Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

10-2013

On effects of visual query complexity

Jialie SHEN
Singapore Management University, jlshen@smu.edu.sg

CHENG ZHIYONG
Singapore Management University, zy.cheng.2011@phdis.smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons

Citation

SHEN, Jialie and CHENG ZHIYONG. On effects of visual query complexity. (2013). *The era of interactive media: Proceedings of the 12th IEEE Pacific Rim Conference on Multimedia: Sydney, December 20-22, 2011*. 531-541.

Available at: https://ink.library.smu.edu.sg/sis_research/3536

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Published in The Era of Interactive Media: Proceedings of the 12th IEEE Pacific Rim Conference on Multimedia: Sydney, December 20-22, 2011. pp. 531-541.
http://doi.org/10.1007/978-1-4614-3501-3_44

On Effects of Visual Query Complexity

Jialie Shen and Zhiyong Cheng

Abstract As an effective technique to manage large scale image collections, content-based image retrieval (CBIR) has been received great attentions and became a very active research domain in recent years. While assessing system performance is one of the key factors for the related technological advancement, relatively little attention has been paid to model and analyze test queries. This paper documents a study on the problem of determining "visual query complexity" as a measure for predicting image retrieval performance. We propose a quantitative metric for measuring complexity of image queries for content-based image search engine. A set of experiments are carried out using IAPR TC-12 Benchmark. The results demonstrate the effectiveness of the measurement, and verify that the retrieval accuracy of a query is inversely associated with the complexity level of its visual content.

Keywords Image • Performance • Evaluation • CBIR

1 Introduction

The emergence of the Internet and advancements in information technology can provide an economical reproduction and distribution platform for digitized media. As a powerful technique to manage large image collections, content-based image retrieval (CBIR) has been recently received great attentions and became a very active research domain. As a result, many different approaches have been proposed [1–10]. While a large amount of efforts have been invested in the domain, the technologies are still in their infancy. One of the major reasons for this stagnation is

J. Shen (✉) • Z. Cheng
School of Information Systems, Singapore Management University, 80 Stamford Road, Singapore 178902, Singapore
e-mail: jlshen@smu.edu.sg; zy.cheng.2011@smu.edu.sg

due to unavailability of a standard evaluation methodology to make cross method comparison to identify the state-of-the-art. Its importance has been recognized in the information retrieval and multimedia system community. There are four basic steps in the procedure to evaluate performance of CBIR systems [11–14]. They are,

1. construct an image test collection TC;
2. generate a set of visual queries and ground truth information GT;
3. run each image query set IQ through a particular CBIR system; and
4. assess performance of the system via an empirical distribution of particular measurement metric.

As shown above, the testbeds and the query sets involved could greatly influence the robustness and reliability of CBIR system evaluation. And the whole procedure may be strongly biased by human subjectivity. For example, in Step 2, some researchers can select images in favor of their algorithms as test queries. This can directly lead to unreliable system performance assessment. When interacting with CBIR system, users describe their information needs and search intent with an image query example. The central goal of corresponding search process is to rank visual documents according to their relevance/irrelevance to the given query. Based on the view, retrieval problem can be naturally modeled as one of binary classifications (a relevant class vs an irrelevant class). It is not hard to find that for a given CBIR technique, the higher complex visual queries are utilized, the poorer search (or classification) performance can be expected. This suggests that developing a systematical scheme to measure visual query complexity is very useful to gain a deeper structural insight into the behavior of CBIR system. It can enable us to carry out a comprehensive performance evaluation using different test queries. Another direct application domain is to use the measure to refine visual query structure. Motivated by this concern, a novel framework is proposed for quantitatively measuring the image complexity, in which images are represented by a set of visual keyblocks. To validate our proposed scheme, a set of experiments have been carried out using ImageCLEF test collection. The results demonstrate the effectiveness of the measurement, and verify that the retrieval accuracy of a query is inversely associated with the complexity level of its visual content. Further, our empirical study also shows that relatively fine-grained partition and larger codebook size can lead to more effective complexity computation. To the best of our knowledge, this is the first work on modeling visual query complexity quantitatively.

2 Estimating Image Complexity

2.1 Complexity Definition

Design of our framework is motivated by text processing. Each image is evenly segmented into blocks with regular geometric configuration such as rectangles. Each of the blocks is replaced with the most similar *keyblock* in precomputed *codebook*. With the indexes of *keyblocks* in the *codebook*, the image can be viewed as a document comprised by a matrix of indexes. Transforming the matrix using certain order (e.g., the scan order or the zig-zag order) to a one-dimensional (1-D) list, the image is thus syntactically analogous to a text document, which is essentially a list of terms. Based on the concept of *perplexity* in information theory, the measurement of image complexity can be defined as

$$C(I) = 2^{\mathrm{H(I)}} \tag{1}$$

where $H(I)$ is the entropy of an image I, which is estimated by the Shannon theorem [15]. Suppose that $\{w_1, w_2, \ldots, w_n\}$ is a list of indexes for image I in a database D; then the entropy is calculated as

$$H(I) = -\frac{1}{n} \log P(w_1, w_2, \ldots, w_n) \tag{2}$$

where $P(w_1, w_2, \ldots, w_n)$ is the probability of $\{w_1, w_2, \ldots, w_n\}$ over D.

2.2 Codebook Generation and Image Encoding

Blocks in an image are similar to the words in a text document. A codebook, which is similar to the word dictionary, needs to be precomputed. To achieve this, the related computation includes three steps: (1) every image in database is evenly partitioned into multiple blocks. For example, 3 × 3 means cutting an image into nine regions in 3 rows and 3 columns; (2) low-lever features (e.g., color, texture and shape) are then extracted from each block; and (3) a clustering algorithm, such as K-means, is applied to generate the codebook from the blocks in database. The keyblocks in the codebook are defined as the centers of the obtained clusters. The number of the clusters is the codebook size. For each block in an image, find the best match in the codebook and replace the block by the index of the best match code. Each image is then a matrix of indexes, which can be treated as 1-D code vector in the codebook. This property is similar to a text document, which can be viewed as a linear list of keywords.

2.3 N-block Model

This study assumes that the blocks in each image are not independent and they correlate with all the other blocks in the image. However, when the size of the codebook is large, modeling so many relations becomes difficult. So we make a second assumption that the blocks are connected in the order from left to right, and top to bottom. Each block is conditionally dependent on its previous words. In fact, this constraint can be removed by assuming more general models in arbitrary order. We only take this assumption for simplicity.

Given an image *I* encoded with a codebook *C*, let $\{w_1,w_2,\ldots,w_n\}$ denotes the code string of *I*. Based on the chain rule, the probability is written as

$$P(w_1, w_2, \ldots, w_k) = \prod_{k=1}^{n} P(w_k | w_1, w_2, \ldots, w_{k-1}) \tag{3}$$

where $P(w_k|w_1,w_2,\ldots,w_{k-1})$ is the conditional probability of code w_k given previous code $\{w_1,w_2,\ldots,w_{k-1}\}$. In reality, it is a serious problem to accurately compute the conditional probabilities. However, approximation can be possible based on some assumptions. We suppose that each code depends only on its immediate vertical and horizontal neighbors (analogy to the assumptions of *n- gram language model*).

According to the degree of dependency on remote codes, we defined three kinds of N-blocks models, uni-block, bi-block and tri-block. In uni-block model, the codes are considered independent to each other. In bi-block model, the probability of a given code depends only on the previous code, and in tri-block model the codes are assumed to depend on the preceding two codes. The three models can be expressed in Eq. (4–6) respectively.

$$P(w_1, w_2, \ldots, w_n) = \prod_{k=1}^{n} P(w_k) \tag{4}$$

$$P(w_1, w_2, \ldots, w_n) = \prod_{k=1}^{n} P(w_k | w_{k-1}) \tag{5}$$

$$P(w_1, w_2, \ \ldots, w_n) = \prod_{k=1}^{n} P(w_k | w_{k-2k-1}) \tag{6}$$

Notice that the assumptions can be generalized to an arbitrary length *n*. However, for a large *n*, it is not easy to approximate the conditional probabilities.

Let C^+ be the set of all nonempty code string with symbols in *C*, and *N*(*s*) denotes the occurrences of a code string $s \in \ C^+$ in the image database D. Then the estimated conditional probability can be formulated as follows according to uni-block model, for $\forall w_k \in C$

$$P(w_k) = \frac{N(w_k)}{\sum_{w \in C} N(w)} \tag{7}$$

In this model, the appearing of zero probability is avoided spontaneously because of the nature of clustering method.

Bi-blocks and tri-blocks, however, are sparsely distributed in the image. Thus a prior probability is assigned for an unseen bi-block $w_{k\text{-}1}w_k$. Accordingly, the amount of this prior probability should be discounted from the appearing words to meet the condition that the sum of probability is 1. The conditional probability of bi-block model is, for $\forall w_k, w_{k-1} \in C$

$$P(w_k|w_{k-1}) = \begin{cases} \dfrac{N(w_{k-1}w_k)}{N(w_{k-1})}\left(1 - \dfrac{N(w_{k-1})}{\sum_{w\in C} N(w)} \cdot \dfrac{1}{|C|}\right) & \text{if } N(w_{k-1}w_k) > 0 \\ \dfrac{N(w_{k-1})}{\sum_{w\in C} N(w)} \cdot \dfrac{1}{|C|} & \text{otherwise} \end{cases} \tag{8}$$

$|C|$ is the number of keyblock in codebook. With the tri-block model, the conditional probability is, for $\forall w_k, w_{k-1}, w_{k-2} \in C$

$$P(w_k|w_{k-2}w_{k-1}) = \begin{cases} \dfrac{N(w_{k-2}w_{k-1}w_k)}{N(w_{k-2}w_{k-1})}\left(1 - \dfrac{N(w_{k-2}w_{k-1})}{N(w_{k-2})|C|}\right) & \text{if } N(w_{k-2}w_{k-1}w_k) > 0 \\ \dfrac{N(w_{k-2}w_{k-1})}{N(w_{k-2})} \cdot \dfrac{1}{|C|} & \text{if } N(w_{k-2}w_{k-1}) > 0 \\ \dfrac{N(w_{k-2})}{\sum_{w\in C} N_w} \cdot \dfrac{1}{|C|} \cdot \dfrac{1}{|C|} & \text{otherwise} \end{cases} \tag{9}$$

Note that other more sophisticate smoothing methods, such as the one used in [16] which combines back-off and discounting, can be also applied here, instead of just assigning a small constant prior probability to the unseen bi-block or tri-block.

3 Experimental Setup

In the empirical study, we first study the effectiveness of the proposed measurement of image complexity using different N-block models, and then verify the assumption that the query image with higher complexity can lead to lower retrieval accuracy. A reliable measurement should be consistent with human cognitive perspective that an image with multiple objects or heterogeneous visual content is more complex than the one with single object or homogeneous visual appearance.

The visual content of image is closely associated with its complexity. According to the proposed computational method, the complexity is also influenced by the number of partition blocks in an image. Combining both factors, we can have a strong assumption about the characteristics of complexity varying with partition methods for images with different complexity level. For an image with homogenous visual content, its complexity should keep stable with increasing block number partitioned. This is because an image coded by 4 same blocks should have the same complexity as the image consisting of 20 same blocks. While for an image with heterogeneous visual content, there will be more different blocks when partitioning it into more blocks. Thus, its complexity should increase with the increase of block number. However, when the block number increases to some level, the complexity will gradually become stable. The main reason is that when the block size is small enough (maybe in pixel level), the image will be represented by the same number of different blocks, and the distribution of these different blocks will be the same even continuing to increase the block number. So the increasing level and growth rate of complexity with the increase of partition block number should be different for images with different levels of complexity. Here the increasing level means the level of partition granularity when complexity stops increasing. In other words, the more complex image, its complexity will keep increasing till finer partition (more blocks).

To verify these hypothses, two experiments are conducted on a standard test collection. *Experiment 1* is to compute the complexities of images in different concepts, and their complexity variation characteristics with the increase of block number. *Experiment 2* is to confirm the relations between query complexity and retrieval precision.

3.1 Test Collection and Visual Feature Extraction

The IAPR TC-12 Benchmark,[1] is constructed as test collection in our experiments. The collection consists of 20,000 still natural images taken from locations around the world and comprising an assorted cross-section of still natural images. This includes pictures of different sports and actions, photographs of people, animal, cities, landscapes and many other aspects of contemporary life. The benchmark has been used as the evaluation resources in *ImageCLEF* retrieval task for three years. To reduce the influence of image size variation, images with size 480 × 360 are selected as test collection. There are totally 10,971 images in this collection. In our experiments, a 64-D global feature [17] is calculated as composite visual feature representation for each image. The feature vector consists of three different kinds of visual features: 44-D color correlogram, 14-D color texture moment and 6-D RGB color moment. Then the three features are separately normalized into unit length and concatenated into the final 64-D feature. Euclidean distance is used as a similarity measurement.

[1] http://www.imageclef.org/photodata

Table 1 The size of codebooks for different partition methods

Partition	Codebook size	Partition	Codebook size	Partition	Codebook size
2 × 2	100	3 × 3	100	4 × 4	175
5 × 5	274	6 × 6	395	7 × 7	537
8 × 8	702	9 × 9	888	10 × 10	1,097

3.2 Experimental Methodology

Experiment 1: This experiment is to verify the effectiveness of the proposed complexity measure, and demonstrate the variation of complexity with different levels of granular partitions. In order to achieve the goal, 27 topic concepts are carefully selected which are blue sky, night shot of cathedrals, desert, surfing, tennis match, animal swimming, motorcyclist racing, sunset over water, ship, scenes in Inka-Trail, football match, snowcapped building, cycling, beach, accommodation with swimming pool, statue, scenes with building, tourist group, mountain, waterfall, church, wild animal, bedroom, meat dishes, school building, family shot, crowd of people. The partition methods with corresponding codebooks sizes are in Table 1.

Experiment 2: The goal of this experiment is to empirically verify the hypothesis that lower complex query images lead to better retrieval results than the ones with higher complexity. To facilitate the experimental study, three CBIR systems are applied and we use the standard queries in *ImageCLEF* photo retrieval tasks as the test image queries.

CBIR Systems: In order to demonstrate that the relation of query complexity and search performance is independent of the retrieval system, three retrieval methods are applied:

1. Vector Space Model: Since images are represented by an list of keyblock indexes in codebook, VSM approach used in text retrieval can be directly transplanted. The classical TF-IDF weigh is used to denote the block weights.
2. Histogram Model: In this model, the feature vectors **I** and **q** are the keyblock histograms. And the similarity measure is defined as

$$S(\mathbf{q},\mathbf{I}) = \frac{1}{1 + dis(\mathbf{q},\mathbf{I})} \tag{10}$$

$$dis(\mathbf{q},\mathbf{I}) = \sum_{i=1}^{N} \frac{|w_{i,I} - w_{i,q}|}{1 + w_{i,I} + w_{i,q}} \tag{11}$$

where the $\boldsymbol{q}$ and $\boldsymbol{I}$, which are the feature vectors of query and image I, are keyblock histograms. $w_{i,I}$ is the frequency of $c_i (c_i \in C)$ appearing in I. Similarly, $w_{i,q}$ is the frequency of c_i appearing in q.

3. SIMPLIcity: It is a region-based retrieval system proposed by Wang et al. [18]. In the system, an image is represented by a set of regions, roughly corresponding to objects, which are characterized by color, texture, shape, and location. The integrated region matching measure, which integrates properties of all the regions in the images, is used as similarity measurement. For more details, please refer to [18].

Query Topics: The standard query topics in *ImageCLEFphoto 2007* are used in this experiment. There are 60 query topics, and each contains three image examples. The query examples are not included in the test collection. Because the test collection only contains 480 × 360 images in *IAPR TC-12 Benchmark*, we remove some queries that may not have enough relevant images in the test collection. The assessment pool for each query topic are provided by *ImageCLEF*. If the number of relevant images for a query is less than 20 in the pool (after removing the ones whose size are not 480 × 360), then the query is excluded. Besides, the query examples with different size are also removed. There are 48 query topics with 104 query examples left after filtering.

Evaluation Criterion: Relevance assessment of results is based on the visual content of images. Typically a relevant image will have the subject of the topic in the foreground, and the image will not be too dark in contrast. We use *precision at K (P@K)* as performance evaluation measure. *P@K* is the proportion of relevant instances in the top K retrieved results, computed as $P@K = \frac{\text{No. of relevant results in top } K}{K}$. In the empirical study, only the top 20 results are evaluated and thus $K = 20$.

4 Experimental Results

The experimental results are presented and analyze in this section. We intended to apply three N-block models to compute the image complexity. Tri-blocks are sparse in the collection (many tri-blocks only appear once), however, which poses a significant effects on the results. Thus in the following discussion, we only consider the uni-block model and bi-block model.

4.1 Complexity Variation Characteristics

In *Experiment 1*, the complexities of images in different concepts are computed under different levels of partition granularity (namely different numbers of blocks in an image). We define that the complexity of a concept is the average complexity of all images in the concept. In uni-block model, the complexities of all concepts increase dramatically with the number of blocks increases. Besides, the relative complexity of different concepts is not consistent with our perspective either. For example, the obtained complexity of *blue sky* is higher than that of *touristgroup*

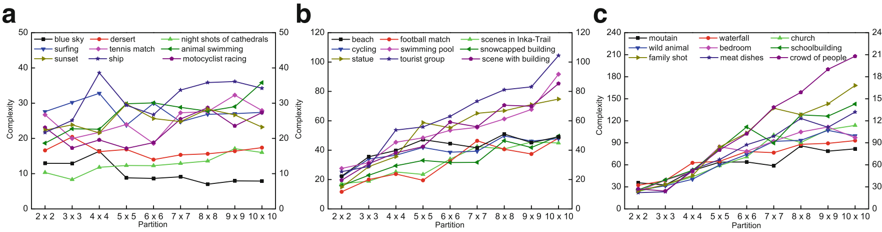

Fig. 1 When the number of partition blocks increases, the complexities of different concepts variation characteristics

in this model. These results are against our original intention described in Sect. 3. The reason for the results is that the blocks are assumed to be independent in uni-block model, and the spatial distribution information of blocks is ignored.

Figure 1 shows the results obtained by bi-block model. Concepts are divided into three groups according to the value of complexity (there are no strict bounds between groups). The concepts in (a) (we use (a) to denote Fig. 1a for simplicity, and it is the same to (b) and (c)) have the lowest complexity and those in (c) have the highest complexity. When the partition is relatively coarse (such as 2 × 2), the difference of complexity between some concepts are not very clear. As the partition become finer (more blocks), the complexities of some concepts increase faster while some grow relatively slower or even keep stable. The complexities of concepts in (a) roughly keep stable as the block number increases, because images in these concepts have simple or homogeneous visual content (e.g., ***bluesky*** and ***desert).*** The concept ***nightshotofcathedrals*** in this group is because the selected images show homogenous visual appearance especially the color. The complexities of concepts in (b) and (c) are increasing as the block number increases, and the complexity growth rates of concepts in (c) are generally larger than those of concepts in (b). Even only comparing the complexities of concepts in (b), the growth rates are different when the block number exceeds 25 (5 × 5). This phenomenon is exactly what we expect. As explained in Sect. 3, image with more complex visual content, the complexity increasing level and growth rate should be deeper and larger. However, the most fine-grained partition is limited to 10 × 10 in current experiment, the results have not shown different increasing levels. But we can anticipate that if the block number keeps increasing, the complexities of concepts in (b) will become stable before the complexities of concepts in (c). If we see the complexities of concepts in (a) in the way that the partition is already fine enough so that they keep stable, then the anticipation is also verified.

4.2 Relation Between Query Complexity and Query Performance

In ***Experiment 2***, query complexity is calculated based on 6 x 6 partition with codebook size 375. The VSM and Histogram Model (HM) adopt the same partition method and codebook. Figure 2a shows the trend of retrieval precision changes

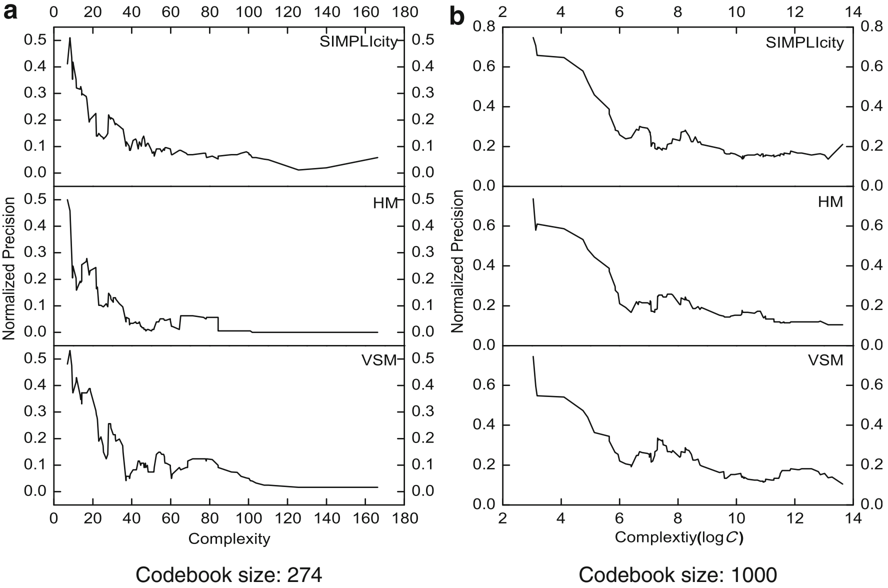

Fig. 2 The relationship between query complexity and query performance

when query complexity is increasing. The precisions in the figure are normalized results of three models, and the curves are processed by adjacent averaging smoothing method. Although there are some differences between results based on three models, the general trends are the same: the retrieval precision decreases as the query complexity increases. Because only the standard queries are used in the experiment, the number of queries is relatively small. It is unavoidable that there are some fluctuations in the statistical results. Besides, the relatively coarse partition and small codebook size which also have some impacts on the results. To demonstrate this point, we increase the codebook size to 1,000 to computer the query complexity, and the results are shown in Fig. 2b. The x axis here is the logarithm value of complexity, because the complexity varies in a very large range in this case. Comparing the two figures, we can see that with larger codebook, the change trends are smoother and more obvious.

5 Conclusion

In this study, a framework is developed to quantitatively measure the complexity of image. With this method, images used as query examples are partitioned into blocks and represented by indexes of keyblocks in the generated codebook. A set of experiments are carried out using a standard image collection. The results demonstrate the effectiveness of the measurement, and verify that the retrieval accuracy of a query is inversely related to the complexity level of its visual content. Besides, the results also show that it is better to use relatively fine-grained partition and larger codebook size to compute the complexity.

References

1. J. Laaksonen, M. Koskela, S. Laakso, and E. Oja. How to complete performance graphs in content-based image retrieval: Add generality and normalize scope. IEEE TPAMI, 2005.
2. J. Huang, S. R. Kumar, and M. Mitra. Combining supervised learning with color correlograms for content-based image retrieval. In ACM Multimedia, 1997.
3. N. Vasconcelos and M. Kunt. Content-based retrieval from image databases: current solutions and future directions. In ICIP (3), 2001.
4. AHH. Ngu, Q. Sheng, D. Huynh, and R. Lei. Combining multi-visual features for efficient indexing in a large image database. The VLDB Journal, 2001.
5. W. Ma and B. Manjunath. Netra: A toolbox for navigating large image database. In Proceedings of IEEE International Conference on Image Processing, 1997.
6. Y. Chen and J. Z. Wang. Image categorization by learning and reasoning with regions. Journal of Machine Learning Research, 2004.
7. M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. ACM TOMCCAP, 2006.
8. Z.-J. Zha, X.-S. Hua, T. Mei, J. Wang, G.-J. Qi, and Z. Wang. Joint multi-label multi-instance learning for image classification. In CVPR, 2008.
9. Z.-J. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T.-S. Chua, and Xian-S. Hua. Visual query suggestion: Towards capturing user intent in internet image search. ACM TOMCCAP, 2010.
10. Y. Gao, M. Wang, H.-B. Luan, J. Shen, S. Yan, and D. Tao. Tag-based social image search with visual-text joint hypergraph learning. In ACM Multimedia, 2011.
11. A. Narasimhalu, M. Kankanhalli, and J. Wu. Benchmarking multimedia databases. Multimedia Tools and Applications, 1997.
12. Proceedings of mira 99: Evaluating interactive information retrieval, glasgow, scotland, uk, april 14–16, 1999. In S. W. Draper, M. D. Dunlop, I. Ruthven, and C. J. van Rijsbergen, editors, MIRA, Workshops in Computing. BCS, 1999.
13. J. Shen and J. Shepherd. Efficient benchmarking of content-based image retrieval via resampling. In ACM Multimedia, 2006.
14. C. D. Manning, P. Raghavan, and H. Schutze. Introduction to Information Retrieval. Cambridge University Press, 2008.
15. C.Shannon. Prediction and entropy of printed english. Bell Syst. Tech. J., 1951.
16. L. Wu, M. J. Li, Z. W. Li, W. Y. Ma, and N. H. Yu. Visual language modeling for image classification. In MIR, 2007.
17. X. Li, C. G. M. Snoek, and M. Worring. Learning tag relevance by neighbor voting for social iimage retrieval. In MIR, 2008.
18. J. Z. Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE TPAMI, 2001.