Proceedings of the 7th WSEAS International Conference on Automation & Information, Cavtat, Croatia, June 13-15, 2006 (pp110-115)

Smart Image Search by Boosted Shape Features

Jiann-Jone Chen, Chia-Jung Hu, Chi-Wen Luo and Che-Kang Chang
Electrical Engineering Dept., National Taiwan Univ. Science & Tech.
43 Keelung Rd., Sec. 4, Taipei Taiwan 10673

Abstract: An one-line image database search method, which utilizes the boosted-shape feature similarities, is proposed. Salient common feature informations provided by the relevance feedback or multi-instance query are boosted for improving retrieval results. Weak classifiers are successively refined to yield a final strong classifier. The similarity between two shape samples was measured in statistic space of features, through which relative instead of absolute similarity was targeted for visual information retrieval. Experiments of query by the boosted features on thirty thousand trademark images showed that the retrieved results meet visual similarity of shape very well. Objective evaluations, precision-recall hit curve and averaged normalized modified retrieval rank, ANMRR, demonstrate improved retrieval performances of the proposed method. It shows that only 5 - 7 boosted features out of 100 or more were enough to represent subjective recognition on shape similarity.

Key–Words: MPEG-7, boosting, retrieval, image database, Zernik moments, rotation and scale invariant

1 Introduction

Content-based similarity retrieval for multimedia data becomes important after international coding standard, such as JPEG, MPEG-1, -2, had been widely used and distributed over Internet. The multimedia content description interface, MPEG-7[1], had been proposed to provide normal descriptors for database search engine. For 2D shapes, contour-based and region-based descriptors, shown in Fig. 1, have been proposed in MPEG-7. Describing image shape contours with Fourier descriptors (FDs) [2] yields size, rotation and transition invariant descriptors for indexing. However, FDs are sensitive to noises. A multiscale, curvature scale space descriptors had been proposed to improve the stability in contour description and matching [3]. For region-based descriptors, Zernike and psuedo-Zernike moments (ZMs and PZMs) also provide size, transition and rotational invariant descriptors for similarity retrieval [4]. The ART is another angular-radial transform adopted by MPEG-7 and it can be generalized [5] to deal with gray-scale images and perspective deformations. The indexing system, though more than one feature sets are accommodated, is with single similarity retrieval target. In [6], visually salient feature is determined using probabilistic distribution model of features from trademark databases, however, feedback control had not been investigated.

For similarity retrieval, each user has his definition on shape similarity that the query system usually provides relevance feedback or multi-instances to learn. In [7], an online relevance feedback learning method for CBIR, by utilizing the region-based representation to describe images in a uniform feature space, was proposed. Other learning approach could be found in [8]. We proposed to boost salient common features among query samples such that user's recognition on similarity is targeted. The boosting algorithm consistently selects weak hypotheses that are slightly better than random guessing, and the error of the final hypothesis would drops exponentially fast [9]. In general, only global shape contents are subjectively recognized for similarity measure [6]. We thus proposed to successively boost one salient common feature under the boosting framework. Similar concept can be found in [10]. In addition, when both positive and negative query images are provided, its proper to actively choose converging positive features while excluding negative ones, which are usually sparse and diverse, passively.

This paper is organized as follows. In section 2, descriptors for shape content in images are presented. Pre-processing for each image is introduced in section 3. The query mechanism by boosting salient common features is described in section 4.2. Simulation study is provided in section 5. Section 6 concludes the paper.

2 Shape Descriptors

Shape descriptors are extracted from image contents according to applications of indexing. For shapes in images, we use region-based shape descriptors, Zernike and pseudo-Zernike moments [11], as the basic feature sets.

Zernike moments are defined inside the unit circle, and the radial polynomials $R_{nm}(\rho)$ are defined as:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!}\rho^{n-2s}, \tag{1}$$

where $n = 0, 1, \cdots, \infty$, $|m| \leq n$ and $n - |m|$ is even. The two-dimensional Zernike moment with order n and repetition m of an image, in polar coordinate $I(\rho, \theta)$, is defined as:

$$A_{nm} = \frac{n+1}{\pi} \sum_{\rho} \sum_{\theta} [V_{nm}(\rho, \theta)]^* \cdot I(\rho, \theta), s.t \rho \leq 1. \tag{2}$$

The Zernike basis polynomials, $V_{nm}(\rho, \theta)$, are defined as

$$V_{nm}(\rho, \theta) = R_{nm} \cdot exp(-jm\theta). \tag{3}$$

Pseudo-Zernike moments are another sets of orthogonal polynomials with similar properties to Zernike polynomials. The psuedo-Zernike radial polynomials are defined as:

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|} (-1)^s \frac{(2n+1-s)!\rho^{n-s}}{s!(n+|m|-s)!(n+|m|+1-s)!}, \tag{4}$$

where $n \geq 0$ and $0 \leq |m| \leq n$. The two-dimensional pseudo-Zernike moments can be defined similarly to eq. 2 and are denoted as B_{nm}.

To make both ZMs and PZMs of query q representing scale invariant shape features, normalized projections vectors are used and the final feature vector $f(q)$ is the union of both ZMs and PZMs, i.e.,

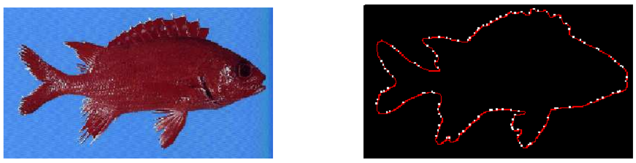
$$\begin{aligned} f_{ZM}(q,n) &= \{A_{ij}/A_{00} | 0 \leq i \leq n, i - |j| \text{ is even}\} \\ f_{PZM}(q,n) &= \{B_{ij}/B_{00} | 0 \leq i \leq n, |j| \leq i\}, \\ f(q,n) &= f_{ZM} \bigcup f_{PZM}, \end{aligned} \tag{5}$$

where $f(q, n)$ is rewritten as $\{f_i\}_{i=1,\cdots,F}$ for simplicity and F is the vector dimension of feature, i.e., $F = size(f_{ZM}) + size(f_{PZM})$.

Though ZMs and PZMs are similar but, in general, PZMs are less sensitive to noises than are ZMs.

3 Preprocessing

Both ZMs and PZMs provide orientation invariant feature for shapes. Both ZMs and PZMs are projections of signals representing shape content on a complete set of complex-values functions orthogonal on



(a)

(a)

Figure 1: Shape description by (a) shape-contour and (b) shape-region

the unit disk whose coordinates are $\{(x, y)|x^2 + y^2 \leq 1\}$. To achieve transition invariance, the coordinate zero are moved to shape centroid, i.e., $(E[x_i], E[y_i])$, where (x_i, y_i)s belong to points of object shape. The scale invariance is accomplished by enlarging or reducing each shape such that its zero-th order moment equals to a predetermined value. Both preprocessing steps are designed for discrimination of different shapes or identification of the same shapes, which are not quite the same when dealing with retrieval of similar shapes from image databases. For this, we intend to accommodate shape content in image with minimum bounding circle (MBC). We have devised a fast algorithm for locating the MBC of shape content, i.e., center and radius, which excludes erroneous noises, when computing ZMs and PZMs, outside the circle. Experiments show that shape features computed with the MBC center do perform better in similarity retrieval of shapes than do with shape centroid. Note that with the MBC, the projections of shape content on Zernike bases could be computed efficiently, i.e., less computation time and less erroneous noises.

4 Similarity Retrieval

Shape feature sets are defined according to specific database and user requirements. They are efficient in similarity retrieval for their applications. Nonetheless, it's clear there's no one set of universal descriptors could satisfy all requirements. In addition, users may need specific combination of features for their retrieval target, either query by multi-instance or relevance feedback. Hence the query mechanism must accommodate plural features sets while being flexible in selecting proper features to reflect user's definitions on shape similarity. Note that the target of visual information retrieval is relative instead of absolute similarity. We intend to measure similarity between samples in the supervised approach, such as those in statistical space [6]. To reflect visual similarity, the on-line learning mechanism should bring up subjectively similar features among query samples for refined query. For this, we propose to choose features by boosting salient common ones to meet visual similarity.

4.1 Distance Measure

To measure relative similarity between shape samples in statistic space, the probability distribution of each shape feature was modelled by gamma distribution function with parameters α and β:

$$p(f_i;\alpha,\beta)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}f_i^{\alpha-1}exp(-f_i)U(f_i), \quad (6)$$

where $\Gamma(\alpha)=\int_a^b x^{\alpha-1}e^{-x}dx$ and $U(f_i)$ is the unit-step function. The parameters, α and β, could be computed from $\alpha=\frac{m_i^2}{\sigma^2}$ and $\beta=\frac{\sigma_i^2}{m_i}$, where m_i and σ_i are the mean and standard deviation of feature f_is of all images in database. Dissimilarity between feature values *a* and *b* thus could be measured by the following probability distance:

$$P(a,b)=\int_{min(a,b)}^{max(a,b)}p(f_i)df_i. \quad (7)$$

4.2 Boosting Algorithm

The basic idea of boosting algorithm is to collect weak hypotheses to form a single highly accurate prediction rule. If weak hypotheses are slightly better than random guessing, the error of the final hypothesis would drops exponentially fast [9]. For shape-based similarity retrieval, visual perception would appreciate global instead of detailed shape contents in images [6]. This boosting algorithm is thus designed to successively select one highly deterministic feature among query samples to output robust global classifier. The boosting algorithm which is capable of dealing with image shape features is described:

1. Inputs: (a) N query images,$q_{i_{i=1,2,\cdots,N}}$, and their relevancy $\{s_i\}_{i=1,2,\cdots,N}$: (q_1,s_1),$\cdots$, (q_N,s_N); (b) weak learning algorithm **WeakLearn**; (c) set $w_i=1$ for $i=1,2,\cdots,F$ and integer T specifying the number of iterations

2. Do for $t=1,2,\cdots,T$:

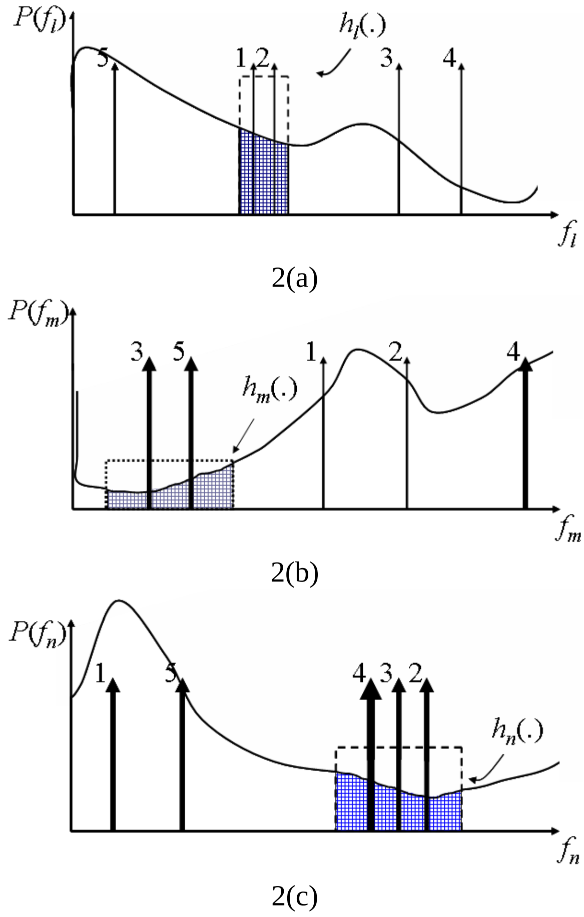


Figure 2: Locate weak classifiers by greedily covering, while excluding negative ones, possible salient common features. The width of line denotes the weighting of the corresponding sample.

- Get the distribution by: $\hat{w}_i^t=\frac{w_i^t}{\sum_{i=1}^N w_i^t}$

- Call function **WeakLearn**($\hat{w^t}$) and return one hypothesis h_j for each feature dimension j and compute its error

$$\varepsilon_j=\sum_{i=1}^N \hat{w}_i^t\cdot|h_j(x_i)-y_i|; \quad (8)$$

- Set $h_t(\cdot)=h_k(\cdot)$, where k is the dimension index of features such that

$$\varepsilon_k=min\{\varepsilon_j\}_{j=1,\ldots,F}; \quad (9)$$

- Let $\beta_t=\frac{\epsilon_t}{1-\epsilon_t}$ and set the new weight vector to be:

$$w_i^{t+1}=w_i^t\cdot\beta_t^{1-|h_t(x_i)-y_i|}. \quad (10)$$

3. Output the hypothesis

$$h(x) = \sum_{t=1}^{T}(log\frac{1}{\beta_t} \cdot h_t(x)) \geq \frac{1}{2}\sum_{t=1}^{T}(log\frac{1}{\beta_t}). \quad (11)$$

The **WeakLearn** function is designed to locate the decision boundaries for each hypothesis by finding the minimum error with function $\epsilon_j(m_i)$ in eq. 7, in which $m_j = \frac{1}{N}\sum_{i=1}^{N} f_j(x_i)$ and

$$h_j(\cdot) = \begin{cases} 0, & \text{if } P(f_j(x_i), m_j) \leq P_T \\ 1, & \text{otherwise} \end{cases} \quad (12)$$

where P_T could be the constant threshold to determine whether features j of two samples are relatively similar or not.

The boosting mechanism could be easily understood with the aids of Fig. 2. With P_T specifying the shaded region, the weaklearn hypotheses are designed to accommodate salient common (converging) features among query samples, such as feature of samples 1 and 2 with the hypothesis $h_l(\cdot)$ in Fig. 2 (a). Excluded samples, such as samples 3, 4 and 5 in Figs. 2 (a)(b), are weighted heavily to make them easily boosted thereafter. As seen in Fig. 2(b), feature m of samples 3 and 5 are boosted and heavily weighted sample 4 is then targeted. Possible hypothesis could be as that shown in Fig. 2(c). The final strong hypothesis is thus a greedy collection of deterministic weak hypotheses. Note that only positive features are plotted in Fig. 2. The weak hypothesis could accommodate negative samples as well.

5 Simulation Study

5.1 Evaluation Methods

When databases and categorization information In general, it uses precision ($\frac{A}{A+B}$) and recall ($\frac{C}{B+C}$) to evaluate the objective retrieval performance, where $(A + B)$ and $(B + C)$ denote the number of relevant samples and top-ranked retrieved samples for evaluations, respectively, for one query. The retrieval performances can be evaluated by displaying precision and recall graphs, which are usually inversely related. As shown in Fig. 5, improved retrieval performances would yield much upper-right recall-precision hit curves.

In addition to indicate how many of the correct samples are retrieved by the recall-precision hit curves, the MPEG-7 retrieval metric, Normalized Modified Retrieval Tank (NMRR) [12], also provide numerical measurements of how top-ranked the correct samples are in the list of retrieved samples. NMRR is defined by

$$NMRR(q) = \frac{\frac{1}{NG(q)}\sum_{k=1}^{NG(q)} Rank(k) - 0.5 \cdot [1 + NG(q)]}{K(q) - 0.5 \cdot [1 + NG(q)]}, \quad (13)$$

where $NG(q)$ denotes the size of the ground-truth set of the query sample q in the image database **IDB**, $Rank(k)$ is the rank of the k-th ground-truth image in the retrieved list and $K(q)$ specify the relevance rank for query q. Since the size of each ground-truth set is different, $K(q)$ is determined by:

$$K(q) = \min(4 \cdot NG(q), 2 \cdot GTM), \quad (14)$$

where $GTM = \max\{NG(q)|q \in IDB\}$. Averaged Normalized Modified Retrieval Rate (ANMRR) is the average of NMRR for all queries:

$$ANMRR = \frac{1}{NQ}\sum_{q=1}^{NQ} NMRR(q), \quad (15)$$

where NQ is the number of queries. ANMRR is another criterion used in measuring retrieval performance of MPEG-7 related multimedia search engine [12]. Experiments [13] showed evidences that the ANMRR measure coincides linearly with subjective evaluation results in regard to the retrieval accuracy.

To obtain the ANMRR data for m instance queries, it performs C_m^{NG} retrieval operations, i.e., evaluate retrieval performances in an exhaustive approach, and evaluates NMRR(q) for each query. The total average ANMRR is computed over all ground-truth set in the database.

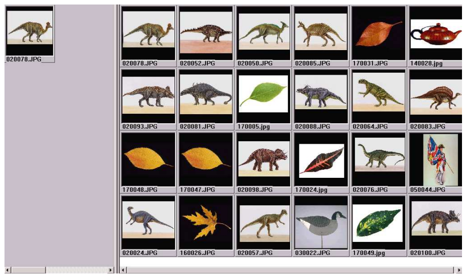
5.2 Performance Evaluations

Two thousand images were collected in the test database. They contain pattern, animal, insects, cup, regular geometric shapes et al. Each image had been processed in advance to segment the image foreground. It then located the MBC of the shape region for normalization and feature extractions. Magnitudes of ZMs and PZMs for each sample image are computed by lookup-tables to speed up processing of database images. With order n=10, the number of ZMs and PZMs are 36 and 66, respectively. Fig. 3 shows the retrieval results of one query image in the upper-left corner. The similarity ranking is from left to right and up to down. In Fig. 3(a), the retrieved shape images are not coherent in subjective similarity since only one query image was given. When one positive sampled were added for refined query, more visually similar shape were retrieved as shown in Fig. 3(b)(c). However, the proposed algorithm (Fig. 3(c)) demonstrates better subjective performances as compared to the mean feature vector one (Fig. 3(b)). Given more positive samples, the salient and common features were boosted by the proposed algorithm and the retrieved results demonstrate convergence toward subjective similarity very well (Fig. 3(e)) as compared to that shown in Fig. 3(d).

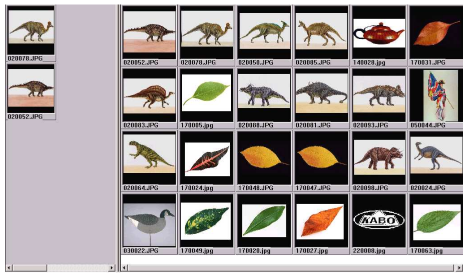
For objective evaluations, the ANMRR performance is demonstrated in Fig. 4. As shown the averaged ranks of ground-truth sets are about 15% to 17% lower than that of the mean-feature vector. The PR hit curves, averaged over all the ground-truth sets in IDB, of the proposed algorithm and the mean vector method are shown in Fig. 5. Note that images of the same category may not all subjectively similar to others such that the precision-recall performance would not outperform the other, as compared to the subjective evaluation. The retrieval performance of the proposed algorithm outperform the other when more query samples are provided. As shown in Fig. 5, the boosting method yield more upper-right curves, as compared to the mean-feature vector ones, when more query samples are provided.

6 Conclusions

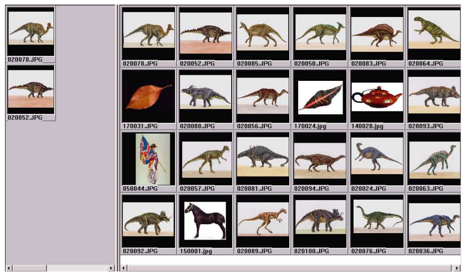
We presented a shape-based similarity retrieval for image databases, which boosts salient common features among query samples successively. The selected weak classifiers actively locate converging features of positive samples while excludes passively negative ones. The most distinguished functionality of the proposed method is that visual similarities could be targeted well from multiple sets of features. In addition, the time complexity had been reduced about 20 times smaller as compared to previous methods. Experiments showed that the retrieved results by the proposed method provide satisfactory subjective retrieval performance. Objective evaluations, ANMRR and precision-recall hit curves, also demonstrate improved retrieval ranks and precisions. With the proposed mechanism, new feature types could be directly accommodated for specific applications to enhance the retrieval performance, which is considered as the future work.



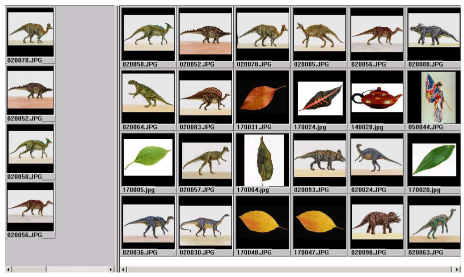
3(a) query by one sample



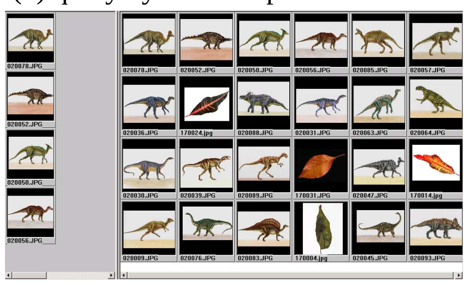
3(b) query by two samples: mean vector



3(c) query by two samples: boosting



3(d) query by four samples: mean vector



3(c) query by four samples: boosting

Figure 3: Retrieved results by the boosting algorithm and mean vector with:(a) one; (b)(c) two; and (d)(e) four query images. As shown, when more shape images were given, more subjectively similar shapes were retrieved by the proposed algorithm as compared to the mean vector ones.

References:

[1] Overview of the MPEG-7 Standard (version 4.0), ISO/IEC JTC1/SC29/WG11 N3753, Oct. 2000.

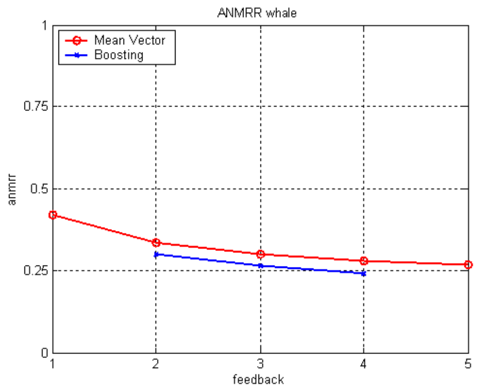
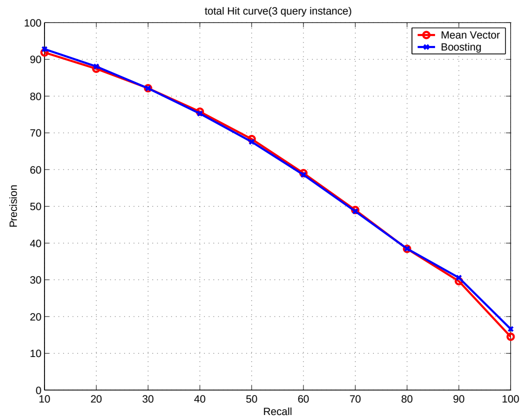
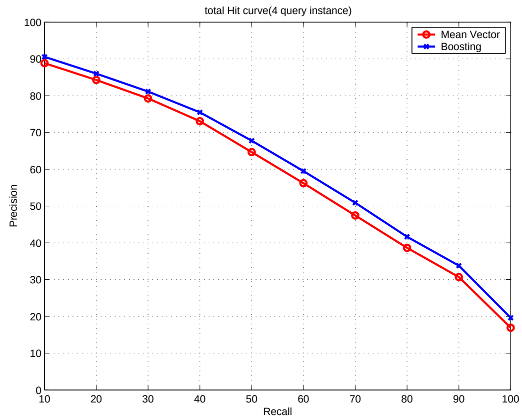


Figure 4: The ANMRR retrieval performances



5(a) query with three samples



5(b) query with four samples

Figure 5: The recall-precision hit curves for multi-instance queries with different number of query samples.(a) three sample query (b) the four sample query

[2] C. T. Zhan et al., "Fourier descriptors for plane closed curves, *IEEE Trans. Computer*, vol. 21, pp. 269-281, 1972.

[3] F. Mokhtarian, et al., "A theory of multi-scale, curvature-based shape representation for planar curves," *IEEE Trans. PAMI*, vol. 14, no. 8, pp. 789-805, 1992.

[4] A. Khotanzad and Y. H. Hong, "Rotational invariant image recognition using features selected via a systematic method," *Pattern Recognition*, vol. 23, no. 10, pp. 1089-1101, 1990.

[5] J. Ricard et al., "Generalization of angular radial transform," *IEEE Conf. Image Processing, vol. 4, pp. 2211-2214, 2004.*

[6] Y. S. Kim et al., "Content-based trademark retrieval using visually salient feature," *IEEE Proc. Comput. Vision and Pattern Recog.*, pp. 307-312, 1997.

[7] W. Jiang et al., "Similarity-Based online feature selection in content-based image retrieval," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 702-712, 2006.

[8] I. El-Naqa et al, "Image retrieval based similarity learning," *IEEE Conf. Image Processing*, vol. III, pp. 722-725, 2000.

[9] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of online learning and an application to boosting, *Journal of Comp. & Sys. Sci.*, 55(1): 119-139, 1997.

[10] K. Tieu and P. Viola, "Boosting image retrieval," *IEEE Conf. Comput. Vision and Pattern Recog.*, vol. 1, pp. 228-235, 2000.

[11] C. H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10. No. 4. pp. 496-513, July 1988.

[12] B. S. Manjunath et al., "Color and texture descriptors," *IEEE Trans. Circuits & Systems Video Tech.*, vol. 11, no. 6, pp. 703 -715, June 2001.

[13] P. Ndjiki-Nya et al., "Subjective evaluation of the MPEG-7 retrieval accuracy measure (ANMRR)," ISO/WG11 MPEG Meeting, Geneva, Switzerland, Doc. M6029, May 2000.