Comparing Global and Interest Point Descriptors for Similarity Retrieval in Remote Sensed Imagery

Shawn Newsam
Computer Science and Engineering
University of California
Merced, CA 95343
snewsam@ucmerced.edu

Yang Yang
Computer Science and Engineering
University of California
Merced, CA 95343
yyang6@ucmerced.edu

ABSTRACT

We investigate the application of a new category of low-level image descriptors termed interest points to remote sensed image analysis. In particular, we compare how scale and rotation invariant descriptors extracted from salient image locations perform compared to proven global texture features for similarity retrieval. Qualitative results using a geographic image retrieval application and quantitative results using an extensive ground truth dataset show that interest point descriptors support effective similarity retrieval in large collections of remote sensed imagery.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Image retrieval

Keywords

Interest points, similarity search, remote sensed imagery

1. INTRODUCTION

Remote sensed imagery continues to accumulate at an increasing rate. Exciting new geographic information systems such as Google Earth and Microsoft Virtual Earth are allowing more and more people to access this imagery. However, these systems only allow users to *view* the raw image data. A much richer interaction would be enabled by the integration of automated techniques for annotating the image content. Services such as land-use classification, similarity retrieval, and spatial data mining would not only satisfy known demands but would also spawn new unthought-of applications.

Automated remote sensed image analysis remains by-and-large an unsolved problem. There has been significant effort over the last several decades in using low-level image descriptors, such as spectral, shape and texture features, to make sense of the raw image data. While there has been noted successes for specific problems, ample opportunities remain.

In this paper, we investigate the application of a new category of low-level image descriptors, termed interest points, to remote sensed image analysis. Interest point descriptors have enjoyed surprising success for a range of traditional computer vision problems. There has been little research, however, on applying them to remote sensed imagery.

Our investigation is done in the context of similarity retrieval. In particular, we compare interest point descriptors to global texture features which have been shown to be particularly effective for remote sensed image retrieval. Similarity retrieval is not only an interesting application but also serves as an excellent platform for evaluating the overall descriptiveness of a descriptor.

Our comparison confirms that interest point descriptors show great promise for remote sensed image analysis. Even a straight forward application to similarity retrieval performs comparably to the proven global texture features. This finding opens the door to further investigation.

2. RELATED WORK

Content-based image retrieval (CBIR) has been an active research area in computer vision for over a decade with IBM's Query by Image Content (QBIC) system from 1995 [2] being one of the earliest successes. A variety of image descriptors have been investigated including color, shape, texture, spatial configurations, and others. A recent survey is available in [5].

Similarity based image retrieval has been proposed as an automated method for managing and interacting with the the growing collections of remote sensed imagery. As in other domains, a variety of descriptors have been investigated including spectral [3, 4], shape [14], texture [11, 7, 10, 15, 20], and combinations such as multi-spectral texture [19]. While the most effective descriptor is problem dependent, texture features have enjoyed success since, unlike spectral features, they incorporate spatial information which is clearly important for remote sensed imagery but avoid the difficult pre-processing step of segmentation needed to extract shape features.

The recent emergence of interest point descriptors has revitalized several research areas in computer vision. A number of different techniques have been proposed which have two fundamental components in common. First, a method for finding the so-called interesting or salient locations in an image. Second, a descriptor for describing the image patches

ACMGIS'07 , November 7-9, 2007, Seattle, WA

at these locations. Interest point detectors and descriptors have shown to be robust to changes in image orientation, scale, perspective and illumination conditions as well as to occlusion, and, like global features, do not require segmentation. They are very efficient to compute which allows them to be used in real-time applications. They have been successfully applied to a number of problems including image stereo pair matching, object recognition and categorization, robot localization, panorama construction, and, relevant to this work, image retrieval. Excellent comparisons of different interest point detectors and descriptors can be found in [18] and [17] respectively.

The application of interest point detectors and descriptors to image retrieval has focused primarily on *retrieving images of the same object or scene under different conditions*. Examples include finding additional appearances of a given object in scenes or shots in a video [22], finding images of 3D objects acquired from different viewpoints [21, 23, 24] or against different backgrounds [27], finding images belonging to distinct, homogeneous semantic categories [8, 23], finding frames of the same scene in a video [25], and finding images of the same indoor scene for localization [9]. There has been little application to finding *similar* images or image regions. In particular, to the best of our knowledge, interest point detectors and descriptors have not been applied to the problem of similarity retrieval in large collections of remote sensed imagery.

3. METHODOLOGY

3.1 Global Features

We consider Gabor texture features as the global features. Texture features, and in particular Gabor texture features, have proven to be effective for performing content-based similarity retrieval in remote sensed imagery [11, 19, 20, 10, 15, 7]. The MPEG-7 Multimedia Content Description Interface [16] standardized Gabor texture features after they were shown to outperform other texture features for similarity retrieval. One of the evaluation datasets used in the competitive standardization process consisted of remote sensed imagery.

Gabor texture analysis is accomplished by applying a bank of scale and orientation selective Gabor filters to an image. Gabor functions are Gaussians functions modulated by a sinusoid. Two dimensional spatial filters based on Gabor functions can be made orientation and scale selective by controlling this modulation. While the choice of the number of orientations and scales 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.

A Gabor texture feature vector is formed from the filter outputs as follows [26]. Applying a bank of Gabor filters with R orientations and S scales to an image results in a total of RxS filtered images:

$$f'_{11}(x,y), \ldots, f'_{RS}(x,y) \ . \tag{1}$$

A single global feature vector for the original image is formed by computing the first and second moments of the filtered images. That is, a $2RS$ dimension feature vector, h_{GABOR}, is formed as

$$h_{GABOR} = [\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}, \ldots, \mu_{1S}, \sigma_{1S}, \ldots, \mu_{RS}, \sigma_{RS}] \ , \tag{2}$$

where μ_{rs} and σ_{rs} are the mean and standard deviation of $f'_{rs}(x,y)$. Finally, to normalize for differences in range, each of the $2RS$ components is scaled to have a mean of zero and a standard deviation of one across the entire dataset.

The (dis)similarity between two images is measured by computing the Euclidean distance between their texture features

$$d(h1,h2) = \|h1 - h2\|_2 \ = \sqrt{\sum_{i=1}^{2RS} (h1_i - h2_i)^2} \ . \tag{3}$$

This results in an orientation (and scale) sensitive similarity measure. Orientation invariant similarity is possible by using the modified distance function

$$d_{RI}(h1,h2) = \min_{r \in R} \|h1_{<r>} - h2\|_2 \tag{4}$$

where $h_{<r>}$ represents h circularly shifted by r orientations:

$$\begin{aligned} h_{<r>} = \ & \\ & [(h_{r1}, h_{r2}, \cdots, h_{rS}), (h_{(r+1)1}, h_{(r+1)2}, \cdots, h_{(r+1)S}), \\ & \cdots, (h_{R1}, h_{R2}, \cdots, h_{RS}), (h_{11}, h_{12}, \cdots, h_{1S}), \\ & \cdots, (h_{(r-1)1}, h_{(r-1)2}, \cdots, h_{(r-1)S})] \ . \end{aligned} \tag{5}$$

Parentheses have been added for clarity. Conceptually, this distance function computes the best match between rotated versions of the images without repeating the feature extraction. The granularity of the rotations is of course limited by the filter bank construction.

3.2 Interest Points

We choose David Lowe's Scale Invariant Feature Transform (SIFT) [12, 13] as the interest point detector and descriptor. SIFT-based descriptors have been shown to be robust to image rotation and scale, and to be capable of matching images with geometric distortion and varied illumination. An extensive comparison with other local descriptors found that SIFT-based descriptors performed the best in an image matching task [17]. Like most interest point based analysis, there are two components to SIFT-based analysis. First, a detection step locates points that are identifiable from different views. This process ideally locates the same regions in an object or scene regardless of viewpoint, illumination, etc. Second, these locations are described by a descriptor that is distinctive yet also invariant to viewpoint, illumination, etc. In short, SIFT-based analysis focuses on image patches that can be found and matched under different image acquisition conditions.

The detection step is designed to find image regions that are salient not only spatially but also across different scales. Candidate locations are initially selected from local extrema in Difference of Gaussian (DoG) filtered images in scale space. The DoG images are derived by subtracting two Gaussian blurred images with different σ

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \ . \tag{6}$$

where $L(x,y,\sigma)$ is the image convolved with a Gaussian kernel with standard deviation σ, and k represents the different sampling intervals in scale space. Each point in the three

dimensional DoG scale space is compared with its eight spatial neighbors at the same scale, and with its nine neighbors at adjacent higher and lower scales. The local maximum or minimum are further screened for minimum contrast and poor localization along elongated edges. The last step of the detection process uses a histogram of gradient directions sampled around the interest point to estimate its orientation. This orientation is used to align the descriptor to make it it rotation invariant.

A feature descriptor is then extracted from the image patch centered at each interest point. The size of this patch is determined by the scale of the corresponding extremum in the DoG scale space. This makes the descriptor scale invariant. The feature descriptor consists of histograms of gradient directions computed from a 4x4 spatial grid. The interest point orientation estimate described above is used to align the gradient directions to make the descriptor rotation invariant. The gradient directions are quantized into eight bins so the final feature vector has dimension 128 (4x4x8). This histogram-of-gradients descriptor can be roughly thought of a summary of the edge information in the image patch centered at the interest point.

Rather than work with the full 128 dimension SIFT feature vectors, we clustered a large sampling of the features and labeled the full SIFT feature set with the id of the closest cluster center. Representing the features using the cluster labels has been shown to be effective in other image retrieval tasks [22]. The clustering was performed using the standard k-means algorithm.

The final interest point descriptor used to compute the similarity between two images is composed of the frequency counts of the labeled SIFT feature vectors. That is, h_{INT} for an image, is

$$h_{INT} = [t_0, t_1, \ldots, t_{k-1}] \ , \tag{7}$$

where t_i is number of occurrences of SIFT features with label i in the image. h_{INT} is similar to a term vector in document retrieval. The cosine distance has shown to be effective for comparing documents represented by term vectors [6] so we use it here to compute the similarity between images:

$$d(h1, h2) = \frac{\sum_{i=0}^{k-1} h1_i h2_i}{\sqrt{\sum_{i=0}^{k-1} h1_i^2 \sum_{i=0}^{k-1} h2_i^2}} \ . \tag{8}$$

The cosine distance measure ranges from zero (no match) to one (perfect match). To make it compatible with the distance function used for comparing the global Gabor texture features, for which zero is a perfect match, we use one minus the cosine distance to perform similarity retrieval using interest point descriptors.

3.3 Similarity Retrieval

The above features and associated distance measures are used to perform similarity retrieval as follows. Let T be a collection of M images; let h^m be the feature vector extracted from image m, where $m \in 1, \ldots, M$; let $d(\cdot, \cdot)$ be a distance function defined on the feature 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:

$$m^* = \arg\min_{1 \le m \le M} d(h^{query}, h^m) \ . \tag{9}$$

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 (kNN) query.

Given a ground-truth dataset, there are a number of ways to evaluate retrieval performance. One common method is to plot the precision of the retrieved set for different values of k. Precision is defined as the percent of the retrieved set that is correct and can be computed as the ratio of the number of true positives to the size of the retrieved set. It is straight forward and meaningful to compute and compare the average precision for a set of queries when the ground truth sizes are the same. (It is not straight forward to do this for precision-recall curves.)

Plotting precision versus the size of the retrieved set provides a graphical evaluation of performance. A single measure of performance that not only considers that the ground-truth items are in the top retrievals but also their ordering can be computed as follows [16]. Consider a query q with a ground-truth size of $NG(q)$. The $Rank(k)$ of the kth ground-truth item is defined as the position at which it is retrieved. A number $K(q) \ge NG(q)$ is chosen so that items with a higher rank are given a constant penalty

$$Rank(k) = \begin{cases} Rank(k), & \text{if } Rank(k) \le K(q) \\ 1.25K(q), & \text{if } Rank(k) > K(q) \end{cases} \ . \tag{10}$$

$K(q)$ is commonly chosen to be $2NG(q)$. The *Average Rank* (AVR) for a single query q is then computed as

$$AVR(q) = \frac{1}{NG(q)} \sum_{k=1}^{NG(k)} Rank(k) \ . \tag{11}$$

To eliminate influences of different $NG(q)$, the *Normalized Modified Retrieval Rank* (NMRR)

$$NMRR(q) = \frac{AVR(q) - 0.5[1 + NG(q)]}{1.25K(q) - 0.5[1 + NG(q)]} \tag{12}$$

is computed. $NMRR(q)$ takes values between zero (indicating whole ground truth found) and one (indicating nothing found) irrespective of the size of the ground-truth for query q, $NG(q)$. Finally, the *Average Normalized Retrieval Rate* (ANMRR) can be computed for a set NQ of queries

$$ANMRR = \frac{1}{NQ} \sum_{q=1}^{NQ} NMRR(q) \ . \tag{13}$$

4. EVALUATION

This section describes the datasets and techniques used to perform the comparisons.

4.1 Datasets

A collection of IKONOS 1-m panchromatic satellite images of the United States was used to compare the descriptors. Separate datasets were created for the qualitative and quantitative analyses. For the qualitative analysis, two large IKONOS images of the Phoenix and Los Angeles areas were partitioned into non-overlapping 64-*by*-64 pixel tiles. The Phoenix image measures 21,248-*by*-11,328 pixels for a total

Figure 1: Image patches corresponding to two of the 50 clusters used to label the SIFT features. The top row shows a cluster that has captured corner-like patches. The bottom row shows a cluster that has captured grid-like patches.

of 58,764 tiles and the Los Angeles image measures 10,560-*by*-10,624 pixels for a total of 27,390 tiles (86,154 tiles in total). A single Gabor texture feature was extracted from each tile using a filterbank tuned to $R = 6$ orientations and $S = 5$ scales. The interest points descriptors were extracted and assigned to the tiles as follows. First, interest points and SIFT features were extracted from the complete images, resulting in 4,880,415 features for the Phoenix image and 2,406,787 features for the Los Angeles image. 100,000 features were sampled from the combined set and clustered into 50 clusters using k-means clustering. Figure 1 shows sample image patches for two of the 50 clusters. Each of the 7,287,202 features in the large images was labeled based on the clustering results and assigned to the tile containing the interest point location. Thus, the 86,154 tiles contained 84.6 labeled SIFT features on average. Finally, a single interest point descriptor consisting of the label counts was assigned to each tile.

The quantitative analysis required a ground-truth dataset. Ten sets of 100 64-*by*-64 pixel images were manually extracted from 22 large IKONOS images for the following land-use/cover classes: aqueduct, commercial, dense residential, desert chaparral, forest, freeway, intersection, parking lot, road, and rural residential. Figure 2 shows examples from each of these ten classes. A single Gabor texture feature was extracted for each of the 1,000 ground-truth images, again using a filterbank tuned to $R = 6$ orientations and $S = 5$ scales. Interest points and SIFT features were also extracted from each image and labeled using the clustering from the larger dataset above (thus the clustering and labeling was not tuned to the ground-truth dataset). A single interest point descriptor consisting of the label counts was assigned to each image. The images here contained an average of 59.1 labeled features (fewer than above since the SIFT features were extracted from the small images placing an upper bound on the scale of the interest points). Figure 3 shows the locations of the detected interest points for the sample images in figure 2.

It is worth pointing out the different feature extraction times for the ground-truth dataset. It took approximately 51 seconds to extract and label the interest points and approximately 353 seconds to extract the Gabor texture features for the 1,000 images in the ground-truth dataset (on a typical desktop workstation). While the extraction software was not optimized and the timing measurements were not scientific, we believe this order-of-magnitude difference between the two features is to be expected. Efficient extraction is a noted strength of SIFT features.

4.2 Qualitative Analysis

A Geographic Image Retrieval (GIR) demonstration application was used for the qualitative analysis. The GIR demo allows a user to navigate large IKONOS images and select 64-*by*-64 pixel tiles as query images. The user can then perform a k-nearest neighbor query using either the interest points or the global Gabor texture features. The most similar k tiles in the result set is displayed in order of decreasing similarity. This demo turns out to be a valuable tool for evaluating the descriptive power of a feature. Figure 4 shows a screen capture of the GIR demo in which the user has selected a tile from a dense residential region in the center of the displayed IKONOS image of Phoenix. The user is now ready to perform a 128-nearest neighbor query in the 86,154 tile Los Angeles and Phoenix image dataset. Figure 5 shows the top 32 retrievals in order of decreasing similarity for this query tile for each of the three approaches. An earlier version of this demo that uses the global Gabor texture features to perform similarity retrieval on a large collection of aerial images is available online at [1].

4.3 Quantitative Analysis

The quantitative analysis involved a comprehensive set of similarity retrievals using each of the 1,000 images in the ground-truth dataset as a query. Precision was computed for each query as a function of retrieved set size from 1 to 1,000. These precision values were then averaged over the 100 queries from each of the ten ground-truth classes. This was performed three times: 1) for the interest point descriptors; 2) for the global Gabor texture features using the standard orientation sensitive distance measure; and 3) for the global Gabor texture features using the modified rotation invariant (RI) distance measure. Figure 6 shows the averaged precision curves for the ground truth datasets. The optimal case is also plotted for comparison.

The *Average Normalized Modified Retrieval Rate* (ANMRR) described in section 3.3 was also computed for each of the three similarity retrieval methods, for each of the ten ground-truth classes. Table 1 shows these values which range from zero for all the ground-truth items retrieved in a result set the size of the ground-truth to one for none of the ground-truth items retrieved.

Table 1: *Average Normalized Modified Retrieval Rate* (ANMRR). Lower value is better.

Ground-truth	Interest pts	Global	Global RI
Aqueduct	0.494	0.417	0.243
Commercial	0.604	0.432	0.385
Dense residential	0.413	0.314	0.280
Desert chaparral	0.023	0.015	0.020
Forest	0.188	0.327	0.368
Freeway	0.458	0.761	0.430
Intersection	0.438	0.358	0.420
Parking lot	0.358	0.502	0.460
Road	0.637	0.623	0.485
Rural residential	0.463	0.413	0.454
Average	0.408	0.416	0.354

Again, the interest point descriptors were more computationally efficient, this time in terms of how long it took to perform all 1,000 queries. On average, using the interest point descriptors took only two seconds, using the global Gabor texture features took 12 seconds, and using the texture features with the rotation invariant distance measure took

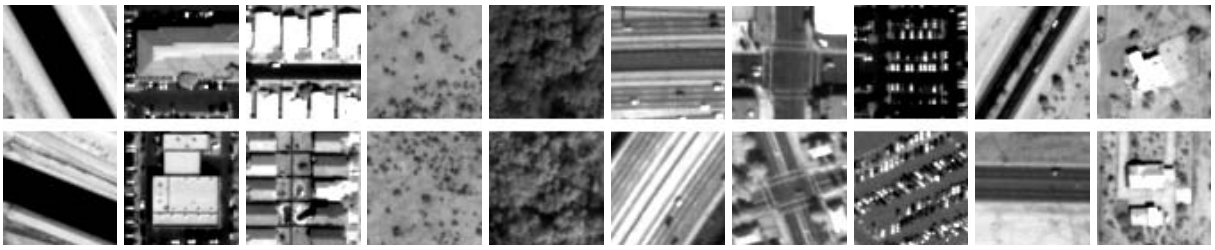


Figure 2: Two examples from each of the ground-truth classes. (a) Aqueduct. (b) Commercial. (c) Dense residential. (d) Desert chaparral. (e) Forest. (f) Freeway. (g) Intersection. (h) Parking lot. (i) Road. (j) Rural residential.

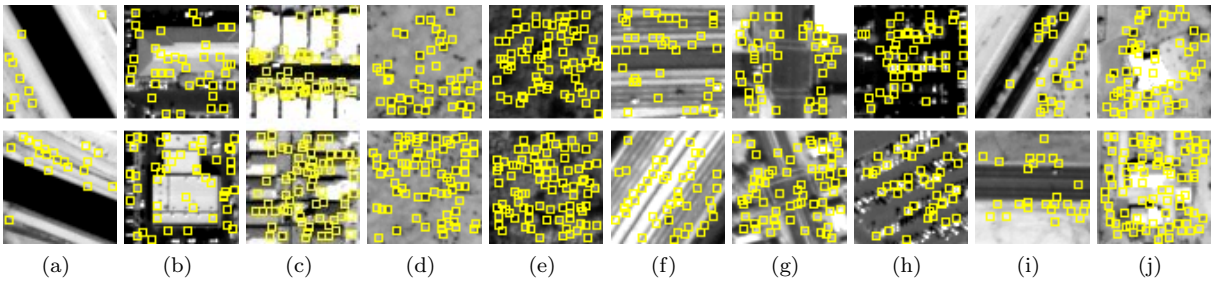


Figure 3: The interest point locations for the images in figure 2.

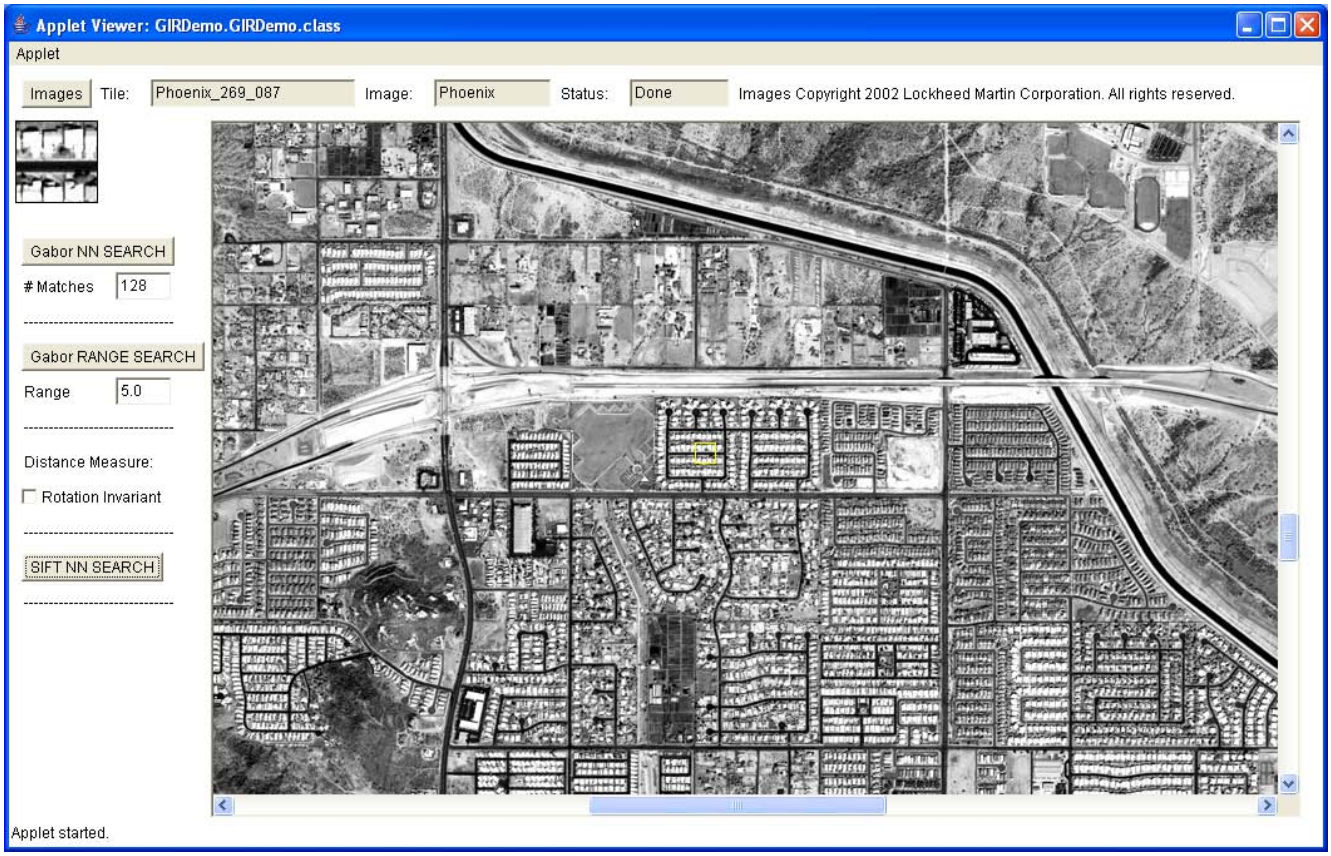
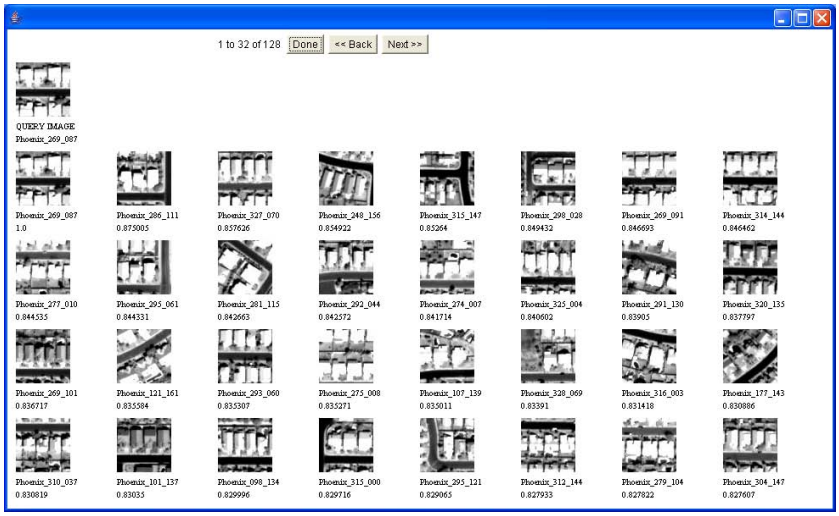
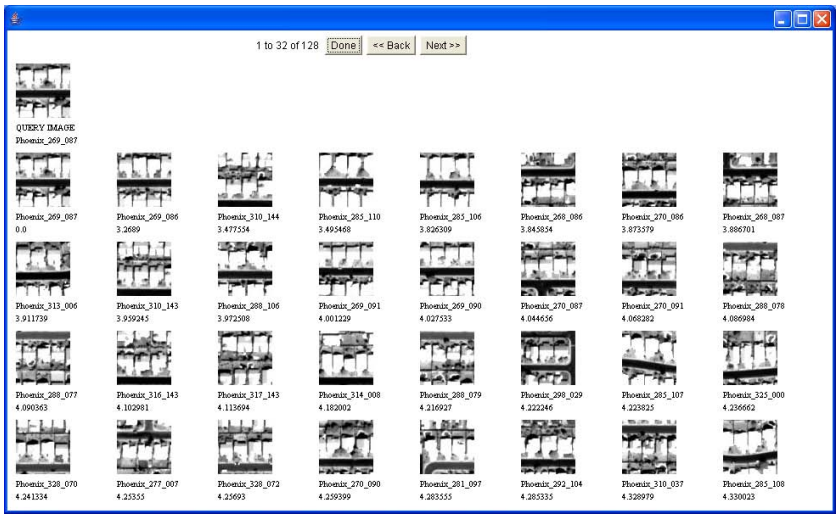


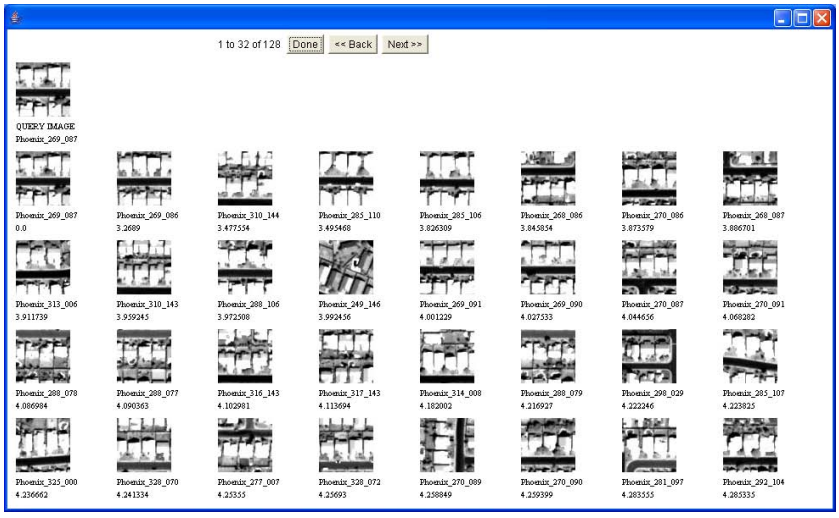
Figure 4: The Geographic Image Retrieval demo which allows users to perform similarity retrieval in remote sensed imagery.



(a)



(b)



(c)

Figure 5: Examples of similarity retrieval using the GIR demo. The query tile (top left) and the top 32 retrieved images in order of decreasing similarity for (a) interest point descriptors, (b) global Gabor texture features, and (c) global Gabor texture features using the rotation invariant similarity measure.

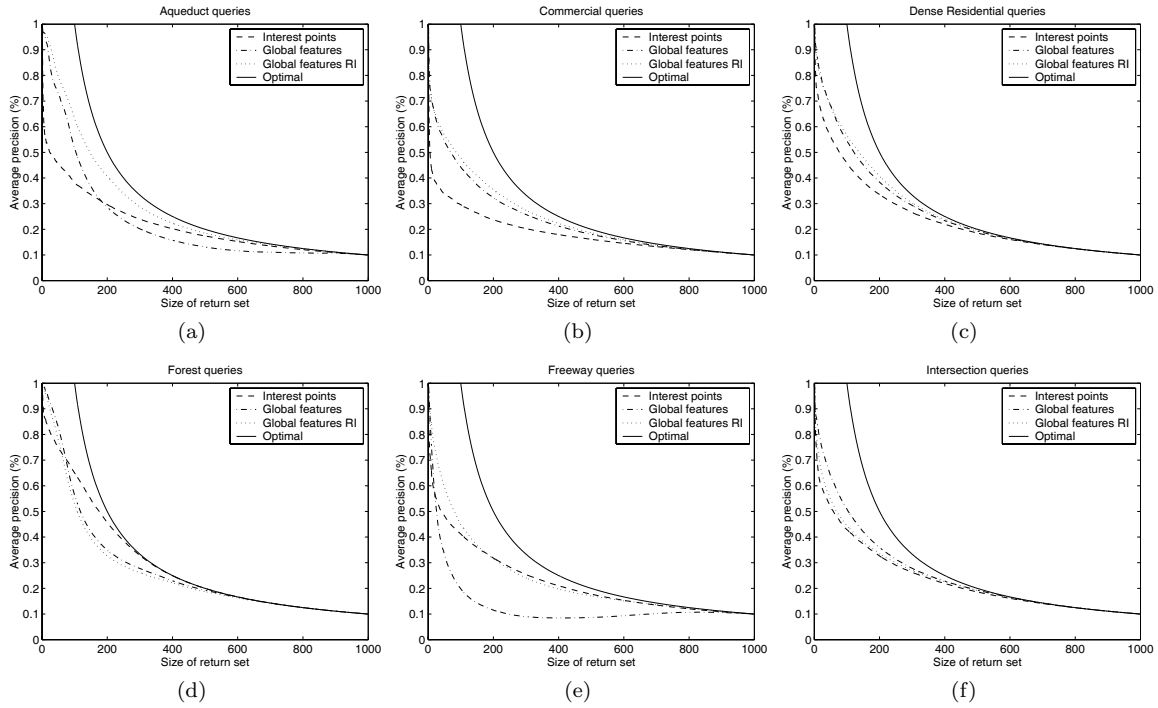


Figure 6: Precision as a function return set size for the three similarity retrieval methods for the ground-truth classes. (RI=rotation invariant) (a) Aqueduct. (b) Commercial. (c) Dense residential. (d) Forest. (e) Freeway. (g) Intersection. Not shown are desert chaparral (methods perform comparably), parking lot (curves are similar to forest), road (curves are similar to aqueduct), and rural residential (curves are similar to commercial).

60 seconds. This variation could be critical for supporting interactive similarity retrieval.

5. DISCUSSION

The qualitative analysis provided by the GIR demo showed the interest point descriptors support effective similarity retrieval. Retrieval results for the interest points, such as the example in figure 5, are rotation invariant, and when compared to the global features, are less sensitive to differences in scale. This makes sense because the interest points are normalized for scale during the detection step. We also observed that they are more robust to variation in the spatial configurations of the ground truth classes. Notice that the retrieved set for the interest point descriptors in figure 5 exhibits greater variability in the arrangement of the houses and streets than the retrieved sets for the global texture features.

None of the approaches was shown to clearly outperform the others in the quantitative analysis. Both the precision curves and the ANMRR values indicate that different descriptors are better for different ground truth classes. The following general observations can be made from the precision curves in figure 6. The interest points descriptors have difficulty with the aqueduct, commercial, and road classes (the precision curves for the road class are not shown but are very similar in shape to those for the aqueduct class). These classes tend to be very structured which presents a challenge for the interest point descriptors. The interest point descriptors perform the best for the forest and parking lot classes (the curves for parking lot are similar to those for forest). Again, these classes exhibit less structure–forest is a stochastic rather than a regular pattern, and the parking lots vary in how full they are. The rotation invariance of the interest point descriptors makes them perform comparable to the rotation invariant global texture approach for the freeway class. Finally, the rotation sensitive global texture approach performs the best for the intersection and rural residential classes (the curves for rural residential are similar to those for forest). Due to the nature of the IKONOS images, these classes tend to be similarly oriented thus providing an advantage to an approach that exploits this.

The ANMRR values in table 1 are in agreement with these observations. The ANMRR averaged over all ground-truth classes indicates the rotation invariant global texture approach performs the best overall, followed by the interest points, and last is the rotation sensitive global texture features.

This work represents an initial investigation into using interest point descriptors for content-based analysis of remote sensed imagery. This new category of low-level features was shown to perform comparably to proven approaches to similarity retrieval. There is plenty of future work to be done. We plan to incorporate the spatial arrangement of the interest points into the descriptor. This should improve the performance for ground-truth classes such as aqueduct and road. The challenge will be to do this in a computationally efficient manner. We also plan on performing a comparison using a ground-truth dataset containing class exemplars at varying scales. This should further validate the scale invari-

ance of the interest point descriptors.

6. REFERENCES

[1] MPEG-7 homogeneous texture descriptor demo. http://faculty.ucmerced.edu/snewsam/MPEG7Demo.html.

[2] J. Ashley, M. Flickner, J. Hafner, D. Lee, W. Niblack, and D. Petkovic. The query by image content (QBIC) system. In *ACM SIGMOD International Conference on Management of Data*, 1995.

[3] T. Bretschneider, R. Cavet, and O. Kao. Retrieval of remotely sensed imagery using spectral information content. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pages 2253–2255, 2002.

[4] T. Bretschneider and O. Kao. A retrieval system for remotely sensed imagery. In *International Conference on Imaging Science, Systems, and Technology*, volume 2, pages 439–445, 2002.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. In *Penn State University Technical Report CSE 06-009*, 2006.

[6] D. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining*. The MIT Press, 2001.

[7] Y. Hongyu, L. Bicheng, and C. Wen. Remote sensing imagery retrieval based-on Gabor texture feature classification. In *International Conference on Signal Processing*, pages 733–736, 2004.

[8] C.-T. Hsu and M.-C. Shih. Content-based image retrieval by interest points matching and geometric hashing. In *SPIE Photonics Asia Conference*, volume 4925, pages 80–90, 2002.

[9] L. Ledwich and S. Williams. Reduced SIFT features for image retrieval and indoor localisation. In *Australasian Conference on Robotics and Automation*, 2004.

[10] C.-S. Li and V. Castelli. Deriving texture feature set for content-based retrieval of satellite image database. In *IEEE International Conference on Image Processing*, 1997.

[11] Y. Li and T. Bretschneider. Semantics-based satellite image retrieval using low-level features. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, volume 7, pages 4406–4409, 2004.

[12] D. G. Lowe. Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157, 1999.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[14] A. Ma and I. K. Sethi. Local shape association based retrieval of infrared satellite images. In *IEEE International Symposium on Multimedia*, 2005.

[15] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996.

[16] B. S. Manjunath, P. Salembier, and T. Sikora, editors. *Introduction to MPEG7: Multimedia Content Description Interface*. John Wiley & Sons, 2002.

[17] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[18] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005.

[19] S. Newsam and C. Kamath. Retrieval using texture features in high resolution multi-spectral satellite imagery. In *SPIE Defense and Security Symposium, Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI*, 2004.

[20] S. Newsam, L. Wang, S. Bhagavathy, and B. S. Manjunath. Using texture to analyze and manage large collections of remote sensed image and video data. *Journal of Applied Optics: Information Processing*, 43(2):210–217, 2004.

[21] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[22] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *IEEE International Conference on Computer Vision*, volume 2, pages 1470–1477, 2003.

[23] Q. Tian, N. Sebe, M. Lew, E. Loupias, and T. Huang. Content-based image retrieval using wavelet-based salient points. In *SPIE International Symposium on Electronic Imaging, Storage and Retrieval for Media Databases*, 2001.

[24] J. Wang, H. Zha, and R. Cipolla. Combining interest points and edges for content-based image retrieval. In *IEEE International Conference on Image Processing*, pages 1256–1259, 2005.

[25] C. Wolf, W. Kropatsch, H. Bischof, and J.-M. Jolion. Content based image retrieval using interest points and texture features. *International Conference on Pattern Recognition*, 04:4234, 2000.

[26] 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):33–43, 2000.

[27] H. Zhang, R. Rahmani, S. R. Cholleti, and S. A. Goldman. Local image representations using pruned salient points with applications to CBIR. In *Proceedings of the 14th Annual ACM International Conference on Multimedia*, 2006.