Human Perception-Driven, Similarity-Based Access to Image Databases

M. Emre Celebi and Y. Alp Aslandogan

Department of Computer Science and Engineering
University of Texas at Arlington
Arlington, TX 76019-0015 USA
{celebi,alp}@cse.uta.edu

Abstract

Similarity-based access to image databases assumes one or more similarity models. Although this assumption affects the retrieval precision of a system considerably, it is rarely described explicitly. Furthermore, because the similarity model is typically hard-coded into the system, it is very difficult if not impossible to use such a system for applications that do not fit the same similarity model. In this work, we develop a framework for designing similarity-based image access systems that are driven by human perception, and hence can be tailored for multiple, diverse applications. The driving components of the approach are Principal Components-based feature selection, perception modeling via psychophysical experiments and Genetic Algorithm-driven distance function optimization. While our framework is general and flexible, we demonstrate the application in a particular image access scenario: Shape-based retrieval of skin lesion images. The experimental results show that, by incorporating human perception of similarity into the system, retrieval performance may be significantly improved.

Introduction

Most similarity-based image retrieval or querying systems in literature assume a particular similarity model. While the choice of the similarity model affects the system retrieval performance significantly, it is rarely described explicitly. It is the authors' experience that unless there is a strong resemblance between the similarity model of an access system and a new application, the system may be unusable.

Current content-based retrieval systems use low-level image features based on color, texture, and shape to represent images. However, another aspect that is as important as the features themselves has been neglected: The processing and interpretation of those features by human cognition. For this reason, except in some constrained applications such as human face and fingerprint recognition, these low-level features do not capture the high-level semantic meaning of images (Rui, Huang, and Chang 1999).

Although the ultimate goal of all image similarity metrics is to be consistent with human perception, little work has been done to systematically examine this consistency. Commonly, the performance of similarity metrics is evaluated based on anecdotal accounts of good and poor matching results (Frese, Bouman, and Allebach 1997).

In this work, we develop a system for retrieving medical images with focus objects incorporating models of human perception. The approach is to guide the search for an optimum similarity function using human perception.

Figure 1 shows an overview of the system. First, the images are segmented using an automated segmentation tool. Then, 15 shape features are computed for each image to obtain a feature matrix. Principal component analysis is performed on this matrix to reduce its dimensionality. The principal components obtained from the analysis are used to select a subset of variables that best represents the data. A human perception of similarity experiment is designed to obtain a perceptual distance matrix. Finally, an optimum weighted (city-block) distance function is designed using a genetic algorithm utilizing a matrix correlation procedure (the Mantel test) as a fitness function.

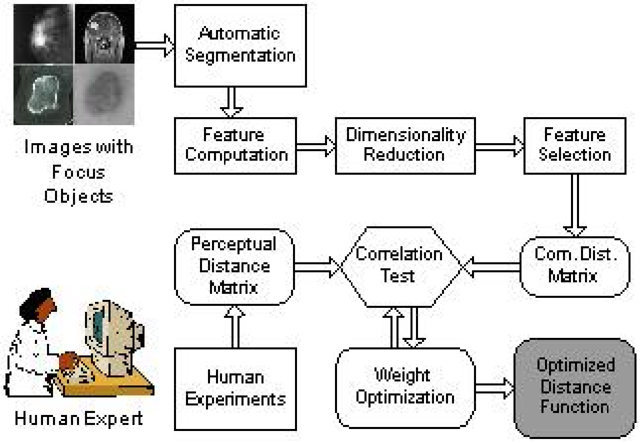

Figure 1. System Overview

The system is tested for shape-based retrieval of skin lesion images. However, it can be used in other image domains where the images have focus objects, such as brain tumors (Figure 2, lower right) and bone tumors (Figure 2, lower left).

Segmentation and Feature Computation

Our database contains 500 clinical skin lesion images. The images have a resolution of about 500 pixels per centimeter.

Segmentation

Segmentation is an extremely important step in image retrieval since accurate computation of shape features depends on good segmentation (Rui, Huang, and Chang 1999). For segmentation of lesion images we have used an automated tool, SkinSeg, described in (Xu et al. 1999). Two examples of segmented lesion images are given in the upper half of the Figure 2.

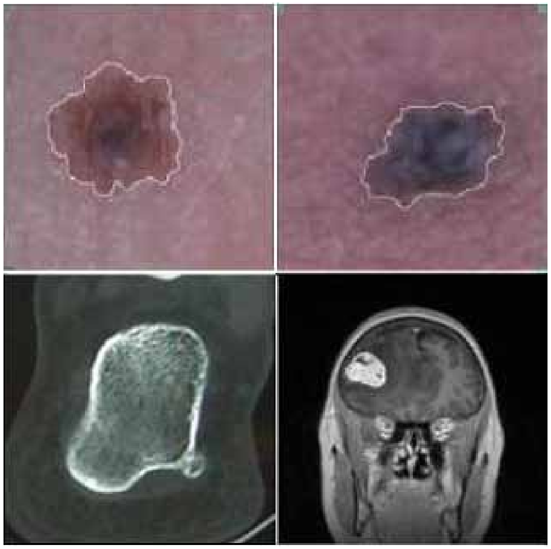

Figure 2. Biomedical images with focus objects

Feature Computation

The ABCD rule of dermatoscopy (Argenziano et al. 2000), recommended by the American Cancer Society, summarizes the clinical features of pigmented lesions suggestive of melanoma (a deadly form of skin cancer) by: asymmetry (A), border irregularity (B), color variegation (C) and diameter greater than 6 mm (D). Interestingly, three of these features are shape related. For each image in the database, we compute 15 shape features: Diameter, area, bending energy, contour sequence moments (Gupta and Srinath 1987), solidity, compactness, eccentricity, orientation, asymmetry (Costa and Ceesar 2001), fractal dimension (Costa and Ceesar 2001), and border irregularity (Xu et al. 1999).

We choose weighted city-block (L1) distance as the distance metric due to its ease of computation and robustness in the presence of outliers (Rousseeuw and Leroy 1987). Following (Aksoy and Haralick 2001), we normalize the features by transforming them to Uniform [0,1] random variables using their cumulative distribution functions (CDFs), since given a random variable x with CDF $F_x(x)$, the random variable $\tilde{x}$ resulting from the transformation $\tilde{x} = F_x(x)$ will be uniformly distributed in the [0,1] range.

Dimensionality Reduction and Feature Selection

After the feature computation step, we have data with 15 dimensions to be analyzed. Well known problems associated with high dimensionality include (a) high computational cost, (b) classifier accuracy degradation, known as Hughes phenomenon, and (c) difficulty in visualization.

We apply Principal Component Analysis (PCA) to this data to reduce its dimensionality. As a result of the analysis we choose to retain 5 principal components (PCs) that account for 96.15% of the variation in the data. The correlations between the PCs and the variables constitute the principal component loading matrix. An examination of this matrix shows that 5 PCs explain more than 90% of the variation in each variable.

Substantial dimensionality reduction can be achieved using a small number of PCs instead of the original variables, but usually the values of all of the original variables are still needed to calculate the PCs, since each PC is a linear combination of all of the original variables (Jolliffe 1986). Some variables may be difficult or expensive to measure therefore, collecting data on them in future studies may be impractical. Furthermore, while the original variables are readily interpretable, the constructed PCs may not be easy to interpret. Therefore, it might be preferable if, instead of using PCs, we could use a subset of the original variables, to account for most of the variation in the data (Jolliffe 1986).

The procedure to find n most important variables in the original set of variables is as follows: Starting with the largest PC, select the variable with the largest coefficient (loading) on that PC to represent that component, unless it has been chosen to represent a larger PC. Then, select the representative variable in the next largest PC. Repeat this procedure until you have a total of n variables (Jolliffe 1986). We use this method (a.k.a. Jolliffe's B4 method) to retain the following 5 variables: diameter, compactness, asymmetry, first contour sequence moment, and eccentricity.

The total amount of variation the selected variables account for can be used as a criterion to evaluate the efficiency of a particular subset of variables in representing the original set of variables (Jolliffe 1986). The total amount of variation that a subset of variables explains is the sum of the variation they explain in each of the discarded variables plus the sum of the variances for the variables comprising the subset. Each discarded variable is regressed on the retained variables and the corresponding squared multiple correlations are summed. If we add to that the variances of the retained variables, in our case 1.0 for each variable since the variables are normalized, we obtain a measure of the total amount of variation that a

$$n+\sum_{\mathrm{i}=1}^{m-n}\mathrm{R}^2(\mathrm{i})$$

subset of variables explains. This can be formulated as: where n and m are the number of variables in the subset and the original set, respectively, and $R^2(i)$ is the squared multiple correlation of the i^{th} discarded variable with the retained variables.

In our case, the subset of 5 variables retained by the Jolliffe's B4 method explains 87.54% of the variation in the data. Note that this is significantly lower than the percentage of variation explained by the 5 PCs (96.15%). In fact, no subset of 5 variables can explain more variation than 5 PCs, because, PC coordinate axes are defined as those directions, which maximize the total variation accounted for in the data. Therefore, we include a 6th variable (solidity), which has the highest loading on the 6th largest PC, in this subset to increase the percentage of variation it explains from 87.54% to 96.72%. In the rest of the study, feature vectors representing the images will consist of the following 6 features: diameter, compactness, asymmetry, first contour sequence moment, eccentricity, and solidity.

Human Perception of Similarity Experiment

Since the ultimate user of an image retrieval system is human, the study of human perception of image content from a psychophysical perspective is crucial (Rui, Huang, and Chang 1999). However, few content-based image retrieval systems have taken into account the characteristics of human visual perception and the underlying similarities between images it implies (Guyader et al. 2002). In (Payne and Stonham 2001), the authors argue that if perceptually derived criteria and rank correlation are used to evaluate textural computational methods, retrieval performances are typically 25% or less, unlike the 80-90% matches often quoted.

We conduct a psychophysical experiment to measure the perceived similarity of each image with every other image in the experimental database.

Experiment Description

Ten subjects (7 male and 3 female) participated in the experiment. They were graduate students, ranging in age from 22 to 25. The subjects, except one of the authors of this work, had no background on image retrieval and were not familiar with the images. All subjects had normal or corrected-to-normal vision.

Figure 3 shows a snapshot of the graphical user interface of the experiment. The image on the left is the reference image, and the one on the right is the test image. In order to focus the subjects only on shape similarity we converted the images to black and white so that there is no color or texture information in them.

In each trial, the subjects were asked to rate the similarity between a pair of images on a scale of four: Extremely similar, considerably similar, partially similar, and not similar. This scale is adapted from an earlier psychophysical study (Kurtz, White, and Hayes 2000). There were no instructions concerning the criteria on which the similarity judgments were to be made.

With a database of n images, this type of experimental design requires $n(n-1)/2$ comparisons, which, in our case, means 124750 trials. Therefore, in order to keep the experiment duration within reasonable limits, we randomly selected 100 images from our original database of 500 images. With this experimental database size the number of trials is 4950. The time required to complete the experiment in a single session is too long. Therefore, the experiment is divided into 5 sessions each containing 4950/5=990 trials.

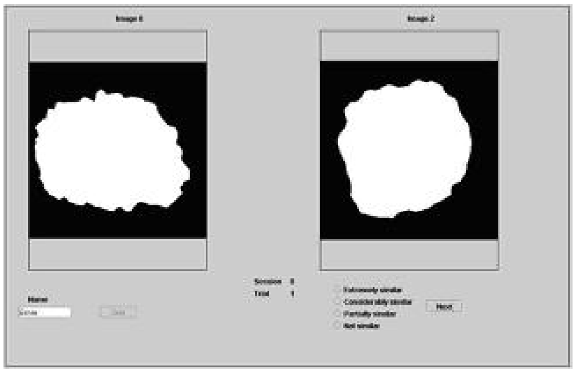

Figure 3. Experiment GUI

A warm-up session, which precedes the real sessions, was performed with each subject in order to make him/her comfortable with the procedure. Ten images (which are not included in the experimental database) were used in this session.

The sessions were self-paced and the subjects were free to take breaks whenever they wanted. Each session took about 41 minutes on the average.

Computation of Dissimilarity Matrix

We compute the aggregate similarity matrix S_T from the responses of human subjects following the approach described in (Guyader et al. 2002). For each subject, a sample similarity matrix S is computed as a weighted average of the elementary similarity matrices $S_K = \{S_K(i,j)\}$, one for each level of similarity judgment (matrix S_1 for "Extremely similar", S_2 for "Considerably similar", S_3 for "Partially similar", and S_4 for "Not similar"). Each time a subject associates a test image j to a reference image i, the appropriate $S_K(i,j)$ is set to one. The weights in the computation of S are determined based on a psychological model of similarity (Ashby and Perrin 1988). According this model, for two stimuli A and B, if $s(A,B)$ is the perceived similarity between the two, then the judged similarity is given by:

$$\sigma(A,B) = g[s(A,B)] \tag{1}$$

where g is a suitable monotonic non-decreasing function of its argument. Based on this model, the following relation between σ (judged similarity) and s (perceived similarity) is assumed:

$$\sigma = s^{1/3} \tag{2}$$

which agrees with the observation that humans are good at distinguishing short perceptual distances while they have the tendency to mix large and very large perceptual distances in their judgment. Based on (2) the following formulation for S is chosen:

$$S(i,j) = S_1(i,j) + \frac{1}{8}S_2(i,j) + \frac{1}{27}S_3(i,j) + \frac{1}{64}S_4(i,j)$$

Note that the entries of the sample similarity matrix S fall into the [0,1] range since the elementary similarity matrices S_K are boolean and for a given entry (i,j) only one of the $S_K(i,j)$ can be 1.

Now, the aggregate similarity matrix S_T is calculated as the average of the sample similarity matrices, which makes its entries fall into the [0,1] range, as well. We convert the aggregate similarity matrix S_T to a dissimilarity matrix P by $P(i,j) = 1 - S_T(i,j)$.

Optimization of the Distance Function Using a Genetic Algorithm

In the previous steps, we have obtained the distances between pairs of images using a computational approach (feature computation) and a perceptual approach (subjective experiments). Now, in order to incorporate human perception of similarity into the system, we determine the weights of the distance function (city-block distance) that maximizes the correlation between the outputs of the computational and perceptual approaches using a genetic algorithm (GA).

GAs are stochastic and adaptive optimization methods. They can handle high-dimensional, nonlinear optimization problems and are especially effective in situations where the search space is mathematically uncharacterized and not fully understood (Eiben 2002). The randomness in the Mantel's procedure (which will be described in the following subsection) suggests the use of GA as an appropriate optimization technique.

We used Parallel Genetic Algorithm Library (PGAPack) (Levine 1996) to implement the GA. In order to minimize the effects of random starting points in the optimization process, 20 different runs of the GA were carried out.

The Mantel Test

The Mantel test (Manly 1991) is a statistical method for finding the correlation between two symmetric distance matrices. It involves calculation of a suitable correlation statistic between the elements of the matrices. The null hypothesis is that entries in the matrices are independent of each other. Testing of the null hypothesis is done by a Monte Carlo randomization procedure in which the original value of the statistic is compared with the distribution found by randomly relocating the order of the elements in one of the matrices. The significance of the association is estimated by counting the number of randomizations in which the test statistic is lower than or equal to that obtained from the original matrices, and dividing this number by the number of randomizations.

The statistic used in the Mantel test for measuring the correlation between two n-by-n distance matrices, A and B, is the classical Pearson correlation coefficient:
where $\bar{A}$ and S_A are the mean and standard deviation of

$$r_{AB} = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{A_{ij}-\overline{A}}{s_A}\right]\left[\frac{B_{ij}-\overline{B}}{s_B}\right] \tag{3}$$

elements of A, respectively.

The procedure for the Mantel test is as follows (Bonnet and Peer 2002):
1) Compute the correlation coefficient r_{AB} using (3).
2) Permute randomly rows and the corresponding columns of one of the matrices, creating a new matrix A'.
3) Compute the correlation coefficient $r_{A'B}$ using (3).
4) Repeat steps 2 and 3 a great number of times (>5000). The number of repeats determines the overall precision of the test (≈1000 for α= 0.05; ≈5000 for α= 0.01; ≈10000 for greater precision (Manly 1991)).

We used the software described in (Bonnet and Peer 2002) to perform the Mantel test.

Optimization Procedure

In the optimization procedure, to evaluate the goodness of a particular set of weights, we determine the correlation between the perceptual distances and the computational distances. This correlation is computed by the simple Mantel test, which requires two symmetric distance matrices as input. The P matrix, which represents the perceptual distances, is already symmetric. In each generation, we form a C matrix, which represents the computational distances, by taking the weighted city-block (L1) distances between feature vectors of all pairs of images using the gene values of the fittest individual in that generation as weights. Since city-block distance function is a metric, C matrix is always a symmetric distance matrix.

When the GA terminates, the gene values of the fittest individual in the population give the optimum set of weights of the distance function, which maximizes the correlation between the perceptual distance and computational distance matrices.

Results of the Optimization

Initially (i.e., with all weights equal to 1.0), the correlation between the C (computational distance matrix) and P (perceptual distance matrix) matrices is 32%. After the optimization, the correlation becomes 50%. To test the impact of this correlation improvement on the actual retrieval performance, we perform queries with 100 images chosen randomly from the remaining 400 images in the original database that are not included in the experimental database.

We determine the retrieval performance of the system for the case of optimized distance function as follows: A matrix C_f is computed using the final set of weights, which were obtained as a result of the optimization. This matrix represents the computational distances (between pairs of images) calculated using the optimized distance function. For each query image, the 10 most similar images to it are determined using the C_f matrix. Then, these 10 images are marked as relevant (extremely or considerably similar) or non-relevant (partially or not similar). For each rank position K, where K ranges from 1 to 10, the number of relevant images among the K images is counted and averaged over 100. The retrieval performance of the system for the case of unoptimized distance function is determined similarly with the only difference that the weights used in the computation of C_f matrix are taken as 1.0. Figure 4 shows the average precision values calculated for both cases for N (number of images retrieved) ranging from 1 to 10. As can be observed from this figure, the optimization of the distance function has a great impact on the average precision.

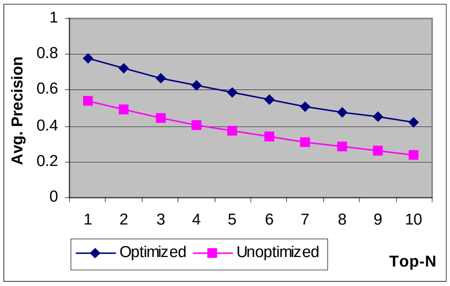

Figure 4. Performance comparison graph

Related Work

To the authors' knowledge relatively little work has been done to incorporate human perception of similarity in CBIR systems in a systematic manner.

Scassellati et al. (Scassellati, Alexopoulos, and Flickner 1994) have used shape similarity judgments from human subjects to evaluate the performance of several shape distance metrics.

Frese et al. (Frese, Bouman, and Allebach 1997) have developed an image distance metric based on a multiscale model of the human visual system, which is systematically optimized for maximum consistency with human perception of the similarity of images of natural scenes.

Rogowitz et al. (Rogowitz et al. 1998) have studied how human observers judge image similarity. They conduct two psychophysical scaling experiments aimed at uncovering the dimensions human observers use in rating the similarity of photographic images and compare the results with two algorithmic image similarity methods.

Mojsilovic et al. (Mojsilovic et al. 2000) have developed a perceptually based image retrieval system based on color and texture attributes. They perform subjective experiments and analyze the results using multidimensional scaling to extract relevant dimensions. They also design distance metrics for color and texture matching.

In (Chang, Li, and Li 2000), several perceptual characteristics are described to argue that using Euclidean distance may not be appropriate in all circumstances. Furthermore, these perceptual characteristics are used in designing image filters that support customizable queries.

Guyader et al. (Guyader et al. 2002) have developed a natural scene classification system using Gabor features. They optimize the feature weights by analyzing the results of a psychophysical experiment using a multidimensional scaling technique.

Conclusions and Future Work

Content-based image retrieval has been an active research area in the past 10 years. Since the early 90s numerous image retrieval systems, both research and commercial, have been developed (Niblack et al. 1993; Pentland, Picard , and Sclaroff 1996; Smith and Chang 1996). The main contribution of this work is the incorporation of human perception into this task in a systematic and generalizable manner.

In this work, we used human perception of similarity as a guide in optimizing an image distance function in a content-based image retrieval system. A psychophysical experiment was designed to measure the perceived similarity of each image with every other image in the database. The weights of the distance function were optimized by means of a genetic algorithm using the distance matrix obtained from the subjective experiments.

Using the optimized distance function, the retrieval performance of the system is significantly improved.

In this study we focus on shape similarity. However, the same approach can be used to develop similarity functions based on other low-level features such as color or texture. Also, for general similarity based retrieval, another content-based image retrieval system powerful in color or texture aspects can be combined with our system.

The retrieval performance of the system can be further improved by using better shape features, such as those adopted by the MPEG-7 consortium (Manjunath, Salembier, and Sikora 2002).

Acknowledgments

This research is supported in part by NSF grant 0216500-EIA and by Texas Workforce Commission grant #3204600182.

References

Aksoy S. and Haralick R.M. 2001. Feature Normalization and Likelihood-Based Similarity Measures for Image Retrieval. *Pattern Recognition Letters* 22(5):563-582.

Argenziano G., Soyer H.P., Giorgi V., and Piccolo D. 2000. *Dermoscopy: A Tutorial*. Milan, Italy: EDRA Medical Publishing & New Media.

Ashby F.G. and Perrin N.A. 1988. Toward a Unified Theory of Similarity and Recognition. *Psychological Review* 95(1):124-150.

Bonnet E. and Peer Y. 2002. zt: A Software Tool for Simple and Partial Mantel Tests. *Journal of Statistical Software* 7:1-12.

Chang E.Y., Li B., and Li C. 2000. Towards Perception-Based Image Retrieval. In *Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries*, 101-105.

Costa L.F. and Ceesar R.M. 2001. *Shape Analysis and Classification: Theory and Practice*. Boca Raton, Florida: CRC Press.

Eiben A.E. 2002. Evolutionary Computing: The Most Powerful Problem Solver in the Universe ?. *Dutch Mathematical Archive* 5(2):126-131.

Frese T., Bouman C.A., and Allebach J.P. 1997. A Methodology for Designing Image Similarity Metrics Based on Human Visual System Models. In *Proceedings of the SPIE Conference on Human Vision and Electronic Imaging II* 3016:472-483.

Gupta L. and Srinath M.D. 1987. Contour Sequence Moments for the Classification of Closed Planar Shapes. *Pattern Recognition* 20(3):267-272.

Guyader N., Herve L.B., Herault J., and Guerin A. 2002. Towards the Introduction of Human Perception in a Natural Scene Classification System. In *Proceedings of the IEEE International Workshop on Neural Network for Signal Processing*, 385-394.

Jolliffe I.T. 1986. *Principal Component Analysis*. Springer-Verlag Inc.

Kurtz D.B., White T.L., and Hayes M. 2000. The Labeled Dissimilarity Scale: A Metric of Perceptual Dissimilarity. *Perception & Psychophysics* 62:152-161.

Levine D. 1996. Users Guide to the PGAPack Parallel Genetic Algorithm. Available at: ftp://ftp.mcs.anl.gov/pub/pgapack/user_guide.ps

Manjunath B. S., Salembier P., and Sikora T. eds. 2002. *Introduction to MPEG 7: Multimedia Content Description Language*. John Wiley & Sons Inc.

Manly B.F.J. 1991. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Boca Raton, Florida: CRC Press.

Mojsilovic A., Kovacevic J., Hu J., Safranek R.J., and Ganapathy S.K. 2000. Matching and Retrieval Based on the Vocabulary and Grammar of Color Patterns. *IEEE Transactions on Image* Processing 9(1):38-54.

Niblack W., Barber R., Equitz W., Flickner M., Glasman E.H., Petkovic D., Yanker P., Faloutsos C., and Taubin G. 1993. The QBIC Project: Querying Images by Content, Using Color, Texture, and Shape. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases* 1908:173-187.

Payne J.S. and Stonham T.J. 2001. Can Texture and Image Content Methods Match Human Perception ?. In *Proceedings of the International Symposium on Intelligent Multimedia, Video, and Speech Processing.*

Pentland A., Picard R.W., and Sclaroff S. 1996. Photobook: Content-based Manipulation of Image Databases. *International Journal of Computer Vision* 18:233-254.

Rogowitz B.E., Frese T., Smith J.R., Bouman C.A., and Kalin E. 1998. Perceptual Image Similarity Experiments. In *Proceedings of the SPIE Conference on Human Vision and Electronic Imaging* 3299:26-29.

Rousseeuw P.J. and Leroy A.M. 1987. *Robust Regression and Outlier Detection*. John Wiley & Sons Inc.

Rui Y., Huang T.S., and Chang S. 1999. Image Retrieval: Current Techniques, Promising Directions and Open Issues. *Journal of Visual Communication and Image Representation* 10:39-62.

Scassellati B., Alexopoulos S., and Flickner M. 1994. Retrieving Images by 2D Shape: A Comparison of Computation Methods with Human Perceptual Judgments. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases Conference II* 2185:2-14.

Smith, J.R. and Chang S.–F. 1996. VisualSEEk: A Fully Automated Content-Based Image Query System. In *Proceedings of ACM Multimedia Conference*, 87-98.

Xu L, Jackowski M., Goshtasby A., Roseman D., Bines S., Yu C., Dhawan A., and Huntley A. 1999. Segmentation of Skin Cancer Images. *Image and Vision Computing* 17(1):65-74.