USING DEFORMABLE SURFACE MODELS TO DERIVE A DCT-LIKE 2D TRANSFORM

M. Krinidis, N. Nikolaidis and I. Pitas

Department of Informatics
Aristotle University of Thessaloniki
Box 451, 54124 Thessaloniki, GREECE
Email: {mkrinidi,nikolaid,pitas}@aiia.csd.auth.gr

ABSTRACT

This paper introduces a $2D$ discrete, non-separable transform for image processing, which can be regarded as a combination of the well known Discrete Cosine Transform (DCT) with an analytically derived quantization table that includes a compression ratio selection parameter. A $3D$ deformable surface model is used to approximate the image intensity and the introduced discrete transform is an intermediate step of the explicit surface deformation governing equations. The proposed transform is applied to lossy image compression and the obtained results are compared to those of a DCT-based compression scheme.

1. INTRODUCTION

Intensive research has been carried out during the last decades on signal transforms. Transforms have been used in several signal and image processing applications, such as enhancement, restoration, filtering and data compression, the latter being perhaps the most important application domain [1, 2].

The overall idea behind transform-based compression is to use a transform that decorrelates the input signal and packs its total energy into few coefficients. The Karhunen-Loeve Transform (KLT) [3] can be considered to be optimal for compression under certain circumstances, i.e. for a Gaussian source at any bit rate and bit allocation strategy. However, the computation of the KLT is computationally expensive and time consuming. Other transforms, such as Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), Hadamard Transform, Slant Transform, which are computationally faster than the KLT, while exhibiting slightly worse performance in terms of energy compaction and decorrelation, have emerged. The transform that has been mostly used in image compression, e.g. in various image/video compression standards, such as JPEG (Joint Photographic Experts Group) [4], MPEG 1/2 (Moving Picture Experts Group) [5], H.261, and H.263, is DCT.

In this paper, we introduce a discrete transform which embeds a compression ratio selection mechanism. The proposed transform which will be referred to as Discrete Modal Transform (DMT), was motivated by the technique presented in [6], which aims at analyzing non-rigid object motion, with application to medical images. The basic idea is to warp a $3D$ physics-based deformable surface model onto the intensity surface of a target $2D$ image. We, then, utilize an intermediate step of the deformation procedure to introduce a non-separable $2D$ transform that can decompose the image into a number of basis images. The resulting transform is proven to be a generalization of the DCT that includes an "embedded" quantization procedure. More specifically, the proposed transform can be considered as a combination of DCT with an analytically derived quantization matrix that includes a compression selection parameter. The fact that DCT can be derived starting from a deformable model that tries to approximate the intensity surface of an image is a significant outcome of this study.

The remainder of the paper is organized as follows. The $3D$ physics-based deformable surface model [6] is reviewed in Section 2 and the proposed transform is derived. Section 3 compares a lossy image compression scheme based on the derived transform with a DCT-based compression scheme. Conclusions are drawn in Section 4.

2. DERIVATION OF DISCRETE MODAL TRANSFORM (DMT) FROM 3D PHYSICS-BASED DEFORMABLE SURFACES

Let $I(x, y)$ denote the intensity (grayscale value) of the pixel at position (x, y) on an image. By combining both the spatial (x, y) and grayscale $I(x, y)$ components of an image one can obtain a $3D$ surface representation $(x, y, I(x, y))$ of the image [7] that will be subsequently called intensity surface. An elastic $3D$ physics-based deformable model (Figure 1) [6] consisting of a mesh of connected springs comprising of $N = N_h N_w$ nodes (assumed to be equal to the image height and width), can be used to model this surface.

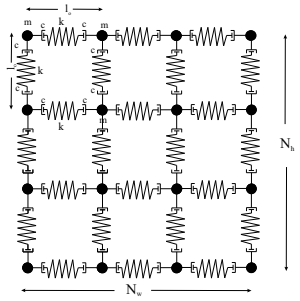

Fig. 1. *The elastic* $3D$ *physics-based deformable model consisting of* $N_h \times N_w$ *nodes.*

The deformable surface model is ruled by Lagrangian dynamics [8]:

$$\mathbf{M}\ddot{\mathbf{u}}^{\tau} + \mathbf{C}\dot{\mathbf{u}}^{\tau} + \mathbf{K}\mathbf{u}^{\tau} = \mathbf{f}^{\tau}, \tag{1}$$

where $\mathbf{K}$ is the stiffness matrix, $\mathbf{u} = [\underline{u}_1, \ldots, \underline{u}_N]^T$ is defined as the vector comprising of the vectors of nodal displacements, and $\mathbf{f} = [\underline{f}_1, \ldots, \underline{f}_N]^T$ is the external force vector comprising of the external forces vectors applied to each node. The forces in this vector have zero x and y components whereas their z component is taken to be equal to the Euclidean distance between the point

$(x, y, I(x, y))$ of the intensity surface and the corresponding node of the model in its initial configuration $(x, y, 0)$, i.e. equal to the intensity $I(x, y)$ of pixel (x, y): $f_z(x, y) = f_{(x-1)N_w+y,z} = I(x, y)$, where $f_{(x-1)N_w+y,z}$ is defined as the z component of the force vector $\underline{\mathrm{f}}_{(x-1)N_w+y}$ in vector $\mathbf{f}$.

Instead of finding directly the equilibrium solution of (1), one can transform it by a basis change [9]:

$$\mathbf{u}^\tau = \mathbf{\Psi}\tilde{\mathbf{u}}^\tau, \tag{2}$$

where $\mathbf{\Psi}$ is a square nonsingular transformation matrix of order N to be determined and $\tilde{\mathbf{u}}^\tau$ is referred to as the *generalized displacements* vector. One effective way of choosing $\mathbf{\Psi}$ is setting it equal to matrix $\mathbf{\Phi}$ whose entries are the eigenvectors $\boldsymbol{\phi}_i$ (called vibration modes) of the generalized eigenproblem:

$$\mathbf{K}\boldsymbol{\phi}_i = \omega_i^2\mathbf{M}\boldsymbol{\phi}_i, \tag{3}$$

$$\mathbf{u}^\tau = \mathbf{\Phi}\tilde{\mathbf{u}}^\tau = \sum_{i=1}^{N=N_hN_w} \underline{\tilde{\mathbf{u}}}_i^\tau \boldsymbol{\phi_i}. \tag{4}$$

Equation (4) is referred to as the *modal superposition equation*. $\underline{\tilde{\mathbf{u}}}_i^\tau$ is the i-th component of $\tilde{\mathbf{u}}^\tau$ and ω_i is the corresponding eigenvalue (also called *frequency*).

A significant advantage of the formulations described so far, is that the vibration modes (eigenvectors) $\boldsymbol{\phi}_i$ and the frequencies (eigenvalues) ω_i of a plane topology do not have to be computed using eigen-decomposition techniques but have an explicit formulation [6]:

$$\begin{aligned}\omega^2(j, j') &= \omega^2_{(j-1)N_w+j'} = \\ &= \frac{4\mathrm{k}}{m}\left(\sin^2\left(\frac{\pi j}{2N_h}\right) + \sin^2\left(\frac{\pi j'}{2N_w}\right)\right), \end{aligned} \tag{5}$$

$$\begin{aligned}&\text{where } j \in \{0, 1, \ldots, N_h - 1\},\ j' \in \{0, 1, \ldots, N_w - 1\} \text{ and} \\ \boldsymbol{\phi}_{n,n'}(j, j') &= \boldsymbol{\phi}_{(j-1)N_w+j'}((n-1)N_w + n') = [\ldots, \\ &\quad \cos\frac{\pi j(2n-1)}{N_h}\cos\frac{\pi j'(2n'-1)}{N_w}, \ldots]^T, \end{aligned} \tag{6}$$

where $n \in \{1, 2, \ldots, N_h\}$, $n' \in \{1, 2, \ldots, N_w\}$ and $\boldsymbol{\phi}_{(j-1)N_w+j'}$ $((n-1)N_w + n')$ is the $((n-1)N_w + n')$-th element of vector $\boldsymbol{\phi}_{(j-1)N_w+j'}$.

We consider that no deformations occur along the x and y axes, i.e., deformations occur only along the intensity z axis, driven by the intensity (grayscale value) of the image under examination. Thus, for each component $[\tilde{u}_{x_i}, \tilde{u}_{y_i}, \tilde{u}_{z_i}]$ of vector $\tilde{\mathbf{u}}$ in (4), we have $\tilde{u}_{x_i} = \tilde{u}_{y_i} = 0$ and $\tilde{\mathbf{u}}$ is simplified to:

$$\tilde{\mathbf{u}} = [\tilde{u}_1, \ldots, \tilde{u}_{N_hN_w}]^T, \tag{7}$$

where $\tilde{u}_i \triangleq \tilde{u}_{z_i}$.

In our case, where the initial and the final (desirable) deformable surface states, i.e. the initial model configuration and the image intensity surface, are known, it is assumed that a constant force load $\mathbf{f}$ is applied to the surface model. Thus, equation (1) reduces to the following equilibrium governing equation that corresponds to the static problem:

$$\mathbf{K}\mathbf{u} = \mathbf{f}, \tag{8}$$

or in the modal space:

$$\tilde{\mathbf{K}}\tilde{\mathbf{u}} = \tilde{\mathbf{f}}, \tag{9}$$

where $\tilde{\mathbf{K}} = \mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi}$ and $\tilde{\mathbf{f}}^\tau = \mathbf{\Phi}^T\mathbf{f}^\tau\mathbf{\Phi}$, $\mathbf{f}^\tau$ being the external force vector.

In the new basis, equation (9) is simplified to the following scalar equations:

$$\omega_i^2\tilde{u}_i = \tilde{f}_i, \tag{10}$$

where $\tilde{f}_i$ is the z component of the i-th element of $\tilde{\mathbf{f}}$.

Using (4) and (10) it can be found that the deformations u_{xy} along the intensity axis of the node of the deformable surface model that corresponds to pixel (x, y) based on modal analysis for a plane topology can be described by:

$$\begin{aligned}u_{xy} &= \sum_{n=0}^{N_h-1}\sum_{n'=0}^{N_w-1}\sum_{i=1}^{N_h}\sum_{j=1}^{N_w} I(i, j)\phi_{i,j}(n, n') \cdot \\ &\quad \frac{1}{(1+\omega^2(n, n'))\sqrt{\sum_{i=1}^{N_h}\sum_{j=1}^{N_w}\phi^2_{i,j}(n, n')}}\phi^2_{x,y}(n, n'), \end{aligned} \tag{11}$$

where u_{xy} is the z component of vector $\underline{\mathbf{u}}_{(x-1)N_w+y}$ and $I(i, j)$ is the image intensity of pixel (i, j). One can see that deformations are directly related to the eigenvalues $\omega^2(n, n')$ and the eigenvectors $\boldsymbol{\phi}^2(n, n')$ of the model.

The deformations $\mathbf{u}$ of the $3D$ deformable surface model can be rewritten as:

$$u_{xy} = \sum_{n=0}^{N_h-1}\sum_{n'=0}^{N_w-1} \mathrm{F}(n, n')\phi^2_{x,y}(n, n'), \tag{12}$$

where

$$\mathrm{F}(n, n') = \frac{\sum_{i=1}^{N_h}\sum_{j=1}^{N_w} I(i, j)\phi_{i,j}(n, n')}{(1+\omega^2(n, n'))\sqrt{\sum_{i=1}^{N_h}\sum_{j=1}^{N_w}\phi^2_{i,j}(n, n')}}. \tag{13}$$

Equation (12) is applied to each node (n, n') of the deformable surface model independently.

The normalization factor $\left(\sum_{i=1}^{N_h}\sum_{j=1}^{N_w}\phi^2_{i,j}(n, n')\right)$ in (13) can be further analyzed [10]:

$$\sum_{i=1}^{N_h}\sum_{j=1}^{N_w}\phi^2_{i,j}(n, n') = a(n)a(n'), \tag{14}$$

where

$$a(n) = \begin{cases} N, & n = 0 \\ \frac{N}{2}, & n \neq 0 \end{cases},\ n \in \{0, 1, \ldots, N-1\}. \tag{15}$$

Thus, $\mathrm{F}(n, n')$ in (13) can be rewritten as:

$$\mathrm{F}(n, n') = \sum_{i=0}^{N_h-1}\sum_{j=0}^{N_w-1} I(i, j)v_{n,n'}(i, j), \tag{16}$$

$$\begin{aligned}v_{n,n'}(i, j) = &\quad \cos\tfrac{\pi n(2i+1)}{2N_h}\cos\tfrac{\pi n'(2j+1)}{2N_w} \cdot \\ &\frac{1}{1+\lambda\left[\sin^2\left(\frac{\pi n}{2N_h}\right)+\sin^2\left(\frac{\pi n'}{2N_w}\right)\right]}\frac{1}{\sqrt{a(n)a(n')}}, \end{aligned} \tag{17}$$

where $n = 0, 1, \ldots, N_h - 1$, $n' = 0, 1, \ldots, N_w - 1$, $\lambda = 4\frac{k}{m}$, k being the stiffness of the springs (which have natural length l_0) that form the surface model and m the mass of each node of the surface model (Figure 1). When k increases and/or m decreases, the surface model tends to behave as a rigid one, thus the intensity surface model can hardly deform. On the other hand, when m increases and/or k decreases, the intensity surface model tends to be a fully deformable one, meaning that each force affects only the node (mass), where it

is applied to. Equation (16) defines the proposed $2D$ Discrete Modal Transform (DMT).

It is easy to show that the proposed Discrete Modal Transform coefficients $\mathrm{F}(\mathrm{n},\mathrm{n}')$ are related with DCT coefficients $\mathrm{C}(n,n')$ as follows:

$$\mathrm{F}(n,n') = \frac{1}{\left[1+\lambda\left(\sin^2\left(\frac{\pi n}{2N_h}\right)+\sin^2\left(\frac{\pi n'}{2N_w}\right)\right)\right]}\mathrm{C}(n,n'). \tag{18}$$

One can easily notice that when $\lambda = 0$, $\mathrm{F}(n,n') = \mathrm{C}(n,n')$. Thus, we can claim, that the proposed transformation is a generalization of the DCT that includes a compression ratio selection mechanism. Indeed, the denominator of (18) plays a role similar to that of the quantization matrix used in DCT-based coding in the JPEG standard. Therefore, DMT can be seen as a DCT combined with a new analytically computed quantization matrix which includes the parameter λ that provides a physically meaningful compression ratio/quantization selection mechanism.

The inverse Discrete Modal Transform is expressed as:

$$f(i,j) = \sum_{n=1}^{N_h}\sum_{n'=1}^{N_w}\mathrm{F}(n,n')w_{n,n'}(i,j), \tag{19}$$

where $i = 0,1,\ldots,N_h-1$, $j = 0,1,\ldots,N_w-1$ and $w_{n,n'}(i,j)$ is given by:

$$w_{n,n'}(i,j) = \cos\frac{\pi n(2i+1)}{2N_h}\cos\frac{\pi n'(2j+1)}{2N_w}\cdot \frac{1+\lambda\left[\sin^2\left(\frac{\pi n}{2N_h}\right)+\sin^2\left(\frac{\pi n'}{2N_w}\right)\right]}{\sqrt{a(n)a(n')}}. \tag{20}$$

$w_{n,n'}(i,j)$ comprise the basis images of the Discrete Modal Transform.

The proposed transform, is linear, real and has excellent energy compaction properties. Due to the factor that appears in the denominator of (18), DMT is non-separable and non-orthonormal. However, the forward and inverse transform form an orthonormal transform pair. The proofs of these properties are omitted due to lack of space.

3. APPLICATION OF DMT TO IMAGE COMPRESSION

An application of the DMT to lossy image compression is presented in this Section. The proposed transform is compared with DCT (combined with appropriate quantization tables) in terms of compression and quality of the compressed images. We chose to use DCT, because of its close relation with the proposed transform and because it is the most widely used transform in many image compression algorithms e.g. in JPEG.

The first set of experiments dealt with the evaluation of the quality of the compressed images when DMT and DCT are applied to the target images. To perform image compression, the same procedure is used for both transforms. Both $2D$ transforms are applied to an image, by using 8×8 blocks. In order to discard high-frequency details and achieve compression, the DCT output must be quantized. In this experiment, the quantization table found in the Annex of the JPEG standard [2], multiplied by a scaling coefficient Q to enable variable compression, was used for all image blocks. No quantization table was used for DMT, since the coefficient λ along with the scaling factor in the denominator of (18) essentially act as quantization coefficients. Once the frequency coefficients of DCT were

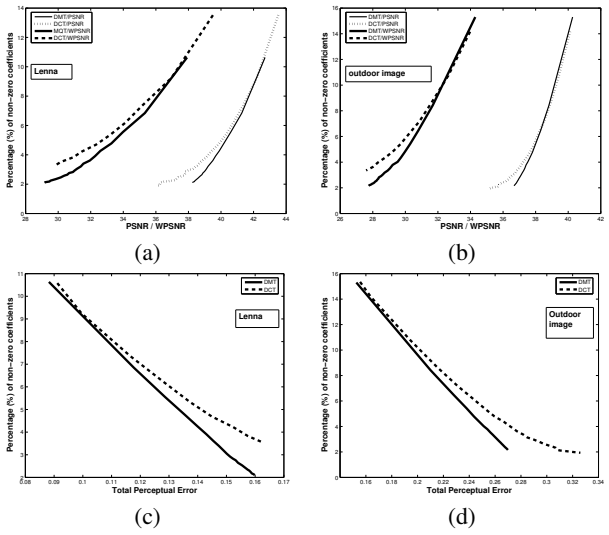

Fig. 2. *Image quality versus the percentage of non-zero coefficients plots for different values of λ and Q for DMT and DCT respectively, for two test images (Lenna and an outdoor image): in plots **(a)**, **(b)** the image quality is measured in terms of PSNR and WPSNR whereas in plots **(c)**, **(d)** the total perceptual error of Watson metric is used.*

divided by the values of the quantization table, both the DMT and DCT outputs were rounded to their closest integers. The frequency components that either have a small coefficient or a large divisor in the quantization table will likely round to zero. By sorting the frequency components (e.g. zig-zag scanning), one will typically end up with a run of zeros at the end of the coefficient vector, which can be discarded for compression purposes. In order to acquire the compressed image, first, the output for DCT is multiplied with the quantization table and then, the inverse transforms are applied to the frequency vectors, for both transforms.

The above procedure was applied to various images of different sizes and content, facial images, studio images, images depicting humans, indoor and outdoor scenes etc. By varying the coefficients λ and Q for DMT and DCT respectively, different levels of compression were achieved. In this experimental setup, compression was measured as the percentage of non-zero frequency coefficients in the compressed images. Plots of image quality versus the number of non-zero frequency coefficients can be seen in Figure 2. The metrics used for comparing the results in terms of image quality were the PSNR (peak signal-to-noise ratio) and the weighted PSNR (WPSNR) [11] between the original (initial) and the compressed (final) image as well as the total perceptual error (TPE) of the Watson metric [12].

Results prove that the DMT can achieve approximately the same or better image quality at the same levels of compression. In general, in this simple compression setup, the DMT achieves better image quality than DCT for high compression levels (small percentage of non-zero coefficients), whereas in low compression levels the two algorithms perform almost the same. Analogous results were obtained for other images.

In the second set of experiments, instead of using standard quantization tables that are based on general experiments for DCT, a technique based on bit-rate control was exploited. More specifically, an

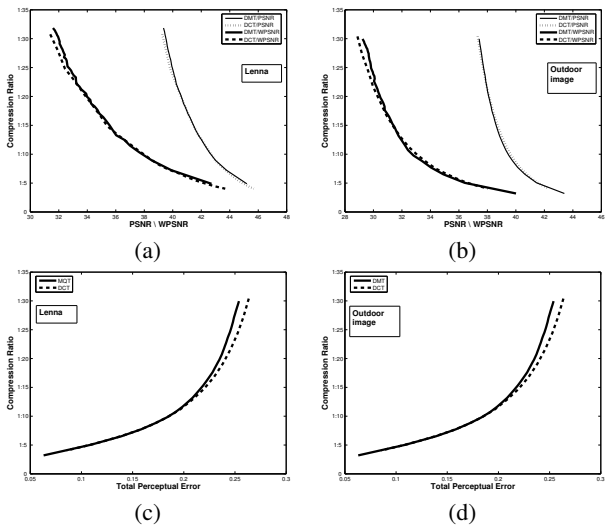

Fig. 3. *Image quality versus the compression ratio plots for different values of λ and Q for DMT and DCT respectively, for two test images (Lenna and an outdoor image): in plots* ***(a)****,* ***(b)*** *the image quality is measured in terms of PSNR and WPSNR whereas in plots* ***(c)****,* ***(d)*** *the total perceptual error of Watson metric is used.*

optimal bit allocation procedure, reviewed in [13], based on the statistical properties of an image was applied. The same procedure, as in the previous experiments was used. The image was partitioned in 8×8 non-overlapping blocks and both transforms were applied to each block. The DCT output was scaled by the quantization coefficients that were derived from the statistical properties of the image using the bit allocation procedure in [13] and the selected bit rate. Then, both DMT and DCT coefficients were rounded to their closest integer and entropy coding [13] was performed. During the entropy coding, DCT coefficients were ordered by using the zig-zag scan, whereas the DMT coefficients were sorted in ascending order based on the denominator (18):

$$Z(k,l) = 1 + \lambda \left[\sin^2 \left(\frac{\pi k}{2N_h} \right) + \sin^2 \left(\frac{\pi l}{2N_w} \right) \right], \qquad (21)$$

where N_h, N_w are the dimensions of the $2D$ transform. The DMT coefficients are divided by this term, thus, sorting based on this term means that the coefficients which are divided with small values of Z have larger information and must be kept and the coefficients which are divided with large values of Z may be discarded. In Figure 3, the compression ratio-distortion curves for two test images are depicted. In the case of DMT, compression was controlled by varying λ. In this experiment, compression ratio was measured as the number of bits in the original image divided by the number of bits in the compressed image, after quantization and entropy coding, whereas distortion was measured using PSNR, WPSNR and the total perceptual error of the Watson metric. One can see, that DMT achieves in most cases, better image quality than DCT for high compression ratios, while at lower compression ratios the two transforms have almost the same performance. Similar results were obtained for all other test images.

4. CONCLUSION

A novel $2D$ discrete, non-separable, signal transform was introduced in this paper. The proposed transform results as an intermediate step of the deformation procedure of a $3D$ physics-based deformable model that deforms to adapt to the intensity surface of the image and it was proven to be equivalent to DCT combined with a novel quantization scheme that embeds a compression ratio selection mechanism. We applied the proposed transform to lossy image compression and compared it with DCT, since DCT is widely used in image compression. The results show that the proposed transform can achieve comparable or better image quality to DCT (quantized by the quantization tables of JPEG, or quantization tables derived from an optimal bit allocation procedure) at the same level of compression.

5. ACKNOWLEDGMENT

This work has been conducted in conjunction with the 'SIMILAR' European Network of Excellence on Multimodal Interfaces of the IST Programme of the European Union (www.similar.cc).

6. REFERENCES

[1] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[2] W. B. Pennebaker and J. L. Mitchell. JPEG*: Still Image Data Compression Standard*. Van Nostrand Reinhold, New York, 1993.

[3] M. Effros, H. Feng, and K. Zeger. Suboptimality of the Karhunen Loeve Transform for transform coding. *IEEE Transactions on Information Theory*, 50(8):1605–1619, Aug. 2004.

[4] G. K. Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, Apr. 1991.

[5] Didier Le Gall. Mpeg: A video compression standard for multimedia applications. *Communications of the ACM*, 34(4):46–58, Apr. 1991.

[6] C. Nastar and N. Ayache. Frequency-based nonrigid motion analysis: Application to four dimensional medical images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1069–1079, 1996.

[7] B. Moghaddam, C. Nastar, and A. Pentland. A bayesian similarity measure for direct image matching. In *International Conference on Pattern Recognition (ICPR 1996)*, pages 350–358, Vienna, Austria, August 1996.

[8] K. J. Bathe. *Finite Element Procedure*. Prentice Hall, Englewood Cliffs, New Jersey, 1996.

[9] A. Pentland and S. Sclaroff. Closed-form solutions for physically-based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):715–729, Jul 1991.

[10] M. R. Spiegel. *Mathematical Formulas*. McGraw-Hill, New York, 1996.

[11] A. Netravali and B. Haskell. *Digital Pictures Representation and Compression*. Plenum Press, New York, 1998.

[12] A. Mayache, T. Eude, and H. Cherifi. A comparison of image quality models and metrics based on human visual sensitivity. In *Proceedings of the International Conference on Image Processing (ICIP 1998)*, volume 3, pages 409–413, Chicago, IL, USA, October 1998.

[13] V. Bhaskaran and K. Konstantinides. *Image and Video Compression Standards*. Kluwer Academic Publishers, London, 1995.