9-2007

Cross-Language and Cross-Media Image Retrieval: An Empirical Study at ImageCLEF2007

Steven C. H. HOI
Singapore Management University, chhoi@smu.edu.sg

Cross-Language and Cross-Media Image Retrieval: An Empirical Study at ImageCLEF2007

Steven C.H. Hoi

School of Computer Engineering
Nanyang Technological University
Singapore 639798
{chhoi}@ntu.edu.sg

Abstract. This paper summarizes our empirical study of cross-language and cross-media image retrieval at the CLEF image retrieval track (ImageCLEF2007). In this year, we participated in the ImageCLEF photo retrieval task, in which the goal of the retrieval task is to search natural photos by some query with both textual and visual information. In this paper, we study the empirical evaluations of our solutions for the image retrieval tasks in three aspects. First of all, we study the application of language models and smoothing strategies for text-based image retrieval, particularly addressing the short text query issue. Secondly, we study the cross-media image retrieval problem using some simple combination strategy. Lastly, we study the cross-language image retrieval problem between English and Chinese. Finally, we summarize our empirical experiences and indicate some future directions.

1 Introduction

Digital image retrieval has attracted a surge of research interests in recent years. Most existing Web search engines usually search images by text only. They have yet to solve the retrieval tasks very effectively due to unreliable text information. Until now, general image retrieval is still a challenging research task. In this paper, we study the methodology of cross-language and cross-media retrieval techniques to attack some open challenges at ImageCLEF.

In this participation, we offer major contributions in three aspects. Firstly, we study an empirical evaluation of language models and smoothing strategies for cross-language image retrieval. Secondly, we conduct an evaluation of cross-media image retrieval, i.e., combining text and visual contents for image retrieval. The last contribution is the empirical evaluation of a methodology for bilingual image retrieval spanning English and Chinese sources.

The rest of this paper is organized as follows. Section 2 reviews some methodology of the TF-IDF retrieval model and the language model for information retrieval. Section 3 presents our implementation for this participation, and outlines our empirical study on cross-language and cross-media image retrieval. Section 4 set out our conclusions.

C. Peters et al. (Eds.): CLEF 2007, LNCS 5152, pp. 538–545, 2008.

2 Review of Language Models and Smoothing Techniques

In our approaches, we have conducted an extensive set of experiments to evaluate the performance of state-of-the-art language models and smoothing techniques with applications to text-based image retrieval tasks. Specifically, two retrieval models are studied: (1) the TF-IDF (Term Frequency-Inverse Document Frequency) model, and (2) the KL-divergence language model. Three smoothing strategies [1] are evaluated: (1) the Jelinek-Mercer (JM) method, (2) Bayesian smoothing with Dirichlet priors (DIR), and (3) Absolute discounting (ABS).

2.1 TF-IDF Retrieval Model

The TF-IDF retrieval model is a well-known method for text-based retrieval [2]. In general, a document and a query can be represented as a term frequency vector $\mathbf{d} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{q} = (y_1, y_2, \ldots, y_n)$ respectively, where n is the number of total terms, x_i and y_i are the frequency (counts) of term t_i in the document vector $\mathbf{d}$ and query vector $\mathbf{q}$, respectively. In a retrieval task, given a document collection $\mathcal{C}$, the IDF of a term t is defined by $log(N/n_t)$, where N is the total number of documents in $\mathcal{C}$, and n_t is the number of documents that contain the term t. For the TF-IDF representation, all terms in the query and documents vectors are weighted by the TF-IDF weighting formula, i.e., $\mathbf{d}' = (tf_d(x_1)idf(t_1), tf_d(x_2)idf(t_2), \ldots, tf_d(x_n)idf(t_n))$ and $\mathbf{q}' = (tf_q(y_1)idf(t_1), tf_q(y_2)idf(t_2), \ldots, tf_q(y_n)idf(t_n))$. For a simple TF-IDF retrieval model, one simply takes $tf_d(x_i) = x_i$. One can also define some other heuristic formula for the TF function. For example, the Okapi retrieval approach is a special case of TF-IDF model by defining the document TF formula [3] as: $tf_d(x) = \frac{k_1 x}{x + k_1(1 - b + b\frac{l_d}{l_{\mathcal{C}}})}$, where k_1 and b are two parameters for the document TF function, l_d and $l_{\mathcal{C}}$ are the lengths of the given document and collection, respectively. Similarly, a query TF function can be defined with parameters k_1 and b as well as l_q representing the average length of queries. In TF-IDF retrieval models, cosine similarity is often adopted as similarity measure.

2.2 Language Modeling for Information Retrieval

Language model, or the statistical language model, employs a probabilistic mechanism to generate text. The earliest serious approach for a statistical language model may be tracked to Claude Shannon [4]. To apply his newly founded information theory to human language applications, Shannon evaluated how well simple n-gram models did at predicting or compressing natural text. In the past, there has been considerable attention paid to using the language modeling techniques for text document retrieval and natural language processing tasks [5].

The KL-Divergence Measure. Given two probability mass functions $p(x)$ and $q(x)$, $D(p||q)$, the Kullback-Leibler (KL) divergence (or relative entropy) between p and q is defined as $D(p||q) = \sum_x p(x)log\frac{p(x)}{q(x)}$. One can show that $D(p||q)$ is always non-negative and is zero if and only if $p = q$. Even though it is not a true distance between distributions (because it is not symmetric and does not satisfy the triangle inequality), it is often still useful to think of the KL-divergence as a "distance" between distributions [6].

The KL-Divergence Based Retrieval Model. In the language modeling approach, we assume a query q is generated by a generative model $p(q|\theta_Q)$, where θ_Q denotes the parameters of the query unigram language model. Similarly, we assume a document d is generated by a generative model $p(q|\theta_D)$, where θ_Q denotes the parameters of the document unigram language model. Let $\hat{\theta}_Q$ and $\hat{\theta}_D$ be the estimated query and document models, respectively. The relevance of d with respect to q can be measured by the negative KL-divergence function [5]:

$$-D(\hat{\theta}_Q||\hat{\theta}_D) = \sum_w p(w|\hat{\theta}_Q) logp(w|\hat{\theta}_D) + (-\sum_w p(w|\hat{\theta}_Q) logp(w|\hat{\theta}_Q)) \qquad (1)$$

In the above formula, the second term on the right-hand side of the formula is a query-dependent constant, i.e., the entropy of the query model $\hat{\theta}_Q$. It can be ignored for the ranking purpose. In general, we consider the smoothing scheme for the estimated document model as follows:

$$p(w|\hat{\theta}_D) = \begin{cases} p_s(w|d) & \text{if word } w \text{ is present} \\ \alpha_d p(w|\mathcal{C}) & \text{otherwise} \end{cases} \qquad (2)$$

where $p_s(w|d)$ is the smoothed probability of a word present in the document, $p(w|\mathcal{C})$ is the collection language model, and α_d is a coefficient controlling the probability mass assigned to unseen words, so that all probabilities sum to one [5]. We discuss several smoothing techniques in detail below.

2.3 Three Smoothing Techniques

In the context of language modeling, the term "smoothing" can be defined as the adjustment of the maximum likelihood estimator of a language model so that it will be more accurate [1]. As the maximum likelihood estimator often underestimates the probabilities of unseen words in the given document, it is important to employ smoothing methods that usually discount the probabilities of the words seen in the text and assign the extra probability mass to the unseen words according to some model [1]. Specifically, three representative smoothing methods are used in our scheme:

Jelinek-Mercer (JM) smoothing: a linear interpolation of the maximum likelihood model with the collection model, using a coefficient λ to control the influence: $p_\lambda(\omega|d) = (1-\lambda)p_{ml}(\omega|d) + \lambda p(\omega|\mathcal{C})$, which is a simple mixture model [7].

Bayesian smoothing with Dirichlet Priors (DIR): the model is represented as: $p_\mu(\omega|d) = \frac{c(\omega;d)+\mu p(\omega|\mathcal{C})}{\sum_\omega c(\omega;d)+\mu}$, where μ in the is a DIR parameter that is estimated empirically from training sets [1].

Absolute discounting Smoothing (ABS): the model is represented as: $p_\delta(\omega|d) = \frac{\max(c(\omega;d)-\delta,0)}{\sum_\omega c(\omega;d)} + \sigma p(\omega|\mathcal{C})$, where $\delta \in [0,1]$ is a discount constant and

$\sigma = \delta|d|_\mu/|d|$, so that all probabilities sum to one. Here $|d|_\mu$ is the number of unique terms in document d, and $|d|$ is the total count of words in the document, i.e., $|d| = \sum_\omega c(\omega; d)$.

3 Cross-Language and Cross-Media Image Retrieval

The goal of the photographic retrieval task is to find as many relevant images as possible from an image collection given a multilingual statement describing a user information need. This task intends to simulate the text-based retrieval from photographs with multilingual captions, meanwhile queries for content-based image retrieval will also be offered. In this section, we study techniques to address several open challenges in this retrieval task, including (1) short text query problem, (2) cross-media image retrieval, and (3) cross-language retrieval. In the following section, we first describe the experimental testbed and setup at the ImageCLEF 2007, in which we have participated in the photo retrieval task. We will then conduct the empirical evaluations to address the above challenges and summarize our empirical experiences.

3.1 Experimental Testbed and Setup

The experimental testbed contains 20,000 color photographs with semi-structured captions in English, German and Spanish. For performance evaluations, there are 60 queries, each of them describes the user's information needs by short text in a range of languages including English, Italian, Spanish, French, German, Chinese, Japanese and Russian, and sample images.

For the photographic retrieval task, we have studied the query tasks in English and Chinese (simplified). Both text and visual information are used in our experiments. To evaluate the language models correctly, we employ the *Lemur* toolkit [1]. A standard list of stopwords provided by the Lemur toolkit is used in the parsing step.

To evaluate the influence on the performance of using the different schemes, we have evaluated the methods by trying a variety of different configurations in order to examine every aspects of the solutions. In particular, three groups of performance evaluations will be studied in the subsequent parts.

3.2 Evaluation of Language Models and Smoothing Techniques

In our experiments, we study several retrieval methods by language models with different smoothing techniques for the text-based image retrieval tasks. Table 1 shows the results of a number of our submissions with respect to the text based retrieval approaches by Language Models. The listed methods are ranked by the MAP (mean average precision) score. From the results, we can observe that the best approach is the "Eng-kl-dir-fb2" solution, which is based on the KL-divergence language model with the Dirichlet priors smoothing technique. We

[1] http://www.lemurproject.org/.

Table 1. Evaluation of language models for text-based image retrieval tasks

Run ID	Method	Query	Source	Modality	RunType	QE/RF	MAP	P10	REL_RET	REL
Eng-kl-dir-fb2	KL-DIR	English	English	TEXT	AUTO	FB	0.1660	0.2217	1827	3416
Eng-kl-jm-fb1	KL-JM	English	English	TEXT	AUTO	FB	0.1641	0.2017	1788	3416
Eng-tf-idf-fb3	TF-IDF	English	English	TEXT	AUTO	FB	0.1641	0.2150	1955	3416
Eng-kl-jm-fb2	KL-JM	English	English	TEXT	AUTO	FB	0.1640	0.2033	1870	3416
Eng-kl-abs-fb2	KL-ABS	English	English	TEXT	AUTO	FB	0.1635	0.2017	1757	3416
Eng-okapi-fb2	OKAPI	English	English	TEXT	AUTO	FB	0.1612	0.2333	1674	3416
Eng-kl-abs-fb1	KL-ABS	English	English	TEXT	AUTO	FB	0.1611	0.1950	1700	3416
Eng-kl-dir-fb1	KL-DIR	English	English	TEXT	AUTO	FB	0.1603	0.2117	1682	3416
Eng-kl-abs-fb3	KL-ABS	English	English	TEXT	AUTO	FB	0.1593	0.2000	1797	3416
Eng-kl-dir-fb3	KL-DIR	English	English	TEXT	AUTO	FB	0.1571	0.1867	1823	3416
Eng-kl-jm-fb3	KL-JM	English	English	TEXT	AUTO	FB	0.1566	0.1917	1860	3416
Eng-tf-idf-fb2	TF-IDF	English	English	TEXT	AUTO	FB	0.1560	0.2117	1842	3416
Eng-okapi-fb3	OKAPI	English	English	TEXT	AUTO	FB	0.1540	0.1950	1733	3416
Eng-tf-idf-fb1	TF-IDF	English	English	TEXT	AUTO	FB	0.1540	0.2133	1750	3416
Eng-okapi-fb1	OKAPI	English	English	TEXT	AUTO	FB	0.1492	0.2000	1726	3416
Eng-kl-abs	KL-ABS	English	English	TEXT	AUTO	NOFB	0.1455	0.1883	1570	3416
Eng-okapi	OKAPI	English	English	TEXT	AUTO	NOFB	0.1437	0.1850	1556	3416
Eng-kl-jm	KL-JM	English	English	TEXT	AUTO	NOFB	0.1428	0.1850	1547	3416
Eng-kl-dir	KL-DIR	English	English	TEXT	AUTO	NOFB	0.1419	0.1850	1554	3416
Eng-tf-idf	TF-IDF	English	English	TEXT	AUTO	NOFB	0.1341	0.1900	1539	3416

"TF-IDF" and "OKAPI" are two typical retrieval methods, "KL" denotes Kullback-Leibler divergence based model, "DIR" denotes the smoothing technique using the Dirichlet priors, "ABS" denotes the smoothing using the absolute discounting, and "JM" denotes the Jelinek-Mercer smoothing approach.

also found that the retrieval methods by KL-divergence language models do not always outperform the traditional TF-IDF and Okapi approaches, while the language models tend to outperform the TF-IDF and Okapi approaches on average. Further, we found that the retrieval methods with pseudo-relevance feedback (FB) consistently outperform the ones without any feedback. For example, the "Eng-kl-dir" approach is the KL-divergence language model approach using the Dirichlet priors smoothing technique without feedback, which achieved only a MAP score of 0.1419. However, by using relevance feedback, the MAP performance will be importantly improved, such as the "Eng-kl-dir-fb2" solution, which achieved a MAP score of 0.1660. Moreover, comparing several different smoothing techniques, there is no a clear evidence that which smoothing technique significantly outperform the others, though the Dirichlet priors smoothing approach achieved the best MAP performance among all runs. Finally, by examining the results of previous years [8], we found that the search tasks in this year seem to be more challenging for the text-based solutions.

3.3 Cross-Language Image Retrieval

In this part, we study the bilingual image retrieval using Chinese queries and English sources. To this purpose, the first step is to translate the Chinese queries into English. In our experiment, we simply test an online translation tool offered by Google [2].

Given the translated results, we conducted the experimental evaluations to examine the retrieval performance. Table 2 shows the experimental results of

[2] http://www.google.com/language_tools

Table 2. Evaluation for cross-language image retrieval tasks between Chinese (simplified) queries and English sources (#REL=3416)

Run ID	Method	Query	Source	Modality	RunType	QE/RF	MAP	P10	REL_RET
Chn-tf-idf-fb3	TF-IDF	Chinese S	English	TEXT	AUTO	FB	0.1574	0.2000	1874
Chn-kl-dir-fb3	KL-DIR	Chinese S	English	TEXT	AUTO	FB	0.1429	0.1650	1709
Chn-tf-idf-fb2	TF-IDF	Chinese S	English	TEXT	AUTO	FB	0.1413	0.1783	1790
Chn-kl-abs-fb3	KL-ABS	Chinese S	English	TEXT	AUTO	FB	0.1406	0.1667	1713
Chn-kl-abs-fb2	KL-ABS	Chinese S	English	TEXT	AUTO	FB	0.1385	0.1500	1732
Chn-kl-dir-fb2	KL-DIR	Chinese S	English	TEXT	AUTO	FB	0.1382	0.1600	1763
Chn-kl-jm-fb2	KL-JM	Chinese S	English	TEXT	AUTO	FB	0.1380	0.1533	1801
Chn-kl-jm-fb3	KL-JM	Chinese S	English	TEXT	AUTO	FB	0.1378	0.1600	1748
Chn-kl-jm-fb1	KL-JM	Chinese S	English	TEXT	AUTO	FB	0.1345	0.1533	1696
Chn-kl-dir-fb1	KL-DIR	Chinese S	English	TEXT	AUTO	FB	0.1333	0.1650	1672
Chn-okapi-fb3	OKAPI	Chinese S	English	TEXT	AUTO	FB	0.1312	0.1517	1646
Chn-kl-abs-fb1	KL-ABS	Chinese S	English	TEXT	AUTO	FB	0.1309	0.1417	1675
Chn-tf-idf-fb1	TF-IDF	Chinese S	English	TEXT	AUTO	FB	0.1286	0.1767	1553
Chn-okapi	OKAPI	Chinese S	English	TEXT	AUTO	NOFB	0.1268	0.1417	1404
Chn-kl-dir	KL-DIR	Chinese S	English	TEXT	AUTO	NOFB	0.1265	0.1467	1410
Chn-kl-abs	KL-ABS	Chinese S	English	TEXT	AUTO	NOFB	0.1264	0.1483	1411
Chn-kl-jm	KL-JM	Chinese S	English	TEXT	AUTO	NOFB	0.1252	0.1450	1415
Chn-okapi-fb1	OKAPI	Chinese S	English	TEXT	AUTO	FB	0.1237	0.1350	1654
Chn-tf-idf	TF-IDF	Chinese S	English	TEXT	AUTO	NOFB	0.1223	0.1567	1388
Chn-okapi-fb2	OKAPI	Chinese S	English	TEXT	AUTO	FB	0.1177	0.1383	1540

cross-language retrieval evaluation. From the experimental results, we found that the average retrieval performance of the bilingual retrieval tasks is less than the results of the single language image retrieval as shown in Table 1. For example, for a same retrieval method by the KL-divergence language model with the Dirichlet priors smoothing technique, the scheme “Chn-kl-dir-fb3” achieved only the MAP of 0.1429 in the bilingual retrieval task, while the same approach “Eng-kl-dir-fd3” can achieve the MAP of 0.1571 in the single langauge retrieval tasks. Nonetheless, the overall performance of the bilingual approach is quite impressive. In the future work, we will study other translation techniques to improve the results [9].

3.4 Cross-Media Image Retrieval

In this task we study the combination of text and visual information for cross-media image retrieval. We consider a simple combination scheme to combine the information from both the textual and visual modalities. Specifically, for a given query, we first rank the images using the language modeling techniques. We then measure the similarity of the top ranked images with respect to the sample images of the query. Finally, we combine the similarity values from both textual and visual modalities and re-rank the retrieval results based on the overall similarity scores.

In our experiment, three types of low-level visual features are engaged: color, shape, and texture [10,11]. For color features, we use the grid color moment. Each image is partitioned into 3×3 grids and three types of color moments are extracted for representing color content of each grid. Thus, an 81-dimensional color moment is adopted for the color feature. For shape features, we employ the edge direction histogram. A Canny edge detector is used to acquire the edge images and then the edge direction histogram is computed from the edges. Each histogram is quantized into 36 bins of 10 degrees each. An additional bin is used to count the number of pixels without edge information. Hence, a 37-dimensional edge direction histogram is used for the shape feature. For texture features, we adopt the Gabor feature [12]. Each image is scaled to 64×64. Gabor wavelet transformation is applied on the scaled image with 5 scale levels and 8 orientations, which results in 40 subimages. For each subimage, three moments are computed: mean, variance, and skewness. Thus, a 120-dimensional feature vector is adopted for the texture feature. In total, a 238-dimensional feature vector is employed to represent each of images in the testbed.

Table 3 [3] shows the cross-media retrieval results, in which we evaluate the influence of fusion coefficient. Specifically, the runs with ID from "Eng-kl-dir-fb2-tv1" to "Eng-kl-dir-fb2-tv9" represent the cross-media solution with the fusion coefficient from 0.1 to 0.9, respectively. The fusion coefficient here is the weight for the visual modality. From the experimental results, we can draw several observations. Firstly, we can see that the cross-media solutions improve the retrieval performance of the text-based approach for most cases with different fusion coefficients. Secondly, we found that the best MAP performance tends to be obtained when setting the fusion coefficient to 0.4. Moreover, we found that when the fusion coefficient increases, the precision of TOP 10 returned results tends to increase. This shows that when the visual modality accounts for more, the retrieval results become more accurate and relevant. This result again verifies the effectiveness of the proposed cross-media solutions.

Table 3. Evaluation for cross-media image retrieval tasks with queries of both textual and visual information (#REL=3416)

Run ID	Query	Method	Source	Modality	RunType	QE/RF	MAP	P10	REL_RET
Visual	Euclidean	Visual	Visual	VISUAL	AUTO	NO	0.0511	0.2067	883
Eng-kl-dir-fb2-tv1	KL-DIR	English	English	MIXED	AUTO	FB	0.1748	0.2317	2036
Eng-kl-dir-fb2-tv2	KL-DIR	English	English	MIXED	AUTO	FB	0.1789	0.2350	2018
Eng-kl-dir-fb2-tv3	KL-DIR	English	English	MIXED	AUTO	FB	0.1805	0.2400	1990
Eng-kl-dir-fb2-tv4	KL-DIR	English	English	MIXED	AUTO	FB	0.1811	0.2567	1954
Eng-kl-dir-fb2-tv5	KL-DIR	English	English	MIXED	AUTO	FB	0.1794	0.2583	1900
Eng-kl-dir-fb2-tv6	KL-DIR	English	English	MIXED	AUTO	FB	0.1776	0.2883	1807
Eng-kl-dir-fb2-tv7	KL-DIR	English	English	MIXED	AUTO	FB	0.1709	0.3183	1691
Eng-kl-dir-fb2-tv8	KL-DIR	English	English	MIXED	AUTO	FB	0.1483	0.3350	1534
Eng-kl-dir-fb2-tv9	KL-DIR	English	English	MIXED	AUTO	FB	0.0902	0.3000	1223

In future work, we will study more advanced combination methods. For example, we can train SVM classifiers with labeled images and then apply the classifiers to re-rank the top images from text retrieval. We can also study semi-supervised learning to exploit the unlabeled data for the retrieval task [13].

4 Conclusions

In this paper we reported our empirical study at the ImaegCLEF 2007 photo track. We have conducted three parts of empirical evaluations for three different purposes. One is to evaluate the language models and smoothing techniques

[3] This table has been updated by fixing some bug after the official evaluation.

with applications to text image retrieval. We found that the language models approaches did not achieve significantly promising results compared as we did in the ImageCLEF2005 campaign. The main reason is that the testbed in this year is totally different from 2005. In this year, images are only associated with very short text captions, which makes the text retrieval models less effective. The second evaluation is the cross-media image retrieval by combining both textual and visual information. Promising improvements were observed in our experiments. Finally, we also examined a commercial language translation tool for the cross-language retrieval tasks and found good retrieval results. In future work, we will study more effective techniques to improve current approaches.

Acknowledgements

The work in this paper was fully supported by a university startup grant from School of Computer Engineering, Nanyang Technological University, Singapore.

References

1. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to ad hoc information retrieval. In: ACM SIGIR Conference, pp. 334–342 (2001)
2. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley, Reading (1999)
3. Robertson, S., Walker, S., Jones, S., Hancock-Beaulieu, M.M., Gatford, M.: Okapi at trec-3. In: The Third Text REtrieval Conference (TREC-3), NIST
4. Shannon, C.E.: Prediction and entropy of printed English. Bell Sys. Tech. Jour. 30, 51–64 (1951)
5. Zhai, C., Lafferty, J.: Model-based feedback in the kl-divergence retrieval model. In: Proc. of Tenth International Conference on Information and Knowledge Management (CIKM 2001), pp. 403–410 (2001)
6. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, Chichester (1991)
7. Jelinek, F., Mercer, R.: Interpolated estimation of markov sourceparameters from sparse data. Pattern Recognition in Practice, 381–402 (1980)
8. Hoi, S.C.H., Zhu, J., Lyu, M.R.: CUHK at imageclef 2005: Cross-language and cross-media image retrieval. In: Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., Giampiccolo, D. (eds.) CLEF 2005. LNCS, vol. 4022, pp. 602–611. Springer, Heidelberg (2006)
9. Trujillo, A.: Translation Engines: Techniques for Machine Translation. Springer, London (1999)
10. Hoi, S.C.H., Lyu, M.R., Jin, R.: A unified log-based relevance feedback scheme for image retrieval. IEEE Trans. on Knowledge and Data Engineering 18(4), 509–524 (2006)
11. Wong, Y.M., Hoi, S.C.H., Lyu, M.R.: An empirical study on large-scale content-based image retrieval. In: Proc. IEEE Int. Conference on Multimedia & Expo. (ICME 2007) (2007)
12. Wu, P., Manjunath, B., Newsam, S., Shin, H.: A texture descriptor for browsing and similarity retrieval. Journal of Signal Processing: Image Communication 16(1-2), 33–43 (2000)
13. Hoi, S.C.H., Lyu, M.R.: A semi-supervised active learning framework for image retrieval. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005), New York, US, June 20–25 (2005)