Region-of-Interest Data Compression with Prioritized Buffer Management (III)

Sam Dolinar,
Aaron Kiely,
Matt Klimesh,
Ryan Mukai,
Shervin Shambayati
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California
{{sam,aaron,klimesh,shervin}
@shannon, rmukai@arcadia}
.jpl.nasa.gov

Antonio Ortega,
Baltasar Beferull-Lozano,
Sang-Yong Lee,
Phoom Sagetong,
Hua Xie
Department of Electrical Engineering
Signal and Image Processing Institute
Integrated Media Systems Center
University of Southern California
Los Angeles, California
{ortega,huaxie,sagetong}
@sipi.usc.edu

Roberto Manduchi,
Ajoy Frank
Computer Engineering Department
University of California, Santa Cruz
Santa Cruz, California
{manduchi@soe,ajoy@cse}.ucsc.edu

baltasar.beferull@epfl.ch,
sangyong@ti.com

Abstract — **We describe the third and final year of work on an integrated system for intelligent compression and transmission of copious data acquired by spaceborne instruments. Our system contains a wavelet-based progressive image compression algorithm, ROI-ICER, that accepts input priorities measuring the relative importance of various "regions of interest" in the source data, and organizes its output packets to reflect both the regional priorities and the wavelet bit layer priorities. The output of the data compression module is supervised by an intelligent buffer manager that shuffles the prioritized packets from many different source images and tries to select packets for transmission that will maximize the total science value received on the ground.**

We have developed new classification and prioritization algorithms for a specialized application to identify regions of interest from aerial images of wildfires. This application uses multispectral image data with a handful of spectral bands. The classification works by analyzing radiance ratios from long- and mid-wavelength bands.

This year we have continued our research into onboard feature detection and compression with distributed classification algorithms based on both onboard data and a ground database. The distributed classification algorithms are based on joint optimization of two pruned tree functionals. We apply steerable transforms to achieve rotational invariance, which enforces the same rotation angle across transform levels and allows alignment of images by steering features. Performance is measured as tradeoffs among bit rate, distortion, and complexity, using experiments with complete feature vectors.

Our prioritized compression and buffer management software has been ported to a web-accessible testbed, which allows users at remote computers to run end-to-end simulations of system capabilities using images and priority maps supplied by the user. In the paper, we briefly describe the capabilities of the testbed and its underlying software.

This work was funded by the ESTO Technology Program and performed at the Jet Propulsion Laboratory, California Institute of Technology, at the Signal and Image Processing Institute, Integrated Media Systems Center, University of Southern California, and at the Computer Engineering Department, University of California, Santa Cruz. Sang-Yong Lee is now at Texas Instruments, Dallas, Texas. Baltasar Beferull-Lozano is now at the Swiss Federal Institute of Technology-EPFL, Lausanne, Switzerland.

I. Introduction

We are in the final year of a three-year project to develop integrated data compression and buffer management algorithms to maximize the science value of data returned from spacecraft instruments [1, 2]. Our approach is to adapt existing progressive compression algorithms to make use of identified "regions of interest" (ROIs) in the data, and to develop buffer strategies for prioritizing, storing, and delivering the most valuable compressed segments, and later reconstituting the original data. Our system incorporates ROI considerations across many images or different data types. The algorithms are subject to practical limits on the onboard computer's speed, memory, and storage. In this paper we describe third-year results in four general areas: (1) new classification and prioritization algorithms for a specialized application to identify regions of interest from aerial images of wildfires; (2) efficient onboard feature detection and compression with distributed classification algorithms; (3) steerable transforms for identifying rotationally invariant features; and (4) a web-accessible testbed for evaluating our ROI compression and buffer management software in scenarios of interest to scientist-users.

II. Classification and Prioritization for Aerial Wildfire Imagery

Wildfire detection and monitoring via airborne sensor platforms and real-time data telemetry is a promising and viable technology for mitigating loss of life, property and natural habitat from wildland fires. Early detection, continuous monitoring of fire fronts, and fast deployment of fire fighting resources all rely on the ability to collect, process and transmit data from airborne systems in real time. Imagery data is collected in several spectral bands, each providing useful information for fire analysis. Visible wavelengths are suitable for monitoring smoke plumes and for distinguishing surface cultural and vegetative features not obscured by smoke or clouds. Various infrared bands are suitable for analysis of vegetative composition, and for analysis of distinct fire temperatures, while penetrating the associated smoke column. The longest infrared bands can collect thermal data on earth ambient temperatures and on the lower temperature soil heating conditions behind fire fronts, as well as the minute temperature differences in pre-heating conditions.

Current implementations of multispectral aerial imaging systems are severely hampered by the downlink data rate bottleneck of current telemetry systems. Our ROI-prioritized compression system has

the capability to provide a leap in the state-of-the-art for wildfire monitoring, by enabling real-time fire management using aerial images of active wildfires. To this end, we developed new classification and prioritization algorithms to identify regions of interest in aerial wildfire images.

Our ROI-prioritized compression system uses a priority map to identify the image areas that should be reproduced more faithfully (by allocation of more bits in the compression process), versus areas with lower priority, for which a coarser rendering is acceptable. Active fires, fire fronts and very hot surfaces should receive the highest priority, in order to provide a high resolution view of those critical areas to the fire fighting officials and rescue crews. Identifying such image segments from thermal multispectral imagery is relatively simple, by analyzing the ratio of radiance values in a long-wavelength and in a mid-wavelength band. Figure 1 shows an example of a priority map obtained by spectral ratioing. As shown in the figure, our prioritized compression and buffer management system uses this priority map to reconstruct the corresponding image with greater fidelity in the regions identified as high-priority.

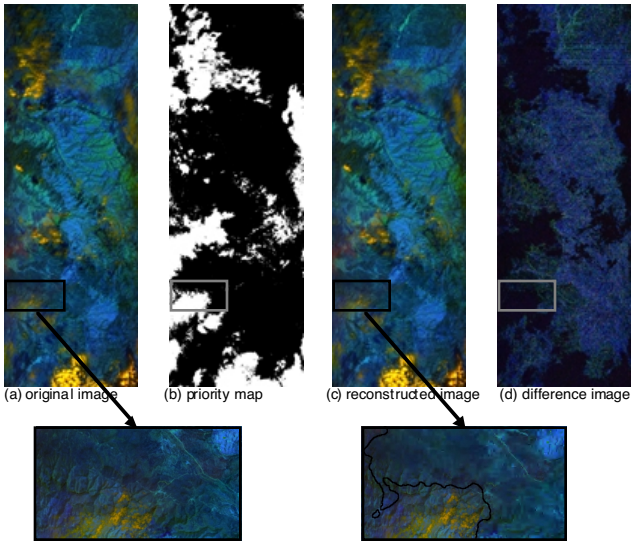


Fig. 1: Example of ROI-prioritized compression of aerial wildfire images. The three color planes in the original image (a) correspond to three bands of data (450 nm, 2.15 mm, and 8.5 mm) collected by an MAS imager flown over an Arizona wildfire in July, 2000. The original data resolution is 16 bits/pixel in each spectral band. The second image (b) is the priority map obtained by band ratioing, with white (high priority) pixels indicating areas of high temperature. This priority map was used for ROI compression to an average rate of 0.68 bits per pixel per band (including overhead needed to specify the priority map), as shown in the reconstructed image (c). The zoomed-in panels at the bottom of the figure illustrate the effects of prioritized compression. The black line superimposed on the right panel shows the boundary of the high priority region. Pixels to the left of the line are in the high-priority region, and are faithfully reproduced. Pixels to the right of the line belong to the low-priority region, where artifacts due to compression are visible. The uneven error distribution of ROI compression is also evident in image (d), which shows the difference between the decompressed and the original image (stretched by a factor of 8 to highlight errors). Darker areas indicate smaller reconstruction error, and mostly correspond to high priority regions.

More sophisticated prioritization mechanisms may also be required for effective use of the compressed images. Firefighters and wildfire managers would like to see not only the burning or hot areas, but also contextual neighboring portions of the scene. Contextual areas should also be given high, or at least intermediate, priority. Prominent landmarks such as roads or buildings should be preserved in order to facilitate scene georectification and registration. These landmarks can be extracted (and given high priority) by standard image processing operations such as edge, line and corner detectors. Our ROI-prioritized compression system accepts priority maps with multiple prioritization levels, and it is well equipped to handle intermediate priorities assigned to important contextual features.

III. Efficient Feature Compression for Distributed (Remote) Image Classification

We continued our research into onboard feature detection and compression with distributed classification algorithms based on both onboard data and a ground database. We represent regions of the captured image by a feature set and compress this feature set. By doing this we are able to establish a prioritization feedback loop between spacecraft and ground. Features are compressed and transmitted through the downlink to the ground. A priority is assigned to each region by performing a search in the database using the compressed features. Figure 2 shows such a remote image classification system. By doing this, we are able to reduce the onboard storage requirement and processing complexity. Since the database and classifier are kept on the ground where memory is cheap, we can obtain accurate priority assignments based on the entire history of received information.

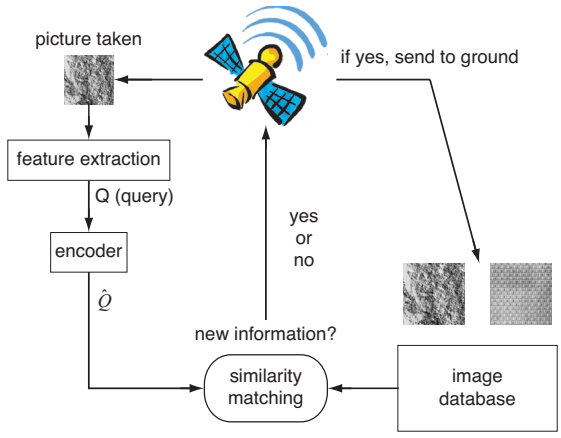


Fig. 2: Priority feedback in remote image classification.

We assume a vector model for image classification. Each image is represented by a feature vector $X \in \mathcal{R}^N$. Similarity matching is performed between the query vector Q and each of the entries X in the database to predict the classification label Y of the query. One commonly used matching method is the K-NN (K Nearest Neighbor) classifier [6] where K closest objects to the query vector are searched based on some distance measure $d(Q, X)$, then the majority class in these K retrieved objects is taken as the predicted classification label of the query vector Q. The K-NN method gained its popularity in the field of pattern recognition due to its simplicity and asymptotic property that it achieves the minimum Bayesian risk as the training set becomes arbitrarily large.

Vector quantization (VQ) is a natural way to compress the feature vector under this scenario because the classifier operates on vectors. But the complexity of VQ is exponential with the vector dimension. This is undesirable in our application since the onboard processing

power is limited and the dimensionality of the feature vector is usually very high (up to hundreds). This motivates us to look for simpler coding schemes. The performance of these schemes is measured in terms of tradeoffs among bit rate, distortion, and complexity, using experiments with complete feature vectors.

A. Entropy- and Complexity-constrained classified quantizer design

In previous work by Xie and Ortega [17, 2], we addressed the design of a classified quantizer to optimally trade off rate, distortion and complexity. Figure 3 shows a block diagram of a classified encoding system. We assume simple uniform scalar quantization and separate entropy coding of each element of the vector. A Decision Tree Classifier (DTC) is applied to classify the compressed data. We assume that the pre-classifier is a pruned subtree of the full decision tree.

The Generalized Breiman, Freidman, Olshen, and Stone (GBFOS) algorithm [4, 5] is employed to jointly search for the optimal pre-classifier and quantization parameters for each of the classes. The optimization is based not only on the rate budget, but also on a coding complexity constraint. The goal is to find the optimal subtree $S^* \preceq \mathcal{T}$ and the set of stepsizes $\{\Delta^*_{i,j},\ j = 1, ..., N\}$ for each class i, such that the overall distortion is minimized subject to the rate budget R_b and complexity constraint C_b.

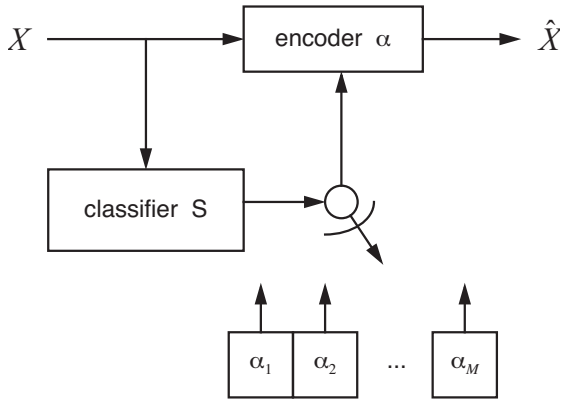


Fig. 3: Classified encoding system. Separate encoders $\{\alpha_i\}$ are designed to exploit the local statistics of the data.

$$D^* = \min_{S^*, \{\Delta^*_{i,j}\}} \sum_{i=1}^{|\tilde{S}|} P_i \times D_i(\Delta_{i,1}, \Delta_{i,2}, ..., \Delta_{i,N}) \quad (1)$$
$$\text{such that } R(S, \{\Delta_{i,j}\}) \le R_b \text{ and } C(S) \le C_b$$

Instead of solving the constrained problem (1), we use Lagrange multipliers and solve the dual problem:

$$\min_{S^*}[\min_{\{\Delta^*_{i,j}\}} \{D(S, \{\Delta_{i,j}\}) + \lambda \times R(S, \{\Delta_{i,j}\}) + \mu \times C(S)\}] \quad (2)$$

Now we need to find the optimal multipliers λ and μ such that the rate and complexity constraints are satisfied with equality. We developed a nested optimization algorithm to jointly search for the optimal subtree S^* and the set of quantization stepsizes $\{\Delta_{i,j}\}$ for a given rate budget and complexity constraint. The basic idea is: Initialize the multiplier λ, then for this fixed multiplier, prune the tree until the complexity constraint C_b is satisfied. With the resulting operating rate R, adjust the multiplier λ using a bisection method [7] and repeat the process. Figure 4 gives a geometric interpretation of this nested optimization. Pruning with fixed λ finds $(S^*(\lambda), \{\Delta^*_i(\lambda)\})$ with complexity constraint C_b satisfied. Adjusting λ and pruning until the rate meets the constraint R_b gives us the optimal point with both constraints satisfied.

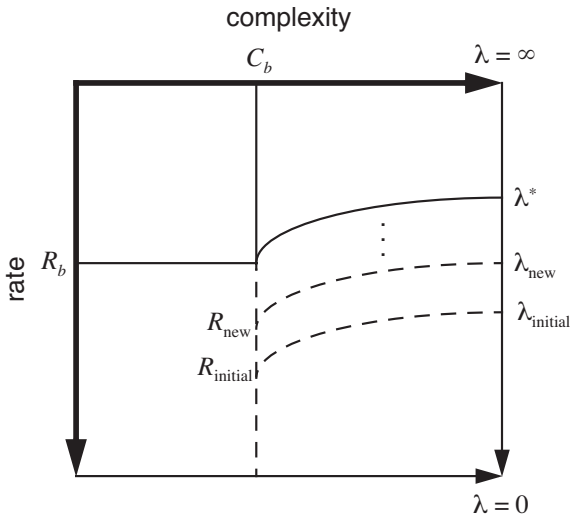


Fig. 4: We start with multiplier λ_{initial} and prune the tree until C_b is satisfied. Then update λ to λ_{new} using the bisection method. Repeat the process until R_b is satisfied.

We performed texture classification [8] based on a 1-NN classification rule using compressed data. Figure 5 shows that a lower classification error rate was achieved by using the classified encoder instead of a single encoder without pre-classification.

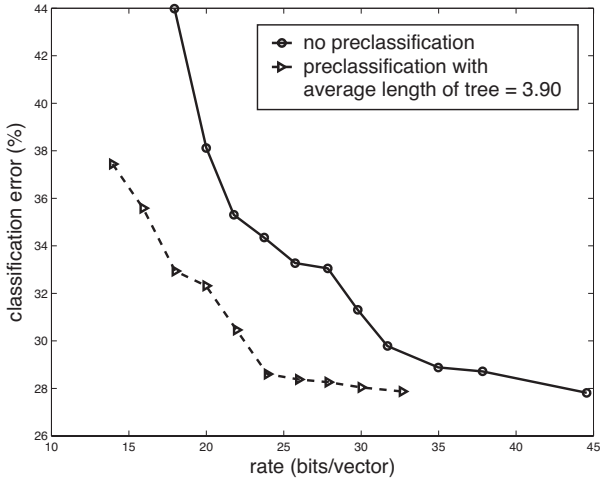


Fig. 5: Classification performance for encoders with and without pre-classification.

Experimental results on natural images with complete feature vectors are shown in Figs. 6 and 7. In this experiment, we tested 15 classes[1] from the Corel image database [9], with 100 images for each class. We used 80% of the images in each class for training

[1]The classes include Flowers II, Exotic cars, Sunrise and sunsets, Religious stained glass, Ski scenes, Painting, English country garden, Land of the

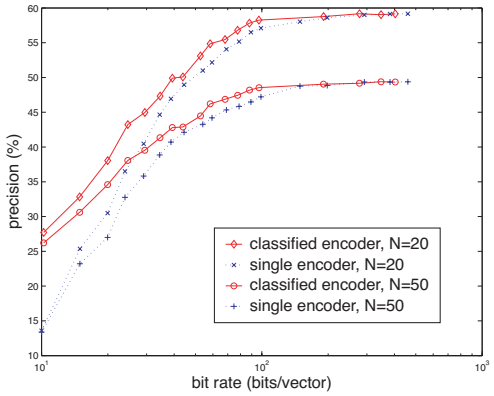


Fig. 6: Retrieval precision versus rate, comparing encoders with and without pre-classification.

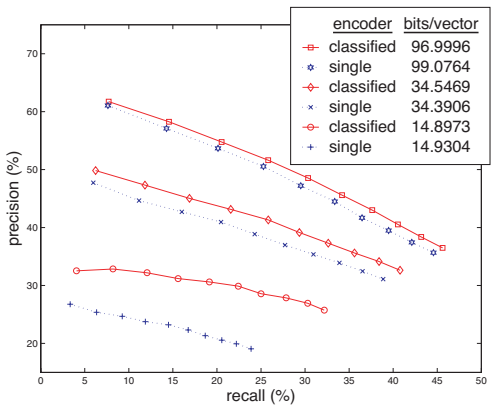


Fig. 7: Retrieval precision versus recall, comparing encoders with and without pre-classification at different operating rates.

and 20% for testing. The feature vectors consist of color histogram, texture histogram and edge histogram in La^*b^* color space [10]. The dimensionality of the feature vectors in our experiment was 99. We performed 20-NN and 50-NN similarity searching using l_1 norm as the distance measure. The quality of the retrieval result is measured by two quantities: precision and recall. *Precision* is the percentage of objects in the retrieved set that are relevant to the query image; it measures the purity of the retrieval. *Recall* is a measurement of completeness of the retrieval, computed as the percentage of retrieved relevant objects in the total relevant set in the database. Again we see that substantial gain is achieved by employing the classified encoder, especially at low operating bit rates. At bit rate around 0.2 bits/sample, our proposed encoding scheme achieves about 7% higher retrieval precision than a single encoder.

B. Discriminant analysis for transform coding[2]

State-of-the-art source coding yields a good compromise between complexity and coding performance. The motivating principle of transform coding is that simple coding is more effective in the transform domain. By simple coding we mean scalar quantization followed by entropy coding of each dimension separately.

pyramids, Mayan and Aztec ruins, Divers and diving, Glacier and mountains, Owls, Arabian horses, Coasts and Fireworks.

[2]Work also submitted in part for publication in [18].

Without loss of generality, we assume a linear transform T followed by a bank of uniform quantizers $\{\Delta_i\}$. The quantization indexes of each dimension are entropy-coded independently. We seek the optimal linear transform T^* and quantization stepsizes $\{\Delta_i^*\}$, such that the probability of classification error P_e based on the compressed data is minimized.

$$\min_{T^*,\{\Delta_i\}} P_e(\hat{X}) \quad \text{such that} \quad R \le R_b \tag{3}$$

Unfortunately, an explicit mathematical expression for the probability of error has is not available except for a very few special cases. So here we consider an alternative class separability criterion called the *scatter measure* [11]. The scatter measure is a ratio $S_w^{-1}S_b$ between the within-class scatter S_w and the between-class scatter S_b:

$$\begin{aligned} S_w &= \sum_{i=1}^{L} P_i \Sigma_i \\ S_b &= \sum_{i=1}^{L} P_i (M_i - M)(M_i - M)^t \end{aligned} \tag{4}$$

where Σ_i and M_i are the covariance matrix and mean vector, respectively, for the ith class, P_i is its a priori probability, and M is the overall mean vector.

Linear discriminant analysis (LDA) is a method used in pattern recognition to solve the "curse-of-dimensionality" problem. Classification procedures that are analytically or computationally manageable in low-dimensional spaces can become totally impractical when the dimensionality reaches 50 or higher. LDA looks for an optimal linear transform such that most of the class separability information is preserved in a small number of dimensions in the transform domain. Thus, classes become well separated after projecting to the most discriminant directions, and the representation entropy of the class separability information is reduced in the LDA transform domain as compared to the original domain. This nice property enables efficient quantization.

It was shown in [11] that the optimal transform T^* for "compacting" the class discrimination information measured by the scatter matrix $S_w^{-1}S_b$ is a matrix whose columns consist of the eigenvectors of $S_w^{-1}S_b$. The intuition here is that, to obtain good class separation, different classes should be far apart from each other (corresponding to large S_b) while at the same time samples belonging to the same class should be closely clustered together (corresponding to small S_w). Figure 8 shows the discriminating power of projecting feature vectors of two Brodatz texture classes $D1$ and $D2$ onto the eigensystem of the scatter matrix $S_w^{-1}S_b$. We see that most of the class discrimination information is preserved in the direction that corresponds to the eigenvector with the largest eigenvalue.

We performed a linear discriminant transform, followed by uniform scalar quantization with a greedy bit allocation. The greedy bit allocation starts with finest quantization at each dimension, and then a dimension chosen for coarser quantization such that the magnitude of the slope of the increase in class entropy to the decrease in entropy rate is minimized. Let $\alpha^t = \{\Delta_i^t, i = 1, ..., N\}$ denote the set of stepsizes at time t, and $\mathcal{A}^t$ be the resulting partition of the sample space using quantizer α^t. At time $t+1$, let $\alpha_k^{t+1} = \{\Delta_1^t, \Delta_2^t, ..., \Delta_k^{t+1}, ..., \Delta_N^t\}$ be the quantization with a larger stepsize applied to the kth dimension, leaving the stepsizes

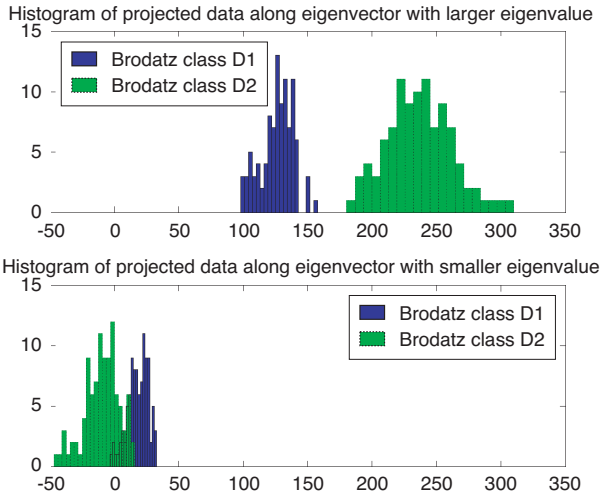


Fig. 8: Histograms of features projected onto the eigensystem of the scatter matrix.

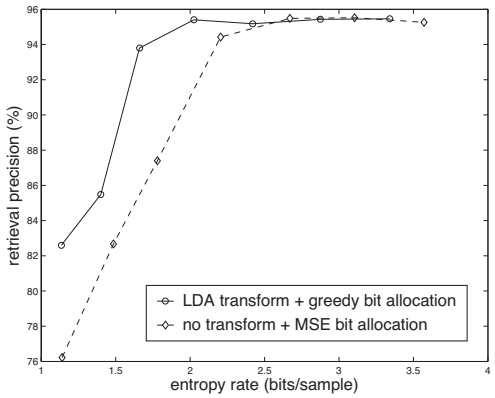


Fig. 9: Textural retrieval performance comparing LDA transform coding with traditional MSE coding.

on the other dimensions unchanged. We denote the resulting partition by $\mathcal{A}_k^{t+1}$. The entropy of the class distribution for a partition $\mathcal{A}$ is computed as:

$$H(\mathcal{A}) = \sum_{c \in \mathcal{A}} \Pr(c) \times H(c) \tag{5}$$

$$H(c) = -\sum_{i=1}^{C} p_i \times \log(p_i) \tag{6}$$

where p_i is the probability of class i in partition cell c, and C is the total number of classes. Class entropy is a measure of purity of the partition. We need to find at time $t+1$ the dimension k^* along which to apply a coarser quantization as compared to time t, such that $\frac{H(\mathcal{A}_k^{t+1}) - H(\mathcal{A}^t)}{R(\mathcal{A}^t) - R(\mathcal{A}_k^{t+1})}$ is minimized. Here $R(\mathcal{A})$ is the entropy rate associated with partition $\mathcal{A}$. This process is continued until the rate constraint is satisfied.

Applying the above quantization scheme to a four-class (D1, D2, D3, D4) Brodatz texture example, we show in Fig. 9 the retrieval performance of the LDA transform coding scheme compared with traditional encoding where bit allocation is performed in the original domain minimizing mean squared error (MSE). From the figure we see clearly that the proposed LDA compression scheme achieves more accurate retrieval.

IV. Steerable Transforms for Identifying Rotation-Invariant Features[3]

Texture information is useful for image classification, and several approaches have been proposed to extract texture-related features based on various linear transforms, such as the wavelet transform. Most well-known texture feature extraction methods measure the energies of the subbands obtained from a wavelet transform as texture discriminating features. However, one drawback of using critically sampled transforms for this purpose is that the features are not rotation- or shift-invariant. We address the problem of designing efficient rotation-invariant texture features and demonstrate their use in the context of decision tree classifier. Our goal here is to enable locating similar images in the database, even if the captured image is rotated with respect to those most similar to it in the database. We apply steerable transforms to achieve rotational invariance. This enforces the same rotation angle across transform levels and allows alignment of images by steering features.

Two main previous approaches have been proposed for rotation-invariant texture matching. In the first one [12], a stochastic model (hidden Markov model) is assumed for features derived from a wavelet transform, and the training is performed using samples with different orientations and identified as belonging to the same class. In the second one [13], given the outputs of a transform, some specific rotation-invariant quantities are defined. In our work, instead, we achieve rotation-invariance by using the concept of angular alignment, i.e., the features obtained from two images are aligned before being compared. To achieve this, we define a set of features which are *steerable* in the sense that, given the features of an image sample, it is possible to obtain the features corresponding to any rotated version of it. These features are obtained from the subbands of a steerable pyramid [14, 15]. We also define a new similarity measurement which measures the distance between two feature vectors only after they have been aligned. This angular alignment can be performed efficiently using simple constrained steepest descent algorithms. Although several features obtained from a steerable pyramid have been proposed in previous work [16], the property of achieving steerability in the feature space, which is essential for the angular alignment, has not been considered.

One drawback of steerable transforms in many image representation applications comes from the fact that they are oversampled and thus result in a significant storage penalty with respect to critically sampled transforms. We show that steerable features can be efficiently compressed using the classified encoding system proposed by Xie and Ortega [17], to demonstrate that no rate penalty exists and that features based on steerable representations outperform wavelet-based features when operating at the same rate.

In order to decrease the complexity of the retrieval, a Decision Tree Classifier (DTC) is often used in practical applications. Images having similar features are clustered together in the nodes of the tree-structured classifier. We show how the rotation-invariance can be incorporated in the DTC-based retrieval by performing angular alignment at each node in the tree and defining an appropriate distance between an image and a tree node, which ensures that a

[3] Work also submitted in part for publication in [19].

best-first-search method works correctly. Our experimental results show that our method achieves a substantial gain in retrieval precision versus rate compared to a retrieval system based on a wavelet transform.

Section A below describes the feature extraction process, and Section B describes the basic similarity measurement, which is the main novelty of our work. In Section C, we briefly describe the quantization schemes that are considered and also the DTC-based retrieval system incorporating the angular alignment. Finally, Section D shows some experimental results.

A. Feature Extraction

Since we are interested in achieving rotation invariance, the feature extraction we consider is based on the subbands obtained from a steerable pyramid [14]. We then should choose features that are as "steerable" as possible, that is, given the features of an image oriented at an angle ϕ, it should be possible to obtain the features corresponding to the same image but oriented at an angle ϕ', by direct manipulation of the features at angle ϕ, i.e., without actually having to recalculate the features after rotating the image. In our work, we try to achieve good retrieval performance using energy-based features which are simple to manipulate.

Let $c(\mathbf{x}_o, \phi)$ represent the value of a transform coefficient corresponding to the output of a rotated steerable filter with orientation ϕ for a certain image spatial location $\mathbf{x_0}$. In a steerable pyramid with J basic orientations and L levels, at each level l, given the J basic coefficients $\{c^l(\mathbf{x}_o, \phi_1), c^l(\mathbf{x}_o, \phi_2), \ldots, c^l(\mathbf{x}_o, \phi_J)\}$, the transform coefficient $c^l(\mathbf{x}_o, \phi)$ for an angle (orientation) ϕ of that same spatial location will be given by:

$$c^l(\mathbf{x}_o, \phi) = \sum_{i=1}^{J} \alpha_i(\phi) c^l(\mathbf{x}_o, \phi_i) \quad \forall \phi, \quad l = 1, \ldots, L \tag{7}$$

where $\{\alpha_1(\phi), \alpha_2(\phi), \ldots, \alpha_J(\phi)\}$ is the set of J steering functions which can be used for (exact) interpolation of a transform coefficient for any angle ϕ and at any level $l = 1, \ldots, L$.

Let $E^l(\phi)$ represent the average energy of a subband oriented at an arbitrary angle ϕ in a level l, that is, $E^l(\phi)$ is given by $E^l(\phi) = \left(\frac{1}{N_l}\right) \sum_{k=1}^{N_l} (c^l(\mathbf{x}_k, \phi))^2$, where N_l is the number of pixels of each of the subbands in level l and the subscript k goes through all the spatial locations of the subband. It is very simple to show that $E^l(\phi)$ can be calculated from the energies (sampled autocorrelations) of the basic J subbands and all the sampled cross-correlations between each pair of basic subbands:

$$E^l(\phi) = \alpha^T(\phi) \mathbf{C}^l \alpha(\phi), \quad \alpha(\phi) = (\alpha_1(\phi) \cdots \alpha_J(\phi))^T \tag{8}$$

where $\mathbf{C}^l$ is the (symmetric) sampled correlation matrix with elements $C^l_{ij} = \left(\frac{1}{N_l}\right) \sum_{k=1}^{N_l} c^l(\mathbf{x}_k, \phi_i) c^l(\mathbf{x}_k, \phi_j) = C^l_{ji}$, $l = 1, \ldots, L$. Each diagonal element of $\mathbf{C}^l$ corresponds to $C^l_{ii} = E^l(\phi_i)$, that is, the average energy at the basic angle ϕ_i, while the off-diagonal elements correspond to sampled cross-correlations between the subbands corresponding to each pair of basic angles.

Notice that since $c(\mathbf{x}_o, \phi + \pi) = -c(\mathbf{x}_o, \phi)$, clearly, $E^l(\phi + \pi) = E^l(\phi)$, that is, $E^l(\phi)$ is a periodic function with period equal to π. Given a perfectly homogeneous image I with energy profile $E^l_I(\phi)$ at level l, if this image is rotated counter-clockwise by an angle θ, obtaining an image I_θ, then, we will have that $E^l_{I_\theta}(\phi) = E^l_I(\phi - \theta)$, that is, a rotation of an image corresponds to a shifted version of the energy profile.

Based on this we choose the correlation matrices $\{\mathbf{C}^l\}_{l=1}^L$ as the energy-based texture features in our system. Notice that since each matrix $\mathbf{C}^l$ is symmetric, the total number of features will be $J(J+1)L/2$. Therefore, the interdependencies between different orientations in terms of cross-correlations are necessary in order to characterize the energy profile of an arbitrary rotation of a given image. We do not consider the use of the energy of the low-pass residual subband as a feature in our proposed system. Obviously, as the number J of basic orientations increases, the resolution in angle (angular bandwidth of basic filters) increases and the energy profile $E^l(\phi)$ will be therefore more accurate, but on the other hand, the number of raw features may become substantially larger than in the case of a wavelet-based texture representation.

B. Similarity Measurement

In the similarity measurement, we are interested in making use of the steerability property present in the features in order to identify equivalent features, where equivalency means having different rotated versions of a unique image.

The next proposition shows that the sampled correlation matrix $\mathbf{C}^l_I$ for an image at a given level l is related in a simple way to the sampled correlation matrix $\mathbf{C}^l_{I_\theta}$ for the same image but rotated counter-clockwise by an angle θ.

Proposition 1 *Given a steerable representation with J basic angles, the correlation matrices $\mathbf{C}^l_{I_\theta}$ and $\mathbf{C}^l_I$, both evaluated with respect to the same set of basic angles $\{\phi_1, \ldots, \phi_J\}$, are related as follows:*

$$\mathbf{C}^l_{I_\theta} = \mathbf{R}(\theta) \mathbf{C}^l_I \mathbf{R}^T(\theta),$$

$$\mathbf{R}(\theta) = \begin{pmatrix} \alpha_1(\phi_1 - \theta) & \alpha_2(\phi_1 - \theta) & \cdots & \alpha_J(\phi_1 - \theta) \\ \alpha_1(\phi_2 - \theta) & \alpha_2(\phi_2 - \theta) & \cdots & \alpha_J(\phi_2 - \theta) \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1(\phi_J - \theta) & \alpha_2(\phi_J - \theta) & \cdots & \alpha_J(\phi_J - \theta) \end{pmatrix} \tag{9}$$

In the particular case where the J basic angles are taken to be equally spaced, then $\mathbf{R}(\theta)$ becomes an orthogonal matrix for any θ, and therefore, $\mathbf{C}^l_{I_\theta}$ and $\mathbf{C}^l_I$ become orthogonally equivalent.

Proof: The proof is given in [20].

This property holds for every level independently. However, notice that when an image is rotated, all the decomposition levels will be equally rotated. This means that given an image I and a rotated version I_θ of it, the Frobenius norms:

$$\| \mathbf{C}^l_I - \mathbf{R}(-\theta) \mathbf{C}^l_{I_\theta} \mathbf{R}^T(-\theta) \|_F, \quad l = 1, \ldots, L$$

(same rotation angle for all the levels), will tend to be small.

Taking all this into account, the similarity measurement $D(I_1, I_2)$ between 2 different images I_1 and I_2 that we propose is the following:

$$D(I_1, I_2) = \min_\theta \left(\sum_{l=1}^{L} \| \mathbf{C}^l_{I_1} - \mathbf{R}(-\theta) \mathbf{C}^l_{I_2} \mathbf{R}^T(-\theta) \|_F \right) \tag{10}$$

Clearly, those levels containing more energy will influence more in the minimization of (10) and those levels with small energy will have little influence.

Notice that when I_1 and I_2 are two rotated versions of the same image, the angle θ^* for which the minimum is achieved in (10)

should be close to the relative angle between I_1 and I_2, that is, the angle one needs to rotate (clockwise) I_1 in order to get I_2. Thus, one way to see the goodness of our similarity measurement (10) is to check whether the estimated angle θ^* is actually close to the real relative angle between two physically rotated versions of the same image. Moreover, it might also be useful in some practical applications to find out approximately this relative angle. Fig. 10 illustrates this by showing the function $D(\theta) = \sum_{l=1}^{L} \|\mathbf{C}^l_{I_1} - \mathbf{R}(-\theta)\mathbf{C}^l_{I_2}\mathbf{R}^T(-\theta)\|_F$, for the case where I_1 and I_2 are rotated versions of "bark" texture from the Brodatz set [21].

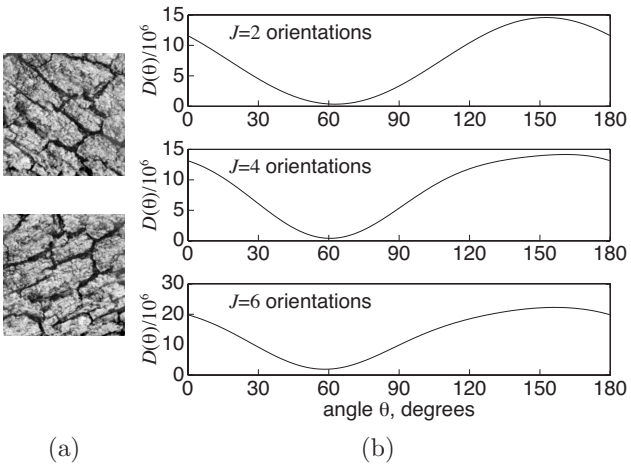


Fig. 10: (a) "bark" physically rotated at 60 and 120 degrees; (b) $D(\theta)$ for $J = 2, 4, 6$. Notice how in all three cases, the minimum is achieved for $\theta = 60$ degrees, which is the exact relative angle between the two texture image samples.

As explained below, this angular alignment has to be performed many times in the retrieval process and thus it is important to devise fast algorithms to find the minimizing angle θ^* in (10). In [20], it is shown that θ^* can be found analytically for $J = 2$, and for $J > 2$ it is possible to design low-complexity constrained steepest descent algorithms. This is because it can be proved that the number of stationary points of the function being minimized in (10) is upper bounded and, at the same time, the angular distance between any two contiguous stationary points is lower bounded [20], making it simple to search for these points in a few non-overlapping angular intervals.

C. Quantization and Retrieval Process

We tested our proposed scheme when the feature vectors are quantized using a set of scalar quantizers with three different quantization algorithms: (a) simple uniform quantization (same stepsize); (b) non-uniform quantization with optimal bit allocation in a rate-distortion sense [22]; (c) classified quantization optimized in a rate-distortion-complexity sense, as proposed by Xie and Ortega [17] and depicted earlier in Fig. 3.

The retrieval process is always performed with the DTC $\mathcal{T}$ using the best-first-search and branch-and-bound.[4] Let $\underline{D}(Q, t)$ denote the distance of the query Q with node t in the tree. In order to ensure that this search algorithm finds the correct closest matches, we need to define a distance $\underline{D}(Q, t)$ satisfying the property that $\underline{D}(Q, t)$ is a lower bound of the distances of Q to all the images at node t, and we need to take into account the angular alignment process. Notice that:

$$\begin{aligned} \underline{D}(Q, I) &= \min_{\theta} d(Q_\theta, I) \\ &\geq \min_{\theta} d(Q_\theta, I_c) - d(I_c, I) \text{ (triangle inequality)} \\ &\geq \min_{\theta} d(Q_\theta, I_c) - R(t) \quad \text{(upper bound)} \end{aligned} \tag{11}$$

where I_c is the centroid at node t, $R(t)$ is the radius of node t given by $R(t) = \max_{I \in t} d(I_c, I)$ and $d(Q_\theta, I)$ is given by the expression inside the parentheses in (10). Thus, defining $\underline{D}(Q, t)$ as $\underline{D}(Q, t) = \min_\theta d(Q_\theta, I_c) - R(t)$, then, it is guaranteed that the best-first-search method will find the correct closest match.

Thus, we see that a crucial differential point in our work is that in the retrieval process using the DTC, at each node of the tree, alignments between the query (quantized) feature vector and each of the two representing vectors (corresponding to the two branches) have to be performed using (10). After these two alignments, two distance measurements are performed and a branch is chosen.

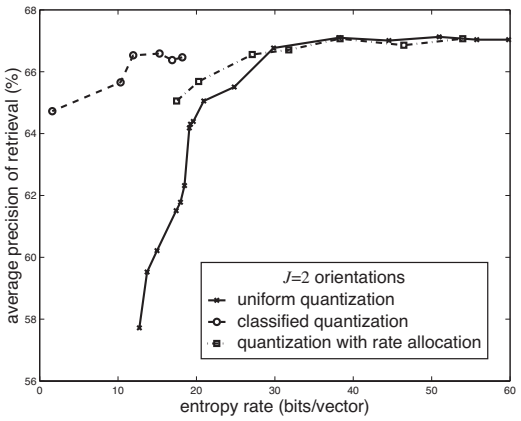
D. Experimental results

We evaluated the performance of our steerable transform applied to the Brodatz texture images [8] and compared it to that of a standard wavelet transform. The features we considered in the wavelet case are also correlation matrices obtained from the corresponding four wavelet subbands.

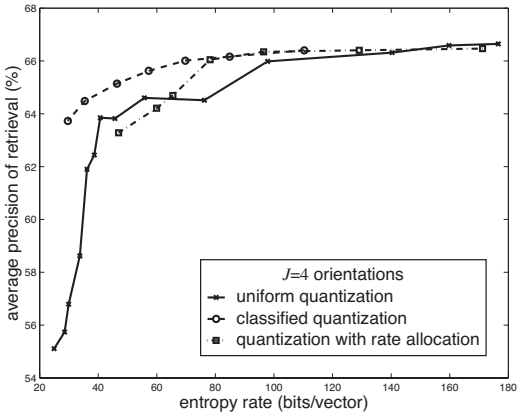
We used two collections of texture samples of size 128×128. The first collection, which forms the *non-rotated* image database, is obtained by partitioning each of the 13 Brodatz (512×512) non-rotated texture images [21] into 16 non-overlapping texture subimages of size 128×128 with a total of 208 texture samples. This set is used in training of the DTC for retrieval. The second collection, which forms the *rotated* set, is obtained by partitioning (for each of the 13 texture classes) 4 large texture images oriented at $30, 60, 90$ and 120 degrees also into non-overlapping subimages of size 128×128 and taking the 4 central subimages. In this way, in the second database, there are also 16 textures for each class and therefore, also the same total number of 208 textures. A query texture sample is taken from the *rotated* set and the feature vector is extracted and quantized using the three quantization schemes described in Section C. We assume that each quantized component of the feature vector is independently entropy coded. The $M = 16$ closest textures from the *non-rotated* set are obtained and the average retrieval precision over all the rotated texture samples is measured.

Without compressing the features, the average retrieval precision for the steerable transform is 67.03% and 66.55% for $J = 2$ and $J = 4$, respectively,, whereas for the wavelet transform the average precision is only 41.85%. Fig. 11 shows the retrieval precision of compressed steerable feature vectors for $J = 2$ and $J = 4$. We see clearly that the classified quantizer achieves the best performance among the three quantization schemes. By using the classified quantizer with expected tree length $l = 2$ (complexity constraint), the retrieval performance degrades very gracefully. Even with the bit rate reduced to around 1 bit/element, we can still achieve nearly the same precision as that obtained by using uncompressed features. Fig. 12 shows a comparison of the retrieval precision with compressed features between the steerable transform with $J = 4$ and a standard wavelet transform. Our reason for comparing these two cases is that

[4] This algorithm has a complexity of $O(\log M)$, where M is the number of feature vectors in the database, as compared to $O(M)$ in a linear search.



(a)



(b)

Fig. 11: Average Retrieval Performance using a 3 level steerable pyramid for the three different quantization algorithms: (a) $J = 2$ and (b) $J = 4$.

the dimension of the feature vector is the same for both ($N = 48$), so they will yield comparable bit rates. We see that the steerable transform achieves much better retrieval precision than the wavelet transform over all bit rates.

We also illustrate the reduction in retrieval complexity obtained by employing a DTC instead of a linear search, by tallying the number of distance computations that need to be performed to find the $M = 16$ closest matches, Instead of 208 distance computations in the case of linear search, the DTC requires on average 121.97 distance computations for $J = 2$ and 39.82 for $J = 4$.

V. The Web-Accessible Software Testbed

Our prioritized compression and buffer management software has been ported to a web-accessible testbed,[5] which allows users at remote computers to run end-to-end simulations of system capabilities using images and priority maps supplied by the user. The web site offers a front-end interface to the ROI-ICER simulator developed at JPL. Here we briefly describe the capabilities of the testbed and its underlying software.

At its core, our data compression system contains a wavelet-based progressive image compression algorithm, ICER [3], that is being

[5]Contact the authors for access privileges.

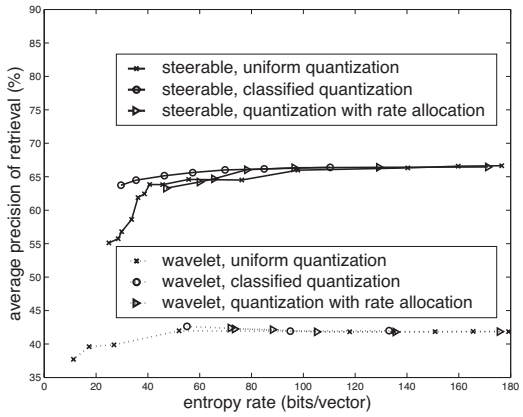


Fig. 12: Comparison between a standard 3 level wavelet pyramid ('daub16' filter bank) and a 3 level steerable pyramid with $J = 4$ for the three different quantization algorithms.

used on the Mars Exploration Rover (MER) mission. The ICER algorithm applies a wavelet decomposition and prioritizes the compressed bit layers from the wavelet subbands so as to progressively transmit the layer that gives the largest estimated improvement in image quality per transmitted bit. Our modified version, ROI-ICER, accepts additional input priorities in the form of a data prioritization map that gives the relative importance of different regions of interest in the source data. Then ROI-ICER produces output packets of compressed data along with priority labels that reflect both the input regional priorities and the wavelet bit layer priorities.

The output of the data compression module is supervised by an intelligent buffer manager that receives prioritized packets from many different source images and tries to select packets for transmission that will maximize the total science value received on the ground. Just as importantly, it attempts to discard only the least valuable packets when the buffer overflows (which is inevitable if the average data transmission rate is lower than the average data collection rate). Our buffer manager uses a simple form of double-valued prioritization: admissions and discards are determined by priorities established by ROI-ICER, while transmissions are first-in, first out (FIFO) among packets that survive the admission/discard process during their time of residency in the buffer. The FIFO protocol for transmissions keeps intact the chains of compressed data packets that are later used to progressively reconstruct each image or image segment, yet the prioritized decisions on admissions and discards ensure that the scarce downlink resource is not clogged by less valuable data. Using a FIFO transmission priority eliminates the need to unshuffle the packets received on the ground, because successive (truncated) packet chains can be used to reconstruct the source images in the same order in which they were acquired (but to different levels of distortion depending on how many packets from each chain survived the prioritized admission/discard process).

The operation of the testbed simulator is as follows. The user chooses a set of images and classification maps, either samples provided on the web page or files uploaded to the server by the user.[6] The user may declare that the classification maps should be directly interpreted as priority maps, or alternatively may specify a table or rule defining the priority of each class identified in the classification

[6]Users who upload their own images and classification maps should follow the instructions on the web site.

maps. The user also specifies the capacity of the prioritized buffer in kilobytes, the average rate at which image data arrive in bytes per second, and the downlink capacity of the channel. The order in which images are presented to the simulator can be determined by the user or chosen randomly by the server. The simulator's operations on each image in the submitted sequence are outlined below:

1. During the course of simulation, each arriving 3-color image first undergoes a Y-Cr-Cb transform to facilitate efficient compression. The three components are independently compressed using ROI-ICER.
2. The user's classification map is transformed into a priority map.
3. The priorities contained in the priority map are sent along with the image component (Y, Cr, or Cb) to the prioritized ICER compression system. The image component is compressed and packetized, and each packet is tagged with a priority value from the compressor. High priority regions are typically mapped to higher priority packets while lower priority regions are typically mapped to lower priority packets. The result is three separate packet streams for Y, Cr, and Cb.
4. The priority map, which is needed for both compression and decompression, is separately compressed and packetized at the highest priority.
5. The four packet streams, from the priority map and from the Y, Cr, and Cb components, are sent to the prioritized buffer system. The prioritized FIFO buffer maintains packet ordering while dropping or rejecting low priority packets when the buffer overflows and transmitting the surviving higher priority packets at the rate permitted by the downlink channel capacity.
6. The simulator reconstructs images based upon the packets that are successfully transmitted from the prioritized FIFO buffer. Given the original and reconstructed image, a difference image is also created and the pSNR evaluated to permit users to evaluate overall performance.

After the simulation finishes, the user clicks a link to view the results page. A series of image thumbnails is displayed along with the resulting overall pSNR for each image. By clicking on the thumbnail, the user can go to the image's page of results. This page displays:

1. The original, received, and difference images. There are links that permit a user to download these images in PPM format for further review and analysis.
2. The pSNR for each regional priority level. This permits the user to evaluate the performance of the system according to regional priorities. This is in addition to overall pSNR.
3. The number of bits per pixel for each of four data streams: priority map, Y component, Cr component, and Cb component. In addition, the overall number of bits per pixel received and used in reconstruction is given to permit evaluation of overall performance.

REFERENCES

[1] S. Dolinar, A. Kiely, M. Klimesh, R Manduchi, A. Ortega, S. Lee, P. Sagetong, H. Xie, G. Chinn, J. Harel, S. Shambayati, and M. Vida, "Region-of-Interest Data Compression with Prioritized Buffer Management," 2001 Earth Science Technology Conference, College Park, MD, Aug. 28–30, 2001.

[2] S. Dolinar, A. Kiely, M. Klimesh, R Manduchi, A. Ortega, S. Lee, P. Sagetong, H. Xie, and S. Shambayati, "Region-of-Interest Data Compression with Prioritized Buffer Management (II)," 2002 Earth Science Technology Conference, Pasadena, CA, June 11–13, 2002.

[3] A. Kiely, M. Klimesh, and J. Maki, "ICER on Mars: Wavelet-Based Image Compression for the Mars Exploration Rovers," IND Technology and Science News, Issue 15, to appear, 2002.

[4] L. Breiman, J. H. Freidman, R. A. Olsehn, and C. J. Stone, *Classification and Regression Trees*, Wadsworth, 1984.

[5] P. A. Chou, T. Lookabough, and R. M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Trans. on Info. Theory*, vol. IT-35, pp. 299–315, March, 1989.

[6] T. M. Cover and P. Hart, "Nearest neighbor pattern classification," *Proc. IEEE Trans. Inform. Theory*, vol. 13, pp. 21–27, 1967.

[7] K. Ramchandran and M. Vetterli, "Best Wavelet Packet Bases in a Rate-Distortion Sense," *IEEE Trans. on Image processing*, vol. 2, pp. 160–175, April, 1993.

[8] P. Brodatz, "Textures: A photographic album for artists and designers," Dover, New York, 1966.

[9] Corel Corp, "Corel stock photo library," Ontario, Canada.

[10] Charles A. Bouman Jau-Yuen Chen and John Dalton, "Similarity pyramids for browsing and organization of large image databases," *Proc. of SPIE Conf. on Human Vision and Electronic Imaging III*, vol. 3299, Jan. 1998.

[11] Keinosuke Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, 1991.

[12] J. L. Chen and A. Kundu, "Rotational and gray-scale transform invariant texture identification using wavelet decomposition and hidden markov model," *IEEE Trans. on Patt. Recog. and Mach. Intell.*, vol. 16, pp. 208–214, 1994.

[13] W. R. Wu and S. C. Wei, "Rotation and gray-scale transform-invariant texture classification using spiral resampling," *IEEE Trans. on Image Processing*, vol. 5, pp. 1423–1434, 1996.

[14] E. P. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *IEEE Trans. on Information Theory*, vol. 38, pp. 587–607, 1992.

[15] B. Beferull-Lozano and A. Ortega, "Coding techniques for oversampled steerable transforms," *Int. Asilomar Conf. on Signals, Systems and Computers*, 1999.

[16] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes," *ICIP*, 1998.

[17] H. Xie and A. Ortega, "Entropy and complexity constrained classified quantizer design for distributed image classification," *Proc. of Multimedia Signal Processing (MMSP)*, 2002.

[18] H. Xie and A. Ortega, "Exploration of Linear Discriminant Analysis for Transform Coding in Distributed Image Classification," 37th Annual Asilomar Conference on Signals, Systems, and Computers, Asilomar Hotel and Conference Grounds, Pacific Grove,California. Nov 9-12, 2003.

[19] B. Beferull-Lozano, H. Xie and A. Ortega, "Rotation-Invariant Features based on Steerable Tranforms with an Application to Distributed Image Classification," IEEE International Conference on Image Processing, Barcelona, Spain, September 14-17, 2003.

[20] B. Beferull-Lozano, "Quantization design for structured overcomplete expansions," *Ph. D. thesis, University of Southern California*, 2002.

[21] University of Southern California, "Brodatz rotated textures," http://sipi. usc. edu/services/database/Database. html.

[22] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. on Acoust. Speech and Signal Proces.*, vol. 36, pp. 1445–1453, 1988.