0.5037	31.725
SADCT	0.7336	0.0574	0.7910	33.117
	0.4415	0.0574	0.4989	31.309

Table 2: Comparison of Coding Results of the Image "Peppers" by Block-Based Techniques

5 Discussion

The contributions of this work to image segmentation and shape-adaptive coding are discussed in sections 2.1 and 4.2. Here we mainly discuss potential directions for further study.

Segmentation-based image coding is a relatively new research field. To achieve the expected better reconstruction quality over conventional techniques such as JPEG, related techniques still need to be improved. This research work achieved some positive results, and we suggest here some directions for further studies. The discussion is divided into two parts concerning extensions in both segmentation and shape-adaptive coding.

In segmentation, the decision of whether to introduce the region refinement step for further segmentation is presently done by subjective judgment after the initial segmentation. We hope that an objective criterion could be found in the future to decide when is region refinement needed. Besides, the region refinement itself could be further exploited. In our scheme, the small regions or low contrast regions are assumed to be embedded in the background which, in general, is the biggest segment obtained after the first level of the region growing step. But further studies are needed to devise a scheme which will detect other regions which need to be split further. Moreover, since histogram equalization increases the system's sensitivity to noise, the question of denoising is another problem that one needs to cope with. If these problems are successfully solved, the proposed algorithm can be a promising candidate for a generic still image segmentation algorithm. Another more general extension is to video sequence segmentation. Since the amount of computation time of the segmentation algorithm is greatly reduced, it can be used for the intra-frame image segmentation. Furthermore, by exploiting the connection between segments in consecutive frames, an image sequence segmentation system could be generated.

In shape-adaptive image coding of still images, the improvements obtained are not significant. The techniques that we discussed, SADCT, LPE and minimal l_1, are actually more suitable for image-sequence (video) coding. The reasons are: In video coding, each frame in the sequence (except for an intra-mode frame) is separated into foreground and background segments, so that only the coding of foreground segments needs to be considered. Therefore, applying shape-adaptive techniques can achieve better results. Particularly, by using shape-adaptive methods to code the residual image generated after motion compensation, even more significant rate reduction is expected. This claim relies on the fact that the *relative* number of boundary blocks in the set of blocks that need to be coded is much larger, as compared with still image coding.

References

[1] "Information technology – Digital compression and coding of continuouse-tone still image: Requirements and guidelines," tech. rep., ISO/IEC JTC1 10918-1, (JPEG standard), 1994.

[2] M. Gilge, T. Engelhardt, and R. Mehlan, "Coding of arbitrarily shaped image segments based on a generalized orthogonal transform," *Signal Process. Image Commun.*, vol. 1, pp. 153–180, Oct. 1989.

[3] W. Philips, "Fast coding of arbitrarily shaped image segments using weakly separable bases," *Opt. Eng.*, vol. 35, pp. 177–186, Jan. 1996.

[4] T. Ebrahimi, "MPEG-4 video verification model version 10.0," tech. rep., ISO/IEC JTC1/SC29/WG11, Feb. 1998.

[5] T. Sikora and B. Makai, "Shape-adaptive DCT for generic coding of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 59–62, Feb. 1995.

[6] Y. Zhao, "Segmentation-based shape-adaptive image coding," Master's thesis, Technion–Israel Institute of Technology, June 1999.

[7] P. Salembier, L. Torres, F. Meyer, and C. Gu, "Region-based video coding using mathematical morphology," *Proc. IEEE*, vol. 83, pp. 843–857, June 1995.

[8] D. Wang, C. Labit, and J. Ronsin, "Segmentation-based motion-compensated video coding using morphological filters," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 549–555, June 1997.

[9] J. Pandel, "Variable bit-rate image sequence coding with adaptive quantization," *Signal Process. Image Commun.*, vol. 3, pp. 123–128, 1991.

[10] A. K. Jain, *Fundamentals of Digital Image Processing*. Prentice Hall International, Inc., 1989.

[11] C. Gu and M. Kunt, "Contour simplification by a new nonlinear filter for region-based coding," in *VCIP-94*, vol. 2308, pp. 1180–1191, Sept. 1994.

[12] S. S. Chen, *Basis Pursuit*. PhD thesis, Stanford University, 1995.

[13] F. Marques, J. Sauleda, and A. Gasull, "Shape and location coding for contour images," *Picture Coding Symp.*, 1993.