A Fuzzy Rank-Based Late Fusion Method for Image Retrieval

Savvas A. Chatzichristofis[1,2], Konstantinos Zagoris[1], Yiannis Boutalis[1,3], and Avi Arampatzis[1]

[1] Department of Electrical and Computer Engineering
Democritus University of Thrace, Xanthi 67100, Greece
[2] Informatics and Telematics Institute (ITI)
Centre for Research and Technology Hellas (CE.R.T.H.),Greece
[3] Department of Electrical, Electronic and Communication Engineering
Chair of Automatic Control, Friedrich-Alexander University of Erlangen-Nuremberg
Erlangen 91058, Germany
{schatzic,kzagoris,ybout,avi}@ee.duth.gr

Abstract. Rank-based fusion is indispensable in multiple search setups in lack of item retrieval scores, such as in meta-search with non-cooperative engines. We introduce a novel, simple, and efficient method for rank-based late fusion of retrieval result-lists. The approach taken is rule-based, employs a fuzzy system, and does not require training data. We evaluate on an image database by fusing results retrieved by three MPEG-7 descriptors, and find statistically significant improvements in effectiveness over other widely used rank-based fusion methods.

Keywords: Image Retrieval, Rank-Based Late Fusion, Fuzzy Systems, Heterogeneous Databases.

1 Introduction

Fusion in image retrieval is critical for the future of image retrieval research [5] and is not trivial [15]. Two main approaches to fusion have been taken: *early fusion*, where multiple image descriptors are composed to form a new one before indexing, and *late fusion*, where result rankings from individual descriptors are fused during query time. In general, late fusion approaches concern every technique for combining outputs of distinct systems [12] and can be accomplished either as a function of retrieval scores, or as a function of the position in which the results appear in each rank-list. In most cases, score-based late fusion is a better performer [2], but since in some practical situations scores are unknown, the use of rank-based fusion is necessary. A typical need for rank-based fusion arises in meta-search setups with non-cooperative search engines. Additionally, the score-based strategies, require a normalization among all systems in order to balance the importance of each of them, which is not the case of the rank-based strategies [12].

A commonly used method for rank-based fusion is Borda Count (BC), which originates from social theory in voting back in 1770. The image with the highest rank on each rank-list gets n votes, where n is the collection size. Each subsequent rank gets

K. Schoeffmann et al. (Eds.): MMM 2012, LNCS 7131, pp. 463–472, 2012.

one vote less than the previous. Votes across rank-lists are summed. Borda count is strictly equivalent to combSUM on ranks [12]. The literature about the Borda rule is very extensive (see [6] for references).

Alternatively, in methods such as Borda Count - Max (BC-MAX) and Borda Count - Min (BC-MIN) the final rank-list does not originate from the sum of the votes. In BC-MAX, the images are rated with the highest vote they get across rank-lists while in BC-MIN with the lowest. Another method often used is the Inverse Rank Position (IRP), which merges rank-lists in the decreasing order of the inverse of the sum of inverses of individual ranks. More details about IRP as well as about Borda Count derivatives are given in [7].

In [16], the traditional Borda method is extended by using the Ordered Weighted Averaging (OWA) operator to consider the risk-attitudinal characteristics. This new approach, entitled Borda-OWA, solves the group decision making problem in a more intelligent procedure. Classic BC does not consider the optimistic/pessimistic view of the system, which has a great effect on group decisions. Fusing several rank-lists, a system faces various types of uncertainty so the decision making process will be under risk. If the system strongly avoids the risk of making bad decisions, it will consider more rank-lists in the decision process. However, this will result in conservative decisions which are different than the decisions of a neutral or optimistic decision maker. In common words, the authors are using the terms 'Most of Them' and 'Few of Them' which could be modeled by fuzzy linguistic quantifiers and are used to characterize the aggregation inputs in an OWA operator. The term 'Most of Them' corresponds to the 'Pessimism' optimistic nature, while 'Few of Them' corresponds to the 'Optimism' optimistic nature.

In this paper we introduce a novel, simple, and efficient, rank-based late fusion method. The approach utilizes a Mamdani-type rule-based fuzzy system, and it does not require training data. We evaluate the effectiveness of the proposed method by fusing image rankings of a benchmark database for three MPEG-7 descriptors [10]: the Scalable Color Descriptor (SCD), the Edge Histogram Descriptor (EHD), and the Color Layout Descriptor (CLD). As illustrated from the experimental results, the proposed method provide statistically significant improvements in retrieval quality over other widely used rank-based fusion techniques such as IRP, Borda Count and derivatives.

The rest of the paper is organized as follows: Section 2 provides some details about fuzzy inference systems while Section 3 describes the proposed fuzzy rank-based late fusion technique. The experimental results are depicted in Section 4 and finally the conclusions are drawn in Section 5.

2 Fuzzy Inference Systems

Fuzzy inference is the process of determining the response of a system to a given input by using fuzzy logic and fuzzy linguistic rules for expressing the system's i/o relation. Its main characteristic is that it inherently performs an approximate interpolation between "neighboring" input and output situations [14]. The process comprises of four parts: Initially, (1) the fuzzification of the inputs using appropriately defined membership functions, (2) the designation of the type of the linguistic connection (fuzzy operator AND or OR) in the input variables, (3) the determination of the fuzzy output variables consequences using the fuzzy inference engine and a preset set of rules, and finally, (4) the defuzzification process. Fuzzy inference systems have been successfully applied in fields such as automatic control, decision analysis, expert systems, and computer vision.For more details, see [8].

Two main types of fuzzy modeling schemes are the Takagi-Sugeno model and the fuzzy relational model. The Mamdani scheme is a type of fuzzy relational model where each rule is represented by an *IF-THEN* fuzzy relationship which is numerically built by considering the linguistic rule, the type of the participating fuzzy membership values and the appropriate implication operator. Mamdani scheme is also called a linguistic model because both the antecedent and the consequent are fuzzy propositions [1]. Mamdani fuzzy rule-based systems are among the most popular approaches used in classification problems.

3 Fuzzy Rank-Based Late Fusion

In this section we are describing a Mamdani fuzzy rule-based system for rank-based late fusion. The parameters of the proposed fuzzy inference system are given in Table 1.

Table 1. Parameters of the fuzzy inference systems

Fuzzy modeling scheme	Mamdani
Inputs	3 membership functions for each input (Fig. 1)
Fuzzy operator in the input variables	AND
Fuzzy output variables	7 membership functions (Fig. 2) with 27 Rules
Defuzzification process	Centroid defuzzification method

Initially, we assume that the results of each rank-list can be divided into 3 fuzzy clusters according to their probability degree of similarity, i.e. **High**, **Medium** and **Low**. The membership functions (MF) of each class are illustrated in Fig. 1. The horizontal axis corresponds to the total number of results in the rank-list (in percentage) while the vertical one represents the membership value for each class. Position A defines the center of the class **Medium**, as well as the lower limits of the other 2 classes. A can be moved to the left or to the right of the position shown in Fig. 1 according to design preferences.

When dividing a rank-list in this manner, we assume that each result participates in all 3 classes but with a membership value. In the example outlined in Fig. 1, the result, activates the first membership function by 0.7 and the second by 0.3. This means that this result participates in the first class by 0.7, the second by 0.3 and the third by 0.0.

In each of the rank-lists of the 3 descriptors we employed, there is a corresponding fuzzy system which classifies the results into the 3 classes, with a participation value in each. The shape of all 3 systems is the same. The principle of the system operation is as follows:

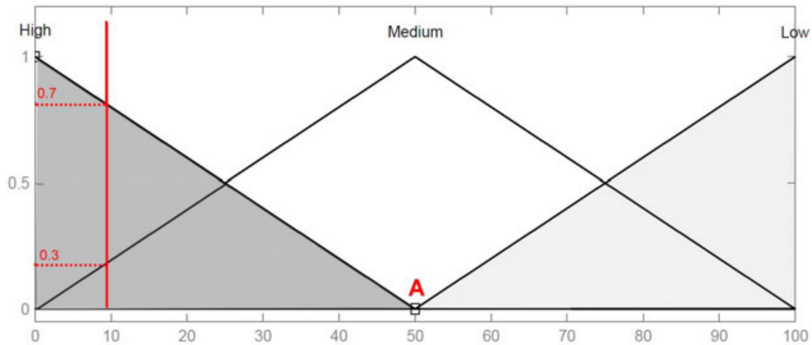

Fig. 1. The fuzzy input of the proposed system

The position of each result in each rank-list is defined as R_i^j, where i is the id of the result and j is the rank-list from which it originates. The value of R_i^1 interacts with the fuzzy membership functions of system 1 to get a membership degree in each class of the system. Similarly, R_i^2 and R_i^3 interact with the membership functions of systems 2 and 3 respectively.

The system output consists of $(2 \times i) + 1 = 7$ triangular membership functions, which are illustrated in Fig. 2. The fuzzy system employs 27 rules. These rules are given in Table 2.

Next, we explain how we build the rules using a "voting" concept. Activation (to any degree) of the **High** membership function (MF) of each input contributes with +1 vote, activation of the **Medium** MF contributes 0 votes and activation of the **Low** MF contributes with -1 votes. The output that corresponds to a particular input combination depends on the summation of the votes carried by the three inputs and is determined in respect to the central output MF which is the **MM** in Fig. 2.

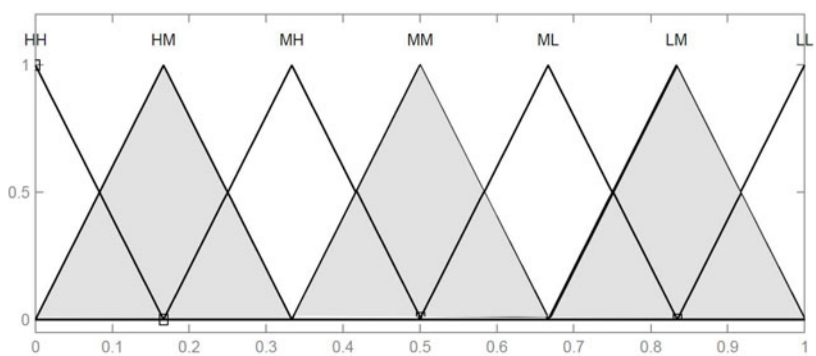

Fig. 2. The fuzzy output of the proposed system

Table 2. Fuzzy Inference Rules

RULE	IF **Input 1** is	AND **Input 2** is	AND **Input 3** is	THEN **Output** is
1	HIGH	HIGH	HIGH	LOW-LOW (LL)
2	HIGH	HIGH	MEDIUM	LOW-MEDIUM (LM)
3	HIGH	HIGH	LOW	MEDIUM-LOW (ML)
4	HIGH	MEDIUM	HIGH	LOW-MEDIUM (LM)
5	HIGH	MEDIUM	MEDIUM	MEDIUM-LOW (ML)
6	HIGH	MEDIUM	LOW	MEDIUM-MEDIUM (MM)
7	HIGH	LOW	HIGH	MEDIUM-LOW (ML)
8	HIGH	LOW	MEDIUM	MEDIUM-MEDIUM (MM)
9	HIGH	LOW	LOW	MEDIUM-HIGH (MH)
10	MEDIUM	HIGH	HIGH	LOW-MEDIUM (LM)
11	MEDIUM	HIGH	MEDIUM	MEDIUM-LOW (ML)
12	MEDIUM	HIGH	LOW	MEDIUM-MEDIUM (MM)
13	MEDIUM	MEDIUM	HIGH	MEDIUM-LOW (ML)
14	MEDIUM	MEDIUM	MEDIUM	MEDIUM-MEDIUM (MM)
15	MEDIUM	MEDIUM	LOW	MEDIUM-HIGH (MH)
16	MEDIUM	LOW	HIGH	MEDIUM-MEDIUM (MM)
17	MEDIUM	LOW	MEDIUM	MEDIUM-HIGH (MH)
18	MEDIUM	LOW	LOW	HIGH-MEDIUM (HM)
19	LOW	HIGH	HIGH	MEDIUM-LOW (ML)
20	LOW	HIGH	MEDIUM	MEDIUM-MEDIUM (MM)
21	LOW	HIGH	LOW	MEDIUM-HIGH (MH)
22	LOW	MEDIUM	HIGH	MEDIUM-MEDIUM (MM)
23	LOW	MEDIUM	MEDIUM	MEDIUM-HIGH (MH)
24	LOW	MEDIUM	LOW	HIGH-MEDIUM (HM)
25	LOW	LOW	HIGH	MEDIUM-HIGH (MH)
26	LOW	LOW	MEDIUM	HIGH-MEDIUM (HM)
27	LOW	LOW	LOW	HIGH-HIGH (HH)

Assuming the **MM** is the starting MF, each positive vote denotes a transition by one MF to the left, while a negative vote denotes a transition to the right. This way, if the inputs contribute with a sum of 1 vote, the output MF of the rule will be the **MH**. A +3 votes contribution means that the output MF of the rule will be the **HH**. On the contrary, -3 votes determines that the output MF of the rule is the **LL**.

Let that R_i^1 activates the input MF **High** by an activation degree $AV_{i,1} = 0.1$, R_i^2 activates the input MF Medium by $AV_{i,2} = 0.2$, and R_i^3 activates the input MF Low by $AV_{i,3} = 0.7$. Then the total votes will be:

$$(+1) + (0) + (-1) = 0$$

This means that the output MF will be the **MM** and the rule will be:

"*If R_1^i is **High** and R_2^i is **Medium** and R_3^i is **Low**, then the output is **MM***".

This procedure can be followed in all possible input combinations deriving 27 rules. In this particular example the rules' degree of fullfilment (DOF) is given by:

$$min(AV_{i,1}, AV_{i,2}, AV_{i,3}) = 0.1$$

An input combination may normally activate more than one rule, each one by a different DOF. The final crisp output is produced by using a conventional fuzzy inference procedure for Mamdani type systems, employing the min implication operator and the centroid defuzzification method. Centroid defuzzification method is also known as center of gravity or center of area defuzzification. This technique can be expressed as:

$$x* = \frac{\int \mu_i(x) x \partial x}{\int \mu_i(x) \partial x}$$

where $x*$ is the defuzzified output, $\mu_i(x)$ is the aggregated membership function and x is the output variable.

The 3 rank-lists are fused into a new one, with their results being sorted based on the values (in the range [0,1]) provided by the fuzzy system.

4 Experimental Results

In this study, we evaluate the retrieval effectiveness of the proposed late fusion technique which enable the combined use of the Scalable Color Descriptor (SCD), Edge Histogram Descriptor (EHD), and Color Layout Descriptor (CLD)[1], on a heterogeneous database suggested in [2]. This database consist of 20230 images; 9000 grayscale images are from the IRMA 2005 database[2]; 10200 are natural color images from the NISTER[13] database and 1030 artificially generated images are from the Flags database [4]. The database includes 40 fully-judged queries. The first 20 are natural color image queries from the NISTER database and the second 20 are grayscale queries of the IRMA 2005 database.

A detailed description of the experiment is demonstrated in the following steps and illustrated in Figure 3.

Initially, a query image interacts with the image retrieval system. The three MPEG-7 descriptors are calculated and the application executes the searching procedure using each one of the descriptors. For every descriptor the similarity matching technique recommended for this descriptor is employed.

For each descriptor, when the procedure is complete, the application arranges the images contained in the database according to their proximity to the query image, generating a ranking list. Overall, the system generates three individual ranking lists. Then, using either the proposed method, or a method from the literature, these three result lists are fused in order to generate the final ranking list.

For the evaluation of the performance of the proposed image retrieval method one of the metrics we employed is the Averaged Normalized Modified Retrieval Rank (ANMRR) [11]. The average rank AVR(q) for query q is:

[1] The source code for the MPEG-7 Descriptors is a modification of the implementation that can be found in the LIRe[9] retrieval library.

[2] IRMA is courtesy of TM Deserno, Dept. of Medical Informatics, RWTH Aachen.

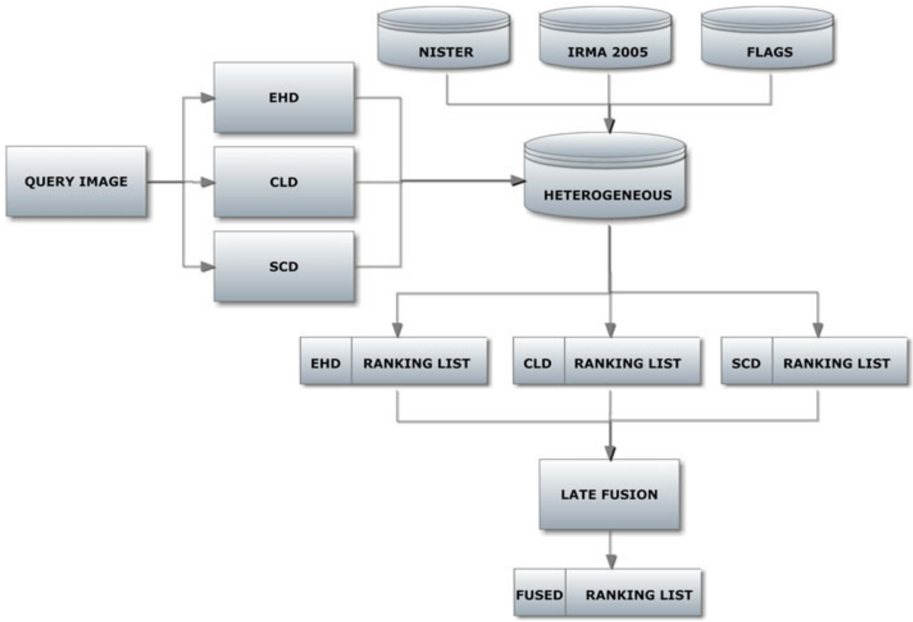

Fig. 3. Late fusion implementation process

$$\mathrm{AVR}(q) = \sum_{k=1}^{NG(q)} \frac{\mathrm{Rank}(k)}{NG(q)} \tag{1}$$

- $NG(q)$ is the number of ground truth images for query q
- $K = \min(X_{NG} \times NG(q), 2 \times \mathrm{GTM})$
- $\mathrm{GTM} = \max(NG)$.
- If $NG(q) > 50$ then, $X_{NG} = 2$ else $X_{NG} = 4$.
- $\mathrm{Rank}(k)$ is the retrieval rank of the ground truth image. Consider a query and assume that the kth ground truth image for this query q is found at position R. If this image is in the first K retrievals then $\mathrm{Rank}(k) = R$ else $\mathrm{Rank}(k) = (K + 1)$.

The modified retrieval rank is:

$$\mathrm{MRR}(q) = \mathrm{AVR}(q) - 0.5 \times [1 + NG(q)] \tag{2}$$

The normalized modified retrieval rank is defined as:

$$\mathrm{NMRR}(q) = \frac{\mathrm{MRR}(q)}{1.25 \times K - 0.5 \times [1 + NG(q)]} \tag{3}$$

and finally the average of NMRR over all queries is computed as:

$$\mathrm{ANMRR}(q) = \frac{1}{Q} \sum_{q=1}^{Q} \mathrm{NMRR}(q) \tag{4}$$

where Q is the total number of queries. The ANMRR has a range of 0 to 1 with the best matching quality defined by the value 0 and the worst by 1.

Apart from the ANMRR metric, we also evaluated the performance of the method using the Mean Average Precision (MAP) metric:

$$\text{Percision} = P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{5}$$

$$\text{Recall} = R = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{6}$$

The average precision AP is:

$$\text{AP}(q) = \frac{1}{N_R} \sum_{n=1}^{N_R} P_Q(R_n) \tag{7}$$

where R_n is the recall after the nth relevant image retrieved and N_R the total number of relevant documents for the query. MAP is computed by:

$$\text{MAP} = \frac{1}{Q} \sum_{q \in Q} \text{AP}(q) \tag{8}$$

where Q is the set of queries q.

The last evaluation metric that we employ is the Precision at 10 (P@10) and Precision at 20 (P@20) metrics that describe the system's capability to retrieve as many relevant results as possible in the first 10 and 20 ranked positions, respectively. This evaluation of the system's performance is critical for web based retrieval systems where the users are particularly interested in the credibility of the first results.

Additionally, we calculate how significant is the performance deviation between the methods. Significance test tell us whether an observed effect, such as a difference between two means, or a correlation between two variables, could reasonably occur just by chance in selecting a random sample. This application uses a bootstrap test, one-tailed, at significance levels 0.05, 0.01, and 0.001, against a baseline run.

The results are outlined in Table 3. As a baseline we assumed the Borda Count, which is one of the most commonly used methods in the literature for rank-based fusion.

All fusion methods beat the single descriptor performance. The best effectiveness overall is achieved by the proposed method; it beats BC by wide margins for all the A levels. BC-OWA (Neutral) results are the same with BC. This is inline with [16] which shows that BC is a special case of the Borda-OWA approach.

In Table 3, we also present the significance test results at significance levels of 0.05 ($^{\vartriangle\triangledown}$), 0.01 ($^{\blacktriangle\triangledown}$), and 0.001 ($^{\blacktriangle\blacktriangledown}$) against the BC baseline. The proposed fuzzy method significantly improves the results, for all the three A levels we experimented with, and in all 4 evaluation measures. MAP value improved by 6.25% comparing to BC, 21.7% comparing to EHD, 25% comparing to CLD and 93.9% comparing to SCD. ANMRR value improoved by 10.2% comparing to BC, 37.89% comparing to EHD, 34.33% comparing to CLD and 85.15% comparing to SCD.

Table 3. Experimental Results

	MAP	P@10	P@20	ANMRR
CLD	0.5046	0.4600	0.3837	0.4198
EHD	0.5183	0.5225	0.4525	0.4309
SCD	0.3254	0.2625	0.1875	0.5786
Borda Count (BC)	0.5973	0.5175	0.4237	0.3444
Fuzzy Fusion, $A = 5\%$	**0.6308**▲	0.5300▲	**0.4450**▲	**0.3125**▼
Fuzzy Fusion, $A = 10\%$	0.6147▲	0.5325△	0.4337△	0.3253▼
Fuzzy Fusion, $A = 50\%$	0.6123▲	**0.5350**▲	0.4350▲	0.3244▼
BC-OWA (Pessimism)	0.5540 ▼	0.4825 ▼	0.3962 ▼	0.3947▲
BC-OWA (Optimism)	0.5802-	0.4650 ▼	0.3837 ▽	0.3453 -
BC-OWA (Neutral)	0.5973-	0.5175-	0.4237-	0.3444-
BC-MAX	0.5552 ▼	0.4875 ▼	0.3962 ▼	0.3935▲
BC-MIN	0.5263 ▼	0.4375 ▼	0.3600 ▼	0.3849△
IRP	0.5574 ▼	0.4550 ▼	0.3687 ▼	0.3574-

5 Conclusions

We proposed a new, simple, and efficient, rank-based late fusion method, employing a fuzzy rule-based system with no need of training data. The method was found to provide statistically significant improvements in retrieval quality over other widely used rank-based fusion techniques such as IRP, Borda Count and derivatives. Although we evaluated on an image database, the method can be directly applied to other media as well. In order to have the proposed method to make sense, we assume that all the rank-lists in the group are considered to contribute equally to the final fused ranking. For the future, we suggest the dynamic calculation of both the number and limits of the Membership Functions of fuzzy system, based on training data.

The proposed method is implemented in the image retrieval system img (Rummager)[3] and is available online[3] along with the image database and the queries.

References

1. Babuska, R.: Fuzzy modeling for control. Kluwer Academic Publishers (1998)
2. Chatzichristofis, S.A., Arampatzis, A., Boutalis, Y.S.: Investigating the behavior of compact composite descriptors in early fusion, late fusion, and distributed image retrieval. Radioengineering 19(4), 725–733 (2010)
3. Chatzichristofis, S.A., Boutalis, Y.S., Lux, M.: Img(rummager): An interactive content based image retrieval system. In: Skopal, T., Zezula, P. (eds.) SISAP, pp. 151–153. IEEE Computer Society (2009)
4. Chatzichristofis, S.A., Boutalis, Y.S., Lux, M.: Spcd - spatial color distribution descriptor - a fuzzy rule based compact composite descriptor appropriate for hand drawn color sketches retrieval. In: Filipe, J., Fred, A.L.N., Sharp, B. (eds.) ICAART (1), pp. 58–63. INSTICC Press (2010)

[3] www.img-rummager.com

5. Datta, R., Joshi, D., Li, J., Wang, J.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40(2), 1–60 (2008)
6. Garcia-Lapresta, J., Martínez-Panero, M., Meneses, L.: Defining the borda count in a linguistic decision making context. Information Sciences 179(14), 2309–2316 (2009)
7. Jović, M., Hatakeyama, Y., Dong, F., Hirota, K.: Image Retrieval Based on Similarity Score Fusion from Feature Similarity Ranking Lists. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 461–470. Springer, Heidelberg (2006)
8. Lee, C.: Fuzzy logic in control systems: fuzzy logic controller. I. IEEE Transactions on systems, man and cybernetics 20(2), 404–418 (1990)
9. Lux, M., Chatzichristofis, S.A.: Lire: lucene image retrieval: an extensible java cbir library. In: El-Saddik, A., Vuong, S., Griwodz, C., Bimbo, A.D., Candan, K.S., Jaimes, A. (eds.) ACM Multimedia, pp. 1085–1088. ACM (2008)
10. Manjunath, B., Salembier, P., Sikora, T.: Introduction to MPEG-7: multimedia content description interface. John Wiley & Sons Inc. (2002)
11. MPEG-7. Subjective Evaluation of the MPEG-7 Retrieval Accuracy Measure (ANMRR). ISO/WG11, Doc. M6029 (2000)
12. Muller, H., Clough, P., Deselaers, T.: ImageCLEF: Experimental Evaluation in Visual Information Retrieval, vol. 32. Springer, Heidelberg (2010)
13. Nistér, D., Stewénius, H.: Scalable recognition with a vocabulary tree. In: CVPR (2), pp. 2161–2168. IEEE Computer Society (2006)
14. Ross, T.: Fuzzy logic with engineering applications. Wiley (2004)
15. van Leuken, R.H., Pueyo, L.G., Olivares, X., van Zwol, R.: Visual diversification of image search results. In: Quemada, J., León, G., Maarek, Y.S., Nejdl, W. (eds.) WWW, pp. 341–350. ACM (2009)
16. Zarghami, M.: Soft computing of the borda count by fuzzy linguistic quantifiers. Appl. Soft Comput. 11(1), 1067–1073 (2011)