GRAPH-CUT RATE-DISTORTION OPTIMIZATION FOR SUBBAND IMAGE COMPRESSION

Maria Trocan and Béatrice Pesquet-Popescu

GET - Télécom Paris
Signal and Image Processing Dept.
46, rue Barrault
75634 Paris, France
Email: {trocan, pesquet}@tsi.enst.fr

ABSTRACT

We present a new rate-distortion optimization algorithm based on graph-cuts that can encode efficiently the coefficients of a critically sampled or even redundant non orthogonal transform. The basic idea is to construct a specialized graph such that its minimum cut minimizes an energy functional. We propose to use the graph-cut mechanism for the minimization of the rate-distortion Lagrangian function. To this aim, we have designed a graph able to represent the decomposition subbands and take into consideration their correlations in a biorthogonal multiresolution representation. The method yields good compression results compared to the state-of-art JPEG2000 codec, as well as a general improvement in visual quality.

1. INTRODUCTION

The compression of natural images is still a challenge for the research and industry. Indeed, the geometric features of images, such as edges, characterized by abrupt changes in pixel intensity, are difficult to represent. The wavelet transform has been succesfully used for image representation [1], due to its energy compaction capacities and compression efficiency [2]. The drawback of wavelets is the orientation selectivity because they provide local frequency representation of image regions over a range of spatial scales, and therefore, they do not represent two-dimensional singularities effectively.

In order to solve this problem, several families of geometrical wavelets able to represent the sharp transitions in images have been proposed. It has been shown in [3] that ridgelet representations solve the problem of sparse approximation of smooth objects with straight edges. In [4], an attempt has been made for ridgelet image compression. However, in image processing, edges are typically curved rather than straight and ridgelets alone cannot yield efficient representation. But, if one uses a sufficient fine scale to capture curved edges, such contours get almost straight, therefore ridgelets are deployed in a localized manner. In consequence the curvelet transform [5] has been introduced. However, for discrete images sampled on a rectangular grid, the discrete implementation of the curvelet transform is very challenging. Therefore a new method was introduced: the contourlet transform [6]; initially described in the discrete-domain, the authors proved its convergence to an expansion in the continuous-domain. Thus, a discrete-domain multiresolution and multidirectional expansion is constructed, in the same way as wavelets are derived from filter banks, but using non-separable ones. Due to the fast-iterated filter bank algorithm, the construction results in a flexible multiresolution, local and directional image expansion using contour segments. However, the countourlet decomposition has the adverse property of showing other types of artifacts.

In this paper we present a rate-distortion optimization method based on graph cuts, which can compress efficiently the coefficients of a spatial transform. As described in [7, 8, 9], problems that arise in computer vision can be naturally expressed in terms of energy minimization. Each of these methods consists in modelling a graph for an energy type, such that the minimum cut minimizes globally or locally that functional. Usually, these graph constructions are dense and complex, designing the energy function at pixel level. For example, in [10, 11] the graph cut provides a clean, flexible formulation for image segmentation. With a 4-connected grid design, the graph provides a convenient manner to represent simple local segmentation decisions and provides a set of powerful computational mechanisms to extract global segmentation from these simple local (pairwise) pixel similarities. Good graph-cut based energy optimization results have been obtained in image restoration [12, 13], as well as in stereo [14], motion segmentation [15], texture synthesis in image and video [16, 17] etc. We propose to use the graph-cut mechanism for the minimization of the rate-distortion Lagrangian function. To this aim, we have designed a specialized graph able to represent the decomposition subbands and take into consideration their correlations in a biorthogonal multiresolution representation. The Lagrangian functional was discretized such that it sums the contribution of each subband in terms of rate and the distortion is computed as the direct as well as the cross-correlation impact of quantization. Moreover, the graph model is planar [18] and the energy function we optimize is convex, so the minimum graph cut can be found in polynomial time. As it is shown by the experimental results, the method gives good compression results compared to the state-of-art JPEG2000 codec, as well as a general improvement in visual quality.

This paper is organised as follows: Section 2 describes the graph-cut rate-distortion algorithm used for the coding. The interest of contourlets for image compression is presented in Section 3. Some experimental results are presented in Section 4. Finally, conclusions and future work directions are given in Section 5.

2. GRAPH-CUT RATE-DISTORTION LAGRANGIAN OPTIMIZATION

As mentioned in the introduction, the max-flow/min-cut algorithm has been successfully used in computer vision for solving different energy minimization problems. In this paper we propose to apply this technique for rate-distortion Lagrangian optimization in subband image coding.

Generally, for a graph $G=(V,E,W)$, where $V/E/W$ is the set of vertices/edges and W represents the edges weights (i.e. capacities) and which have two special vertices (terminals), $q_1,q_2\in V$, a q_1-q_2 cut is defined as a partition of the vertices in V into two disjoint sets Q_1 and Q_2 such that $q_1\in Q_1$ and $q_2\in Q_2$. The cost of the cut is given by the sum of weights w of all edges linking Q_1 to Q_2, i.e:

$$C(Q_1,Q_2)=\sum_{u\in Q_1,v\in Q_2,(u,v)\in E} w(u,v) \qquad (1)$$

The minimum cut is found thus as the cut with minimal cost. There are polynomial-time methods to solve the min-cut problem, notably the Ford-Fulkerson algorithm [19].

Now consider the graph $G=(V,E,W)$ with positive edge weights W, which have not only two, but a set of terminal nodes, $Q\in V$. A subset of edges $\mathscr{E}_C\in E$ is called a *multi-way cut* if the terminal nodes are completely separated in the induced graph $G(\mathscr{E}_C)=(V,E-\mathscr{E}_C,W)$ and no proper subset of $\mathscr{E}_C$ separates the terminals in $\mathscr{E}_C$. If C is the cost of the multiway cut, then the multi-terminal min-cut problem is equivalent to finding the minimum cost multiway cut.

In [13], Y. Boykov *et al.* propose to find the minimal multiway cut by succesively finding the min-cut between each terminal and the rest of them. This approximation approach guarantees a local minimization of the energy function within a close factor from the optimal solution for concave energies and gives a global minimization solution for convex functionals. As the rate-distortion Lagrangian lies on a convex decreasing curve (i.e. $D(R)$), we propose to use in the following this method for its optimization.

Consider the problem of coding an image at a maximal rate R_{max} with a minimal distortion D. Each image consists of a fixed number of coding units, X (e.g., in our case, the contourlet spatial subbands), each of them coded with a different quantizer q_i, $q_i\in Q$, where Q is the quantizer set. Let $D_i(q_i)$ be the distortion of subband i when quantized with q_i, and let $R_i(q_i)$ be the number of bits required for coding it. The problem can now be formulated as: find $\min\sum_i D_i(q_i)$, such that $\sum_i R_i(q_i)=R\le R_{max}$.

In the Lagrange-multiplier framework, this constrained optimization problem can be written as the equivalent problem:

$$\min\sum_{i=1}^{X}\left(D_i(q_i)+\lambda R_i(q_i)\right), \quad R\le R_{max} \qquad (2)$$

where the choice of λ measures the relative importance of distortion, respectively rate for the optimization and whose optimal value can be determined using a binary search. The advantage of the problem formulation in Eq. (2) is that the sum and the minimum operator can be exchanged to:

$$\sum_{i=1}^{X}\min\left(D_i(q_i)+\lambda R_i(q_i)\right), \quad R\le R_{max} \qquad (3)$$

This formulation obviously reveals that the global optimization can now be carried out independently for each spatial subband, making an efficient implementation feasible.

The distortion D between the original image x and the quantized one, $\widehat{x}$ can be written as the L^2 norm, i.e. $D=\|x-\widehat{x}\|^2$. For orthonormal transforms, this norm can be equivalently estimated in the transform domain. However, for arbitrary transforms (biorthogonal, redundant, non-linear *etc.*) this property does not hold any more. In the following we focus on this more complicated case and show how the distortion can be approximated and then estimated in the spatial domain, allowing us a graph modelling of the subband interactions. If in the reconstructed image $\widehat{x}$ we highlight the contribution of each subband, $\widehat{x}=\sum_{i=1}^{X}\widehat{x}_i$, where $\widehat{x}_i$ is the contribution of the i^{th} subband, then we can also write the image in a similar way, $x=\sum_{i=1}^{X}x_i$. However, here x_i is completely arbitrary. In the case of a linear basis, it may become $x_i=\sum_k\left\langle x,\widetilde{e}_{k,i}\right\rangle e_{k,i}$, where $\widetilde{e}_{k,i}$, $e_{k,i}$ are the analysis, respectively synthesis elements of the biorthogonal basis. Then we have:

$$D=\left\|\sum_{i=1}^{X}(\widehat{x}_i-x_i)\right\|^2=\sum_i\sum_{i'}\langle\widehat{x}_i-x_i,\widehat{x}_{i'}-x_{i'}\rangle \qquad (4)$$

In a first approximation, we can consider only the diagonal terms, i.e.:

$$D_I\cong\sum_{i=1}^{X}\|x_i-\widehat{x}_i\|^2 \qquad (5)$$

which amounts at estimating the distortion between the contribution to the image and to the quantized image only of the i^{th} subband. This means we can reconstruct the image only from the i^{th} subband coefficients (the others being set to zero).

In a second approximation, one can also consider "cross-correlation" terms, i.e.:

$$D\cong D_I+\sum_i\sum_{i'\in N(i)}\underbrace{\langle\widehat{x}_i-x_i,\widehat{x}_{i'}-x_{i'}\rangle}_{D_{i,i'}} \qquad (6)$$

where $N(i)$ is a neighborhood of i, containing closely correlated subbands. Indeed, given the limited support of the wavelets, the closer in space are the subbands, the higher the correlation. In practice, this neighborhood is described by the geometrical position of the subbands in a multiresolution decomposition, where only the vertical and horizontal directions are considered.

The second term involves the highest complexity (inverse transforms plus inner products between images), which can however be divided by two, noting that $D_{i,i'}=D_{i',i}$ and therefore:

$$D\cong\sum_i D_i+2\sum_i\sum_{i',i'>i}D_{i,i'} \qquad (7)$$

For $D_{i,i'}$ we need to calculate the error between the image reconstructed from the i^{th} subband (x_i) and its equivalent recontructed from the quantized i^{th} subband ($\widehat{x}_i$), the same from a neighboring subband i' and then compute the inner product.

The minimization of the energy function defined above is equivalent to the best repartition of quantizers per subbands. The graph we have designed for solving this problem has as vertices the set of spatial subbands and the set of quantizers as terminal nodes, where the subbands are linked following the neighborhood system $\mathscr{N}$. Each terminal node is connected to all terminal nodes, considering all quantization possibilities for the spatial subbands. (i.e. $G=(V,E)$, where $V=X\cup Q$ and $E=E_N\cup E_Q$, E_N denoting the regular

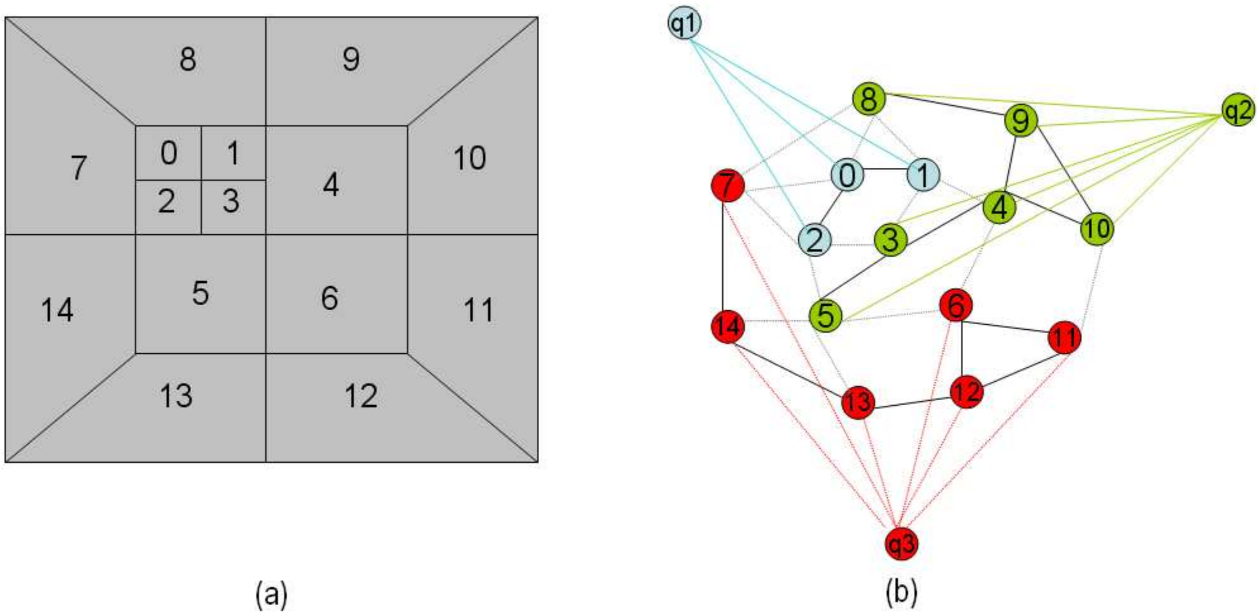


Figure 1: Contourlet decomposition with three levels (a) and three-way graph-cut repartition (b) (q_1 partition in blue, q_2 partition in green, q_3 partition in red, where the regular edges are with full black lines, terminal links in colors and the cut-edges in gray dash-lines).

edges between subband vertices in the neighbourhood system $\mathscr{N}$ and E_Q the terminal links between subband nodes and quantizers). One can distinguish two connexion types: E_N and E_Q. We define the weights for the quantizers links E_Q in terms of the rate-distortion cost; so, the weight associated to the edge connecting subband x to quantizer q is defined as $w_{x,q} = D_x(q) + R_x(q)$. For a E_N link, the associated weight is given by the cross-correlation distortion, i.e.: $w_{x_i,x_{i'}} = \langle \widehat{x}_i - x_i, \widehat{x}_{i'} - x_{i'} \rangle$, $i' \in \mathscr{N}(i)$. So the function we want to minimize can be written as:

$$\min \underbrace{\sum_{i=1}^{X} (\|x_i - \widehat{x}_i\|^2 + \lambda R(i))}_{E_{data}} + \underbrace{\sum_{i=1}^{X} \sum_{i' \in \mathscr{N}(i)} \langle \widehat{x}_i - x_i, \widehat{x}_{i'} - x_{i'} \rangle}_{E_{smooth}} \tag{8}$$

Now we establish the correspondence between our graph and the multiway cut. In Fig. 1 is illustrated an induced graph $G(\mathscr{E}_C) = (V, E - \mathscr{E}_C)$ corresponding to a three-way cut $\mathscr{E}_C$ on G. One can remark that it should be exactly one terminal link to each subband node in the induced graph. There exists a fast approximation algorithm that can minimize our energy functional [13]. Once the graph construction and the energy function to be minimized have been defined, the algorithm starts with an initial (random) partitioning f, where f is a set of quantizers, $f : Q \rightarrow \mathbb{R}$, of the graph. For each quantizer $q \in Q$ finds $\widehat{f}$ as the quantizers repartition which minimizes $E(f')$, i.e. $\widehat{f} \leftarrow \min E(f')$, among f' within one α-expansion [13] of f, where f' denotes the possibilites of linking the terminal node q to the planar nodes that are not linked to it in the initial f partitioning. This operation is repeated this until $E(\widehat{f})$ no longer decreases. Thus $\widehat{f}$ is efficiently found as being the best quantizer repartition, because its cost corresponds to a minimal-cut over the constructed graph.

3. APPLICATION TO CONTOURLET SUBBAND IMAGE CODING

In [6, 20] a double filter bank approach for obtaining sparse expansions for typical images with smooth contours is proposed. For its construction, the Laplacian pyramid [21] is first used to capture the point sigularities, then a directional biorthogonal filter bank [22] is applied for linking the point discontinuities into linear structures. The result is an image expansion using elementary images like contour segments, named the contourlet transform or the pyramidal directional filterbank (PDFB). Due to this cascade structure, multiscale and directional decomposition stages in the contourlet transform are independent one of other. At each scale, one can decompose into any arbitrary power of two number of directions and the number of decomposition directions can vary at different scales (Fig.2). This feature makes contourlets a unique transform that can achieve a high level of flexibility in decomposition while being close to critically sampled. Its redundancy factor has an upper limit of $4/3$, which makes the scheme more appropriate for compression than other geometrical transforms. Another reason for which we have considered this scheme is that contourlets can be approximated with less coefficients than the wavelets; that is, for a contourlet basis, the approximation error for keeping only the M most significant coefficients is:

$$\| f - f_{M_{contourlet}} \| = O((logM)^3 M^{-2}) \tag{9}$$

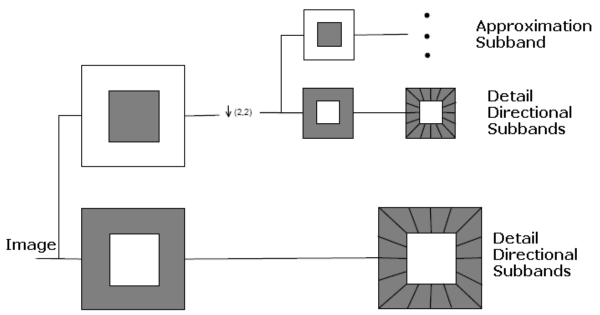


Figure 2: Contourlet filter bank

which is smaller than the one obtained on the wavelet basis:

$$\left\| f - f_{M_{wavelet}} \right\| = O(M^{-1}). \tag{10}$$

As shown in [23], the efficiency of the pyramidal directional filter bank with respect to classical wavelets tends to decrease on natural images when the number of coefficients increases.

4. EXPERIMENTAL RESULTS

For our simulations, we have considered two representative test images: "Circles" (512x512 pixels) and "Mandrill" (512x512 pixels), which have been selected for the difficulty to encode all their texture characteristics.

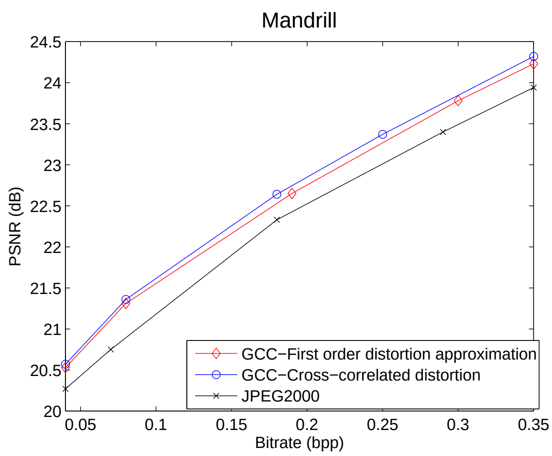


Figure 3: Rate-distortion comparison for Mandrill image

We have used dead-zone scalar quantization, with $q \in \{2^0, \ldots, 2^{10}\}$ and a 5-level contourlet decomposition, where the coarsest three decomposition levels consists in a 9/7 separable wavelet transform (i.e. 3 directions) and the finest two levels are represented with a 16 and respectively 32 bands biorthogonal directional filter. The efficiency of this hybrid scheme has been proved in [23] and in [24]. One can remark that the algorithm can also be used with vector quantizers and the coefficient space be further partitioned into blocks. In a first approach, we have considered a fixed, quantizer dependent, weighting function for the regular links (i.e. edges between neighbour subband nodes). Thus, the cost of the edge $(u,v) \in E$ is 0 if the nodes $u,v \in V$ are linked to the same quantizer and $\beta \in \mathbb{R}^+$ otherwise (where β's magnitude enforces or diminishes the smoothing). This assumption is coherent, as for two strong correlated subbands the same quantizer is imposed. The results obtained for this approach are denoted by "first-order distortion approximation", whereas the "cross-correlated distortion" means that the distortion model in Eq. (8) has been considered.

As shown in Fig. 4, both the numerical and visual quality are improved; for the same coding rate (e.g. 0.1 bpp), one can remark almost 1 dB improvement, even though our method employs a redundant transform. Similar results are also depicted in Fig. 3. Note that for rate estimation in the allocation algorithm we have used a simple (non-contextual) arithmetic coder [25], while JPEG2000 codec uses a highly optimized contextual coder.

5. CONCLUSION

In this paper we have presented a graph-cut method for rate-distortion optimization in image coding of not necessarily orthogonal decompositions. As shown by experimental results, it can efficiently encode the contourlet coefficients at low bitrates, improving both the visual and numerical quality. Moreover, the proposed method can be further used with vector quantizers and the graph design could be developed to model the coding units at a finer level of representation.

REFERENCES

[1] "Information technology – JPEG 2000 image coding system," Tech. Rep., ISO/IEC 15444-1, 2000.

[2] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.

[3] M.N. Do and M. Vetterli, "The finite ridgelet transform for image representation," *IEEE Transactions on Image Processing*, vol. 12, pp. 16–28, Jan. 2003.

[4] M. N. Do and M. Vetterli, "Orthonormal finite ridgelet transform for image compression," *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, Canada, Sept. Oct.2000.

[5] M. N. Do and M. Vetterli, "Pyramidal directional filter banks and curvelets," *Proceedings of the IEEE International Conference on Image Processing*, Thessaloniki, Greece, Oct.2001.

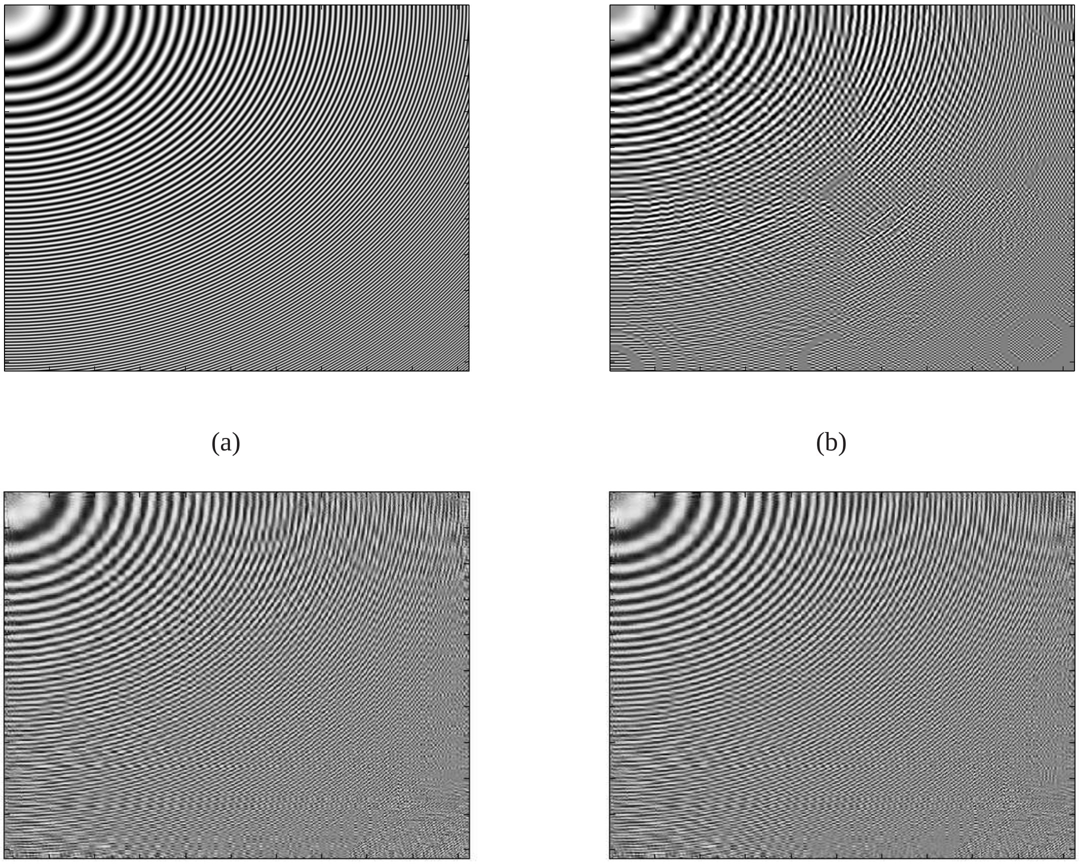
[6] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. on Image Proc.*, vol. 14, pp. 2091–2106, 2005.

[7] V. Kolmogorov and R Zabin, "What energy functions can be minimized via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 147 – 159, 2004.

[8] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1124 – 1137, 2004.

[9] Y. Boykov and O. Veksler, "Graph cuts in vision and graphics: Theories and applications," *Handbook of Mathematical Models in Computer Vision, Springer-Verlag*, pp. 79–76, 2006.

[10] Y. Boykov and V.Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts," *Proc. of IEEE International Conference on Computer Vision*, vol. 1, pp. 26–33, 2003.



(a) (b) (c) (d)

Figure 4: "Circles" (512x512)image: (a) original, (b) JPEG2000 compression at 0.1 bpp (PSNR=14.19 dB), (c) first-order distortion graph-cut method at 0.1 bpp (PSNR=14.64 dB) , (d)cross-correlation distortion graph-cut method at 0.1 bpp (PSNR=15.13 dB).

[11] Y. Boykov and G.Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Springer-Verlag, International Journal of Computer Vision*, vol. 70, pp. 109–131, 2006.

[12] A. Raj and R. Zabih, "A graph cut algorithm for generalized image deconvolution," *Proc. of IEEE International Conference on Computer Vision*, pp. 1048–1054, 2005.

[13] Y. Boykov, O. Veksler, and R Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222 – 1239, 2001.

[14] Y. Boykov, O. Veksler, and R. Zabih, "Markov random fields with efficient approximations," *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.

[15] T. Schoenemann and D. Cremers, "Near real-time motion segmentation using graph cuts," *K. Franke et al.(Eds.): DAGM 2006, Springer-Verlag*, pp. 455–464, 2006.

[16] V. Kwatra, A. Schodl, I. Essa, and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," 2003.

[17] Y. Zhang, X. Ji, D. Zhao, and W. Gao, "Video coding by texture analysis and synthesis using graph cut," *In Proceedings of Pacific-Rim Conference on Multimedia, China*, 2006.

[18] M.-C.Costa, L. Letocart, and F. Roupin, "Minimal multicut and maximal integer multiflow: A survey," *In ELSEVIER European Journal of Operational Research*, pp. 55–69, 2005.

[19] L. Ford and D. Fulkerson, "Flows in networks," *Princeton University Press*, 1962.

[20] M.N. Do and M. Vetterli, *Contourlets*, J. Stoeckler, G. V. Welland (Eds.), Beyond Wavelets, Academic Press, 2002.

[21] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, pp. 532–540, 1983.

[22] R.H. Bamberger and M.J.T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Transactions on Signal Processing*, vol. 40, pp. 882–893, 1992.

[23] V. Chappelier, C. Guillemot, and S. Marinkovic, "Image coding with iterated contourlet and wavelet transforms," *Proc. of IEEE International Conference on Image Processing*, vol. 5, pp. 3157– 3160, 2004.

[24] A.N. Belbachir and P.M. Goebel, "The contourlet transform for image compression," *Proc. of Physics in Signal and Image Processing Conference, PSIT*, 2005.

[25] M. Nelson and J.-L. Gailly, *The Data Compression Book*, M&T Books, New York, 1995.