Adaptive Spectral Transform for Wavelet-Based Color Image Compression

Ulug Bayazit, *Member, IEEE*

***Abstract*—Since different regions of a color image generally exhibit different spectral characteristics, the energy compaction of applying a single spectral transform to all regions is largely inefficient from a compression perspective. Thus, it is proposed that different subsets of wavelet coefficients of a color image be subjected to different spectral transforms before the resultant coefficients are coded by an efficient wavelet coefficient coding scheme such as that used in JPEG2000 or color set partitioning in hierarchical trees (CSPIHT). A quadtree represents the spatial partitioning of the set of high frequency coefficients of the color planes into spatially oriented subsets which may be further partitioned into smaller directionally oriented subsets. The partitioning decisions and decisions to employ fixed or signal-dependent bases for each subset are rate-distortion (R-D) optimized by employing a known analytical R-D model for these coefficient coding schemes. A compression system of asymmetric complexity, that integrates the proposed adaptive spectral transform with the CSPIHT coefficient coding scheme yields average coding gains of 0.3 dB and 0.9 dB in the Y component at 1.0 b/p and 2.5 b/p, respectively, and 0.9 dB and 1.35 dB in the U and V components at 1.0 b/p and 2.5 b/p, respectively, over a reference compression system that integrates the single spectral transform derived from the entire image with the CSPIHT coefficient coding scheme.**

***Index Terms*—Adaptive coding, color, image coding, Karhunen–Loeve transform, quadtree, wavelet transform.**

I. Introduction

TYPICALLY, a single global spectral transform is applied to decorrelate the color planes of a color image prior to compression. The trivial method of independently coding each of the resulting spectral planes with high performance wavelet-based grayscale image codecs, such as set partitioning in hierarchical trees (SPIHT) [1], JPEG-2000 [2], or subband block hierarchical partitioning (SBHP) [3], is highly inefficient since nonlinear dependencies at high transition regions (such as edges) do remain among the spectral planes.

The technical note in [4] presents a benchmark method that processes the sets of wavelet coefficient bits in an interleaved order for the three spectral planes by employing SPIHT and thereby achieves efficient bit allocation among the spectral planes. However, this method does not exploit the nonlinear dependencies among the spectral planes. A similar method that employs set-partitioning embedded block coder (SPECK), the predecessor of SBHP, is also suggested in [3] under the name color SPECK (CSPECK).

As a step toward exploiting the nonlinear dependencies, color embedded zerotree wavelet (CEZW) of [5] extends the spatial orientation tree (SOT) of embedded zerotree wavelet (EZW) by linking the chrominance nodes to their corresponding luminance nodes as their children. However, too many chrominance nodes or sets rooted at them are isolated too early in the coding process and poor low rate performance results due to coding their significance in many passes.

The CSPIHT algorithm, formulated in [6], links the SOTs for the three spectral planes of coefficients by defining the nodes in the lowest frequency subbands of the chrominance planes [second and third eigenvector component planes for Karhunen–Loeve transform (KLT)] as the children of the spatially corresponding nodes in the lowest frequency subband of the luminance plane (principal eigenvector component plane for KLT). Due to the structure of the overall tree, only luminance nodes and sets rooted at them are tested for significance at the initial passes. This is unlike CEZW of [5], where the chrominance root nodes or sets rooted at them are unnecessarily separately tested for significance at the initial passes even though they are almost always insignificant. Thereby, a savings in bit expenditure is achieved with respect to CEZW at low rates. However, for a discrete wavelet transform (DWT) with a large number of levels, the performance gain of CSPIHT over [4] is significant only at very low rates. At higher rates, the gain is negligible since the difference between the number of SOT roots for CSPIHT and the number of SOT roots for the benchmark method is small and the SOT roots are only visited in the bitplanes visited early in the coding process. The SOT of CSPIHT was later extended to link spatially related blocks of coefficients in [7].

By revealing that DWT yields correlations between luminance nodes in the spatial vicinity of each chrominance node, [8] suggests the use of a context covering these luminance nodes for coding the significances of chrominance nodes.

Vector SPIHT [9] is an extension of SPIHT that builds a SOT on vectors of DWT coefficients at corresponding locations in the three color planes to exploit the statistical dependencies among them. For video coding, [9] reports decent gains over H.263 at medium to high rates.

The works of [10]–[12] primarily address the nonuniformity of spectral content in the image for compression purposes. In

Manuscript received September 8, 2010; revised December 6, 2010; accepted January 16, 2011. Date of publication March 28, 2011; date of current version July 7, 2011. This paper was recommended by Associate Editor W. Zhang.

The author is with the Department of Computer Engineering, Istanbul Technical University, Istanbul 34469, Turkey (e-mail: ulugbayazit@itu.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2133790

[10], a region adaptive 1-D KLT across the color planes is followed by the spatial 2-D discrete cosine transform (DCT) within each spectral plane. A quadtree represents the recursive subdivision of the image is divided into variable size blocks. Each block is transformed with the KLT derived from its covariance matrix. The quadtree and the quantizer for the transform coefficients are jointly R-D optimized.

Unlike the block-based approach of [10], the works of [11] and [12] allow for more arbitrary shaped regions or classes that represent approximately homogeneous land cover, but the sizes of individual regions or classes are not R-D optimized. Moreover, side information resulting from class or region-based segmentation is only noted, but not controlled.

Locally adaptive resolution codec [13] exploits the uniformity in chrominance components by region-based coding. In the proposed semantic scalable two layers of coding, the first layer partitions an image into variable size blocks based on dynamic range and represents small blocks at contours and large blocks of uniform regions with their mean values. The second layer refines local texture with respect to this partition. The region segmentation map, generated at the decoder based on first layer information, facilitates region of interest (ROI) formation and region-based coding of chrominance information at little extra cost.

The DWT-based color image coding system proposed in this paper is similar to that of [10] in that an adaptive spectral transform (KLT) is applied across the spectral planes. The proposed method is more relevant than the one in [10] since the proposed adaptive spectral transform is designed to operate together with the DWT that has a proven superior performance record over the DCT. Unlike [10], where the adaptive spectral transform is applied in the spatial domain, the adaptive spectral transform of the current work is applied in the DWT domain for reasons to be explained in the next section. In this respect, it is similar to the work of [14], where filters for linear prediction of the chrominance transform coefficients are designed and applied in the DCT domain followed by rate-distortion (R-D) optimal bit allocation to the DCT subbands, uniform scalar quantization of the prediction errors, and entropy coding of the quantization levels.

The primary contribution of the current work is the specification of a R-D optimized design procedure for an adaptive spectral transform whose color image compression performance surpasses that of the widely used single global spectral transform at all rates when integrated with zerotree or block-based DWT coefficient coding algorithms.

In its minimization of a R-D cost function, the proposed method is somewhat similar to the coding mode selection process formulated in [15] for video coding. Coding mode selection has been extended to the adaptive selection of a fixed spectral transform from a small set of candidate, fixed spectral transforms for each fixed size block of the frame residual prediction error in a H.264/AVC advanced 4:4:4 profile setting in [16]. Unlike the proposed method, the side information rate for conveying the spectral transform identity is very little and not factored into the cost functions of the candidate, fixed transforms in [16].

In H.264/AVC, an experimentally determined model [17] for the Lagrange multiplier (slope of the distortion–rate curve at operating point) in terms of the quantization step size is employed. The proposed method determines the slope analytically by employing distortion and rate models suitable for DWT-based image coding.

In the next section, the key features of the proposed color image coding system are discussed by emphasizing the application of the proposed adaptive transform in the DWT domain. For coding the spectrally transformed DWT coefficients of the color planes, the SPIHT algorithm that forms the basis of CSPIHT, and JPEG2000, are briefly reviewed in this section. In Section III, the models of distortion, rate, and slope estimates, as well as the efficient encoding of the side information, are presented first. Based on the derived estimates of the costs of representing subsets of spatially oriented coefficients by their bases, the design procedure for the adaptive transform, that consists of iterative R-D optimized decisions to merge small subsets into larger ones, is also described in Section III. Section IV discusses the necessity of performing multiple coding iterations to achieve a specified target rate, and proposes a two iteration solution. The modified forms of the rate and slope estimates optimized for use in the second coding iteration are introduced at the end of this section. Section V mainly presents experimental results with a compression system employing the proposed adaptive spectral transform and compares them against results with reference compression systems on a large set of color images. Concluding remarks are presented in the final section.

II. Features of the Adaptive Transform

The decorrelating transforms such as the signal dependent KLT or the fixed RGB-YCrCb are commonly applied to the color planes of a color image prior to encoding so as to compact most of the signal energy in the three color planes to a single spectral plane (the principal eigenvector component plane for KLT and the Y component plane for YCrCb) and thereby realize a compression performance advantage. At the end of decoding, a corresponding inverse spectral transform yields the reconstructed color planes.

A single KLT basis usually does not offer adequate compression performance advantage since the typical natural images are nonuniform in color content. As suggested by [10], the use of a different KLT basis for each region of uniform color/spectral content within the image can better exploit the nonuniformity in color/spectral content over the entire image. Additionally, a transform like 2-D DCT or 2-D DWT must be spatially applied to compact the energy into a few coefficients along the spatial directions for efficient compression.

A. Order of the Spatial and Spectral Transforms

Unlike [10] that uses the computationally efficient 2-D DCT, the spatial transform employed in the current paper is the performance efficient 2-D DWT that is widely adopted in the state-of-the-art image coding. As shown in Fig. 1, the spatial 2-D DWT transform is not preceded, but followed by one or more spectral 1-D KLT transform(s), each of which is uniquely

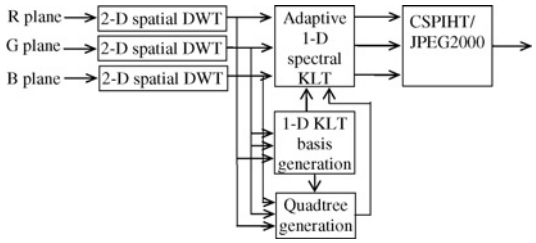

Fig. 1. Block diagram of the proposed system.

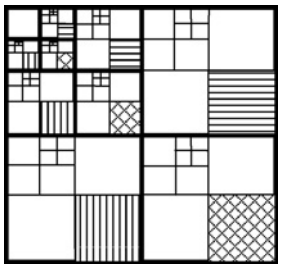

Fig. 2. 2-D DWT decomposition and an example of a single quadtree representing the spatial partitioning of the set of DWT coefficients of all high frequency bands into subsets in a unified way. The thick lines show the boundaries between the subbands, the thin lines show the boundaries between subsets of coefficients. An example spatially oriented subset is shown as the collection of shaded blocks. The second decision specifies whether or not this subset is further partitioned into three directionally subsets with a unique basis representing coefficients of each. If the subset is not partitioned, a single basis is used to represent all coefficients of the subset. All blocks of a particular directionally oriented subset are shown by the same shading texture.

determined for a subset [region(s) or block(s)] of a partition of DWT coefficients of the image. If the spatial 2-D DWT were to follow the 1-D spectral KLT transforms(s) of image pixels as in [10], the subband analysis filter kernels would have to be applied across boundaries between two subsets of spectral coefficients obtained with different KLT bases. Filtering across boundaries may be circumvented by mirror extending the filtered values at the boundaries. However, performing a mirror extension at the boundary between each pair of subsets significantly reduces the performance efficiency of 2-D spatial DWT. The solution proposed in [12] employs shape adaptive DWT and shape adaptive SPIHT for coding spectrally transformed arbitrary regions of the image.

A second reason for the order of transforms is that, in the DWT domain, it is possible to further partition a set of spatially oriented coefficients (corresponding to a region or a block in the image) into smaller subsets of spatially as well as *directionally oriented* coefficients. This way, the KLT bases may be better adapted to the edge directionality.

A final reason is that, in each color plane, each set of spatially and/or directionally oriented high frequency DWT coefficients has near zero mean value. Such mean values do not need to be coded as side information for the application of the 1-D KLT across the color planes of DWT coefficients. Had the 1-D KLT been performed directly on the input image prior to 2-D DWT, the subtraction and subsequent coding of the mean value for each set of image pixels in each color plane would have been necessary.

B. Structure of the Adaptive Transform and the Coded Decisions

A precise description of arbitrary shaped sets of coefficients of uniform content is expensive from a compression viewpoint even at high coding rates. The viable solution proposed in [10], that is also employed in other contexts such as variable block size motion estimation, is to spatially partition the image into variable size blocks of near uniform content. In the current paper, the layout of the subsets of coefficients is concisely represented by a quadtree structure.

When a set of coefficients represented by its KLT basis is partitioned into smaller subsets represented by their own KLT bases, the energy compaction of the resultant coefficients improves, but the R-D performance may or may not improve due to the increase in side information for coding the increased number of KLT bases. This suggests that one should decide to split a quadtree node by partitioning a set or, conversely, prune a subtree of the quadtree by merging small subsets into a large set only if a R-D improvement is expected as a consequence.

Since the spectral 1-D KLT is applied on the 2-D DWT coefficients, it is conceivable that, for each subband, the spatial partition be unique and represented by a different quadtree structure. However, such an approach necessitates a substantial side information rate at low rates. In order to conserve this rate, the strong statistical dependencies among the magnitudes of coefficients of the same spatial orientation across high frequency subbands may be exploited. All high frequency subbands are constrained to be spatially partitioned into blocks of coefficients in the same way. The resultant partitions of all high frequency subbands are represented in a unified way by a single quadtree as shown by an example illustration in Fig. 2. The maximum depth of this quadtree is dictated by the dimensions of the lowest frequency subband.

For each set of spatially oriented blocks of coefficients across all high frequency subbands, a second decision is made as to whether or not to partition it into three directionally (vertically, horizontally, and diagonally) oriented subsets. A unique basis for each of these three subsets can represent and exploit the statistical dependencies among the color components within that subset in a better way. It is favored if the R-D cost is reduced as a consequence.

To improve the coding performance, a final decision is made as to whether to project each subset of coefficients onto the fixed YCrCb basis[1] instead of the signal dependent KLT basis. The fixed basis is favored over the KLT basis if the R-D cost of the fixed basis is smaller due to excessive side information for coding the KLT basis vectors.

While the set of coefficients of the high frequency subbands are partitioned into subsets with each subset represented by its own unique KLT basis, the lowest frequency subband coefficients are not partitioned, but just represented by a single unique basis. The side information for partitioning the lowest frequency subband is hard to justify, since the number of bits coded for its coefficients is typically very small. On the contrary, the mean values of the lowest frequency DWT

[1]Technically, the applied transformation slightly differs from the RGB-YCrCb transformation in that no bias and range limits are applied.

coefficients are coded as side information for the proper application of the 1-D KLT spectral transform since these mean values are generally significantly nonzero.

C. Coding of the Coefficients with CSPIHT or JPEG2000

Both the spatial 2-D DWT and the spectral 1-D KLT compact the signal energy to a small number of transform coefficients. Superior coding performance can be achieved if this nonuniform distribution of the signal energy is exploited by an efficient bit allocation strategy. Toward this end, one may employ a set partitioning-based wavelet coefficient coding scheme for color images [4], [6] that visits (and thereby represents) large magnitude coefficients early in the coding process and explicitly or implicitly allocates rate to the spectral as well as the spatial components.

The well-known grayscale image coding method SPIHT [1] encodes and transmits wavelet coefficients in multiple *coding passes* corresponding to bitplanes (in the order from the most significant to the least significant). In each coding pass, only the wavelet coefficients with magnitudes exceeding a certain threshold (significant coefficients) are encoded. The coefficients are ordered in hierarchies, called SOTs, with roots in the lowest frequency subband, branching successively into higher frequency subbands at the same spatial orientation. The most significant bit of a significant coefficient is encoded by identifying its location via recursively subdividing the spatial orientation tree into smaller subtrees and single coefficient sets. In subsequent passes, the refinement bits (lesser significant bits) of a significant coefficient are also coded by successive approximation quantization. The algorithm has high performance due to representation of large sets of zero bits of coefficients by a single symbol.

Since the performance of the benchmark method [4] is generally inferior to that of CSPIHT [6] for DWT with few levels or for very low coding rates, CSPIHT has been adopted for jointly coding the coefficients of the three spectral planes.

A second alternative that can be employed here is the block-based wavelet coefficient encoding strategy of JPEG-2000. JPEG-2000 determines a sequence of R-D operating points for the coding of each block of coefficients and the coding process for each block is truncated at the operating point with the minimum Lagrangian R-D cost.

III. R-D Optimized Adaptive Transform Design

Let the rate equivalent of the R-D cost estimate at the R-D operating point (R, D) be defined as $\hat{J} = \hat{D}/\hat{s} + \hat{R}$, where $\hat{D}$, $\hat{R}$, and $\hat{s}$ are estimates of the distortion, rate, and the absolute value of the slope at this point, respectively. This section first treats how these estimates are obtained for wavelet-based coding in JPEG-2000 or CSPIHT.

A. Distortion, Rate, and Slope Estimates

The expected distortion of the successive approximation quantizer with a threshold (step size) of Δ can be expressed as a weighted sum of the conditional expected distortions for the insignificant and significant coefficients as follows:

$$\begin{aligned} D_{coef}(\Delta) &= E\left[|X-\hat{X}|^2 \,||X|<\Delta\right] Pr\{|X|<\Delta\} \\ &+ E\left[|X-\hat{X}|^2 \,||X|\geq\Delta\right] Pr\{|X|\geq\Delta\} \end{aligned} \quad (1)$$

where

$$E\left[|X-\hat{X}|^2||X|<\Delta\right] = \int_{-\Delta}^{\Delta} |x|^2 f_X(x||X|<\Delta)dx$$

$$E\left[|X-\hat{X}|^2||X|\geq\Delta\right] = 2\sum_{k=1}^{S/2}\int_{k\Delta}^{(k+1)\Delta} |x-(k+0.5)\Delta|^2 \cdot f_X(x||X|\geq\Delta)dx$$

where X and $\hat{X}$ are the original and reconstructed coefficient values, and S is the number of quantization intervals besides the central bin. The density $f_X(x)$ is commonly modeled as Laplacian. Let M_n be the number of significant coefficients out of a set of n coefficients. By substituting the first integral in (1) with the sum over the square magnitudes of insignificant coefficients, and analytically integrating the second integral, the distortion estimate is formed in [18] as follows:

$$\hat{D}_{coef}(M_n) = \frac{1}{n}\left(\sum_{i=M_n+1}^{n} |x(i)|^2 + \frac{M_n}{4\lambda^2}h(K,S)\right) \quad (2)$$

where $K = \lambda\Delta$, $x(i)$ is the coefficient of ith largest magnitude, λ is the parameter of the Laplacian distribution estimated as the inverse of the mean magnitude of a coefficient, and

$$h(K,S) = \frac{(1-e^{-\frac{KS}{2}})}{e^K-1}(K^2-4K+8-e^{-K}(K^2+4K+8)).$$

Although not stated in [18], the estimate in (2) is valid at operating points on the boundary $M_n = M_{n,\Delta}$ between two coding passes due to all significant coefficients having the same expected distortion of $h(K,S)/(4\lambda^2)$. If not all of n coefficients were coded at threshold value Δ, but say $n_{2\Delta}$ were coded at threshold value 2Δ, and $n_\Delta = n - n_{2\Delta}$ were coded at threshold value Δ, (2) could be modified accordingly if M_{n_Δ} and $M_{n_{2\Delta}}$ were also known. The difficulty here is that the scan order of the coefficients depends on the significant coefficients in previous passes and even the distribution of n to n_Δ and $n_{2\Delta}$ for each set at any operating point is hard to predict before actual coding. Therefore, throughout this paper, (2) is employed only at the coding pass boundaries.

Prior to the actual encoding, the distortion may also be determined at the coding pass boundaries by a computationally more expensive, but also more exact alternative to the estimate provided by (2). Specifically, the quantization strategy of the specific wavelet coefficient coding scheme (SPIHT or JPEG2000) is simulated. On the contrary, the estimate for the rate of the DWT coefficients prior to their coding, given in [19] (or [18]) is as

$$\hat{R}_{coef}(M_n) = \frac{C}{n}M_n \quad (3)$$

is only approximate. Even though the assumption of linearity of the dependence of rate on M_n is valid for most images, the linearity constant C is in a wide range of 3–7. In [18] and [19], the values employed for C are 6.6 and 5.5, respectively.

The total rate estimate for coding a total of N coefficients in the entire image can be expressed as

$$\hat{R}_{tot} = \hat{R}_{coef}(M_N) + R_{sideinfo}.$$

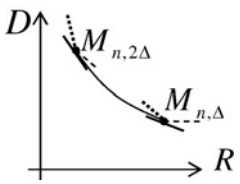

Fig. 3. Distortion-rate characteristic between two adjacent coding pass boundaries (black dots) is shown by the solid thin curve. Symmetric form of the slope estimate at the boundary given in (8) is suggested by the solid thick lines. It is more appropriate for interpolation between two boundaries than the asymmetric form in (4) suggested by the dotted lines above each boundary.

Side information rate, $R_{sideinfo}$, consists of the number of bits necessary to convey to the decoder the quadtree structure (1 bit for each internal node of the quadtree), one flag bit for each subset of the final partition that signals whether a fixed or a signal dependent KLT basis is used, and the bits representing the components of the KLT basis vectors.

For each KLT basis, only two components of two of the basis vectors are coded by representing them in a polar coordinate system. The radial component is coded with 12 bits precision and the angular component is coded with a precision that increases linearly from 6 to 8 bits with radius. The third component of these vectors is derived by applying the unit norm property. The third vector is derived as the cross product of the other two vectors.

The absolute value of slope $s = \left| dD_{coef}/dR_{coef} \right|$ is estimated as

$$\begin{aligned} \hat{s}(M_n) &= \frac{\hat{D}_{coef}(M_n) - \hat{D}_{coef}(M_n - 1)}{\hat{R}_{coef}(M_n - 1) - \hat{R}_{coef}(M_n)} \\ &= C^{-1}[|x(M_n)|^2 - h(K, S)/4\lambda^2] \\ &\approx C^{-1}[(E[|X| \,|\Delta \le |X| < 2\Delta])^2 - h(K, S)/4\lambda^2]. \end{aligned} \tag{4}$$

Unlike (2), the above estimate is valid *for any* operating point M_n belonging to coding pass with threshold value Δ. Note that the substitution of the magnitude of the coefficient with the conditional expectation is legitimate since the coefficients becoming significant within a pass are not ordered by magnitude. It can be shown that $E[|X|| \,\Delta \le |X| < 2\Delta] = \Delta\left(\frac{1}{K} + \frac{1-2e^{-K}}{1-e^{-K}}\right)$. Typically, for even the highest threshold values employed to attain high rates, average coefficient magnitudes are such that $K \le 1$, which confines the conditional expectation to the range $[1.4\Delta, 1.5\Delta]$. Despite this finding, the conditional expectation is well approximated by Δ in [18] and [20] for operating points at (or near) the coding pass boundary. This is due to the fractional bitplane coding feature of JPEG2000 and SPIHT that yields a smaller actual slope than that predicted by (4) for operating points at the coding pass boundary (the solid line through the point $M_{n,\Delta}$ in Fig. 3 has smaller slope than the dotted line).

B. Iterative Cost Minimization Procedure

The proposed transform is the collection of representations at the leaf nodes of a quadtree. Each quadtree leaf node corresponds to a subset of spatially oriented coefficients across all high frequency subbands. For each such subset at a leaf node indexed with letter t, let $\hat{J}_{t,i}$ be the rate equivalent of the R-D cost estimate for employing the ith candidate representation. The best candidate representation that is actually employed for leaf node t satisfies $i^*(t) = \arg\min_i \hat{J}_{t,i}$.

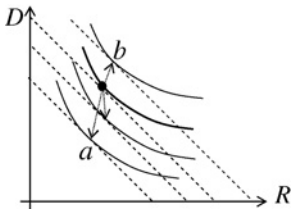

Fig. 4. Illustration of the R-D cost changes as a result of pruning candidate subtrees. Among all candidate subtrees, the best subtree for pruning yields an operating point that falls on the lowest dashed line with slope $\hat{s}$.

In this paper, two candidate representations are obtained by employing a fixed or signal dependent basis for the entire set of coefficients. Seven other candidates are obtained by partitioning this set into three directionally oriented subsets and representing each subset by a fixed or a signal dependent basis. Note that the combination of representing all three directionally oriented subsets with the fixed basis is equivalent to representing the partitioned set with the fixed basis.

A subtree of the quadtree is pruned off by merging the subsets at the leaf nodes of the subtree into a larger set and employing the best of nine candidate representations for this set. In Fig. 4, the thick curve is a distortion versus rate characteristic associated with a transform before pruning the quadtree. The operating point at the target rate is indicated by a black dot. Each thin curve is the distortion versus rate characteristics associated with a candidate transform after pruning off a subtree from the quadtree. Let us assume that $\hat{s}$, the absolute value of the distortion-rate slope estimate at the operating point, does not change with a pruning. Each dashed line is a set of points of equal R-D cost. Pruning of the best candidate subtree by the proposed adaptive transform method yields a new operating point a with the smallest R-D cost. Note that, as suggested by the slopes of dotted light arrows in opposite directions ending at two candidate operating points a and b in Fig. 4, $\delta\hat{D}_{coef}/\delta\hat{R}_{tot}$ for two candidate prunings might be equal despite a large difference in cost change between the two. An acceptable pruning must satisfy $\delta\hat{J} = \hat{J}_{post} - \hat{J}_{prev} = \delta\hat{D}_{coef}/\hat{s} + \delta\hat{R}_{tot} < 0$ where $\hat{J}_{prev}$ and $\hat{J}_{post}$ are the cost estimates before and after the pruning. This requirement is equivalent to the requirements $\delta\hat{D}_{coef}/\delta\hat{R}_{tot} < -\hat{s}$ if $\delta\hat{R}_{tot} > 0$, and $\delta\hat{D}_{coef}/\delta\hat{R}_{tot} > -\hat{s}$ if $\delta\hat{R}_{tot} < 0$ which are satisfied by a transition to a, but not b in Fig. 4. On the contrary, if $\hat{s}$ changes with a pruning, $\delta\hat{D}_{coef}/\delta\hat{R}_{tot}$ cannot be reliably used as a pruning criterion.

For each merge (pruning), a substantial complexity reduction is made possible by approximating $\delta\hat{R}_{coef}(M_{N,\Delta})$, the change in the rate estimate for the entire image with $\delta\hat{R}_{coef}(M_{n,\Delta})$, the change in the rate estimate for only the merged sets containing n coefficients. Note that, the coding of disjoint sets of a partition are loosely coupled in the sense that altering the basis inside a set changes the probability models for the coded symbols which may affect the coding process of other sets. In practice, the basis and partitioning decisions made by using $\delta\hat{R}_{coef}(M_{n,\Delta})$ yield coding gains that are nearly

as good as those made by using $\delta\hat{R}_{coef}(M_{N,\Delta})$. On the contrary, for each merge, the change in the mean square error distortion for the entire image is exactly equal to the change in the the mean square error distortion for only the merged sets, i.e., $\delta D_{coef}(M_{n,\Delta}) = \delta D_{coef}(M_{N,\Delta})$, even though it may be that $\delta\hat{D}_{coef}(M_{n,\Delta}) \neq \delta\hat{D}_{coef}(M_{N,\Delta})$.

A simple design procedure for the adaptive transform can now be prescribed; at the initial step, the entire set of high frequency coefficients are partitioned into the smallest subsets where each subset is a collection of spatially oriented blocks of coefficients. The best of nine representations is employed for each subset. At each following iterative step, the cost change $\delta\hat{J}(v) = \hat{J}_{post}(v) - \hat{J}_{prev}(v)$ for each possible candidate merge (subtree pruning) is evaluated by letting $\hat{J}_{prev}(v) = \sum_{w\in\mathcal{L}(v)} \hat{J}_{w,i^*(w)}$, $\hat{J}_{post}(v) = \hat{J}_{v,i^*(v)}$ where $\mathcal{L}(v)$ is the set of leaf nodes of the subtree rooted at node v. The merge at internal node $v^* = \arg\min_v \delta\hat{J}(v)$ is performed. Iterations continue until the minimum cost change is positive.

IV. Proposed Coding Method that Just Achieves the Target Rate in Two Iterations

One drawback of the method discussed in Sections II and III is that before encoding the coefficient values for a given target coding rate, it is not possible to determine the final threshold value Δ that *just achieves* this rate. Within the context of the successive approximation quantizers of CSPIHT and JPEG-2000, the threshold value Δ is said to just achieve a target coding rate if the maximum coding rate for a quantizer with step size Δ exceeds the target coding rate, but the maximum coding rate for a quantizer with step size 2Δ comes short of the target coding rate. Let the final threshold value that just achieves the target coding rate be designated as Δ_p. The adaptive transform should be designed by taking $\Delta = \Delta_p$ in the distortion, rate, and slope models of (2)–(4). Using smaller Δ values such as $\Delta_p 2^{-1}, \Delta_p 2^{-2}, \ldots$ yields adaptive transforms that are inferior to that for $\Delta = \Delta_p$.

In general, to just achieve a given target coding rate, more than one coding iteration is needed. In a preliminary publication [20] of the method discussed in Sections II and III, a trivial exhaustive search scheme for Δ_p is implemented that exponentially decreases final Δ with each iteration as $\Delta = \Delta_0 2^{-i}$ until $\Delta = \Delta_p$ just achieves the target rate. The number of iterations to attain Δ_p like this could be infeasible for low complexity encoder operation at high coding rates.

A new scheme, that limits the complexity to two coding iterations, is proposed in this paper for CSPIHT-based encoding. For each $i = 0, 1, 2, \ldots, I$ let $\Delta_i = \Delta_0 2^{-i}$. The steps of the proposed scheme are as follows.

1) Initial step.
   a) The target rate R^{tar} is specified by the user.
   b) All 2-D DWT coefficients of the color planes are spectrally transformed by a single KLT basis. Let $R^{tar}_{coef} = R^{tar} - R_{sideinfo}$.
   c) For each $i = 0, 1, 2, \ldots, I$: M_{N,Δ_i}, number of significant coefficients that exceed Δ_i, are determined to yield the rate estimate $\hat{R}_{coef}(M_{N,\Delta_i})$ by using (3).
   d) $\hat{\Delta}_p$ is initialized: $\hat{\Delta}_p = \max_{\{i:\hat{R}_{coef}(M_{N,\Delta_i})\geq R^{tar}_{coef}\}} \Delta_i$.
2) For each of two coding iterations.
   a) The adaptive transform is designed by employing rate and slope estimates based on the previously determined $\hat{\Delta}_p$ and applied to spectrally transform the 2-D DWT coefficients.
   R^{tar} is updated: $R^{tar}_{coef} = R^{tar} - R_{sideinfo}$.
   b) The coefficients are CSPIHT encoded at rate $R_{coef}(\Delta_I)$ in the first and R^{tar}_{coef} in the second iteration.
   c) For each $i = 0, 1, 2, \ldots, I$, the rate $R_{coef}(\Delta_i)$ at the end of the pass for threshold Δ_i is determined.
   d) $\hat{\Delta}_p$ is updated: $\hat{\Delta}_p = \max_{i:R_{coef}(\Delta_i)\geq R^{tar}_{coef}} \Delta_i$.

The bitstream generated in the second CSPIHT encoding of the coefficients is the actual transmitted bitstream. Experiments on all the test images at all the rates tested have shown that $\hat{\Delta}_p$ convergences in two coding iterations to the Δ_p value given by the trivial exhaustive search scheme.

In the first coding iteration of the proposed scheme, the estimates are obtained with (3) and (4) as before. In the second iteration, the estimates are optimized by using data collected in the first iteration. Let p_Δ and $1 - p_\Delta$ be the probabilities of coding and not coding, respectively, a coefficient in the final coding pass at rate R^{tar}_{coef}. The value of p_Δ is estimated as the fraction of the number of coefficients coded in the final coding pass of the first coding iteration. For this purpose, each coefficient coded in this pass is tagged to be counted later on.

In the second coding iteration, the proportionality constant, chosen in an *ad hoc* manner in [18] and [19] (as 6.6 or 5.5), is adjusted adaptively as follows:

$$C = \frac{R^1 - R^1_{sideinfo}}{(p_\Delta M^1_{N,\Delta} + (1 - p_\Delta)M^1_{N,2\Delta})/N}.$$

In this equation, R^1 and $R^1_{sideinfo}$ are the rate and side information rate, respectively, and $M^1_{N,\Delta}$ and $M^1_{N,2\Delta}$ are the number of significant coefficients in the entire image at the end of the final pass and at the end of the pass before the final pass, respectively, in the first coding iteration.

The estimates in (2)–(4), that were employed in [20], are inaccurate especially when p_Δ is small. More accurate estimates are obtained by interpolating between the estimates for the final pass and the estimates for the pass before the final one in the second coding iteration. Specifically, let $M^2_{n,\Delta}$ and $M^2_{n,2\Delta}$ be the number of significant coefficients out of a set of n coefficients at the end of the final pass (designated by Δ) and at the end of the pass before the final pass (designated by 2Δ), respectively, in the second coding iteration (designated by superscript 2). The modified estimates used in the design of the adaptive transform in the second coding iteration are

$$\begin{aligned}\hat{D}^2_{coef}(M^2_{n,\Delta}, M^2_{n,2\Delta}) &= p_\Delta \hat{D}_{coef}(M^2_{n,\Delta}) \\ &+ (1 - p_\Delta)\hat{D}_{coef}(M^2_{n,2\Delta})) \end{aligned} \quad (5)$$

$$\begin{aligned}\hat{R}^2_{coef}(M^2_{n,\Delta}, M^2_{n,2\Delta}) &= p_\Delta \hat{R}_{coef}(M^2_{n,\Delta}) \\ &+ (1 - p_\Delta)\hat{R}_{coef}(M^2_{n,2\Delta}) \end{aligned} \quad (6)$$

$$\hat{s}^2(M^2_{n,\Delta}, M^2_{n,2\Delta}) = p_\Delta \hat{s}(M^2_{n,\Delta}) + (1 - p_\Delta)\hat{s}(M^2_{n,2\Delta}) \quad (7)$$

where the distortion versus rate tradeoff at the end of the coding pass for the threshold value Δ is defined as

$$\begin{aligned}\hat{s}(M^2_{n,\Delta}) &\approx \frac{1}{2}\left[\frac{\hat{D}_{coef}(M^2_{n,\Delta}) - \hat{D}_{coef}(M^2_{n,\Delta} - 1)}{\hat{R}_{coef}(M^2_{n,\Delta} - 1) - \hat{R}_{coef}(M^2_{n,\Delta})}\right. \quad (8)\\ &\quad \left. + \frac{\hat{D}_{coef}(M^2_{n,\Delta} + 1) - \hat{D}_{coef}(M^2_{n,\Delta})}{\hat{R}_{coef}(M^2_{n,\Delta}) - \hat{R}_{coef}(M^2_{n,\Delta} + 1)}\right]\\ &= C^{-1}\left((\left|x(M^2_{n,\Delta})\right|^2 + \left|x(M^2_{n,\Delta} + 1)\right|^2)/2\right.\\ &\quad \left. - (h(K,S) + h(K/2, 2S+1))/8\lambda^2\right)\\ &\approx C^{-1}\left(((E\left[|X|\,|\Delta \le |X| < 2\Delta\right])^2\right.\\ &\quad + (E\left[|X|\,|\Delta/2 \le |X| < \Delta\right])^2)/2\\ &\quad \left. - (h(K,S) + h(K/2, 2S+1))/8\lambda^2\right).\end{aligned}$$

Fractional bitplane encoding feature of SPIHT yields a distortion-rate characteristic that falls below the straight line connecting two adjacent coding pass boundaries. As shown in Fig. 3, a two-sided symmetric form of the slope in (8) better interpolates the slope at intermediate operating points between these boundaries than the asymmetric form in (4).

It must be noted that the p_Δ value used in the interpolations of the rate and slope estimates of each set is common to all the sets rather than uniquely estimated for each set. A set conditional estimate of the probability of a coefficient coded at the final coding pass can also be obtained during the first coding iteration in a manner similar to the one described above, but is not robust enough for reliably predicting the fraction of coded coefficients in a set in the second iteration.

V. Experimental Results and Discussion

In this section, we report and discuss results of simulations conducted to compare the compression performance of compression systems that integrate the adaptive transform and the coefficient coding schemes of CSPIHT or JPEG2000 against the performances of several reference compression systems.

A large set of 32 color images has been used to evaluate the compression performances. Twenty-four of the images are the 768×512 or 512×768 *Kodak true color* 24 bit images, six of them are the 512×512 24 bit images, *Lena*, *Baboon*, *Tiffany*, *Sailboat*, *Airplane*, and *Pepper*, and the other two are the high resolution 2048×3072 24 bit images *artificial* and *hdr*. Throughout the simulations, the DWT filters used are the CDF 9/7 biorthogonal filters.

Let d_γ be the mean squared error of the color or spectral plane γ. By letting $\bar{d} = d_Y$, $\bar{d} = (d_U + d_V)/2$, and $\bar{d} = (d_R + d_G + d_B)/3$, the peak signal-to-noise ratios (PSNRs) of the Y component plane, UV component planes, and the color image are evaluated as

$$PSNR_{dB} = 10log_{10}\left((255)^2/\bar{d}\right). \quad (9)$$

Fig. 5 presents results obtained on several images with an implementation employing 2-D DWT, the adaptive transform introduced in Sections II and III and the JPEG 2000 codec Kakadu [21] for coding the resulting transform coefficients. Since the order of the transform operations proposed here does not agree with the order of the transform operations inside Kakadu (multi-component transform followed by DWT), these operations are performed separately outside of Kakadu and disabled inside Kakadu by defining the multi-component transform as an identity matrix and setting Clevels=0 for DWT. The *i*th distortion-rate point is attained by setting the Qstep value as $Q_0 2^{-i}$. A single coding iteration is run to attain the maximum coding rate for the choice of the quantization step size used. For comparison, corresponding results are also presented for the JPEG 2000 Kakadu implementation with the multi-component transform and DWT enabled inside Kakadu by defining the multi-component transform as the single KLT transform matrix derived from the entire image. The Qderived option that relates the quantization step sizes to the LL subband's step sizes was enabled in both implementations. On certain images, the coding gain with the use of the adaptive transform in place of the single KLT transform is observed to increase progressively with rate upto as much as 2 dB.

The next set of results have been collected with the proposed method (system) described in Sections II–IV that achieves precise rate control in two coding iterations and several other methods for comparison. All methods employ CSPIHT for coding the transform coefficients. "Single spatial domain KLT basis" refers to the application of a single spectral KLT transform before the spatial DWT and is commonly employed in other papers such as [4] and [6]. In "single DWT domain KLT basis," the spatial DWT precedes the spectral KLT transform. The "two DWT domain KLT bases" applies a basis each to the low and high frequency subbands and is equivalent to a simplified version of the proposed method with a quadtree depth of zero. The proposed method, which is termed as "adaptive transform," applies more than one bases to the high frequency coefficients by spatially and directionally partitioning them, and allows fixed basis as well as KLT basis use. The performances at various rates ranging from 0.05 b/p to 2.5 b/p are reported in Table I as the average of the PSNRs of Y and UV components of the 32 images. In Table II, the number of times the proposed method yields a performance higher than each of the other methods is reported for the same set of rates considered.

The performance gain of the "two DWT domain KLT bases" over the "single spatial domain KLT basis" increases with rate upto as much as an average of 0.65 dB in Y and an average of 0.95 dB in U and V components at 2.5 b/p. Most natural images contain significant high frequency content that is distributed across the color planes differently than the low frequency content. Therefore, it is important to capture the different spectral dependencies among the color components of the coefficients of the low and high frequency subbands with different bases.

The performance gain of the "adaptive transform" over the "single spatial domain KLT basis" increases with rate upto as much as an average of 0.9 dB in the Y component and an average of 1.35 dB in the U and V components at 2.5 b/p. In this case, the large side information rate due to the transmission of a large number of bases vectors for a complex partition is

TABLE I

PERFORMANCES (AVERAGE OVER 32 IMAGES) OF THE PROPOSED AND REFERENCE METHODS AT VARIOUS RATES.

Rate (b/p)	Average PSNR(dB) of Y Plane, UV Planes					
	Single Spatial Domain KLT Basis + CSPIHT[6]	Single DWT Domain KLT Basis + CSPIHT	Single Spatial Domain KLT Basis + JPEG2000	Two DWT Domain KLT Basis + CSPIHT	Adaptive Transform + CSPIHT: Sim. of Quantization	Adaptive Transform + CSPIHT: Distortion Estimate (2)
0.05	27.03, 37.19	27.02, 37.20	26.96, 37.19	27.01, 37.22	27.03, 37.31	27.01, 37.32
0.10	28.93, 38.67	28.92, 38.67	28.87, 38.79	28.93, 38.76	28.96, 38.91	28.94, 38.87
0.25	32.01, 40.76	32.01, 40.79	31.98, 40.93	32.04, 41.01	32.09, 41.26	32.08, 41.23
0.50	35.02, 42.66	35.02, 42.70	34.99, 42.82	35.09, 43.08	35.17, 43.35	35.14, 43.36
0.75	37.11, 43.93	37.13, 44.00	37.14, 44.06	37.23, 44.45	37.32, 44.73	37.26, 44.77
1.00	38.79, 44.89	38.82, 44.99	38.84, 45.03	38.96, 45.48	39.08, 45.77	39.01, 45.87
1.50	41.42, 46.39	41.51, 46.56	41.56, 46.50	41.70, 47.11	41.83, 47.42	41.78, 47.58
2.00	43.45, 47.58	43.63, 47.85	43.70, 47.75	43.89, 48.40	44.04, 48.77	43.99, 48.96
2.50	45.21, 48.53	45.53, 48.91	45.60, 48.76	45.86, 49.48	46.10, 49.87	46.05, 50.17

more easily afforded at high rates. On the contrary, the average gain of the "adaptive transform" over the "two DWT domain KLT bases" is limited to 0.25 dB in Y component and 0.4 dB in U and V components at medium to high rates since not all images have local regions of distinct color transitions that can be exploited with the allocation of unique bases to these regions. For example, in Kodak color images number 5 and number 8, the sharp color transitions of contours of motobikes and roofs occur in close proximity and similar high frequency content occurs over most of these images which cannot be isolated distinctively in local regions. Moreover, even when such an isolated local region of distinct color transition does exist in an image, it need not coincide with a block structured subset of the finest quadtree partition. The highest gains with the "adaptive transform" have been observed for the *artificial* image which has sharp color transitions along edges of artificial objects that can be isolated in local regions.

The last two columns of Table I show that the distortion estimate of (2) compromises average PSNR of Y component only slightly if it is used in place of the simulation of the quantization strategy that exactly determines distortion. On the contrary, the slope estimates in (4) or (7) and (8) are even less sensitive to the error in the distortion model of (2) and the use of exact distortion values in forming the slope estimate does not yield an added benefit.

Fig. 6 shows the unified spatial partitioning of the high frequency subband coefficients for a coding rate of 0.5 b/p overlayed on top of the original *Lena* image. A perceptible visual quality advantage is observed in the reconstruction of the *Lena* image obtained with the "adaptive transform" over the reconstruction obtained with the "single spatial domain KLT basis." The advantage is more conspicuous at the fine texture areas on the hat. It is also observed that the texture is slightly better reconstructed with the "adaptive transform" than "two DWT domain KLT bases."

In Fig. 7, the reconstruction of a section of the high resolution image *artificial* coded at 0.1 b/p with the "adaptive transform" displays a marked quality advantage when compared to the reconstructions with the "two DWT domain KLT bases" and especially the "single spatial domain KLT basis." CSPIHT coding of transform coefficients does not appear to perform worse than JPEG2000 coding for this image either.

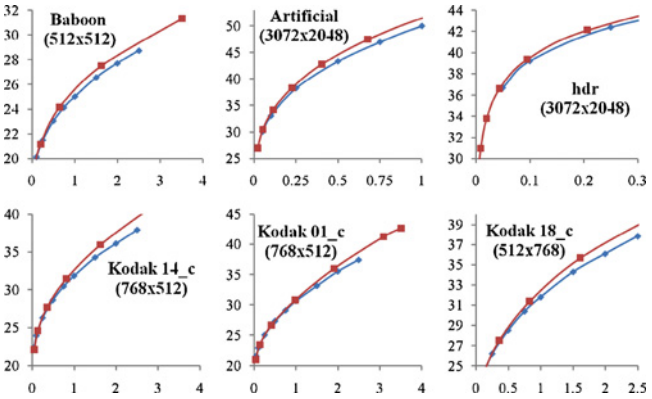

Fig. 5. PSNR (dB) versus rate (b/p) curves for several of the tested images obtained with the adaptive transform introduced in Sections II and III (square) and the single KLT transform derived from the entire image (diamond). The transform coefficients were coded with JPEG-2000 Kakadu implementation in both cases.

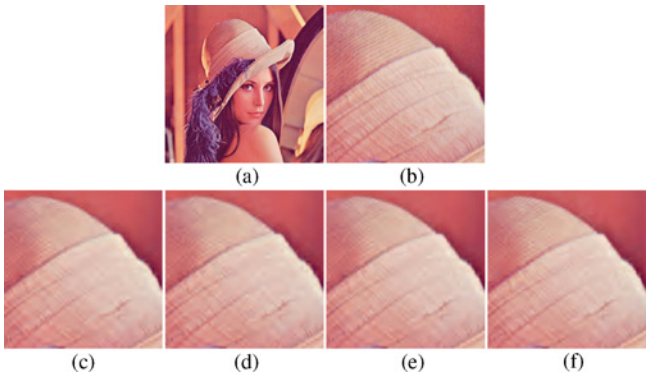

Fig. 6. Top: (a) Spatial partitioning (at 0.5 b/p) of high frequency subband coefficients overlayed on the original *Lena* image. (b) Enlarged part of *Lena*'s hat. Bottom: Corresponding reconstructions at 0.5 b/p. (c) "Single spatial domain KLT basis" [6] (overall image Y-PSNR = 37.13 dB, UV-PSNR = 39.66 dB). (d) Single spatial domain KLT followed by JPEG-2000 coding of coefficients (overall image Y-PSNR = 36.16 dB, UV-PSNR = 39.54 dB). (e) "Two DWT domain KLT bases" (overall image Y-PSNR = 37.26 dB, UV-PSNR = 39.80 dB). (f) Proposed "adaptive transform" (overall image Y-PSNR = 37.31 dB, UV-PSNR = 39.95 dB).

The rationale for the nonuniform partitioning of the coefficients in the "adaptive transform" is seen in Fig. 8. The performances with various number of levels of uniform partitioning fall short of the performance of nonuniform spatial partitioning at either the low or the high rates. Additionally, the number of levels giving the best R-D characteristic depends on the coded image.

Let the fractional error in estimating the rate of the coefficients, $R_{coef} = R_{tot} - R_{sideinfo}$, as $\hat{R}_{coef}$ be denoted as $\epsilon = (R_{coef} - \hat{R}_{coef})/R_{coef}$. The average ϵs obtained with the rate estimation equations of (3) and (6) over the set of 32 images at the nine rates reported in Table I are 0.480 and 0.007, respectively. The modified rate estimate of (6) brings a major improvement over the originally proposed rate estimate.

Fig. 9 illustrates the tradeoff between the performance gain and the encoding complexity. The average runtime figures have been obtained on the six 512×512 resolution images on an Intel Core2 T7700 system with 2 GB memory running Windows Vista. The "two DWT domain KLT bases" yields

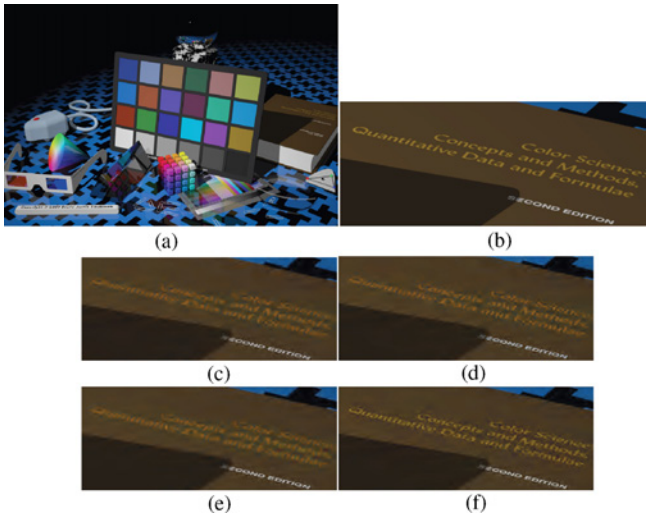

Fig. 7. Top: (a) Original 2048 × 3072 high resolution image *artificial*. (b) 217 × 532 section of the image. Middle and bottom: Corresponding reconstructions of the section at 0.1 b/p. (c) "Single spatial domain KLT basis" [6] (overall image Y-PSNR = 36.47 dB, UV-PSNR = 41.04 dB). (d) "Two DWT domain KLT bases" (overall image Y-PSNR = 36.57 dB, UV-PSNR = 41.12 dB). (e) Single spatial domain KLT followed by JPEG-2000 coding of coefficients (overall image Y-PSNR = 35.72 dB, UV-PSNR = 40.92 dB at 0.1 b/p). (f) Proposed "adaptive transform" (overall image Y-PSNR = 36.92 dB, UV-PSNR = 42.50 dB).

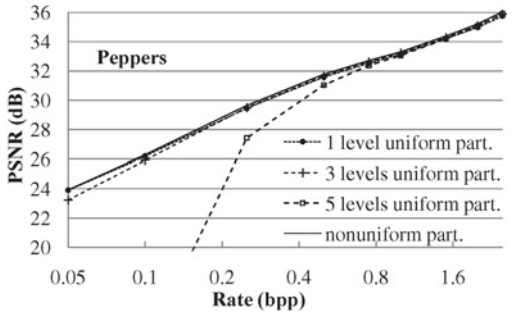

Fig. 8. Comparison of the performances of nonuniform and uniform spatial partitioning. In all cases, each subset is represented by a unique KLT basis and the transform coefficients are coded by CSPIHT.

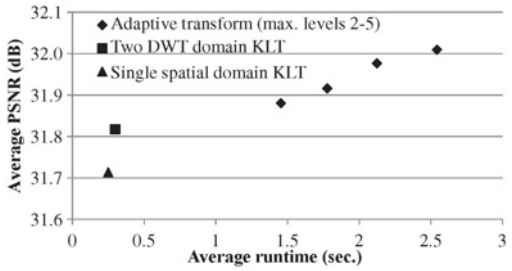

Fig. 9. Performance versus complexity for the "single spatial domain KLT basis," the "two DWT domain KLT bases," and the "adaptive transform" at various maximum quadtree depths.

more than half the gain of the "adaptive transform" over the "single spatial domain KLT basis" with only a slight increase in complexity. With substantially more increase in complexity, "adaptive transform" provides some more performance gain over the "two DWT domain KLT bases." On the contrary, imposing limits on the maximum quadtree depth, and thereby

Fig. 10. Impact of the chrominance gain on subjective quality. (a) 225 × 70 section of Kodak true color image number 18. (b) Corresponding reconstruction with the proposed method (overall image Y-PSNR = 28.117 dB, UV-PSNR = 38.60 dB). (c) Corresponding reconstruction with the "single spatial domain KLT basis" [6] (overall image Y-PSNR = 28.094 dB, UV-PSNR = 37.91 dB).

on the number of precomputed KLT bases, allows one to employ the "adaptive transform" in a complexity scalable way. The average decoder runtime was measured as 0.15 s for all the 512×512 images and methods considered.

Finally, the relatively large gains in chroma components with respect to the luminance component in Table I should not be understated despite the general belief that human visual system is less sensitive to the luma than the chroma. Kodak true color number 18 reconstructions by the "single spatial domain KLT basis" and the proposed method in Fig. 10 illustrate this point. The extra distortion in the UV components with "single spatial domain KLT basis" manifests itself as discolorations of the bright necklace and its ringing artifacts.

Similar coding gains with the "adaptive transform" over the "single spatial domain KLT basis" are noted when 2-D DWT is substituted with a hierarchically applied 2-D 8×8 fast DCT (second level is 8×8 DCT of DC coefficients of the first level). The rate, distortion, and slope models proposed in the previous sections serve the DCT coefficients equally well. However, DCT domain coding is inferior to DWT domain coding in average PSNR by about 1 dB–1.5 dB and insignificantly reduces the overall computational complexity which is largely due to the multitude of KLT transforms in the proposed method.

Binary executable encoder and decoder files for the proposed coding system as well as a demo batch file that encodes an original portable pixel map (.ppm) color image file, decodes back into a reconstructed .ppm image file by appropriate file format conversions, and reports rate and distortion figures may be downloaded from the URL http://web.itu.edu.tr/ulugbayazit/adaptive_transform_demo.rar. The software codec has been successfully tested on images with dimensions that are multiples of 256.

TABLE II
FREQUENCY OF THE AVERAGE PSNR OF THE IMAGE RECONSTRUCTED WITH THE PROPOSED ''ADAPTIVE TRANSFORM'' EXCEEDING THE AVERAGE PSNR OF THE IMAGE RECONSTRUCTED WITH EACH OF THE OTHER METHODS AT VARIOUS RATES

Rate (b/p)	Out of 32 Images, No. of Images with Positive Gain Over			
	Single Spatial Domain KLT Basis + CSPIHT [6]	Single DWT Domain KLT Basis + CSPIHT	Two DWT Domain KLT Bases + CSPIHT	Single Spatial Domain KLT Basis + JPEG2000
0.05	12	13	28	20
0.10	20	20	26	21
0.25	27	26	25	23
0.50	30	30	23	26
0.75	32	31	25	25
1.00	32	31	25	28
1.50	32	32	27	26
2.00	32	32	28	28
2.50	32	32	28	30

VI. CONCLUSION

This paper has demonstrated that a locally adaptive color spectral transform applied in the DWT domain across the color planes of an image yielded a compression performance advantage that increases progressively with rate upto an average of 0.9 dB in the Y and 1.35 dB in the U and V components over a single global transform applied in the spatial domain. This advantage is largely due to the application of separate bases to the low and high frequency DWT coefficients. This subset of the proposed method increased complexity only slightly. The proposed method also employed quadtree-based partitioning of high frequency coefficients into spatially and/or directionally oriented subsets where the decisions to form the partition as well as to employ fixed or signal dependent basis within each subset are R-D optimized by the distortion and rate models of [18] and a slope model derived from these models. The application of a separate basis to each subset is a fine-tuning for further enhancing the performance in return for a considerable increase in computational complexity. The complexity of the encoder that precisely achieved a given target rate can be limited with two iterations of adaptive transform design and subsequent CSPIHT coefficient coding.

The proposed method yielded limited improvement when extended to be applied to multispectral/hyperspectral image data, primarily due to the linear increase of side information rate with the number of coded bands. In the future work, a hierarchical application of the KLT transform for lowering the side information rate for this kind of data will be investigated.

REFERENCES

[1] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 4, pp. 243–250, Jun. 1996.

[2] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, 3rd ed. Norwell, MA: Kluwer, 2004.

[3] W. A. Pearlman, A. Islam, N. Nagaraj, and A. Said, "Efficient, low-complexity image coding with a set-partitioning embedded block coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 11, pp. 1219–1235, Nov. 2004.

[4] SPIHT FAQ. (2002). *Image Compression with Set Partitioning in Hierarchical Trees* [Online]. Available: http://www.cipr.rpi.edu/research/SPIHT/spiht6.html

[5] K. Shen and E. J. Delp, "Color image compression using an embedded rate scalable approach," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1997, pp. 34–37.

[6] A. Kassim and W. S. Lee, "Embedded color image coding using SPIHT With partially linked spatial orientation trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 2, pp. 203–206, Feb. 2003.

[7] A. Moinuddin, E. Khan, and M. Ghanbari, "An efficient wavelet based embedded color image coding technique using block-tree approach," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 1889–1892.

[8] X. San, H. Cai, and J. Li, "Color image coding by using inter-color correlation," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 3117–3120.

[9] D. Mukherjee and S. K. Mitra, "Arithmetic coded vector set-partitioning with classified tree-multistage VQ for color image coding," in *Proc. IEEE Workshop Multimedia Signal Process.*, Dec. 1998, pp. 444–449.

[10] J. Lee, "Optimized quadtree for Karhunen-Loeve transform in multispectral image coding," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 453–461, Apr. 1999.

[11] G. Gelli and G. Poggi, "Compression of multispectral images by spectral classification and transform coding," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 476–489, Apr. 1999.

[12] M. Cagnazzo, R. Gaetano, S. Parrilli, and L. Verdoliva, "Region based compression of multispectral images by classified KLT," in *Proc. 14th EUSIPCO*, Sep. 2006.

[13] O. Deforges, M. Babel, L. Bedat, and J. Ronsin, "Color LAR codec: A color image representation and compression scheme based on local resolution adjustment and self-extracting region representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 8, pp. 974–987, Aug. 2007.

[14] E. Gershikov, E. L. Burlak, and M. Porat, "Correlation-based approach to color image compression," *Signal Process.: Image Commun.*, vol. 22, no. 9, pp. 719–733, Oct. 2007.

[15] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 182–190, Apr. 1996.

[16] D. Marpe, H. Kirchhoffer, V. George, P. Kauff, and T. Wiegand, "An adaptive color transform approach and its application in 4:4:4 video coding," in *Proc. 14th EUSIPCO*, Sep. 2006.

[17] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2001, pp. 542–545.

[18] P. Sagetong and A. Ortega, "Rate-distortion model and analytical bit allocation for wavelet-based region of interest coding," in *Proc. IEEE Int. Conf. Image Process.*, Dec. 2002, pp. III-97–III-100.

[19] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1027–1042, Apr. 1998.

[20] U. Bayazit, "Region adaptive spectral transformation for wavelet based color image compression," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2817–2820.

[21] *Kakadu JPEG2000 Codec* [Online]. Available: http://www.kakadusoftware.com

Ulug Bayazit (S'91–M'96) received the B.S. degree from Bogazici University, Istanbul, Turkey, in 1991, and the M.S. and Ph.D. degrees from Rensselaer Polytechnic Institute, Troy, NY, in 1993 and 1996, respectively, all in electrical engineering.

From 1996 to 2000, he was with the Toshiba America Electronic Components, Inc., San Jose, CA, as a Research Scientist and Systems Architect. He then was an Assistant Professor with the Department of Electronics Engineering, Isik University, Istanbul, in 2000. Since 2007, he has been an Associate Professor with the Department of Computer Engineering, Istanbul Technical University, Istanbul. His current research interests include data compression and image, video, and computer graphics coding.

Dr. Bayazit won the Best Student Paper Award of VCIP in 1997.