Probabilistic vs. Geometric Similarity Measures for Image Retrieval

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

Abstract

Similarity between images in image retrieval is measured by computing distances between feature vectors. This paper presents a probabilistic approach and describes two likelihood-based similarity measures for image retrieval. Popular distance measures like the Euclidean distance implicitly assign more weighting to features with large ranges than those with small ranges. First, we discuss the effects of five feature normalization methods on retrieval performance. Then, we show that the probabilistic methods perform significantly better than geometric approaches like the nearest neighbor rule with city-block or Euclidean distances. They are also more robust to normalization effects and using better models for the features improves the retrieval results compared to making only general assumptions. Experiments on a database of approximately 10,000 images show that studying the feature distributions are important and this information should be used in designing feature normalization methods and similarity measures.

1. Introduction

Image database retrieval has become a very popular research area in recent years [15]. Initial work on content-based retrieval [8, 12, 10] focused on using low-level features like color and texture for image representation. After features are computed for all images in the database, similarity measures are used to find matches between images.

Feature vectors usually exist in a very high dimensional space. Due to this high dimensionality, their parametric characterization is usually not studied. A commonly used assumption is that images that are close to each other in the feature space are also visually similar. In geometric similarity measures like the nearest neighbor rule, no assumption is made about the probability distribution of the features and similarity is based on the distances between feature vectors in the feature space. Given this, Euclidean distance has been the most widely used distance measure [8, 12, 9, 18], as well as the weighted Euclidean distance [4, 16], city-block (L_1) distance [10, 18], the general L_p Minkowsky distance [17] and the Mahalanobis distance [12, 18]. The L_1 distance was also used under the name "histogram intersection" [18]. Polynomial combinations of predefined distance measures were also used to create new distance measures [5].

This paper presents a probabilistic approach for image retrieval. We describe two likelihood-based similarity measures that compute the likelihood of two images, one being the query image and the other one being an image in the database, being similar or dissimilar. First, we define two classes, the relevance class and the irrelevance class, and then the likelihood values are derived from a Bayesian classifier. We use two different methods to estimate the conditional probabilities used in the classifier. The first method uses a multivariate Normal assumption and the second one uses independently fitted distributions for each feature. The performances of these two methods are compared to the performances of geometric approaches that use the city-block (L_1) and Euclidean (L_2) distances as similarity measures.

An important step between feature extraction and distance computation is feature normalization. Complex image database retrieval systems use features that are generated by many different feature extraction algorithms and not all of these features have the same range. Popular distance measures, for example the Euclidean distance, implicitly assign more weighting to features with large ranges than those with small ranges. This paper discusses five normalization methods; linear scaling to unit range, linear scaling to unit variance, transformation to a Uniform[0,1] random variable, rank normalization and normalization by fitting distributions. Experiments are done on a database of approximately 10,000 images and average precision is used to evaluate performances of both the normalization methods and the similarity measures.

The rest of the paper is organized as follows. First, the features that we use in this study are summarized in Section 2. Then, the feature normalization methods are described in Section 3 and are followed by the similarity measures in Section 4. Experiments and results are discussed in Section 5. Finally, conclusions are given in Section 6.

2. Feature Extraction

Textural features that were described in detail in [2, 3] are used for image representation in this paper. The first

set of features are the line-angle-ratio statistics that use a texture histogram computed from the spatial relationships between lines as well as the properties of their surroundings. The second set of features are the variances of gray level spatial dependencies that use second-order (co-occurrence) statistics of gray levels of pixels in particular spatial relationships. Line-angle-ratio statistics result in a 20-dimensional feature vector and co-occurrence variances result in an 8-dimensional feature vector.

3. Feature Normalization

The following sections describe five normalization procedures. The goal is to make all features have approximately the same effect in the computation of similarity by independently normalizing each feature component to the $[0, 1]$ range.

3.1. Linear scaling to unit range

Given a lower bound l and an upper bound u for a feature component x,

$$x' = \frac{x - l}{u - l} \tag{1}$$

results in x' being in the $[0, 1]$ range.

3.2. Linear scaling to unit variance

Another normalization procedure is to transform the feature component x to a random variable with zero mean and unit variance as

$$x' = \frac{x - \mu}{\sigma} \tag{2}$$

where μ and σ^2 are the sample mean and the sample variance of that feature respectively. Under the Normality assumption, an additional shift and rescaling as

$$x' = \frac{\frac{x-\mu}{3\sigma} + 1}{2} \tag{3}$$

guarantees 99% of x' to be in the [0,1] range. We can then round off the out-of-range components to either 0 or 1.

3.3. Transformation to a Uniform[0,1] random variable

Given a random variable x with cumulative distribution function $F_x(x)$, the random variable x' resulting from the transformation $x' = F_x(x)$ will be uniformly distributed in the [0,1] range [11].

3.4. Rank normalization

Given the sample for a feature component for all images as $x_1, \ldots, x_n$, replacing each image's feature value by its corresponding normalized rank, i.e.

$$x'_i = \frac{\underset{x_1,\ldots,x_n}{\text{rank}}\ (x_i) - 1}{n - 1} \tag{4}$$

where x_i is the feature value for the i'th image, uniformly maps all feature values to the [0,1] range. When there are more than one image with the same feature value, especially after quantization, they are assigned the average rank for that value.

3.5. Normalization after fitting distributions

The transformation in Section 3.2 assumed that a feature has a Normal(μ, σ^2) distribution. The Mahalanobis distance [7] also involves normalization in terms of the covariance matrix but is also valid only when the features are Normally distributed. The sample values can be used to find better estimates for the feature distributions.

The following sections describe how to fit Normal, Lognormal, Exponential and Gamma densities to a random sample. After estimating the parameters of a distribution, the cut-off value that includes 99% of the feature values is found and the samples are scaled and truncated so that each feature component has the same range. To measure how well a fitted distribution resembles the sample data (goodness-of-fit), we use the Kolmogorov-Smirnov test statistic [6, 13].

3.5.1. Fitting a Normal(μ, σ^2) density

Let $x_1, \ldots, x_n \in \mathbf{R}$ be a random sample from a population with density $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$, $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$. The maximum likelihood estimators (MLE) of μ and σ^2 can be derived as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2. \tag{5}$$

The cut-off value δ_x that includes 99% of the feature values can be found as $\delta_x = \hat{\mu} + 2.4\hat{\sigma}$.

3.5.2. Fitting a Lognormal(μ, σ^2) density

Let $x_1, \ldots, x_n \in \mathbf{R}$ be a random sample from a population with density $\frac{1}{\sqrt{2\pi}\sigma} \frac{e^{-(\log x - \mu)^2/2\sigma^2}}{x}$, $x \geq 0$, $-\infty < \mu < \infty$, $\sigma > 0$. The MLEs of μ and σ^2 can be derived as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \log x_i \ \text{ and } \ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\log x_i - \hat{\mu})^2. \tag{6}$$

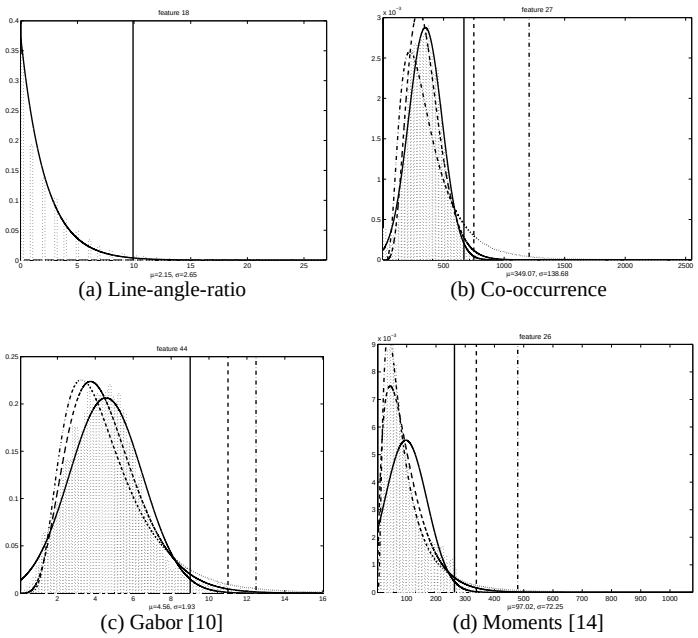
The cut-off value δ_x can be found as $\delta_x = e^{\hat{\mu}+2.4\hat{\sigma}}$.

3.5.3. Fitting an Exponential(λ) density

Let $x_1, \ldots, x_n \in \mathbf{R}$ be a random sample from a population with density $\frac{1}{\lambda} e^{-x/\lambda}$, $x \geq 0$, $\lambda \geq 0$. The MLE of λ is

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{7}$$

The cut-off value δ_x can be found as $\delta_x = -\hat{\lambda} \log 0.01$.



(a) Line-angle-ratio (b) Co-occurrence

(c) Gabor [10] (d) Moments [14]

Figure 1: Feature histograms and fitted distributions for example features. Exponential (solid line) model is used for line-angle-ratio features and Normal (solid line), Lognormal (dash-dot line) and Gamma (dashed line) models are used for the others. The vertical lines show the 99% cut-off point for each distribution.

3.5.4. Fitting a Gamma(α, β) density

Let $x_1, \ldots, x_n \in \mathbf{R}$ be a random sample from a population with density $\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, $x \geq 0$, $\alpha, \beta \geq 0$. The method of moments (MOM) estimators for α and β can be derived as

$$\hat{\alpha} = \frac{\overline{X}^2}{S^2} \quad \text{and} \quad \hat{\beta} = \frac{S^2}{\overline{X}} \tag{8}$$

where $\overline{X}$ and S^2 are the sample mean and the sample variance respectively. The cut-off value δ_x can be found using the Incomplete Gamma function $I_{\delta_x/\hat{\beta}}(\hat{\alpha})$ and can be computed numerically [1, 13].

Histograms and fitted distributions for example features are given in Figure 1. This shows that many features from different feature extraction algorithms can be modeled by the distributions that were presented in Section 3.5.

4. Similarity Measures

After computing and normalizing the feature vectors for all images in the database, given a query image, we have to decide which images in the database are relevant to it and have to retrieve the most relevant ones as the result of the query. In this section, we describe two approaches, probabilistic and geometric approaches to image similarity and retrieval.

4.1. Probabilistic similarity measures

In our previous work [2] we defined two classes, the relevance class $\mathcal{A}$ and the irrelevance class $\mathcal{B}$, and used a Gaussian classifier to measure the relevancy of two images one being the query image and one being a database image so that image pairs which had a high likelihood ratio were classified as relevant and the ones which had a lower likelihood ratio were classified as irrelevant. Given two images with feature vectors x and y, and their feature difference vector $d = x - y,\ x, y, d \in \mathbf{R}^{\mathrm{Q}}$ with Q being the size of a feature vector, the posterior probability that they are relevant is

$$P(\mathcal{A}|d) = P(d|\mathcal{A})P(\mathcal{A})/P(d) \tag{9}$$

and the posterior probability that they are irrelevant is

$$P(\mathcal{B}|d) = P(d|\mathcal{B})P(\mathcal{B})/P(d). \tag{10}$$

Assuming that these two classes are equally likely, the likelihood ratio is defined as

$$r(d) = \frac{P(d|\mathcal{A})}{P(d|\mathcal{B})}. \tag{11}$$

In the following sections, we describe two methods to estimate the conditional probabilities $P(d|\mathcal{A})$ and $P(d|\mathcal{B})$.

4.1.1. Multivariate Normal assumption

We assume that the feature differences for the relevance class have a multivariate Normal density with mean $\mu_{\mathcal{A}}$ and covariance matrix $\Sigma_{\mathcal{A}}$ as

$$f(d|\mu_{\mathcal{A}}, \Sigma_{\mathcal{A}}) = \frac{1}{(2\pi)^{\mathrm{Q}/2}|\Sigma_{\mathcal{A}}|^{1/2}}\, \mathrm{e}^{-(\mathrm{d}-\mu_{\mathcal{A}})'\Sigma_{\mathcal{A}}^{-1}(\mathrm{d}-\mu_{\mathcal{A}})/2}. \tag{12}$$

Similarly, the feature differences for the irrelevance class are assumed to have a multivariate Normal density with sample mean $\mu_{\mathcal{B}}$ and sample covariance matrix $\Sigma_{\mathcal{B}}$ as

$$f(d|\mu_{\mathcal{B}}, \Sigma_{\mathcal{B}}) = \frac{1}{(2\pi)^{\mathrm{Q}/2}|\Sigma_{\mathcal{B}}|^{1/2}}\, \mathrm{e}^{-(\mathrm{d}-\mu_{\mathcal{B}})'\Sigma_{\mathcal{B}}^{-1}(\mathrm{d}-\mu_{\mathcal{B}})/2}. \tag{13}$$

The likelihood ratio in (11) is given as

$$r(d) = \frac{f(d|\mu_{\mathcal{A}}, \Sigma_{\mathcal{A}})}{f(d|\mu_{\mathcal{B}}, \Sigma_{\mathcal{B}})}. \tag{14}$$

$\mu_{\mathcal{A}}, \Sigma_{\mathcal{A}}, \mu_{\mathcal{B}}$ and $\Sigma_{\mathcal{B}}$ are estimated using the multivariate versions of the MLEs given in Section 3.5.1.

To simplify the computation of the likelihood ratio in (14), we take its logarithm, eliminate some constants, and use

$$r(d) = (d - \mu_{\mathcal{A}})'\Sigma_{\mathcal{A}}^{-1}(\mathrm{d} - \mu_{\mathcal{A}}) - (\mathrm{d} - \mu_{\mathcal{B}})'\Sigma_{\mathcal{B}}^{-1}(\mathrm{d} - \mu_{\mathcal{B}}) \tag{15}$$

to rank the database images in ascending order of these values which corresponds to a descending order of similarity.

4.1.2. Independently fitted distributions

We also use the independently fitted distributions to compute the likelihood values. After comparing the Kolmogorov-Smirnov test statistics as the goodness-of-fits, we model the line-angle-ratio features by Exponential densities and the co-occurrence features by Normal densities. Let x and y be two iid. random variables with an Exponential(λ) distribution. The distribution of $z = x - y$ is called Double Exponential(λ) and can be found as

$$f_z(z) = \frac{1}{2\lambda} e^{-|z|/\lambda}, \quad -\infty < z < \infty. \tag{16}$$

Let x and y be two iid. random variables with a Normal(μ, σ^2) distribution. Using moment generating functions, we can easily show that their difference $z = x - y$ has a Normal($0, 2\sigma^2$) distribution

$$f_z(z) = \frac{1}{\sqrt{4\pi\sigma^2}} e^{-z^2/4\sigma^2}, \quad -\infty < z < \infty. \tag{17}$$

Using this Double Exponential model for the 20 line-angle-ratio feature differences and the Normal model for the 8 co-occurrence feature differences, the joint density for the relevance class is given as

$$f(d|\lambda_{\mathcal{A}1}, \ldots, \lambda_{\mathcal{A}20}, \sigma^2_{\mathcal{A}21}, \ldots, \sigma^2_{\mathcal{A}28}) = \prod_{i=1}^{20} \frac{1}{2\lambda_{\mathcal{A}i}} e^{-|d_i|/\lambda_{\mathcal{A}i}} \prod_{i=21}^{28} \frac{1}{\sqrt{4\pi\sigma^2_{\mathcal{A}i}}} e^{-d_i^2/4\sigma^2_{\mathcal{A}i}} \tag{18}$$

and the joint density for the irrelevance class is given as

$$f(d|\lambda_{\mathcal{B}1}, \ldots, \lambda_{\mathcal{B}20}, \sigma^2_{\mathcal{B}21}, \ldots, \sigma^2_{\mathcal{B}28}) = \prod_{i=1}^{20} \frac{1}{2\lambda_{\mathcal{B}i}} e^{-|d_i|/\lambda_{\mathcal{B}i}} \prod_{i=21}^{28} \frac{1}{\sqrt{4\pi\sigma^2_{\mathcal{B}i}}} e^{-d_i^2/4\sigma^2_{\mathcal{B}i}}. \tag{19}$$

The likelihood ratio in (11) becomes

$$r(d) = \frac{f(d|\lambda_{\mathcal{A}1}, \ldots, \lambda_{\mathcal{A}20}, \sigma^2_{\mathcal{A}21}, \ldots, \sigma^2_{\mathcal{A}28})}{f(d|\lambda_{\mathcal{B}1}, \ldots, \lambda_{\mathcal{B}20}, \sigma^2_{\mathcal{B}21}, \ldots, \sigma^2_{\mathcal{B}28})}. \tag{20}$$

$\lambda_{\mathcal{A}i}, \lambda_{\mathcal{B}i}, \sigma^2_{\mathcal{A}i}$ and $\sigma^2_{\mathcal{B}i}$ are estimated using the MLEs given in Sections 3.5.3 and 3.5.1. Instead of computing the complete likelihood ratio, we take its logarithm, eliminate some constants, and use

$$r(d) = \sum_{i=1}^{20} |d_i| \left(\frac{1}{\lambda_{\mathcal{A}i}} - \frac{1}{\lambda_{\mathcal{B}i}} \right) + \frac{1}{4} \sum_{i=21}^{28} \left[\frac{d_i^2}{\sigma^2_{\mathcal{A}i}} - \frac{d_i^2}{\sigma^2_{\mathcal{B}i}} \right] \tag{21}$$

to rank the database images.

4.2. Geometric similarity measures

In the geometric similarity measures for retrieval, similarity between images is measured by computing distances between feature vectors in the feature space. In the well known nearest neighbor decision rule, each image in the database is assumed to be represented by its feature vector y in the Q-dimensional feature space. Given the feature vector x for the input query, the goal is to find the y's which are the closest neighbors of x according to a distance measure. Then, the k-nearest neighbors of x will be retrieved as the most relevant images to x. For the distance metric ρ, we use the city-block distance (Minkowsky L_1 metric)

$$\rho(x, y) = \sum_{q=1}^{\mathrm{Q}} |x_q - y_q| \tag{22}$$

and the Euclidean distance (Minkowsky L_2 metric)

$$\rho(x, y) = \left(\sum_{q=1}^{\mathrm{Q}} (x_q - y_q)^2 \right)^{1/2} \tag{23}$$

where $x, y \in \mathbf{R}^{\mathrm{Q}}$ and x_q and y_q are the q'th components of the feature vectors x and y respectively.

5. Experiments and Results

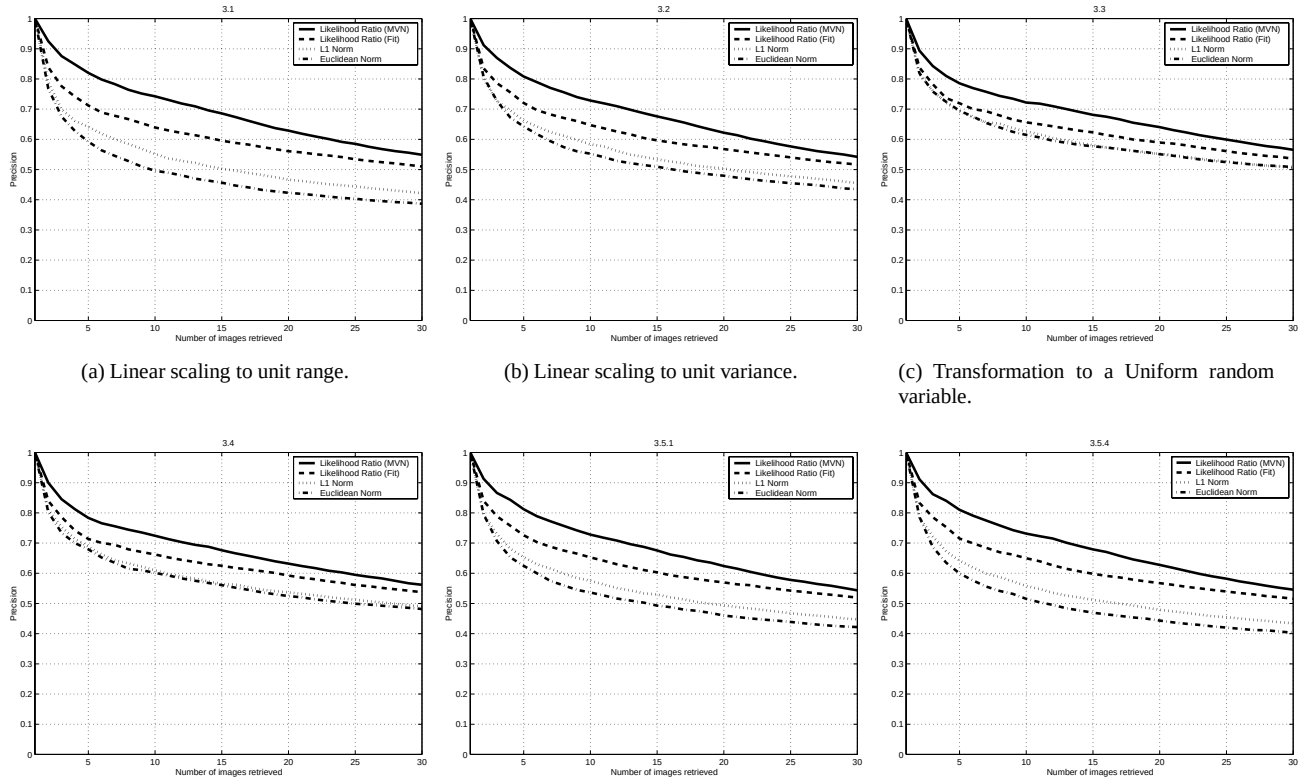
5.1. Database population

Our database contains 10,410 256×256 images that come from the Fort Hood Data of the RADIUS Project and also from the LANDSAT and Defense Meteorological Satellite Program (DMSP) Satellites. The RADIUS images consist of visible light aerial images of the Fort Hood area in Texas, USA. The LANDSAT images are from a remote sensing image collection. For these experiments, 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.

5.2. Retrieval performance

Retrieval results, in terms of precision averaged over the groundtruth images, using the likelihood ratio with multivariate Normal assumption, the likelihood ratio with fitted distributions, the city-block distance and the Euclidean distance with different normalization methods are given in Figure 2. Note that, linear scaling to unit range involves only scaling and translation and it does not have any truncation so it does not change the structures of distributions of the features. Therefore, using this method reflects the effects of using the raw feature distributions while mapping them to the same range.

Example queries using different similarity measures with the same query image are given in Figure 3.



(a) Linear scaling to unit range.

(b) Linear scaling to unit variance.

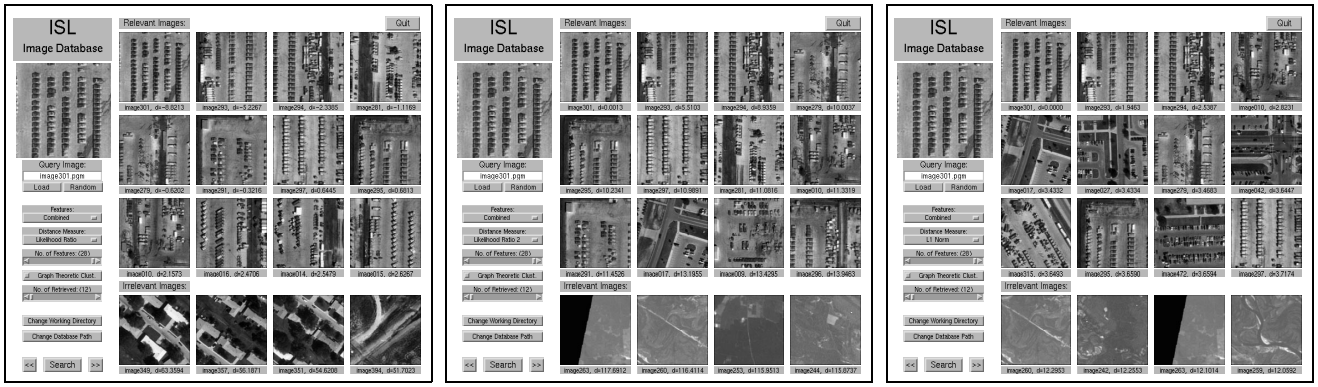
(c) Transformation to a Uniform random variable.

(d) Rank normalization.

(e) Fitting Exponentials to line-angle-ratio and Normals to co-occurrence features.

(f) Fitting Exponentials to line-angle-ratio and Gammas to co-occurrence features.

Figure 2: Precision vs. Number of images retrieved for the similarity measures used with different normalization methods. Fitting Exponentials to line-angle-ratio and Lognormals to co-occurrence features, and fitting Exponentials to all features resulted in similar results but were not included here due to space limitations.



(a) Likelihood ratio (MVN) (12 similar images retrieved)

(b) Likelihood ratio (Fit) (11 similar images retrieved)

(c) City-block (L_1) distance (9 similar images retrieved)

Figure 3: Retrieval examples using the same parking lot image as query with different similarity measures. The upper left image is 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. Euclidean (L_2) distance retrieved 7 similar images for this query. Please note that both the order and the number of similar images retrieved with different measures are different.

5.3. Observations

- Using probabilistic similarity measures always performed better in terms of both precision and recall than the cases where the geometric measures were used. On the average, the likelihood ratio that used the multivariate Normality assumption performed better than the likelihood ratio that used independent features with Exponential or Normal distributions. The covariance matrix in the correlated multivariate Normal usually captured more information than using individually better fitted but independent distributions.

- Probabilistic measures performed similarly when different normalization methods were used. This shows that these measures are more robust to normalization effects than the geometric measures.

- City-block distance performed better than the Euclidean distance. They both performed better with normalization methods like transformation using the cumulative distribution function or the rank normalization, i.e. the methods that tend to make the distribution uniform and spread out the feature values as much as possible.

6. Conclusions

This paper presented two probabilistic similarity measures for image retrieval and compared their retrieval performances to those of the geometric measures. The probabilistic measures used likelihood ratios that were derived from a Bayesian classifier that measured the relevancy of two images, one being the query image and one being a database image, so that image pairs which had a high likelihood value were classified as "relevant" and the ones which had a lower likelihood value were classified as "irrelevant". The first likelihood-based measure used multivariate Normality assumption and the second measure used independently fitted distributions for the features. Experiments on a database of approximately 10,000 images showed that both likelihood-based measures performed significantly better than the commonly used city-block (L_1) and Euclidean (L_2) distances in terms of average precision.

We also discussed the effects of feature normalization on retrieval performance. We described five normalization methods; linear scaling to unit range, linear scaling to unit variance, transformation to a Uniform[0,1] random variable, rank normalization and normalization by fitting distributions to independently normalize each feature component to the [0,1] range. We showed that studying the distributions of the features and using the results of this study significantly improves the results compared to making only general assumptions.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, 1972.

[2] 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, CVPR'98*, pages 45–49, Santa Barbara, CA, June 1998.

[3] S. Aksoy and R. M. Haralick. *Texture Analysis in Machine Vision*, chapter Using Texture in Image Similarity and Retrieval. Series on Machine Perception and Artificial Intelligence. World Scientific, 2000.

[4] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. In *Proceedings of IEEE International Conference on Computer Vision*, 1998.

[5] A. Berman and L. G. Shapiro. Efficient image retrieval with multiple distance measures. In *SPIE Storage and Retrieval of Image and Video Databases*, pages 12–21, February 1997.

[6] K. V. Bury. *Statistical Models in Applied Science*. John Wiley & Sons, Inc., 1975.

[7] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc., 1973.

[8] 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.

[9] C. S. Li and V. Castelli. Deriving texture set for content based retrieval of satellite image database. In *IEEE International Conf. on Image Processing*, pages 576–579, 1997.

[10] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on PAMI*, 18(8):837–842, August 1996.

[11] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, NY, 3rd edition, 1991.

[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.

[13] W. H. Press, B. P. Flannary, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recepies in C*. Cambridge University Press, 1990.

[14] M. Qinghong, B. Cramariuc, and M. Gabbouj. Multidimensional texture characterization using moments. In *Proceedings of Noblesse Workshop on Nonlinear Model Based Image Analysis*, pages 179–182, Glasgow, Scotland, July 1998.

[15] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Past, present, and future. *to appear in Journal of Visual Communication and Image Representation*, 1999.

[16] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, September 1998.

[17] S. Sclaroff, L. Taycher, and M. L. Cascia. Imagerover: A content-based image browser for the world wide web. In *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997.

[18] J. R. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression*. PhD thesis, Columbia University, 1997.