Texture Retrieval Effectiveness Improvement Using Multiple Representations Fusion

Noureddine Abbadeni

Al-Ain University of Science and Technology
College of Engineering and IT
Al-Ain, UAE
noureddine.abbadeni@usherbrooke.ca

Abstract. We propose a multiple representations approach to tackle the problem of content-based image retrieval effectiveness. Multiple representations is based on the use of multiple models or representations and make them cooperate to improve search effectiveness. We consider the case of homogeneous textures. Texture is represented using two different models: the well-known autoregressive model and a perceptual model based on perceptual features such as coarseness and directionality. In the case of the perceptual model, two viewpoints are considered: perceptual features are computed on original images and on the autocovariance function corresponding to original images. Thus, we use a total of three representations (models and viewpoints) to represent texture content. Simple results fusion models are used to merge search results returned by each of the three representations. Benchmarking carried out on the well-known Brodatz database using the recall graph is presented. Retrieval relevance (effectiveness) is improved in a very appreciable way with the fused model.

1 Introduction

Content-based image and multimedia retrieval has became one of the most active research areas in the last two decades and many approaches have been proposed and various results and systems have been carried out since then [5], [8]. In the first years of such Systems, content representation and similarity matching were considered as fundamental issues. More recently, researchers have paid more attention to other approaches including relevance feedback-based image retrieval ([27], [22]) and semantics-based image retrieval ([18], [23]). These approaches allow generally an interesting improvement in search relevance even if they can be criticized at least on the fact that an important effort is asked to users to give relevance judgments or to perform annotations on images.

One approach, which still in the visual CBIR approach, and does not necessarily require the intervention of users, has not received enough attention in our opinion. This approach is Data fusion. Data fusion has been extensively used in the traditional text information retrieval (IR) field, and particularly in distributed IR (DIR) [9], [16], [26]. Data fusion, within DIR, recover three parts:

T. Wada, F. Huang, and S. Lin (Eds.): PSIVT 2009, LNCS 5414, pp. 485–496, 2009.

collection description, collection selection and results fusion. Collection description consists in acquisition of information related to the different (distributed) collections of data used to search information. Collection selection consists to choose which are the most relevant data collections to the user's needs and the queries submission to the corresponding systems. Results fusion, finally, consists in merging returned by different systems (from different selected collections) using appropriate results fusion models.

In content-based image retrieval, among the rare works dealing with data fusion, we cite [10], [14], and [1]. In [10], a data fusion model working on distributed collections of images is proposed based on a normalization procedure of similarities among the various image collections. In [14], a results fusion model working on a centralized image collection is proposed based on multiple representations, called viewpoints or channels, of both the query and the images in the database. They used four channels: the original color images, their corresponding grey-level images and their negatives. Results merging coming from different channels is shown to improve performance in a very important way. In [1], a results fusion approach based on multiple queries was used to tackle the problem of invariant image retrieval.

The work presented in this paper explores the idea of results fusion and applies it in the case of texture retrieval. Texture content is represent by two different models: the autoregressive model and a perceptual model based on a set of perceptual features such as coarseness and contrast. The perceptual model is considered in two viewpoints: the original images viewpoint and the autocovariance function viewpoint. Computational measures are based on these two viewpoints. So we have a total of three models/viewpoints (called representations). Benchmarking presented at the end of the paper show how a multiple representations and results fusion approach to CBIR can improve, in an incredible way, the search effectiveness (relevance) without, necessarily, altering, in an important way, search efficiency.

The rest of this paper is organized as follows: In section 2, we present the multiple representation models considered in this paper and we discuss briefly their capacity to model textures; We also show the benefits from using multiple representations and present the results fusion models used to fuse results returned by different representations; In section 3, benchmarking over the well-known Brodatz database using the recall graph is presented and discussed, and comparison to related works is given; And finally, in section 4, a conclusion is given and further investigations related to this work are briefly depicted.

2 Multiple Representations, Similarity Matching, and Results Fusion

2.1 Multiple Representations

To represent content of textures, we use two different models, the autoregressive model and a perceptual model based on a set of perceptual features[7]. The autoregressive (AR) model used is a causal simultaneous AR model with a non-symmetric half-plan (NSHP) neighborhood with four neighbors. The perceptual model is considered with two viewpoints: the original images viewpoint and the autocovariance function (associated to original images) viewpoint. Each of the viewpoints of the perceptual model used is based on four perceptual features, namely coarseness, directionality, contrast and busyness. So we have a total of three content representations, each having a parameter vector of size four for a total of twelve parameters.

The autoregressive model is characterized, in particular, by a forecasting property that allows to predict the grey-level value of a pixel of interest in an image by using the grey-level values of pixels in its neighborhood. The autoregressive model, when used to model a textured image, allow to estimate a set of parameters (their number corresponds to the number of neighbors considered), each one corresponds to the contribution of its corresponding pixel in the forecasting of the pixel of interest (the total of contributions of all pixels in an image is close to 100%).

The perceptual model, which is perceptual by construction, is based on a set of four computational measures that simulate four perceptual features mentioned above. Briefly, coarseness was estimated as an average of the number of extrema; Contrast was estimated as a combination of the average amplitude of the gradient, the percentage of pixels having the amplitude superior to a certain threshold and coarseness itself; Directionality was estimated as the average number of pixels having the dominant orientation(s); And finally, busyness was estimated based on coarseness since the two features are related to each other. The computational measures proposed for each perceptual textural feature were evaluated by conducting a set of experimentations taking into account human judgments and using a psychometric method. Thirty human subjects were asked to rank a set of textures according to each perceptual feature. Then, for each perceptual feature, we consolidate the different human rankings into one human ranking using the sum of rank values. For each feature, the consolidated human ranking obtained was compared to the ranking given by the corresponding computational measure using the Spearman coefficient of rank-correlation.

Experimental results showed very strong correspondence between the proposed computational measures and human rankings. Values of Spearman coefficient of rank-correlation r_s found are as follows: for coarseness, $r_s = 0.913$; for directionality, $r_s = 0.841$; for contrast, $r_s = 0.755$; and finally, for busyness, $r_s = 0.774$. Comparatively to related works, our results were found better. [7].

The set of features of the perceptual model have a perceptual meaning by construction. The set of features derived from the autoregressive model have no perceptual meaning by construction, however we have proposed in [3] a perceptual interpretation of the set of features derived from the autoregressive model. This perceptual interpretation consists in considering those features as a measure of the randomness/regularity of the texture. For more details on the perceptual model, refer to [7] and for more details on the autoregressive model, refer to [3], [2].

2.2 Similarity Matching

The similarity measure used is based on the Gower coefficient of similarity we have developed in our earlier work [6]. The non-weighted similarity measure, denoted GS, can be defined as follows:

$$GS_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{(k)}}{\sum_{k=1}^{n} \delta_{ij}^{(k)}} \tag{1}$$

Where $S_{ij}^{(k)}$ is the partial similarity between images i and j according to feature k, $\delta_{ij}^{(k)}$ represents the ability to compare two images i and j on feature k ($\delta_{ij}^{(k)} = 1$ if images i and j can be compared on feature k and $\delta_{ij}^{(k)} = 0$ if not. $\sum_{k=1}^{n} \delta_{ij}^{(k)} = n$ if images i and j can be compared on all features $k, k = 1..n$.).

Quantity $S_{ij}^{(k)}$ is defined as follows:

$$S_{ij}^{(k)} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k} \tag{2}$$

Where R_k represents a normalization factor. R_k is computed on the database considered for experimentations and is defined as follows:

$$R_k = Max(x_{ik}) - Min(x_{ik}) \tag{3}$$

The weighed version of the similarity measure can be defined as follows:

$$GS_{ij} = \frac{\sum_{k=1}^{n} w_k S_{ij}^{(k)}}{\sum_{k=1}^{n} w_k \delta_{ij}^{(k)}} \tag{4}$$

Where w_k corresponds to the weight associated with feature k. As mentioned, w_k can be either the inverse of variance of feature k or the Spearman coefficient of rank-correlation. For more details on the similarity measure, please refer to [6].

2.3 Multiple Representations Fusion Benefits

Different representations of the same query or the images in the database, or different search strategies for the same query, etc. return normally different search results. Results fusion is then the merging of the different lists of results returned by the different models, representations, or queries to form a unique fused (merged) list which is, hopefully, more effective (relevant) than the separated lists [9], [16]. Given several list results returned by different representations, there are three important phenomena that can be observed [25], [16]:

- Skimming effect: Each model retrieve a subset of the relevant images and intersection between them is rather low. A relevant image is retrieved, often, by only one model. In this case, results fusion must consider images that are ranked in top positions in different lists.

- Chorus effect: Different models retrieve approximately the same results but with different ranks and similarity values. In this case, a relevant image is ranked by several models in top positions (not necessarily the same position). The fact that several models retrieve an image is a more convincing evidence or proof that this image is relevant to the query compared to the case where this image is retrieved by only one representation. Results fusion, in this case, must take in consideration all the representations used.
- Dark horse effect: Exceptionally, even a good model can return some irrelevant images for a given query. Generally, different models did not return the same irrelevant images. Results fusion, in this case, must consider all the representations and use appropriate techniques to eliminate irrelevant images.

Another important point in image retrieval is retrieval efficiency, which is closely related to the size of feature vectors used to represent the content of images. In fact, the more the size of feature vectors is large, the less the retrieval efficiency is good. Efficiency of fusion-based approaches is quite equivalent to traditional approaches since they use models separately in the matching and retrieval step, which means a reduced size of feature vectors compared to traditional approaches, even if they need to add a fusion step at the end. In general the fusion step is less costly than the matching and retrieval step.

2.4 Results Fusion Models

In literature on results fusion, in particular in the DIR field, many fusion models were presented and experimented including the use of maximum function, average function and other linear combination models [16], [26]. Generally, the proposed models are simple and, even though, they allow sometimes a drastic improvement in retrieval relevance. Results fusion, in our case, is the fusion of results returned by each of the three representations used to represent texture content. Results returned for a query contain mainly two pieces of information that can be used: similarity values (scores) and ranks. Any results fusion model may make use of one or both of these two pieces of information. We have used and experimented three basic results fusion models that are denoted FusMAX (or MAX), FusCL (or CL) and FusComb (or Comb) defined respectively as follows:

$$FusMAX_{ij} = MAX(GS_{M^k_{ij}}) \tag{5}$$

$$FusCL_{ij} = \frac{\sum_{k=1}^{K} GS_{M^k_{ij}}}{K} \tag{6}$$

$$FusComb_{ij} = \Pi_{k=1}^{K} GS_{M^k_{ij}} \tag{7}$$

where M^k represents model/viewpoint k, K represents the number of models/viewpoints used, i represents a given query, j represents images that are found similar to query i according to model M^k and $GS_{M^k_{ij}}$ is the similarity value between query i and image j when using model/viewpoint M^k. These fusion models use only the values of the similarity function returned by the considered model/viewpoint. Ranks can be also used. We have used them as weights. In fact, more an image is ranked at top positions, more is its weight in the fusion models. Thus, we can define a weighted version for each of the FusCL, FusMAX and FusComb model. In such weighted models, each image j is weighted with its rank in the list of results returned for query i using model M^k.

Fusion models FusCL and FusComb, both non-weighted and weighted, exploit the chorus effect since these models give more importance to images that are retrieved and ranked in top positions by different models/viewpoints. They also exploit the dark horse effect since an irrelevant image that is ranked in top positions by one model/viewpoint is not ranked at top positions in the fused list given that this irrelevant image is not ranked at top positions by the other models/viewpoints. The FusMAX model exploits the skimming effect, until some degree, since this model takes images that are classified in top positions in different results lists but it re-ranks them according to similarity values. Generally, when the chorus effect exists in an important way between different lists, the gain that we can obtain by exploiting the skimming effect becomes low and *vice-versa* [25].

3 Experimental Results and Benchmarking

3.1 Experimental Results

We have conducted a large experimentation on Brodatz database [12] [1]. This database contains originally 112 images. We have divided each of the 112 images in 9 tiles to obtain a total of 1008 128x128 images (112 images x 9 tiles per image). Among the 112 original images of Brodatz database, we have counted 29 highly non-homogeneous images. Creating a class of images from an original image by dividing it into tiles and considering them as similar is a questionable procedure. In fact, when the original image is highly non-homogeneous, the resulting tiles are not visually similar. Considering such images can be misleading. For this reason, and for benchmarking purposes, we consider only 83 queries (by excluding the 29 highly non-homogeneous images), each from a different class (we have taken the first image of each class corresponding to the top left corner tile).

Experimental results show that: 1. The autoregressive model in its non-weighted NSHP version perform better that the other versions of the autoregressive model; 2. The weighted version, using Spearman coefficients of rank-correlation, of the perceptual model based on original images performs better than the other versions of this model; 3. And, finally, the weighted version, using the inverse of variances, of the perceptual model based on the autocovariance function performs better than the other versions of this model. For results merging, the FusCL model gives the best results compared to the FusMAX model and gives similar results compared to FusComb model. So, in the following, we will show results for only these best models. Here is the list of notations used to name different models:

[1] We used the version available at http://www.ux.his.no/ tranden/brodatz.html

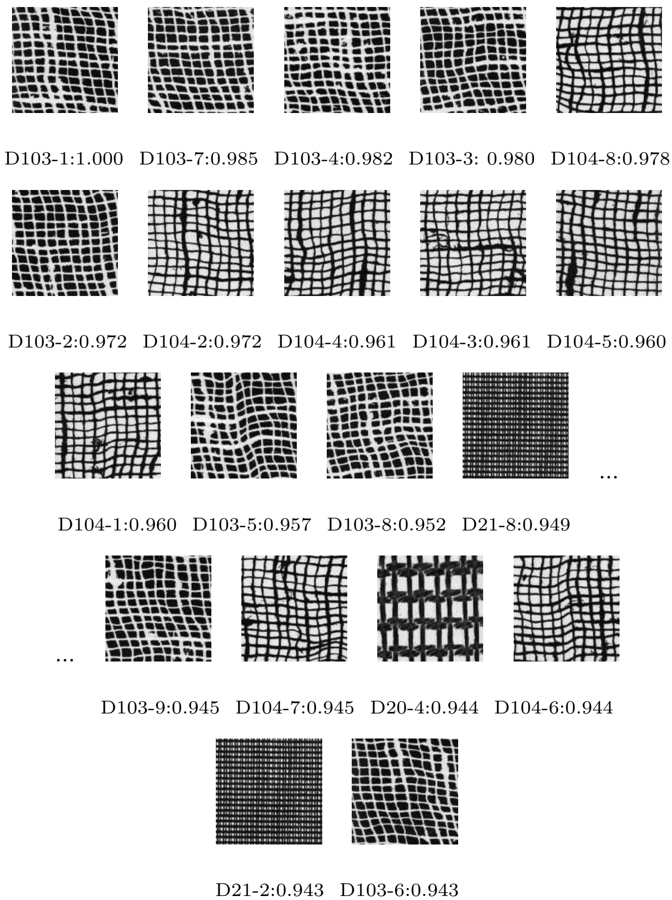
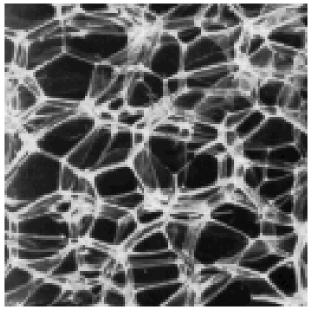


Fig. 1. Results returned for query image D103-1 using the PCP-COV-V model: images and similarities (scores). The results are quite good even if we used only one model (this is not the case always).

- **AR**: The autoregressive model with NSHP neighborhood.
- **PCP-COV-V**: Weighted combination, using the inverse of each feature variance, of the four perceptual features computed on the autocovariance function.
- **PCP-S**: Weighted combination, using the Spearman rank-correlation coefficients, of the four perceptual features computed on original images.
- **CL**: Fusion of **PCP-V**, **PCP-COV-V** and **AR** two by two or all of the three using the **FusCL** data fusion model.

The following figure (Fig. 1) shows an example of results obtained with the $PCP-COV-V$ model taken separately without fusion with other models. The results are quite good even if we used only one model (no fusion here).



D111-1

Fig. 2. Retrieval rate for image query when using separate models is respectively 0.33 with the PCP-S model or the PCP-COV-V model and 0.22 with the AR model. With the fused model, the retrieval rate reaches 0.88.

Of course, this is not the case always and most of retrieval cases will require fusion of multiple models to obtain acceptable results. For example, for image D111-1 (Fig. 2), we found a retrieval rate of 0.33 with the PCP-S model or the PCP-COV-V model and a retrieval rate of 0.22 with the AR model while the retrieval rate for image D111-1 when we fused all these 3 models was improved in an important way and reaches 0.88.

3.2 Recall Graph

Recall is quite a standard technique used to benchmark search relevance (effectiveness) in information retrieval systems in general. Recall, which can be defined as the number of relevant and retrieved images divided by the number of relevant images in the database for the considered query, measures the ability of a model to retrieve all relevant images. Recall is computed for each query at each position. Then, average recall is computed as an average across a set of representative queries.

Figure 3 shows the recall graph. From this figure, we can point out that the overall performance of the different models is as follows (in a decreasing order): **CL**, **AR + PCP-S**, **AR + PC-COV-V**, **PCP-S + PCP-COV-V**, **AR**, **PCP-S** and **PCP-COV-V**. The fused model CL (using all of the three basic representations) gives the best results. The fusion two by two also gives better results than the separated models. The perceptual model using the original images viewpoint (**PCP-S**) performs better than the perceptual model using the autocovariance function viewpoint (**PCP-COV-V**), but when these two viewpoints are fused, the resulting model (**PCP-S + PCP-COV-V** performs better than each of them taken separately. The autoregressive model (**AR**) performs better than the perceptual model (**PCP-COV-V**) based on the autocovariance function viewpoint and have a quite similar performance compared to the perceptual model based on the original images viewpoints (**PCP-S**).

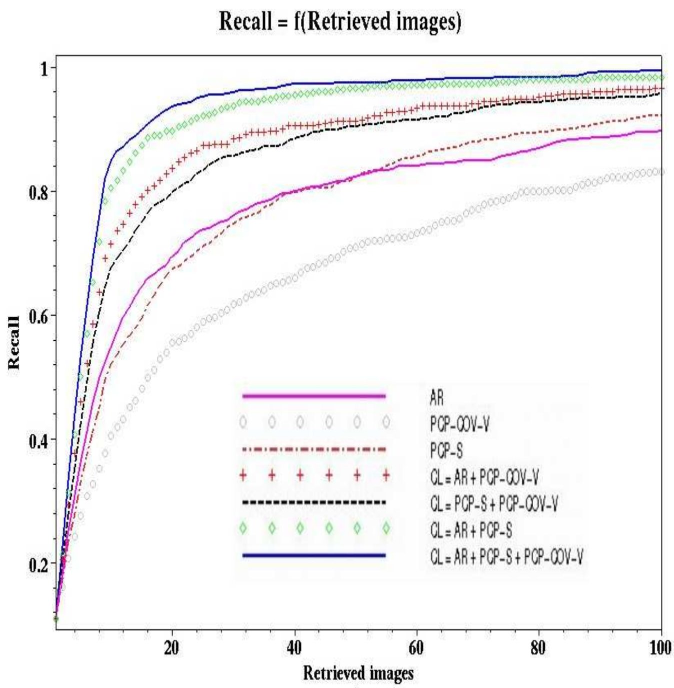


Fig. 3. Recall graph (Recall = f(Retrieved images)) for different separate models as well as Fused models. We can see that fusion of multiple representations, in particular the fusion of all three representation models (AR, PCP-S, and PCP-COV-V), outperforms all the other models.

3.3 Comparison to Related Works

When comparing retrieval performance in terms of recall rate with other works, we can point out the following remarks (see table 1):

Table 1. Average recall rate for different models. We used the rates given by authors of the corresponding model.

Model	**Recall rate**
FusCL (112 classes)	.687
FusCL (83 classes)	.819
MRSAR	.74
Gabor	.74
WOLD	.75
RBF	.737
MARS	.671

– If we consider only 83 classes, our fused model performs better than most of the known works including pure CBIR approaches such as Gabor filters [19], MRSAR [19], [17] and Wold model [17], and relevance feedback-based approaches such as MARS [22] and RBF-based retrieval [20]. Note that for table 1, we give the retrieval rate at the position that corresponds to the number of relevant images for each class. Note that in our approach no relevance feedback from users is used.
– If we consider all of the 112 classes, including highly non-homogeneous images, our model performs better than some and less than some other models. We must mention again that considering the 29 highly non-homogeneous classes may lead to incorrect conclusions since these classes contain images that are not visually similar.

4 Conclusion

An approach to CBIR based on multiple representations and results fusion has been presented in this paper. To demonstrate the power of such an approach, we have considered the case of textures. Texture content are represented by two different content representation models: the autoregressive model and a perceptual model based on a set of perceptual features such as coarseness, directionality, contrast, etc. Two viewpoints were considered in the case of the perceptual models: the original images and the autocovariance function. The similarity model used was based on Gower's coefficient of similarity. Experimental results and benchmarking against the well-known Brodatz database of textures was presented using the recall graph. The fused model is shown to improve in a very appreciable way retrieval performance compared to different single representations.

Extended research related to the work presented in this paper can be done through different directions, in particular the investigation of the possibility to define more representations as well as the possibility to use more complex fusion models.

References

1. Abbadeni, N., Alhichri, H.: Low-level invariant image retrieval based on results fusion. In: Proceedings of the IEEE ICME, Hannover-Germany (June 2008)
2. Abbadeni, N.: Texture Representation and Retrieval Using the Causal Autoregressive Model. In: Qiu, G., Leung, C., Xue, X.-Y., Laurini, R. (eds.) VISUAL 2007. LNCS, vol. 4781, pp. 559–569. Springer, Heidelberg (2007)
3. Abbadeni, N.: Perceptual meaning of the estimated parameters of the autoregressive Model. In: Proceedings of the International Conference of Image Processing, Genova-Italy, pp. 1164–1167. IEEE, Los Alamitos (2005)
4. Abbadeni, N.: Multiple representations, similarity matching, and results fusion for CBIR. Multimedia Systems Journal 10(5), 444–456 (2005)
5. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Transactions on Computing Surveys 40(2), 60 (2008)

6. Abbadeni, N.: A new similarity matching measure: application to texture-based image retrieval. In: Proceedings of the 3^{rd} International Workshop on Texture Analysis and Synthesis. IEEE, Nice-France (2003)
7. Abbadeni, N., Ziou, D., Wang, S.: Computational measures corresponding to perceptual textural features. In: Proceedings of the 15^{th} International Conference on Pattern Recognition, Barcelona-Spain, pp. 3913–3916. IEEE, Los Alamitos (2000)
8. Lew, M., Sebe, N., Djeraba, C., Jain, R.: Content-Based Multimedia Information Retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications, and Applications 26 (2006)
9. Belkin, N.J., Cool, C., Croft, W.B., Callan, J.P.: The effect of multiple query representation on information retrieval performance. In: Proceedings of the 16th International ACM SIGIR Conference, pp. 339–346 (1993)
10. Berretti, S., Del Bimbo, A., Pala, P.: Merging results for distributed content-based image retrieval. Multimedia Tools and Applications 24, 215–232 (2004)
11. Del Bimbo, A.: Visual information retrieval. Morgan Kaufmann Publishers, San Francisco (1999)
12. Brodatz, P.: Textures: A Photographic Album for Artists and Designers. Dover, New York (1966)
13. Dunlop, M.D.: Time, relevance and interaction modeling for information retrieval. In: Proceedings of the International ACM SIGIR Conference, Philadelphia, USA, pp. 206–213 (1997)
14. French, J.C., Chapin, A.C., Martin, W.N.: An application of multiple viewpoints to content-based image retrieval. In: Proceeding of the ACM/IEEE Joint Conference on Digital Libraries, pp. 128–130 (May 2003)
15. Gower, J.C.: A general coefficient of similarity and some of its properties. Biometrics Journal 27, 857–874 (1971)
16. Lee, J.H.: Analysis of multiple evidence combination. In: Proceedings of the ACM SIGIR Conference, Philadelphia, PA, USA, pp. 267–276 (1997)
17. Liu, F., Picard, R.W.: Periodicity, directionality and randomness: Wold features for image modeling and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(7), 722–733 (1996)
18. Lu, Y., Hu, C., Zhu, X., Zhang, H., Yang, Q.: A unified framework for semantics and feature based relevance feedback in image retrieval systems. In: Proceedings of the 8^{th} ACM International Conference on Multimedia, Marina Del Rey, CA, pp. 31–37 (2000)
19. Manjunath, B.S., Ma, W.Y.: Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence, special issue on Digital Libraries 18(8), 837–842 (1996)
20. Muneesawang, P., Guan, L.: An interactive approach for CBIR using a network of radial basis functions. IEEE Transactions on Multimedia 6(5), 703–716 (2004)
21. Payne, J.S., Hepplewhite, L., Stonham, T.J.: Texture, human perception, and information retrieval measures. In: Proceedings of the ACM SIGIR MF/IR Workshop (July 2000)
22. Rui, Y., Huang, T.S., Mehrota, S.: A Relevance feedback architecture for multimedia information retrieval systems. In: IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 82–89 (1997)
23. Sun, Y., Ozawa, S.: Semantic-meaningful content-based image retrieval in wavelet domain. In: Proceedings of the 5^{th} ACM International Workshop on Multimedia Information Retrieval (held in conjunction with ACM Multimedia), Berkeley, CA, pp. 122–129 (November 2003)

24. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. IEEE Transactions on Systems, Man and Cybernetics 8(6), 460–472 (1978)
25. Vogt, C.C., Cottrell, G.W.: Fusion via a linear combination of scores. Information Retrieval Journal 1, 151–173 (1999)
26. Wu, S., Crestani, F.: Data Fusion with Estimated Weights. In: Proceedings of the International ACM Conference on Knowledge and Information Management (CKIM), McLean, Virginie, USA, November 4-9, pp. 648–651 (2002)
27. Zhou, X.S., Huang, T.S.: Relevance feedback for image retrieval: a comprehensive review. ACM Multimedia Systems Journal 8(6), 536–544 (2003)