Retrieval Using Texture Features in High Resolution Multi-spectral Satellite Imagery

Shawn D. Newsam and Chandrika Kamath

Lawrence Livermore National Laboratory
7000 East Avenue, Livermore, CA 94551, U.S.A.

ABSTRACT

Texture features have long been used in remote sensing applications to represent and retrieve image regions similar to a query region. Various representations of texture have been proposed based on the Fourier power spectrum, spatial co-occurrence, wavelets, Gabor filters, etc. These representations vary in their computational complexity and their suitability for representing different region types. Much of the work done thus far has focused on panchromatic imagery at low to moderate spatial resolutions, such as images from Landsat 1-7 which have a resolution of 15-30 m/pixel, and from SPOT 1-5 which have a resolution of 2.5-20 m/pixel. However, it is not clear which texture representation works best for the new classes of high resolution panchromatic (60-100 cm/pixel) and multi-spectral (4 bands for red, green, blue, and near infra-red at 2.4-4 m/pixel) imagery. It is also not clear how the different spectral bands should be combined. In this paper, we investigate the retrieval performance of several different texture representations using multi-spectral satellite images from IKONOS. A query-by-example framework, along with a manually chosen ground truth dataset, allows different combinations of texture representations and spectral bands to be compared. We focus on the specific problem of retrieving inhabited regions from images of urban and rural scenes. Preliminary results show that 1) the use of all spectral bands improves the retrieval performance, and 2) co-occurrence, wavelet and Gabor texture features perform comparably.

Keywords: Image texture, similarity retrieval, high resolution remote sensed imagery

1. INTRODUCTION

Remote sensed imagery continues to accumulate at increased rates due to advances in sensors, storage, and other technologies. The full value of this data is not being realized, however, since similar progress has not been made in providing effective management solutions for the sizable image repositories. In particular need are content-based representations that can be extracted from the data in an automated fashion. Such representations are significant since they can provide the annotation required to support meaningful access to the image data.

Pixel level representations of remote sensed imagery have been successfully derived using spectral information. Such approaches are limited, however, since many land cover types appear similar through such a small aperture. Spatial context must be incorporated, and spatial context at the pixel scale can be considered as image texture. While there has been extensive research over the last few decades into creating quantitative texture representations, little progress has been made in applying these representations to remote sensed imagery, especially to the new classes of high resolution panchromatic (60-100 cm/pixel) and multi-spectral (4 bands for red, green, blue, and near infra-red at 2.4-4 m/pixel) imagery. This paper investigates the use of several popular texture representations for this new imagery. We focus on the specific problem of retrieving inhabited regions from IKONOS images of urban and rural scenes. A query-by-example framework, along with a manually chosen ground truth dataset, allow different combinations of texture representations and spectral bands to be compared.

The rest of the paper is organized as follows. Section 2 describes the four texture features that are compared; Section 3 describes the extension of the representations to multi-spectral imagery; Section 4 describes the similarity retrieval framework used in the comparisons; Section 5 presents the experimental results; and Section 6 concludes with a discussion.

Further author information: (Send correspondence to S.D.N.)
S.D.N.: E-mail: newsam1@llnl.gov, Telephone: 1 925 422 7392

Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI, edited by Belur V. Dasarathy, Proceedings of SPIE Vol. 5433 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.542577

2. TEXTURE FEATURES

This section describes the four texture features compared in this paper, namely, features based on 1) gray level co-occurrence matrices, 2) the Fourier power spectrum, 3) wavelets, and 4) Gabor filters.

2.1. Gray Level Co-occurrence Matrices

Texture features based on the spatial co-occurrence of pixel values are probably the most widely used in the analysis of remote sensed imagery. First proposed by Haralick[1] in 1973, they characterize texture using a variety of quantities derived from second order image statistics. Co-occurrence texture features are extracted from an image in two steps. First, the pairwise spatial co-occurrences of pixels separated by a particular angle and/or distance are tabulated using a gray level co-occurrence matrix (GLCM). Second, the GLCM is used to compute a set of scalar quantities that characterize different aspects of the underlying texture. While these quantities can be interpreted using intuitive notions of texture, their main benefit is providing a quantitative description that can be used for image analysis, such as forming a high-dimensional feature vector to support content based retrieval.

A GLCM is computed for an image as follows. Let $f(x, y)$ be an N_x by N_y image whose pixels take one of L levels, and let $\vec{r}$ be a specified spatial offset. A GLCM entry, $GLCM_{\vec{r}}(i, j)$, is then simply a count of the number of times a pixel with value $j \in 1, \ldots, L$ occurs at offset $\vec{r}$ with respect to a pixel with value $i \in 1, \ldots, L$:

$$\begin{aligned} GLCM_{\vec{r}}(i,j) &= \#\{(x_1,y_1),(x_2,y_2) \in (N_x,N_y) \times (N_x,N_y) | f(x_1,y_1) = i \wedge f(x_2,y_2) = j \\ &\quad \wedge \vec{r} = \overrightarrow{(x_2 - x_1, y_2 - y_1)}\} \, . \end{aligned} \tag{1}$$

The offset $\vec{r}$ can be an angle and/or distance. Intuitively, the diagonal and near-diagonal entries of a GLCM will be larger for images composed of patches with the same or similar pixel values, at least with respect to the offset. The off-diagonal entries will be larger for images in which the pixel values vary locally. Particular spatial patterns, or textures, manifest themselves differently in the distribution of entries both within and between GLCMs computed for a variety of offsets.

The GLCM values are typically normalized to sum to one to remove dependence on the image size. Other practical considerations include choosing the number of quantization levels L for images with large dynamic ranges, and whether the quantization should be uniform or incorporate a transformation such as histogram equalization. These choices are typically dataset dependent.

GLCMs are typically computed for a number of different offsets unless a priori information is available about the underlying texture. A common choice is to compute GLCMs for a distance of one–i.e., adjacency–and four directions, 0, 45, 90, and 135 degrees.

While GLCMs provide a quantitative description of a spatial pattern, they are too unwieldy for practical image analysis. Haralick thus proposed a set of scalar quantities for summarizing the information contained in a GLCM. He originally proposed a total of 14 quantities, or features; however, typically only a subset of these are used. The following five GLCM derived features are now described: angular second moment (ASM), contrast, inverse different moment (IDM), entropy, and correlation. The ASM is computed as

$$\text{ASM} = \sum_{i=1}^{L} \sum_{j=1}^{L} \left(GLCM\,(i,j) \right)^2 . \tag{2}$$

It will have a larger value when the co-occurrence frequencies $g(i, j)$ are concentrated in a few places in the GLCM. This often occurs along the diagonal for a constant image or off diagonal for structured images. A near random or noisy image will have an evenly distributed GLCM with a low ASM. The contrast is computed as

$$\text{CON} = \sum_{n=0}^{L-1} n^2 \left\{ \sum_{|i-j|=n} GLCM\,(i,j) \right\} . \tag{3}$$

The contrast is larger for a GLCM with larger off-diagonal values. Thus, contrast is larger for images with quickly varying intensities which agrees with the intuitive notion of contrast. The IDM is computed as

$$\text{IDM} = \sum_{i=1}^{L}\sum_{j=1}^{L} \frac{GLCM\,(i,j)}{1+(i-j)^2}\ . \tag{4}$$

The IDM will be larger for a GLCM with large diagonal values. Thus, the IDM will be large for images with constant or near-constant patches. The entropy is computed as

$$\text{ENT} = -\sum_{i=1}^{L}\sum_{j=1}^{L} GLCM\,(i,j)\log GLCM\,(i,j)\ . \tag{5}$$

The entropy is larger for an image with an evenly distributed GLCM. Thus, a near random or noisy image will have a larger entropy. Finally, the correlation is computed as

$$\text{COR} = \sum_{i=1}^{L}\sum_{j=1}^{L} \frac{(ij)GLCM(i,j) - \mu_{i'}\mu_{j'}}{\sigma_{i'}\sigma_{j'}}\ , \tag{6}$$

where

$$\mu_{i'} = \sum_{i=1}^{L}\sum_{j=1}^{L} iGLCM(i,j)\ , \tag{7}$$

$$\mu_{j'} = \sum_{i=1}^{L}\sum_{j=1}^{L} jGLCM(i,j)\ , \tag{8}$$

$$\sigma_{i'}^2 = \sum_{i=1}^{L}\sum_{j=1}^{L} GLCM(i,j)(i-\mu_{i'})^2\ , \text{ and} \tag{9}$$

$$\sigma_{j'}^2 = \sum_{i=1}^{L}\sum_{j=1}^{L} GLCM(i,j)(j-\mu_{j'})^2\ . \tag{10}$$

A texture feature vector for an image can be formed from the scalar features derived from GLCMs computed for different offsets. For example, a feature vector composed of the above five features computed at four offsets could be formed as

$$h_{GLCM} = [\text{ASM}_1, \text{CON}_1, \text{IDM}_1, \text{ENT}_1, \text{CORR}_1, \ldots, \text{ASM}_4, \text{CON}_4, \text{IDM}_4, \text{ENT}_4, \text{CORR}_4]\ . \tag{11}$$

2.2. Power Spectrum

Texture is often thought of as being related to periodic image patterns. Since Fourier analysis provides a mathematical framework for the frequency based analysis of images, the Fourier power spectrum has been proposed for characterizing texture in remote sensed imagery. The two dimensional Fourier transform of a continuous image $f(x,y)$ is defined as

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{-2\pi i(ux+vy)} dxdy \;, \tag{12}$$

where $i = \sqrt{-1}$. The Fourier power spectrum is $|F|^2 = FF^*$, where * denotes the complex conjugate.

The radial distribution of values in $|F|^2$ are related to the coarseness of the texture in f. A coarse texture will have high values of $|F|^2$ concentrated near the origin, while a fine texture will have values of $|F|^2$ that are more spread out. Therefore, a useful way to characterize the coarseness of a texture is to take the average of $|F|^2$ over ring-shaped regions centered at the origin

$$\phi_r = \int_0^{2\pi} |F(r,\theta)|^2 \, d\theta \tag{13}$$

for various values of r, the radius of the ring. $F(r,\theta)$ is the polar form of the Fourier transform. Similarly, the angular distribution of the values of $|F|^2$ are related to the direction of the texture in f. A texture with features at a given direction θ will have high values of $|F|^2$ concentrated around the perpendicular direction $\theta + (\pi/2)$, while a non-directional texture will have a $|F|^2$ that is also non-directional. A useful way to characterize the orientation of a texture is to take the average of $|F|^2$ over wedge-shaped regions centered at the origin

$$\phi_\theta = \int_0^{\infty} |F(r,\theta)|^2 \, dr \tag{14}$$

for various values of θ, the angle of the wedge.

The discrete Fourier transform (DFT) is used instead of the continuous Fourier transform for digital images. The DFT of an N_x by N_y pixel digital image $f(x,y)$ is

$$F(u,v) = \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} f(x,y)\, e^{-2\pi i\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)} \;. \tag{15}$$

F is also discrete and the same size as f. The ring-shaped averages are now computed as

$$\phi_{r_1 r_2} = \sum_{\substack{r_1^2 \le u^2+v^2 < r_2^2 \\ 0 \le u \le N_x-1 \\ 0 \le v \le N_y-1}} |F(u,v)|^2 \tag{16}$$

for various values of the inner and outer ring radii r_1 and r_2. Similarly, the wedge-shaped averages are computed as

$$\phi_{\theta_1 \theta_2} = \sum_{\substack{\theta_1 \le \tan^{-1}(v/u) < \theta_2 \\ 0 \le u \le N_x-1 \\ 0 \le v \le N_y-1}} |F(u,v)|^2 \tag{17}$$

for various values of the two angles θ_1 and θ_2 that delimit the wedge.

The averages of the Fourier power spectrum over the ring and wedge shaped regions, $\phi_{r_1 r_2}$ and $\phi_{\theta_1 \theta_2}$, can be used as features characterizing the scale or the orientation of a texture, but not both. Scale and orientation sensitive features can be computed by averaging the Fourier power spectrum over intersections of ring and wedge shaped regions.

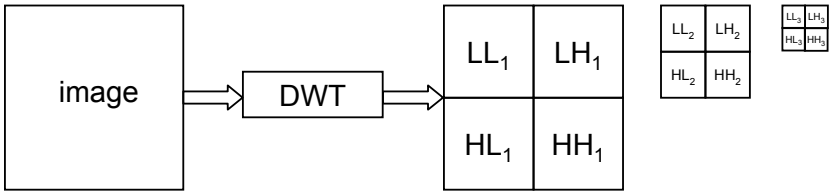


Figure 1. Three level decomposition using the discrete wavelet transform.

A convenient way to approximate the intersections of the wedge and ring shaped regions is to divide the frequency space into a grid of tiles. The averages of the Fourier power spectrum computed over these tiles can then be used as the texture features. If the u and v dimensions are divided into T_u and T_v tiles respectively (assume N_x and N_y are divisible by T_u and T_v respectively) then the power spectrum texture features can be computed as

$$\mathrm{PS}(i,j) = \frac{N_x N_y}{T_u T_v} \sum_{u=i\frac{N_x}{T_u}}^{(i+1)\frac{N_x}{T_u}-1} \sum_{v=j\frac{N_y}{T_v}}^{(j+1)\frac{N_y}{T_v}-1} |F(u,v)|^2 . \tag{18}$$

We supplement these tile averages with four additional quantities computed over the entire frequency space. These are the maximum (MAX), the average (AVG), the energy (ENG), and the variance (VAR) of the magnitude of $F(u,v)$ (see Augusteijn et al.[2] for more details). The final power spectrum texture feature vector is then formed as

$$h_{PSD} = [\mathrm{PS}(0,0), \ldots, \mathrm{PS}(I,J), \mathrm{MAX}, \mathrm{AVE}, \mathrm{ENG}, \mathrm{VAR}] , \tag{19}$$

where $I = N_x/T_u$ and $J = N_y/T_v$.

A major limitation of using texture features based on the Fourier power spectrum is that the Fourier transform is only localized in the frequency domain. It only contains global information about the image in the spatial domain. This constraint has resulted in techniques that provide localization in both space and frequency to be favored over Fourier transform based approaches. Examples of such techniques include those based on the wavelet and Gabor transforms, which are described in the following sections.

2.3. Wavelets

Wavelets are not a single image analysis technique but rather a framework for a variety of related techniques. The common appeal of the techniques is that they allow analysis that is localized in both space and frequency. This is accomplished by decomposing an image using bases of small waves, or wavelets, that have compact, sometimes finite, spatial support. Wavelets are used to characterize texture for the same reason as Fourier based techniques–they provide information about the frequency content of an image which is often associated with texture. However, wavelets, unlike Fourier based approaches, are also localized in space.[3] This makes them more suitable for analyzing texture in non-stationary or non-homogeneous images, such as in remote sensing applications. Wavelets accomplish this not by violating signal processing's analog of the Heisenberg uncertainty principle, that analysis cannot be simultaneously localized in space and frequency, but through a multi-resolution analysis (MRA)[4] that successively trades-off spatial for frequency resolution.[5]

We use the separable discrete wavelet transform (DWT) in an MRA framework to compute the wavelet texture features, as shown in Fig. 1. The DWT applies the Daubechies-4 wavelet to decompose the image into low-low (LL), low-high (LH), high-low (HL), and high-high (HH) components. The MRA framework successively reapplies the DWT to the decimated LL band, resulting in a three level decomposition. The energies and standard deviations of the four components at three resolutions form the texture feature vector:

$$h_{WAVELET} = [energy(LL_1), stdev(LL_1), energy(LH_1), stdev(LH_1), \ldots, energy(HH_3), stdev(HH_3)] \; . \quad (20)$$

Other authors have found that there is very little variation in retrieval performance between texture features derived using different wavelets.[6]

2.4. Gabor Filters

Texture analysis using filters based on Gabor functions falls into the category of frequency-based approaches. These approaches are based on the premise that texture is an image pattern containing a repetitive structure that can be effectively characterized in a frequency domain, such as the Fourier domain. One of the challenges, however, of such an approach is dealing with the tradeoff between the joint uncertainty in the space and frequency domains. Meaningful frequency based analysis cannot be localized without bound. An attractive mathematical property of Gabor functions is that they minimize the joint uncertainty in space and frequency.[7] They achieve the optimal tradeoff between localizing the analysis in the spatial and frequency domains.

Using Gabor filters to analyze texture appeals from a psycho-visual perspective as well. In particular, researchers have found that Gabor functions can be used to model the receptive fields of simple cells in the mammalian visual cortex.[8] Such a finding suggests that Gabor-like filtering takes place early on in the early human visual system.

The texture analysis is accomplished by applying a bank of scale and orientation selective Gabor filters to an image. These filters are constructed as follows.[9] A two-dimensional Gabor function $g(x, y)$ and its Fourier transform $G(u, v)$ can be written as:

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi i W x\right] \quad (21)$$

and

$$G(u,v) = \exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\} \quad (22)$$

where $i = \sqrt{-1}$, $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$ control the tradeoff between spatial and frequency resolution, and W controls the modulation. A class of self-similar functions referred to as Gabor wavelets is now considered. Let $g(x, y)$ be the mother wavelet. A filter dictionary can be obtained by appropriate dilations and translations of $g(x, y)$ through the generating function:

$$\begin{aligned} g_{rs}(x,y) &= a^{-s} g(x', y'), \quad a > 1, \; s \in 0, \ldots, S-1, \; r \in 1, \ldots R \\ x' &= a^{-s}(x\cos\theta + y\sin\theta) \quad \text{and} \\ y' &= a^{-s}(-x\sin\theta + y\cos\theta) \end{aligned} \quad (23)$$

where $\theta = (r-1)\,\pi/R$. The indices r and s indicate the orientation and scale of the filter respectively. R is the total number of orientations and S is the total number of scales in the filter bank. The scale factor a^{-s} in Eq. 23 is meant to ensure that the energy is independent of s.

While the size of the filter bank is application dependent, experimentation has shown that a bank of filters tuned to combinations of five scales, at octave intervals, and six orientations, at 30-degree intervals, is sufficient for the analysis of remote sensed imagery. Fig. 2 displays the real components of a bank of Gabor filters tuned to combinations of three scales and five orientations. See Manjunath and Ma[9] for more details on the filter bank construction, including how the various parameters, such as σ_u, σ_v, a, and W, are derived.

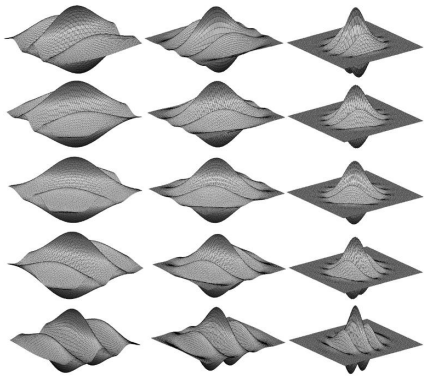


Figure 2. A bank of Gabor filters tuned to combinations of three scales and five orientations.

A Gabor texture feature vector is formed from the filter outputs as follows.[9,10,11] Applying a bank of Gabor filters with R orientations and S scales to an image results in a total of RxS filtered images:

$$f'_{11}(x,y), \ldots, f'_{RS}(x,y) \ . \tag{24}$$

The feature vector is formed by computing the first and second moments of the filtered images. That is, an $2RS$ dimension feature vector, h_{GABOR}, is formed as

$$h_{GABOR} = [\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}, \ldots, \mu_{1S}, \sigma_{1S}, \ldots, \mu_{RS}, \sigma_{RS}] \ , \tag{25}$$

where μ_{rs} and σ_{rs} are the mean and standard deviation, respectively, of $f'_{rs}(x,y)$.

A homogeneous texture descriptor based on Gabor filters was recently standardized by the MPEG-7 Multimedia Content Description Interface[12] with the minor modification that the mean and standard deviation of the filter outputs are computed in the frequency domain for efficiency.

3. MULTI-SPECTRAL TEXTURE

Texture features, such as those described above, are typically applied to panchromatic images. We consider two approaches to extending the texture features to multi-spectral images. In the first approach, we simply append the texture vectors extracted from each of the bands to form a multi-spectral texture feature vector. The advantage of this approach is its simplicity. The disadvantage is the increased size of the feature vector. Furthermore, since there is often significant redundancy between spectral bands, it is likely that the texture feature vectors from different bands are highly correlated. Some ideas of how to address this are presented as future research directions in the conclusion.

The second approach to extending the feature vectors to multi-spectral images is to combine the bands in some appropriate manner and then compute single band texture features. In particular, we compute the normalized difference vegetation index (NDVI) image. Let $f_{RED}(x,y)$ and $f_{NIR}(x,y)$ be the red and near infra-red (NIR) channels of a multi-spectral image respectively. The NDVI image $f_{NDVI}(x,y)$ is then computed as

$$f_{NDVI}(x,y) = \frac{f_{NIR}(x,y) - f_{RED}(x,y)}{f_{NIR}(x,y) + f_{RED}(x,y)} \ . \tag{26}$$

NDVI is a common spectral combination used to determine the amount of green vegetation present at a pixel location since green vegetation reflects the NIR band but absorbs the visible bands. We chose it here to observe how combining the spectral bands to emphasize specific land cover types, here green vegetation, can affect the retrieval performance of the texture features.

4. SIMILARITY RETRIEVAL

Content-based similarity retrieval is a simple yet powerful paradigm for determining the efficacy of perceptual image features, like the texture features described above. The underlying premise is that closeness in the feature space corresponds to perceptual similarity. This allows the similarity between two images to be quantitatively assessed by computing the distance between their feature vectors in the feature space using standard distance functions such as Minkowski norms.

Let T be a collection of M images; let h^m be the d dimension feature vector extracted from image m, where $m \in 1, \ldots, M$; let $dist(\cdot, \cdot)$ be a distance function defined on the d dimensional space; and let h^{query} be the feature vector corresponding to a given query image. Then, the image in T most similar to the query image is the one whose feature vector minimizes the distance to the query's feature vector:

$$m^* = \arg\min_{1 \leq m \leq M} dist(h^{query}, h^m) \; . \tag{27}$$

Likewise, the k most similar images are those that result in the k smallest distances when compared to the query image. Retrieving the k most similar items is commonly referred to as a k-nearest neighbor query.

Typically, the similarity between images is computed using the L1 or L2 (Euclidean) distance measures. The L1 distance between the d dimensional feature vectors h^1 and h^2 is computed as

$$dist_{L1}(h^1, h^2) = \sum_{i=1}^{d} |h_i^1 - h_i^2| \; . \tag{28}$$

5. EXPERIMENTS AND RESULTS

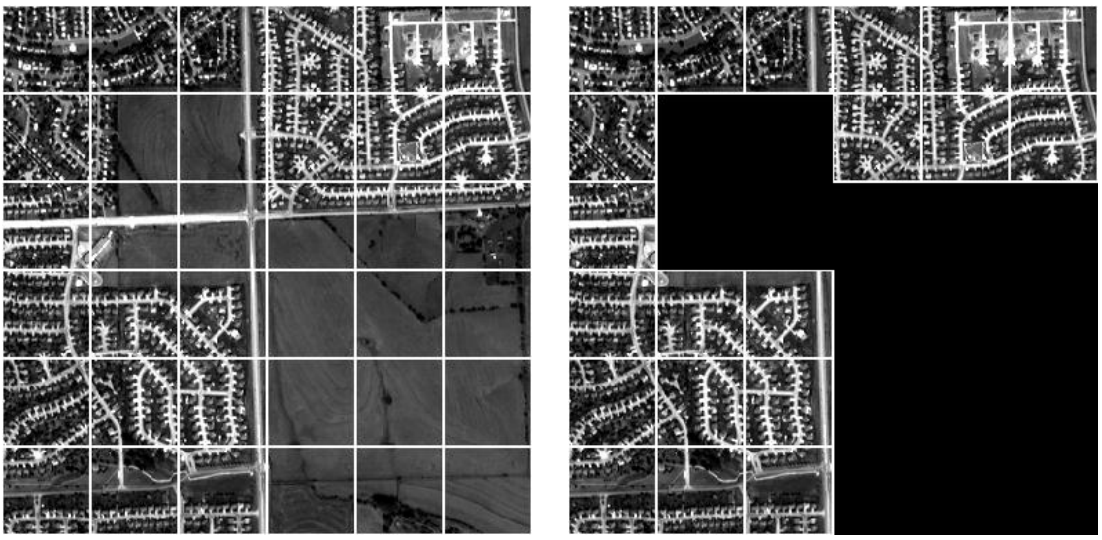
This section describes the experiments performed to compare the texture features described in Section 2. In particular, we consider the problem of identifying inhabited regions in a set of IKONOS satellite images. We divide each image into a set of tiles, and manually identify each tile as being inhabited or not inhabited. This binary labelling provides a ground truth which can be used to evaluate the effectiveness of the different texture features and spectral bands in retrieving inhabited regions.

5.1. The Dataset

The dataset consists of five images from two geographic regions, California and Nebraska. The images measure 384-by-384 pixels and consist of four spectral bands, near infra-red (NIR), red, green, and blue, with a spatial resolution of four meters. Two single band versions of each image are created, an NDVI image as described in Section 3, and a panchromatic image. The panchromatic image is formed from the luminance component of the YCrCb (luminance-chrominance) transformed red, green, and blue bands.

Each 384-by-384 pixel image is partitioned into 36 non-overlapping tiles, each of size 64-by-64 pixels. The total dataset thus consists of 180 tiles. Each tile is manually labelled as being inhabited or not inhabited. Inhabited is here taken to mean a tile with residential buildings. Mixed tiles are labelled based on their majority coverage. An image and its corresponding inhabited tiles are shown in Figs. 3(a) and 3(b). A total of 62 of the 180 tiles are labelled as inhabited.

Each of four texture features–GLCM, power spectrum (PS), wavelet, and Gabor–is extracted from each of the three spectral versions of the dataset tiles–all (NIR, red, green, blue), panchromatic, and NDVI. Table 1 lists the length of each of the texture feature vectors for the four and single band tiles. This table also shows how long it takes to extract the texture features for all 180 of the four band tiles. These times are meant for comparison only and do not represent an optimized extraction procedure. The feature extraction is performed on a Dell PowerEdge 650 Desktop with an Intel Pentium IV processor (2.66 GHz) and 512 MB of memory.



(a) image (b) ground truth

Figure 3. One of the IKONOS images and the corresponding ground truth (Original satellite images by Space Imaging).

	Gabor	**GLCM**	**wavelet**	**PS**
Feature vector length (single band)	60	20	24	20
Feature vector length (all bands)	240	80	96	80
Feature extraction time (seconds)	168	9	11	9

Table 1. Feature vector lengths and extraction times for different texture descriptors.

5.2. Retrieval Results

The texture features are evaluated based on the retrieval rates of similarity searches using the ground truth dataset. Each inhabited tile in the ground truth dataset is used to perform k-nearest neighbor queries for k from 1 to 180. The L1 distance is used to compute similarity. Two quantities are calculated for each query: the true-positive ratio which is the proportion of the inhabited tiles that are in the result set, and the false-positive ratio which is the proportion of the not-inhabited tiles that are in the result set. These values are averaged over the 62 inhabited tile queries, and used to plot receiver operating characteristic (ROC) curves. ROC curves are related to precision/recall curves.

Fig. 4(a) shows the ROC curves for texture descriptors extracted from all four spectral bands. Figs. 4(b) and 4(c) similarly show the ROC curves for the panchromatic and NDVI cases. The Gabor, GLCM, and wavelet features perform comparably in the four band and panchromatic cases, and outperform the PS features. In the NDVI case, the GLCM features are shown to perform the best.

Fig. 5(a) shows the ROC curves for the Gabor features extracted using the four band, panchromatic, and NDVI images. Figs. 5(b), 5(c), and 5(d) similarly show the ROC curves for the GLCM, wavelet, and PS features. In each case, the features extracted using all four spectral bands are shown to perform the best, except for a small section of the PS ROC curves corresponding to high true/false-positive ratios.

The area under an ROC curve provides a single quantity with which to evaluate retrieval performance. Table 2 lists the areas under the ROC curves for different texture feature and spectral band combinations. A higher value indicates better retrieval performance. The GLCM features applied to all four spectral bands is shown to be the best combination.

6. DISCUSSION

This work represents a preliminary investigation, using only a few images, and so the following observations must also be treated as preliminary. The results of the limited experiments indicate that:

- The GLCM, wavelet, and Gabor texture features perform comparably, and outperform the PS features, in retrieving the inhabited regions from the sample dataset.
- The performance of the GLCM features is noteworthy considering their relative simplicity, low extraction cost, and compact feature vector.
- Using all the spectral bands improves the performance but results in significantly larger feature vectors.
- Applying the texture features to a poorly chosen transformation of the spectral bands can reduce the performance. This was observed for using NDVI to retrieve inhabited regions.

Future work includes extending the investigation to consider land cover classes other than just inhabited/not-inhabited, and to include images from other geographic regions. We would also like to see if using more than five of Haralick's fourteen GLCM features can further improve the retrieval performance.

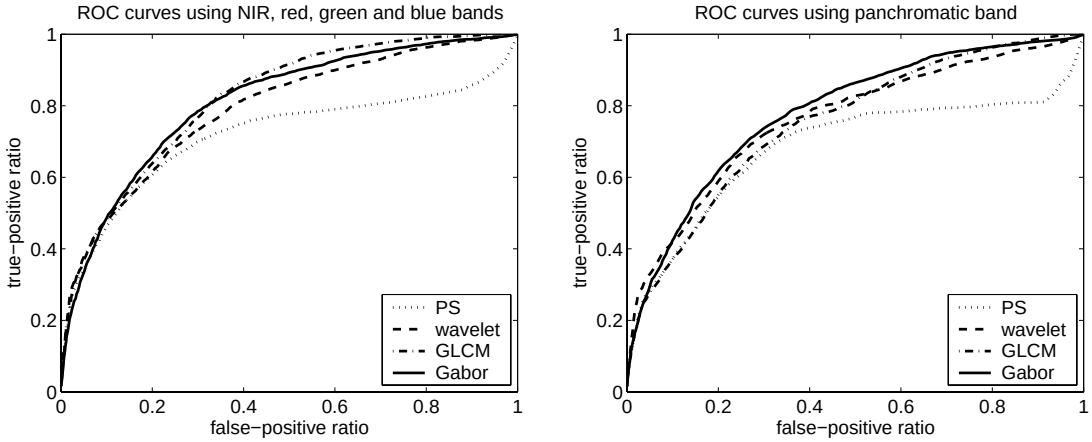
Since using all the spectral bands does appear to improve the retrieval performance, we would also like to investigate ways for dealing with the larger feature vectors. Specifically, we would like to see if dimensionality reduction techniques can be used to make the multi-spectral texture feature vectors more compact without significantly reducing their retrieval performance.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

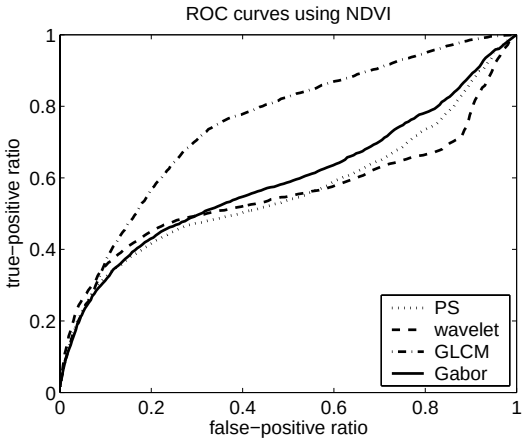
REFERENCES

1. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics* **3**, pp. 610–621, 1973.
2. M. F. Augusteijn, L. E. Clemens, and K. A. Shaw, "Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier," in *IEEE Transactions on Geoscience and Remote Sensing*, **33**(3), pp. 616–626, 1995.
3. A. Graps, "An introduction to wavelets," *IEEE Computational Science and Engineering* **2**(2), 1995.
4. S. Mallat, "A theory for multi-resolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(7), pp. 674–693, 1989.
5. R. Polikar, "The Wavelet Tutorial." Available at http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html.
6. W. Y. Ma and B. S. Manjunath, "A comparison of wavelet features for texture annotation," in *Proceedings of the IEEE International Conference on Image Processing*, **2**, pp. 256–259, 1995.
7. J. Daugman, "Complete discrete 2D Gabor transform by neural networks for image analysis and compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**, pp. 1169–1179, 1988.
8. S. Marcelja, "Mathematical description of the responses of simple cortical cells," *Journal of the Optical Society of America* **70**(11), pp. 1297–1300, 1980.
9. B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(8), pp. 837–842, 1996.
10. P. Wu, B. S. Manjunath, S. D. Newsam, and H. D. Shin, "A texture descriptor for image retrieval and browsing," in *IEEE Conference on Compter Vision and Pattern Recognition: Workshop on Content-Based Access of Image and Video Libraries*, pp. 3–7, 1999.
11. P. Wu, B. S. Manjunath, S. Newsam, and H. D. Shin, "A texture descriptor for browsing and image retrieval," *Journal of Signal Processing: Image Communication* **16**(1), pp. 33–43, 2000.
12. B. S. Manjunath, P. Salembier, and T. Sikora, eds., *Introduction to MPEG7: Multimedia Content Description Interface*, John Wiley & Sons, 2002.



(a) All bands

(b) Panchromatic band

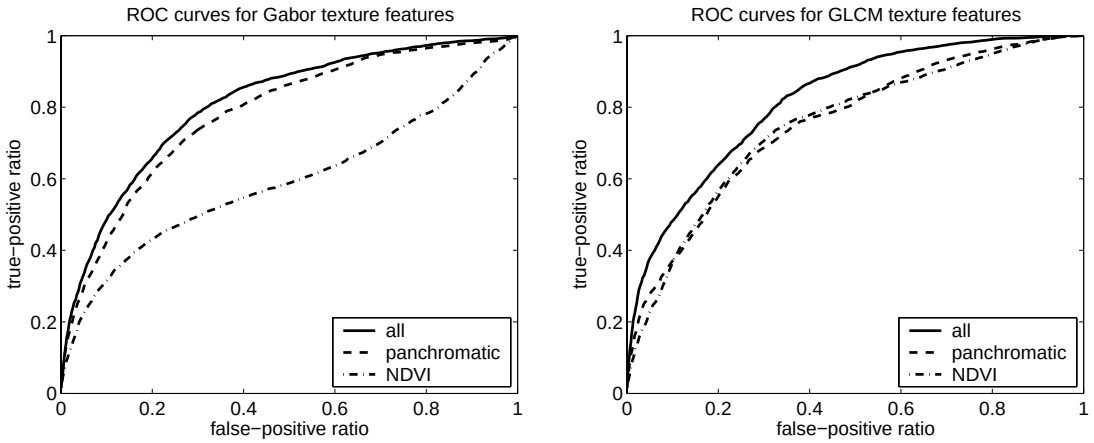


(c) NDVI

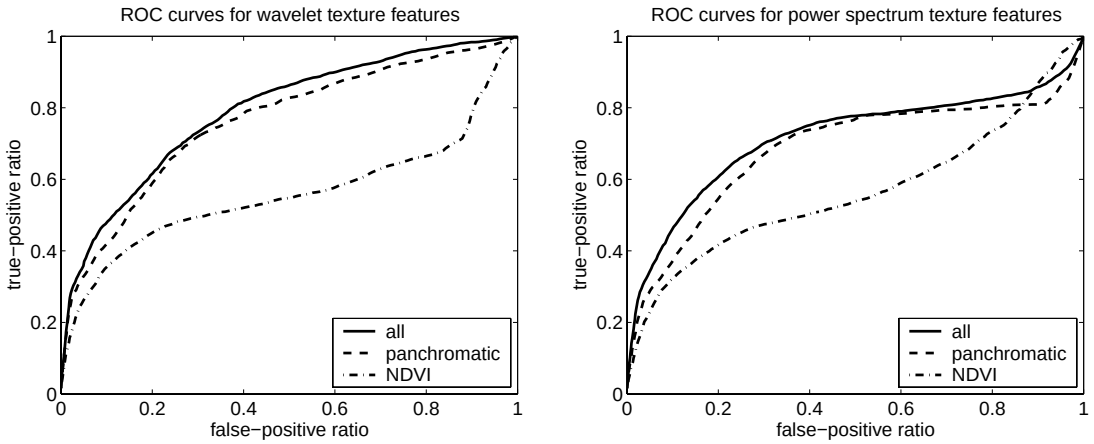
Figure 4. ROC curves comparing different texture descriptors.

	All	**pan**	**NDVI**
Gabor	0.8100	0.7823	0.5942
GLCM	0.8202	0.7532	0.7509
wavelet	0.7898	0.7615	0.5581
PS	0.7155	0.6838	0.5640

Table 2. Area under ROC curves for different combinations of texture features and spectral bands. A higher value indicates better retrieval performance.



(a) Gabor (b) GLCM



(c) wavelet (d) power spectrum

Figure 5. ROC curves comparing different spectral combinations.