Using Texture in Image Similarity and Retrieval

Selim Aksoy and Robert M. Haralick
Intelligent Systems Laboratory
Department of Electrical Engineering
University of Washington
Seattle, WA 98195-2500
{aksoy,haralick}@isl.ee.washington.edu

Abstract. Texture has been one of the most popular representations in image retrieval. Our image database retrieval system uses two sets of textural features, first one being the line-angle-ratio statistics which is a texture histogram computed from the properties of the surroundings and the spatial relationships of intersecting lines, second one being the variances of gray level spatial dependencies computed from co-occurrence matrices. This paper also discusses a line selection algorithm to eliminate insignificant lines and statistical feature selection methods to select the best performing subset of features. Average precision is used to evaluate the retrieval performance in comparative tests with three other texture analysis algorithms. Results show that our method is fast and effective with an average precision of 0.73 when 12 images are retrieved.

1 Introduction

Image databases are becoming increasingly popular due to large amount of images that are generated by various applications and the advances in computer technology. Initial work on content-based retrieval focused on using low-level approaches like color and texture [7, 12]. More recent approaches use region-based methods [4, 9] to infer higher-level information from images but the region segmentation algorithms are still too slow to be used in an image retrieval application.

In this paper we attempt to improve retrieval efficiency using easy-to-compute low-level features that combine macro and micro aspects of the texture in the image. The first feature extraction method is the line-angle-ratio statistics, which is a texture histogram method and a macro texture measure that uses spatial relationships between lines as well as the properties of their surroundings. A statistical line selection algorithm to eliminate insignificant lines is presented. The second feature extraction algorithm, variances of gray level spatial dependencies, which in turn is a micro texture measure that uses second-order (co-occurrence) statistics of gray levels of pixels at particular spatial relationships. Both sets of features are integrated for a multi-scale texture analysis which is crucial for a compact representation, especially for large databases containing different types of complex images.

We use a two-class pattern classification approach to find statistical measures of how well some of the features perform better than others to avoid having less significant or even redundant features that increase computation but contribute very little in the decision process. Retrieval performance is evaluated using average precision computed for a groundtruth data set.

The rest of the paper is organized as follows. First, textural features are presented in Sections 2, 3 and 4. Then, feature selection methods are described in Section 5. Experiments and results are discussed in Section 6. Finally, conclusions are given in Section 7.

2 Line-Angle-Ratio Statistics

Experiments on various types of images showed us that one of the strongest spatial features of an image is its line segments. Edge and line information have been extensively used in both very early and recent approaches to texture. Our algorithm is composed of two stages; pre-processing and texture histogram generation.

2.1 Pre-processing

Each image is processed by an edge detector, an edge linker, a line selection operator and a line grouping operator to detect line pairs to associate with it a set of feature records. Edge detection followed by line detection often results in many false alarms. It is especially hard to select proper parameters for these operators if one does not have groundtruth information as training data. The algorithm we developed to eliminate lines that do not have significant difference between the gray level distributions in the regions on their right and left is given below.

Let the set of N gray levels $x_1, x_2, \ldots, x_N$ be considered as iid $N(\mu_x, \sigma_x^2)$ samples from the region to the right of a line and the set of M gray levels $y_1, y_2, \ldots, y_M$ be considered as iid $N(\mu_y, \sigma_y^2)$ samples from the region to the left of that line. Define $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$, $\bar{y} = \frac{1}{M}\sum_{m=1}^{M} y_m$, and $z = \bar{x} - \bar{y} \quad \sim \quad N(\mu_z, \sigma_z^2) = N\left(\mu_x - \mu_y, \frac{\sigma_x^2}{N} + \frac{\sigma_y^2}{M}\right)$. Define the null hypothesis as $H_0 : \mu_x = \mu_y \,\text{and}\, \sigma_x = \sigma_y$ which means both sets of gray levels come from the same distribution, and the alternative hypothesis as $H_1 : \mu_x \neq \mu_y \,\text{and}\, \sigma_x \neq \sigma_y$. To form the test statistic, define two random variables A and B as

$$A = \left(\frac{z - \mu_z}{\sigma_z}\right)^2 \quad \sim \quad \chi_1^2 \tag{1}$$

and

$$B = \frac{1}{N-1}\sum_{n=1}^{N}\left(\frac{x - \bar{x}}{\sigma_x}\right)^2 + \frac{1}{M-1}\sum_{m=1}^{M}\left(\frac{y - \bar{y}}{\sigma_y}\right)^2 \quad \sim \quad \chi_{N+M-2}^2. \tag{2}$$

Then, define the test statistic $F = \frac{A/1}{B/(N+M-2)}$ which, under the null hypothesis, becomes

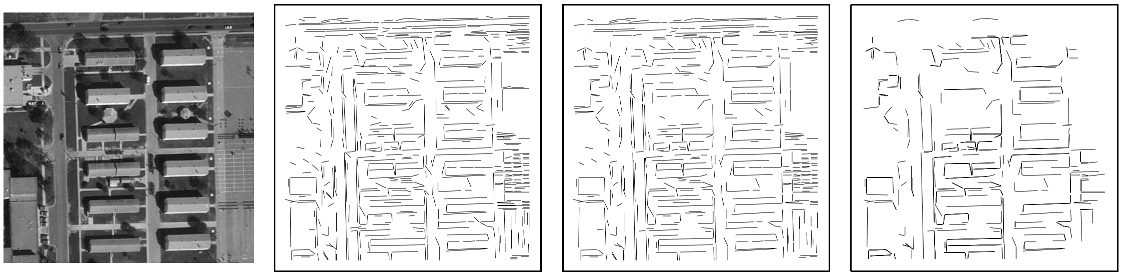
$$F = \frac{z^2 \frac{(N+M-2)}{(\frac{1}{N}+\frac{1}{M})}}{\frac{1}{(N-1)}\sum_{n=1}^{N}(x-\bar{x})^2 + \frac{1}{(M-1)}\sum_{m=1}^{M}(y-\bar{y})^2} \quad \sim \quad F_{1,N+M-2}. \tag{3}$$

Given a threshold for the F-value, if the null hypothesis H_0 is true, the line is rejected, if the alternative hypothesis H_1 is true, the line is accepted as a significant one.

After obtaining relatively significant lines, we use a line grouping operator to find intersecting and/or near-intersecting line pairs. Examples for the pre-processing steps are given in Figure 1.

2.2 Texture Histogram

The features for each pair of intersecting line segments consist of the angle between two lines and the ratio of mean gray level inside the region spanned by that angle to the mean gray level outside that region. Angle values are in the range $[0°, 180°]$. Since the possible range of ratio values is infinite, we restrict them to the range $[0, 1)$ by taking the reciprocal if the inner region is brighter than the outer region.



(a) Grayscale image.

(b) Extracted lines after line detection operator.

(c) Accepted lines after line selection operator.

(d) Resulting lines after line grouping operator.

Figure 1: Line selection and grouping pre-processing steps.

The final features form a two-dimensional space of angles and the corresponding ratios, which is then partitioned into a fixed set of Q cells. The feature vector for each image is then the Q-dimensional vector which has for its q'th component the number of angle-ratio pairs that fall into that q'th cell. These features do not have a uniform distribution so we use vector quantization to form the Q-cell partition. Please refer to [1] for details.

3 Variances of Gray Level Spatial Dependencies

Gray level spatial dependencies combine structural texture analysis algorithms which use the idea that texture is composed of primitives with different properties appearing in particular spatial arrangements and statistical approaches which try to model texture using statistical distributions either in the spatial domain or in a transform domain. This information can be summarized in two-dimensional co-occurrence matrices that are matrices of relative frequencies $P(i, j; d, \theta)$ with which two pixels separated by distance d at orientation θ occur in the image, one with gray level i and the other with gray level j. The initial work on co-occurrence matrices [8] and some comparative studies [6, 11] showed that gray level spatial dependency matrices were very successful in discriminating images with relatively homogeneous textures.

3.1 Pre-processing

Before computing co-occurrence matrices we use equal probability quantization to make the features invariant to distortions resulting in monotonic gray level transformations. We use 64 quantization levels (N_g) which performed the best among 16, 32 and 64 levels in terms of "total cost" that will be defined in Section 5.

3.2 Co-occurrence Variance

In order to use the information contained in the gray level co-occurrence matrices, Haralick *et al.* [8] defined 14 statistical measures. Since many distances and orientations result in a large amount of computation, we decided to use only the variance [2]

$$v(d, \theta) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i-j)^2 P(i, j; d, \theta) \tag{4}$$

which is a difference moment of P and measures the contrast in the image. It will have a large value for images which have a large amount of local spatial variation in gray levels and a smaller value for images with spatially uniform gray level distributions.

4 Multi-Scale Texture Analysis

Line-angle-ratio features capture the global spatial organization in an image by using relative orientations of lines extracted from it; therefore, they can be regarded as a macro-texture measure but are not effective if the image does not have any line content. On the other hand, co-occurrence variances capture local spatial variations of gray levels in the image; therefore, they are effective if the image is dominated by a fine, coarse, directional, or repetitive texture and can be regarded as a micro-texture measure.

In order to take advantage of both methods, we normalize each component of these feature vectors to the $[0,1]$ range by an equal probability quantization and append them to form the final vector. In the rest of the paper size of a feature vector will be denoted by Q.

5 Feature Selection

In many complex feature extraction algorithms there are many parameters that, when varied, result in a large number of possible feature measurements. These high dimensional feature spaces may cause a problem of having less significant or even redundant features that increase computation but contribute very little in the decision process. Most of the times this feature selection process is done heuristically. Only a few researchers presented some form of feature selection in their papers on database retrieval [10, 4].

We use a two-class pattern classification approach to find statistical measures of how well some of the features perform better than others. In doing so, we define two classes, the relevance class $\mathcal{A}$ and the irrelevance class $\mathcal{B}$, in order to classify image pairs as similar or dissimilar. Assume that we are given two sets of image pairs for the relevance and irrelevance classes respectively [3, 1]. Differences of feature vectors for each image pair are assumed to have a normal distribution [1] and sample means $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$ and sample covariance matrices $\Sigma_{\mathcal{A}}$ and $\Sigma_{\mathcal{B}}$ are estimated using the training data.

5.1 Classification Tests

Given a groundtruth image pair (n,m) with Q-dimensional feature vectors $x^{(n)}$ and $y^{(m)}$ respectively, first the difference $d = x^{(n)} - y^{(m)}$ is computed. From Bayes' law, the probability that these images are relevant is $P(\mathcal{A}|d) = P(d|\mathcal{A})P(\mathcal{A})/P(d)$ and that they are irrelevant is $P(\mathcal{B}|d) = P(d|\mathcal{B})P(\mathcal{B})/P(d)$. The image pair is assigned to the relevance class if $P(\mathcal{A}|d) > P(\mathcal{B}|d)$, and to the irrelevance class otherwise. Assuming that two classes are equally likely, taking the natural logarithm of the decision rule and eliminating some constants give

$$(d-\mu_{\mathcal{A}})'\Sigma_{\mathcal{A}}^{-1}(\mathrm{d}-\mu_{\mathcal{A}})/2 < (\mathrm{d}-\mu_{\mathcal{B}})'\Sigma_{\mathcal{B}}^{-1}(\mathrm{d}-\mu_{\mathcal{B}})/2 + \ln\frac{|\Sigma_{\mathcal{B}}|^{1/2}}{|\Sigma_{\mathcal{A}}|^{1/2}}. \tag{5}$$

Therefore, if the difference d of the feature vectors of two images satisfy the inequality in (5), this image pair is assigned to the relevance class, otherwise it is assigned to the irrelevance class.

[1] According to our observations, the line-angle-ratio feature differences follow double-exponential distributions and the co-occurrence feature differences follow normal distributions. Modeling the joint feature differences using a multivariate normal density worked better than using independently fitted double-exponentials or normals because of the covariance matrix that captures the correlation between features.

5.2 Experimental Set-up

Suitable measures for the classification results are misdetection and false alarm. In content-based retrieval we are more concerned with misdetection because we want to retrieve all the images similar to the query image. False alarm rate is also important because the purpose of querying a database is to retrieve similar images only, not all of them. We define total cost as 3 misdetection and 2 false alarm and use it as the criterion for "goodness", i.e. if a subset of features has a small total cost compared to others, it is called "good".

If the dimension of the feature space is large, it is computationally too expensive to do classification tests using all possible subsets of the features. In our work, first, we do tests using only one of the features at a time. The second test, which shrinks down feature sets, is done by first computing the total cost using all Q features. The feature with the worst total cost is discarded and the total cost using the remaining Q-1 features is computed. This procedure continues until one feature is left. A third test, which builds up feature sets, is done by starting with the total cost for each individual feature and selecting the best one. Given this best one, pairs of features are formed using one of the remaining features and this best feature. Total cost is computed for each pair and the one having the smallest cost is selected. Given the best two features, next, triplets of features are formed using one of the remaining features and these two best features. This procedure continues until all or a preselected number of features are used. These tests do not guarantee the optimal subset of features but allow us to select a suboptimal subset without doing an exhaustive search.

6 Experiments and Results

6.1 Feature Selection

The images in our database are obtained from the Fort Hood Data of the RADIUS Project and also from the LANDSAT and Defense Meteorological Satellite Program (DMSP) Satellites. The test database for feature selection contains 10,410 256×256 images with a total of 38,240 groundtruth image pairs. Therefore, experiments for each parameter combination tested consist of classifying approximately 38,000 image pairs.

6.1.1 Line-Angle-Ratio Statistics

The goal of these feature selection tests is to select the quantizer that performs the best. The quantizers with 15, 20 and 25 cells resulted in 30.20%, 30.05% and 30.22% total costs respectively. As a result, we decided to use the quantizer with 20 cells.

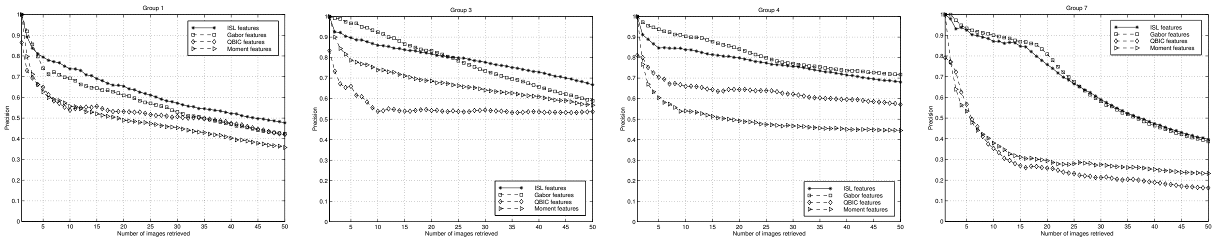
6.1.2 Co-occurrence Variances

The goal of our feature selection tests is to select the set of distances, among distances of 1 to 20 pixels, that perform the best according to the classification criteria. In the experiments, building up feature sets decreased the total cost faster than shrinking down the set of all features. Another observation was that after using approximately 2 or 3 distances, total cost did not decrease much. As a result, using the distances 1 and 20 together had the minimum total cost of 29.36% among all the possible combinations of 2 distances. Although these feature selection tests do not guarantee an optimal solution, they resulted in a suboptimal one in 1,560 classification tests without using exhaustive search which would then require $2^{20} - 1$ classification tests. Details of these experiments can be found in [1].

6.2 Retrieval Performance

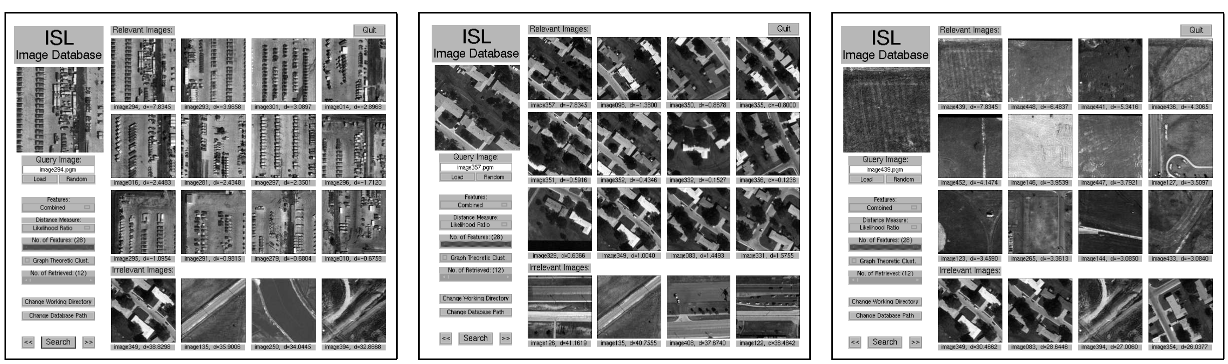
For these tests, we randomly selected 340 images from the total of 10,410 and formed a groundtruth of 7 categories; parking lots, roads, residential areas, landscapes, LANDSAT USA, DMSP North Pole and LANDSAT Chernobyl. Likelihood values [1] which were derived from equation (5) were used to rank the database images. For comparison, IBM's QBIC texture features [7], UCSB's Gabor texture features [10] and TUT's moments texture features [5] were also tested with Euclidean distance as the distance measure. Our features performed similarly to the Gabor features and both of them performed significantly better than others. Precision averaged over all 340 images was 0.73 when 12 images were retrieved. Figure 2 shows the average precision for some of the groundtruth groups. The feature extraction time for our features were approximately 30 times faster than that of the Gabor features.

Some example queries are shown in Figure 3. More examples and our groundtruth data set can be found at http://isl.ee.washington.edu/∼aksoy/research/database.shtml.



(a) Parking lots (b) Residential (c) Landscape (d) Chernobyl

Figure 2: Average precision for some groundtruth groups.



(a) Example query for parking lots (b) Example query for residential (c) Example query for landscape

Figure 3: Retrieval examples with the upper left image as the query. Among the retrieved images, first three rows show the 12 most relevant images in descending order of similarity and the last row shows the 4 most irrelevant images in descending order of dissimilarity.

7 Conclusions

We described easy-to-compute but effective low-level textural features. The first set of features captures the global spatial organization in the image using the edge and line information. The second set of features is effective if the image is dominated by a fine, coarse, directional, or repetitive texture. Some key aspects of this work include a statistical line selection algorithm and feature selection tests to determine the parameters of the feature extraction algorithms.

Retrieval tests showed that our features performed better than the QBIC and moments features and had similar performance as the Gabor features. They can be combined with other features to further improve the performance and make better inferences about the high-level descriptions of the images.

Acknowledgments

The authors would like to thank Tom Minka from MIT Media Laboratory for the code for QBIC texture features, Bogdan Cramariuc from Tampere University of Technology for the code for moment texture features, and Wei-Ying Ma from the University of California, Santa Barbara for the code for Gabor texture features.

References

[1] S. Aksoy. Textural features for content-based image database retrieval. Master's thesis, University of Washington, Seattle, WA, June 1998. Online: http://isl.ee.washington.edu/∼aksoy/thesis.shtml.

[2] S. Aksoy and R. M. Haralick. Content-based image database retrieval using variances of gray level spatial dependencies. In *Proceedings of IAPR International Workshop on Multimedia Information Analysis and Retrieval*, pages 3–19, Hong Kong, August 13–14 1998. as Lecture Notes in Computer Science, vol. 1464.

[3] S. Aksoy and R. M. Haralick. Textural features for image database retrieval. In *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries, in conjunction with CVPR'98*, pages 45–49, Santa Barbara, CA, June 1998.

[4] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Region-based image querying. In *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997.

[5] F. A. Cheikh *et al.* MUVIS: A system for content-based indexing and retrieval in large image databases. In *SPIE Storage and Retrieval of Image and Video Databases VII*, pages 98–106, San Jose, CA, January 1999.

[6] R. W. Conners and C. A. Harlow. A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(3):25–39, May 1980.

[7] M. Flickner *et al.* The QBIC project: Querying images by content using color, texture and shape. In *SPIE Storage and Retrieval of Image and Video Databases*, pages 173–181, 1993.

[8] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, November 1973.

[9] W. Y. Ma and B. S. Manjunath. NETRA: A toolbox for navigating large image databases. In *Proceedings of IEEE International Conference on Image Processing*, 1997.

[10] 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):837–842, August 1996.

[11] P. P. Ohanian and R. C. Dubes. Performance evaluation for four classes of textural features. *Pattern Recognition*, 25(8):819–833, August 1992.

[12] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. In *SPIE Storage and Retrieval of Image and Video Databases II*, pages 34–47, February 1994.