

# Geographic Image Retrieval Using Interest Point Descriptors

Shawn Newsam and Yang Yang

Computer Science and Engineering  
University of California, Merced CA 95344, USA  
snewsam,yyang6@ucmerced.edu

**Abstract.** We investigate image retrieval using interest point descriptors. New geographic information systems such as Google Earth and Microsoft Virtual Earth are providing increased access to remote sensed imagery. Content-based access to this data would support a much richer interaction than is currently possible. Interest point descriptors have proven surprisingly effective for a range of computer vision problems. We investigate their application to performing similarity retrieval in a ground-truth dataset manually constructed from 1-m IKONOS satellite imagery. We compare results of using quantized versus full descriptors, Euclidean versus Mahalanobis distance measures, and methods for comparing the sets of descriptors associated with query and target images.

## 1 Introduction

New geographic information systems such as Google Earth and Microsoft Virtual Earth are providing increased access to geographic imagery. These systems, however, only allow users to *view the raw image data*. Automated techniques for annotating the image content would enable much richer interaction. Solutions for land-use classification, similarity retrieval, and spatial data mining would not only serve existing needs but would also spawn novel 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, plenty of opportunities for improvement 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. In previous work [1], we showed that a straightforward application of interest point descriptors to similarity retrieval performed comparably to state-of-the-art approaches based on global texture analysis. In this paper, we explore the interest point descriptors further.

Our investigation is done in the context of similarity retrieval which is not only a useful application but also serves as an excellent platform for evaluating a

descriptor. We investigate several methods for using interest point descriptors to perform similarity retrieval in a large dataset of geographic images. We compare the results of using quantized versus full-length descriptors, of using different descriptor-to-descriptor distance measures, and of using different methods for comparing the sets of descriptors representing the images.

## 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] 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 [3].

Image retrieval has been proposed as an automated method for accessing the the growing collections of remote sensed imagery. As in other domains, a variety of descriptors have been investigated including spectral [4,5], shape [6], texture [7,8,9,10], and combinations such as multi-spectral texture [11].

The recent emergence of interest point descriptors has revitalized many 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 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 problems such as image stereo pair matching, object recognition and categorization, robot localization, panorama construction, and, relevant to this work, image retrieval. Excellent comparisons of interest point detectors and descriptors can be found in [12] and [13], 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* [14,15,16,17,18]. There has been little application to finding *similar images or image regions*. In particular, there has not been much investigation into using interest point descriptors to perform similarity retrieval in large collections of remote sensed imagery.

## 3 Interest Point Descriptors

We choose David Lowe's Scale Invariant Feature Transform (SIFT) [19,20] as the interest point detector and descriptor. SIFT 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 [13]. Like most interest point based analysis, there are

two components to extracting SIFT descriptors. 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 and illumination. Second, these locations are described by a descriptor that is distinctive yet also invariant to viewpoint and illumination. SIFT-based analysis exploits image patches that can be found and matched under different image acquisition conditions.

The SIFT 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  $\sigma$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1)$$

where  $L(x, y, \sigma)$  is the image convolved with a Gaussian kernel with standard deviation  $\sigma$ , 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 maxima or minima are further screened for low 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 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 over 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 that 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 a scale and orientation normalized image patch centered at the interest point.

## 4 Similarity Measures Using Full Descriptors

This section describes methods for computing the similarity between two images represented by sets of full interest point descriptors. First, we describe the comparison of single descriptors and then extend this to sets of descriptors.

### 4.1 Comparing Single Descriptors

SIFT descriptors are represented by 128 dimension feature vectors. We use standard Euclidean distance to compute the similarity between two SIFT descriptors. Let  $h_1$  and  $h_2$  be the feature vectors representing two SIFT descriptors. The Euclidean distance between these features is then computed as

$$d_{Euc}(h_1, h_2) = \sqrt{(h_1 - h_2)^T (h_1 - h_2)}. \quad (2)$$

We also consider using the Mahalanobis distance to compare single descriptors. The Mahalanobis distance is equivalent to the Euclidean distance computed in a transformed feature space in which the dimensions (feature components) have uniform scale and are uncorrelated. The Mahalanobis distance between two feature vectors is computed as

$$d_{Mah}(h_1, h_2) = \sqrt{(h_1 - h_2)^T \Sigma^{-1} (h_1 - h_2)} \quad (3)$$

where  $\Sigma$  is the covariance matrix of the feature vectors.

## 4.2 Comparing Sets of Descriptors

Since images are represented by multiple interest point descriptors, we need a method to compute the similarity between sets of descriptors. We formulate this as a bipartite graph matching problem between a query and target graph in which the vertices are the descriptors and the edges are the distances between descriptors computed using either the Euclidean or Mahalanobis distance.

We consider two different methods for making the graph assignments. In the first method, we assign each query vertex to the target vertex with the minimum distance, allowing many-to-one matches. Let the query image contain the set of  $m$  descriptors  $H_q = \{h_{q1}, \dots, h_{qm}\}$  and the target image contain the set of  $n$  descriptors  $H_t = \{h_{t1}, \dots, h_{tn}\}$ . Then, we define the *minimum distance measure* between the query and target image to be

$$D_{min}(Q, T) = \frac{1}{m} \sum_{i=1}^m dmin(h_{qi}, T) \quad (4)$$

where

$$dmin(h_{qi}, T) = \min_{1 \leq j \leq n} d(h_{qi}, h_{tj}) \quad (5)$$

and  $d(\cdot, \cdot)$  is either the Euclidean or Mahalanobis distance. The factor of  $1/m$  normalizes for the size of the query descriptor set.

We also consider the optimal complete (perfect) assignment between query and target vertices. In this assignment we allow a query vertex to be assigned to at most one target vertex. In the case where there are fewer target than query vertices, we allow some of the query vertices to remain unassigned. We define the *complete distance measure* between the query and target image to be

$$D_{comp}(Q, T) = \min_f \sum_{i=1}^m d(h_{qi}, h_{tf(i)}) \quad (6)$$

where  $f(\cdot)$  is an assignment which provides a one-to-one mapping from  $(1, \dots, m)$  to  $(1, \dots, n)$ . Again,  $d(\cdot, \cdot)$  is either the Euclidean or Mahalanobis distance. In the case where  $m > n$ , we allow  $m - n$  values not to be mapped and not contribute to the distance summation. We find the optimal mapping using the Hungarian algorithm [21] which runs in polynomial time in  $m$  and  $n$ . Finally, we normalize for the number of descriptors by dividing the distance by  $\min(m, n)$ .

## 5 Similarity Measures Using Quantized Descriptors

As an alternate to using the full 128 dimension descriptors, we investigate quantized features. Quantized features are more compact and support significantly faster similarity retrieval. Quantized interest point descriptors have proven effective in other image retrieval tasks [14].

The 128 dimension descriptors were quantized using the  $k$ -means algorithm. The clustering was performed using features randomly sampled from a large training dataset (the full dataset was too large to cluster). The clustering results were then used to label the features in the test dataset with the ID of the closest cluster centroid. We compared  $k$ -means clustering and labeling using the Euclidean and Mahalanobis distance measures.

A feature vector consisting of the counts of the quantized descriptors was used to compute the similarity between images. That is,  $H_{quant}$  for an image is

$$H_{quant} = [t_0, t_1, \dots, t_{c-1}] \quad (7)$$

where  $t_i$  is number of occurrences of quantized descriptors with label  $i$  and  $c$  is the number of clusters used to quantize the features.  $H_{quant}$  is similar to a term vector in document retrieval. The cosine distance measure has shown to be effective for comparing documents represented by term vectors [22] so we use it here to compute the similarity between images. The similarity between a query image  $Q$  with counts  $[q_0, q_1, \dots, q_{c-1}]$  and a target image  $T$  with counts  $[t_0, t_1, \dots, t_{c-1}]$  is computed as

$$D_{quant}(Q, T) = \frac{\sum_{i=0}^{c-1} q_i t_i}{\sqrt{\sum_{i=0}^{c-1} q_i^2 \sum_{j=0}^{c-1} t_j^2}} \quad (8)$$

The cosine distance measure ranges from zero (no match) to one (perfect match). To make it compatible with the distance measures above, for which zero is a perfect match, we use one minus the cosine distance to perform similarity retrieval.

## 6 Similarity Retrieval

The distance measures above are used to perform similarity retrieval as follows. Let  $Q$  be a query image and let  $T \in \mathbb{T}$  be a set of target images. The image  $T^* \in \mathbb{T}$  most similar to  $Q$  is computed as

$$T^* = \arg \min_{T \in \mathbb{T}} D(Q, T) \quad (9)$$

where  $D(\cdot, \cdot)$  is one of the image-to-image distance measures described above. 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 images is commonly referred to as a  $k$ -nearest neighbor ( $k$ NN) 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 straightforward 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 straightforward 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 [23]. Consider a query  $q$  with a ground-truth size of  $NG(q)$ . The  $Rank(k)$  of the  $k$ th ground-truth item is defined as the position at which it is retrieved. A number  $K(q) \geq NG(q)$  is chosen so that items with a higher rank are given a constant penalty

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

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

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

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

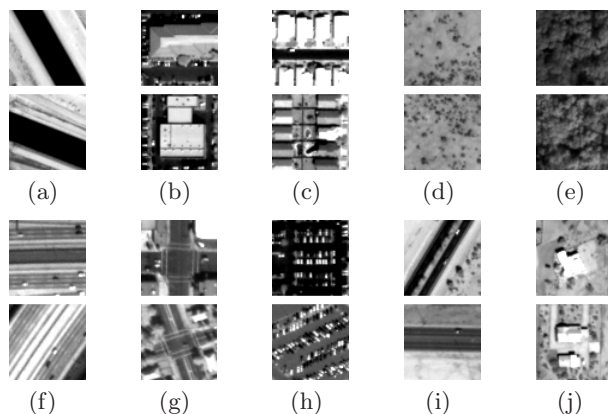
$$NMRR(q) = \frac{AVR(q) - 0.5[1 + NG(q)]}{1.25K(q) - 0.5[1 + NG(q)]} \quad (12)$$

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

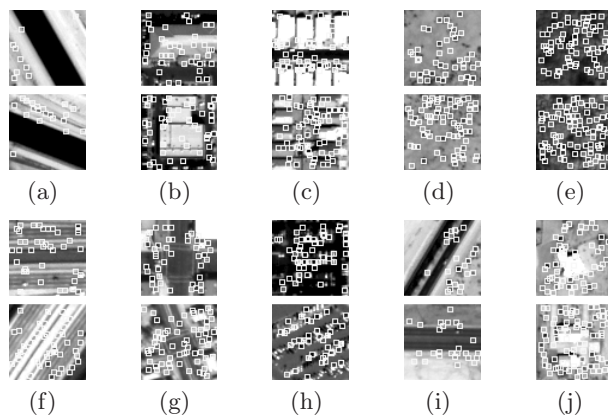
$$ANMRR = \frac{1}{NQ} \sum_{q=1}^{NQ} NMRR(q) . \quad (13)$$

## 7 Dataset

A collection of 1-m panchromatic IKONOS satellite images was used to evaluate the retrieval methods. A ground truth consisting of ten sets of 100 64-by-64 pixel images was manually extracted from the 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 1 shows



**Fig. 1.** 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.



**Fig. 2.** The interest point locations for the ground-truth images in figure 1

examples from each of these ten classes. SIFT interest point descriptors were extracted from each image as described in section 3. Figure 2 shows the locations of the detected interest points for the sample images in Figure 1. Each image contains an average of 59.1 interest points. A large set of features randomly sampled from the the full IKONOS images was clustered using the  $k$ -means algorithm using both the Euclidean and Mahalanobis distance measures. The features in the 1,000 ground-truth images were labeled with the ID of the closest cluster centroid. Each ground-truth image is thus represented by the following:

- A set of full interest point descriptors.
- Quantized feature counts based on clustering using Euclidean distance.
- Quantized feature counts based on clustering using Mahalanobis distance.

## 8 Results

The retrieval performance of the different representations and similarity measures was evaluated by performing a comprehensive set of  $k$ -nearest neighbor similarity searches using each of the 1,000 images in the ground-truth dataset as a query. In particular, the following six methods were compared:

1. Quantized descriptors based on Euclidean clustering. Cosine distance.
2. Quantized descriptors based on Mahalanobis clustering. Cosine distance.
3. Full descriptors. Minimum distance measure using Euclidean distance.
4. Full descriptors. Minimum distance measure using Mahalanobis distance.
5. Full descriptors. Complete distance measure using Euclidean distance.
6. Full descriptors. Complete distance measure using Mahalanobis distance.

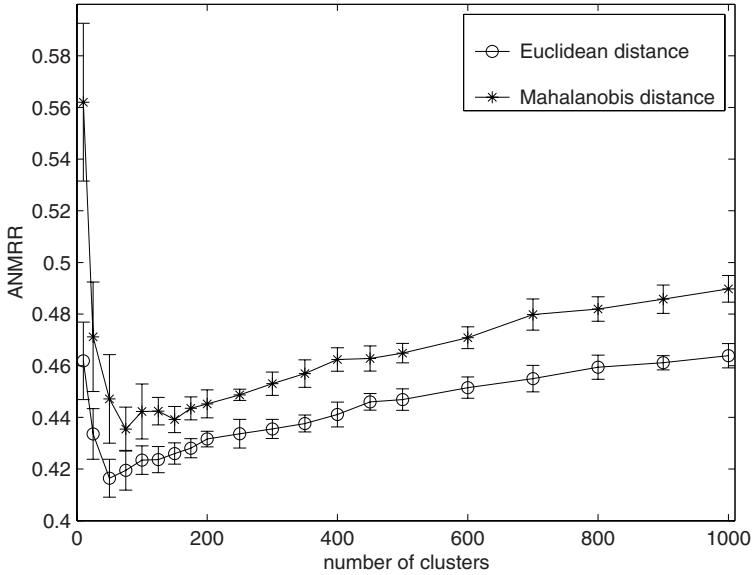
These methods are described in sections 4 and 5 and will be referred to by number in the rest of the paper.

Similarity retrieval using the quantized descriptors was compared for cluster counts  $c$  ranging from 10 to 1000. The clustering was performed on 100,000 points selected at random from the large IKONOS images (a separate dataset from the ground-truth). We computed the average ANMRR over the ten ground-truth classes. This was done ten times for each value of  $c$  since the clustering process is not deterministic (it is initialized with random centroids and is applied to a random set of points). Figure 3 shows the ANMRR values for different numbers of clusters. Error bars show the first standard deviation computed over the ten trials for each  $c$ . Again, ANMRR values 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. We make two conclusions from the results in Figure 3. One, that it is better to quantize the descriptors using Euclidean  $k$ -means clustering; and two, that the optimal number of clusters is 50. We use this optimal configuration in the remaining comparisons.

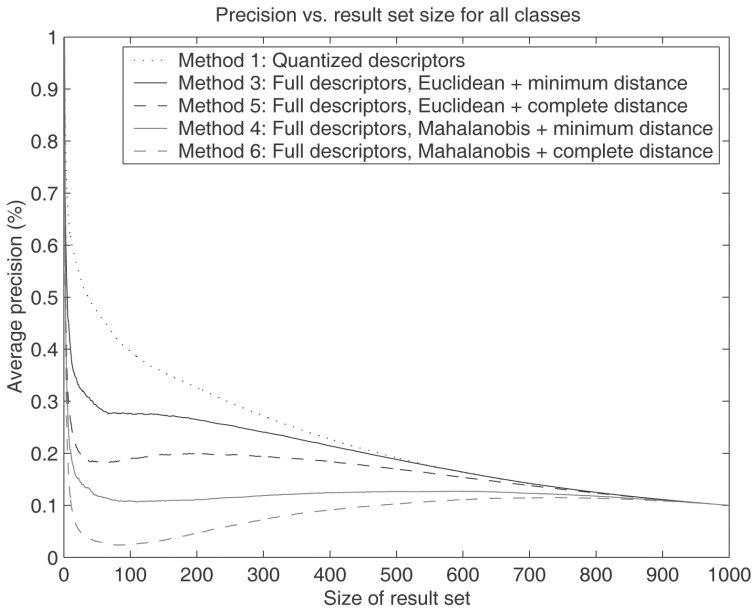
Figure 4 plots precision (the percent of correct retrievals) versus result set size for the different methods. These values are the average over all 1,000 queries. Quantized descriptors are shown to outperform full descriptors for all result set sizes. The minimum distance measure is shown to outperform the complete distance measure for comparing sets of full descriptors. Finally, as above, Euclidean distance is shown to outperform Mahalanobis distance, this time when used for full descriptor-to-descriptor comparison. Table 1 lists the ANMRR values for the specific image categories. The values are the average over all 100 queries in each category. These results confirm that the quantized descriptors outperform the full descriptors on average. It is interesting to note, however, that no single method performs best for all categories.

Finally, it is worth comparing the computational complexity of the different methods. On average, the 1,000 queries took approximately 2 seconds using the quantized descriptors, approximately 10 hours using the minimum distance measure for sets of full descriptors, and approximately 14 hours using the complete distance measure for sets of full descriptors. This significant difference results from the combinatorial expansion of comparing sets of descriptors and the cost





**Fig. 3.** Retrieval performance of descriptors quantized using  $k$ -means clustering for different numbers of clusters  $c$ . Shown for clustering with Euclidean and Mahalanobis distances. Image-to-image similarity is computed using the cosine distance measure.



**Fig. 4.** Retrieval performance in terms of precision versus size of result set

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

Ground-truth	Method 1	Method 3	Method 4	Method 5	Method 6
Aqueduct	<b>0.488</b>	0.655	0.573	0.621	0.577
Commercial	<b>0.575</b>	0.668	0.703	0.761	0.896
Dense residential	0.432	<b>0.412</b>	0.795	0.670	0.959
Desert chaparral	0.015	<b>0.002</b>	0.062	0.003	0.493
Forest	0.166	<b>0.131</b>	0.764	0.338	0.940
Freeway	0.497	0.384	<b>0.290</b>	0.401	0.307
Intersection	<b>0.420</b>	0.435	0.672	0.675	0.953
Parking lot	0.314	0.361	0.526	<b>0.301</b>	0.617
Road	0.680	0.494	<b>0.417</b>	0.660	0.680
Rural residential	<b>0.460</b>	0.592	0.833	0.706	0.943
Average	<b>0.405</b>	0.413	0.563	0.514	0.736

of full descriptor-to-descriptor comparisons in the 128 dimension feature space. Conversely, comparing two images using quantized features only requires a single cosine distance computation. These timings were measured on a typical workstation. No distinction is made between using Euclidean and Mahalanobis distances since the latter is implemented by transforming the feature vectors before performing the queries.

## 9 Discussion

We reach the following conclusions based on the results above. Similarity retrieval using quantized interest point descriptors is more effective and significantly more efficient than using full descriptors. This is true regardless of how the sets of full descriptors for the images are matched—minimum or complete—and how the individual descriptors are compared—Euclidean or Mahalanobis. This finding is initially a bit surprising. One might expect the loss of information from quantizing the descriptors to reduce performance. However, it seems that a binary comparison between quantized descriptors is more effective than an exact (Euclidean or Mahalanobis) comparison between full descriptors. The cosine distance can be viewed as comparing sets of descriptors in which individual descriptors are matched if they are quantized to the same cluster. The exact distance between descriptors does not matter, only that they are in some sense closer to each other than they are to other descriptors. This actually agrees with how interest point descriptors are used to determine correspondences between stereo pairs [20]. It is not the exact distance between a pair of descriptors that is used to assign a point in one image to a point in another but the ratio of this distance to that of the next closest point.

We showed that the optimal number of clusters used to quantize the descriptors seems to be around 50. This is lower than we expected. Other researchers [14] found that a much larger number of clusters, on the order of thousands,

performed better for matching objects in videos. While our application is different it would be interesting to investigate this further. This finding is significant as a coarser quantization supports higher scalability since it results in reduced feature representation and faster similarity comparison.

We found that using the Euclidean distance to compare descriptors is better than the Mahalanobis distance. This is true for using  $k$ -means clustering to construct the quantization space. It is also true for computing individual descriptor-to-descriptor distances when comparing sets of full descriptors. This results from the distribution of the descriptors in the 128 dimension space. This again differs from the findings of other researchers [14] who used the Mahalanobis distance to cluster descriptors. It is not clear, however, if the Euclidean distance was considered or if it was just assumed that removing correlations and scale would improve the quantization induced by the clustering.

We discovered that when comparing sets of full descriptors, it is better to allow many-to-one matches; that is, the minimum distance measure outperformed the complete distance measure. This agrees conceptually with the superior performance of the quantized descriptors. The cosine distance used to compare quantized descriptors “allows” multiple matches.

Finally, we found that no method performed the best for all image classes. This requires additional investigation perhaps with a simpler, more homogeneous ground-truth dataset. Preliminary observations suggest that some methods are better at discriminating visually similar classes than others. In particular, the Mahalanobis distance measure seems better than the Euclidean distance measure at distinguishing the aqueduct, freeway and road classes which are very similar visually. We plan to investigate this further.

## References

1. Newsam, S., Yang, Y.: Comparing global and interest point descriptors for similarity retrieval in remote sensed imagery. In: ACM International Symposium on Advances in Geographic Information Systems (ACM GIS) (2007)
2. Ashley, J., Flickner, M., Hafner, J., Lee, D., Niblack, W., Petkovic, D.: The query by image content (QBIC) system. In: ACM SIGMOD International Conference on Management of Data (1995)
3. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. In: Penn State University Technical Report CSE 06-009 (2006)
4. Bretschneider, T., Cavet, R., Kao, O.: Retrieval of remotely sensed imagery using spectral information content. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, pp. 2253–2255 (2002)
5. Bretschneider, T., Kao, O.: A retrieval system for remotely sensed imagery. In: International Conference on Imaging Science, Systems, and Technology, vol. 2, pp. 439–445 (2002)
6. Ma, A., Sethi, I.K.: Local shape association based retrieval of infrared satellite images. In: IEEE International Symposium on Multimedia (2005)
7. Li, Y., Bretschneider, T.: Semantics-based satellite image retrieval using low-level features. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, vol. 7, pp. 4406–4409 (2004)

8. Hongyu, Y., Bicheng, L., Wen, C.: Remote sensing imagery retrieval based-on Gabor texture feature classification. In: International Conference on Signal Processing, pp. 733–736 (2004)
9. Manjunath, B.S., Ma, W.Y.: Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 18, 837–842 (1996)
10. Newsam, S., Wang, L., Bhagavathy, S., Manjunath, B.S.: Using texture to analyze and manage large collections of remote sensed image and video data. *Journal of Applied Optics: Information Processing* 43, 210–217 (2004)
11. Newsam, S., Kamath, C.: 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)
12. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. *International Journal of Computer Vision* 65, 43–72 (2005)
13. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27, 1615–1630 (2005)
14. Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: *IEEE International Conference on Computer Vision*, vol. 2, pp. 1470–1477 (2003)
15. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19, 530–535 (1997)
16. Wang, J., Zha, H., Cipolla, R.: Combining interest points and edges for content-based image retrieval. In: *IEEE International Conference on Image Processing*, pp. 1256–1259 (2005)
17. Wolf, C., Kropatsch, W., Bischof, H., Jolion, J.M.: Content based image retrieval using interest points and texture features. *International Conference on Pattern Recognition* 4, 4234 (2000)
18. Ledwich, L., Williams, S.: Reduced SIFT features for image retrieval and indoor localisation. In: *Australasian Conference on Robotics and Automation* (2004)
19. Lowe, D.G.: Object recognition from local scale-invariant features. In: *IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157 (1999)
20. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110 (2004)
21. Kuhn, H.W.: The Hungarian Method for the assignment problem. *Naval Research Logistic Quarterly* 2, 83–97 (1955)
22. Hand, D., Mannila, H., Smyth, P.: *Principles of Data Mining*. The MIT Press, Cambridge (2001)
23. Manjunath, B.S., Salembier, P., Sikora, T. (eds.): *Introduction to MPEG7: Multimedia Content Description Interface*. John Wiley & Sons, Chichester (2002)