Implicit Relevance Feedback for Content-Based Image Retrieval by Mining User Browsing Behaviors and Estimating Preference

Wei Dai, Wenbo Li, Zhipeng Mo, and Tianhao Zhao

Abstract—**Nowadays, Content-Based Image Retrieval has been the mainstay of image retrieval both in fields of research and application. To attain optimal retrieval results, relevance feedback (RF) methods are incorporated into CBIR by taking user's feedbacks into account. However, explicit RF methods rely heavily on active user engagement during search sessions, which is unrealistic in real applications. This paper presents an implicit RF method, Preference Estimation-based RF (PERF) for CBIR. PERF utilizes implicit user browsing histories to build a user preference model. The model will be refined iteratively and used to train a preference classifier for the user. In addition, an adaptive mechanism is adopted to realize the personalization of preference model. Our proposed method is tested and the experimental results reveal that PERF can achieve good retrieval precision with scarce explicit engagement from users.**

Index Terms—**CBIR, relevance feedback, implicit, browsing behaviors, preference model, adaptive mechanism.**

I. INTRODUCTION

Multimedia contents are increasing explosively and the need for multimedia retrieval grows quickly. Among researches in this field, image understanding is a hotspot. Multimedia mining becomes an important method of understanding images via extracting valuable knowledge from a large-scale multimedia repository [1]. In the development process of image retrieval system, traditional annotation based image retrieval faces more and more difficulties due to its heavy dependence on related captions [2], [3] and its two sufferings: high-priced manual annotation and inappropriate automated annotation.

Content-Based Image Retrieval (CBIR) is a current active topic in image retrieval studies. The purpose of CBIR is to present image retrieval results which have similar contents to user's query [4]. Most CBIR approaches are based on computation the similarities of low level features between user's query and images via a query by example system [5]. Although some search strategies are very powerful, it is hard to obtain an optimal retrieval result within one query process. Thus relevance feedback (RF) [6]-[9] techniques are incorporated into CBIR by taking into account user's

feedbacks to the results in the next retrieval process. In order to find user's real intention for the image retrieval, conventional RF approaches require users to evaluate retrieval results in multiple iterations. It is apparently unrealistic to require too much users' explicit involvement in retrieval iterations. To our best knowledge, there are two types of strategies to reduce users' burden: reducing iterations and easing users' burden in iterations.

For the former strategy, Su *et al.* [1] proposed Navigation-Pattern-based Relevance Feedback (NPRF) to achieve high retrieval quality of CBIR with RF by using discovered navigation patterns. Their work aims at reducing the redundant browsing and reaching exploration convergence quickly by mining the user query log. They report their method can achieve accuracy of over 90% in 6 iterations. As all RF methods, the retrieval accuracy of NPRF relies on users' active involvement. The faulty operations may lead to performance reduction. As for the latter strategy, the implicit RF methods are proposed. Different with explicit RF, the implicit RF technique [10], [11] is a kind of RF technique by gathering useful data indirectly by monitoring behaviors of the users during the search session instead of requiring much active user engagement. This kind of technique was first applied in retrieving documents and was brought into CBIR several years ago. A number of studies that employ implicit RF have been made to lower the users' burden. Auer *et al.* [12] proposed a system which infers users' intention from eye movements by using machine learning method. Then the system learns a similarity metric of common image features depending on the current interests of the user. Similarly, the system of [13] improves the performance of image retrieval by re-ranking the retrieved images according to color and texture features extracted from the regions where the users pay more attention. The users' interests are found by gazing information collected from an unobtrusive eye tracker. Unfortunately, performances of current implicit RF methods are frustrating compared to those of explicit ones.

To resolve the aforementioned problems, we propose a novel method, PERF to achieve the high retrieval quality of CBIR with implicit RF by monitoring user's behaviors during search sessions, such as browsing time, download, zoom-in and scrolling. In iterations of retrieval, the user browses images according to his/her custom and meanwhile four foresaid data will be collected by the system implicitly. User's preference values for his/her browsed images will be inferred automatically by the system. Then the preference values will be used to train an image classifier to improve the retrieved results for next iteration. Considering that browsing

customs of users vary, PERF incorporates an adaptive mechanism to achieve personalization when calculating user's preference values. PERF is tested on the image collection consisting of Corel image database and web images. The experimental results prove the robustness of PERF under different user engagement and show that through PERF, the optimal results can be attained in very few feedbacks.

The rest of the paper is organized as follows: Section II describes our method framework and technique details. In Section III, we show the experimental results and evaluations. Finally Section IV concludes the paper.

II. RELATED WORK

Relevance feedback [14]-[16], in principle, refers to a set of approaches learning from an assortment of users on image retrieval [17]. By collecting and learning from users' feedbacks in multiple iterations, the search algorithm would provide users with the retrieval result which is most similar to the target image. According to fashion of collecting feedbacks, RF is classified into explicit RF and implicit RF. Explicit RF approaches require users to give explicit feedbacks, e.g. grades and classifications, which have high accuracy and reliability. However, the defect that explicit RF approaches depend heavily on active user engagement becomes a bottleneck for its transformation to large-scale real applications. Implicit RF approaches collect user browsing behaviors as feedbacks. And the system would infer users' behavioral pattern by mining massive users' historical operational records. Despite the advantage of liberating users from heavy burden compared to explicit RF approaches, many implicit RF approaches pays their superiority back by performing poorer in improving the retrieval accuracy.

Due to the defect the implicit RF approaches have in retrieval accuracy improvement, explicit RF approaches are still the current mainstay. Some approaches start to find a new path to reduce users' burden during search sessions. Reference [1] proposed a novel method named NPRF to reduce users' burden by mining users' operational log and further accelerating the speed of convergence of retrieval accuracy. Although NPRF reduces users' burden during search sessions to some extent, it does not tackle the problem that it still relies heavily on active users' engagement. By careful and deep investigation on users' behaviors during search sessions, we find that the reason why some implicit RF approaches [11], [12] perform poorly in improving retrieval accuracy is that they are lacking an effective mining mechanism for users' implicit feedbacks.

III. METHOD

A. Overview

We propose an implicit RF approach to improve CBIR, named PERF. PERF incorporates a preference model to achieve effective implicit feedback mining as well as retrieval personalization. The steps of our method are summarized as follows:

1) Given an initial retrieval image set of user's query, the user is required to browse according to his/her habits. The system will implicitly record user's manipulations.

2) Preference value of each browsed image is estimated by adaptive preference model built on users' manipulation histories.

3) According to preference values, the browsed image set is divided into positive group and negative group, as well as three types of image features of image: color, edge and texture are extracted and used as support vector in training SVM classifier. Then the retrieval image set is reordered by giving priority to positive classified images.

4) The reordered image set is provided to users for further browsing. A stable preference model for each user is achieved after a few iterations.

5) The preference model and classifier can be reused and when meeting the similar queries from the same user.

B. Implicit Adaptive Preference Model

In the implicit adaptive preference model, we assume that each image has a preference value that reflects the degree of user's satisfaction. The retrieval results will be ranked in descending order by preference values. Thus the target of the model is to assign the proper preference values to browsed images.

When building the model, we utilize implicit relevance feedbacks. Four kinds of user's operations are considered: download, browsing time, scrolling and zoom-in. By analyzing the users' browsing behaviors, we find out there is an implied trend that the downloaded images are likely to be most satisfying. Moreover, according to some researches [10], [11] on user's behaviors, the images captioned with long browsing time, multiple scrolling and zoom-in interactive operations mean that they get more attention from the users.

According to users' behaviors during the search session, the initial images given to users can be classified into browsed images and un-browsed images. We can only collect the users' behaviors for browsed images because un-browsed images are totally ignored by the users. Having the records of relevance feedbacks, we calculate the preference value for each browsed image via the following formula.

$$V(I) = H^d [\sum_{k=1}^{3} (w_i \times f_i(I))]^{|d-1|} \qquad (1)$$

In the above formula, $V(I)$ is the preference value of browsed image I. d is a binary value which represents whether the image is downloaded or not. Other three implicit factors, browsing time, scrolling count and zoom-in count are symbolized as $S_i(I)$ ($i = 1, 2, 3$) with corresponding weights w_i ($i = 1, 2, 3$) respectively. H is the ceiling of the preference value of an image. In our method, we stipulate that the preference value of an image ranges from 0.0 to 1.0, thus having $H=1.0$ and $V(I)=0.0$ for those un-browsed images.

For browsed images, their preference values are calculated using weights w_i ($i = 1, 2, 3$) determined by the adaptive mechanism. For users with different browsing custom, the weights vary. In order to estimate w_i ($i = 1, 2, 3$) dynamically, a group of equations are built as follows:

$$F = [F_1, F_2, F_3], \quad F_i = [f_{1i}, f_{2i}, ..., f_{mi}]^T \quad (2)$$

F_i ($i = 1, 2, 3$) represents three implicit feedbacks to m browed images respectively, i.e. browsing time, scrolling count and zoom-in count. To make three aforementioned implicit feedbacks conform to the same criterion, we normalize Fi to F_i' and obtain:

$$F' = [F_1', F_2', F_3'], \quad F_i' = [f_{1i}', f_{2i}', ..., f_{mi}']^T \quad (3)$$

We thus establish the following equation to calculate the preference value of m browsed images:

$$\begin{bmatrix} V_1 \\ V_2 \\ ... \\ V_m \end{bmatrix} = \begin{bmatrix} f_{11}' & f_{12}' & f_{13}' \\ f_{21}' & f_{22}' & f_{23}' \\ ... & ... & ... \\ f_{m1}' & f_{m2}' & f_{m3}' \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \quad (4)$$

Furthermore, we get the simplified form $F'w = V$. The theory of Principal Components Analysis (PCA) is adopted to determine the optimal assignment of w_i ($i = 1, 2, 3$) by analyzing the eigenvalues of the covariance matrix of F'. According to the theory of PCA, the eigenvalue of each dimension denotes the energy of the corresponding dimension, which further reflects the fluctuation degree of data in the same dimension. We assume that the larger fluctuation degree of a dimension has, the more the corresponding implicit feedback contributes to the preference value of an image.

Thus, we first build the covariance matrix, C:

$$C = \frac{F'^T F'}{m-1}, \quad C \in R^{3 \times 3} \quad (5)$$

Then we calculate the eigenvector of C via the following equation:

$$CP = \lambda P, \quad P \in R^{3 \times 1} \quad (6)$$

where P is the eigenvector and λ is the corresponding eigenvalue. Meanwhile, we experimentally prepare a set of weights: $S = \{0.5, 0.3, 0.2\}$ which will be assigned to three foresaid implicit factors according to λ. Assuming that

$$\lambda = [\lambda_1, \lambda_2, \lambda_3]^T, \quad (7)$$

We obtain λ' by sorting λ_i in descending order:

$$\lambda' = [\lambda_1', \lambda_2', \lambda_3']^T \quad (8)$$

Furthermore, we adjust the sorting of w_i ($i = 1, 2, 3$) in w according to the sorting transformation from λ to λ':

$$w' = [w_1', w_2', w_3']^T \quad (9)$$

w_i' ($i = 1, 2, 3$) will be assigned values by corresponding s_i in S. Thus the value of w_i is obtained by conducting inverse sorting transformation from w' to w.

Having estimated w_i ($i = 1, 2, 3$), we use them to calculate the preference values of browsed images according (4). Then

the browsed images are classified into two groups according to the preference values. The threshold of preference value is 0.4 in our paper. Then a preference classifier is trained to classify the images which have not been browsed by the user.

C. Feature Extraction

We extract three image features for training, which are color histogram, edge and texture.

The color histogram is counted in HSV color space with H channel 32 bins, S channel 16 bins, and V channel 16 bins respectively. Therefore, a 64-dimension color feature vector is obtained for each image.

The edge feature is extracted in gray color space. Firstly we convert the color images to grayscale and use canny operators to get the edge of images. Then the edge images are divided into 256 regions, calculating the ratio of edge pixels in each region. This 256-dimension vector will be used as edge feature.

Similarly, the texture feature extraction is performed on grayscale images. We divide the gray images into several part-overlapping regions by 48 × 48 pixels. Then 2D Gabor filter is applied for each region. Basically, the importance of texture is related to the area. In other words, the texture covering more regions is considered more important. Thus in order to extract primary textures, we adopt K-means to cluster similar textures together and select several top representative textures as the texture feature for the image.

For positive images and negative images, all three types of features are extracted as support vectors in SVM training. Then the un-browsed images will be classified by the trained preference classifier. With browsed images, these classified images are reordered and showed to the user for next feedback iteration.

D. Iterative Relevance Feedback

Our proposed PERF is an iterative method. In each iteration, the system presents the reordered retrieval results and receives the further feedbacks from the user. On the basis of refined feedbacks, the adaptive preference model is rebuilt and new preference classifier is created. After several iterations, the stable preference model and the preference classifier are obtained for each user. As there is no special requirements on images such as non-overlapping rule in later iterations, the user tend to repetitively browse very satisfying images and some fresh satisfying images. The classification accuracy will stop improving when no new browsed images are added, thus our method converges in few iterations.

IV. EXPERIMENTS AND RESULTS

To evaluate the effectiveness of PERF, we conduct three experiments. First, the overall performance and convergence of PERF is evaluated. Then, we compare the performance of PERF under different degrees of user engagement.

A. Experimental Data

The experimental data came from the collection of the Corel image database and the web images. We prepared six data sets composed of different categories. The expected retrieval targets of six data sets are beach, earth, cloth, flower,

sail boat and wave. More details of the data can be found in Table I.

B. Evaluation of Implicit Adaptive Preference Model

To measure the accuracy of proposed adaptive mechanism for preference estimation, we ask the user to grade all experimental images, and then the estimated preference values are compared with the ground-truth. As mentioned, the images are classified into positive group and negative group by preference value at threshold 0.4. Furthermore, the estimation is regards correct if it is classified to the same group with value error within 0.05. The accuracy is defined by ratio between correct estimated images and all the images. The preference estimation accuracies of six data sets are shown in Fig. 1, which reveal that our proposed implicit adaptive preference model is reliable.

TABLE I: THE EXPERIMENTAL DATA AND RETRIEVAL PERFORMANCE

Image set	# of images	# of target images	Target ratio	Precision	Recall
beach	275	119	0.4327	0.7586	0.5546
earth	199	95	0.4774	0.8550	0.6941
cloth	200	100	0.5000	0.9901	1.0000
flower	215	72	0.3348	0.7817	0.5278
sail boat	198	78	0.3939	0.8837	0.4871
wave	219	87	0.3973	0.8261	0.7191

C. Evaluation of Performance and Convergence

To evaluate the effectiveness of our proposed approach, the measurement criteria, namely precision and recall, are used. Precision is defined as the ratio of correctly retrieved images to all retrieved images, and recall is defined as the ratio of correctly retrieved images to all relevant images should be retrieved.

Fig. 2 illuminates that the precision rises with the iteration proceeding. The results show that PERF reaches the stable precision in only four iterations. The average precision and recall of PERF for each data set is shown in Table I. As reported, implicit RF method Pinview [11] achieves the average precision 0.224 for a subset of the Corel image database. And an adaptive explicit RF method FARF [13] achieves the average precision over 50%, average recall about 10% for a subset of the Corel database containing 1400 images from 14 categories. By comparing with them, we can conclude that our method can effectively achieve pretty remarkable retrieval performance.

D. Evaluation of User Engagement

In addition to the effectiveness, another issue we concern on is required user engagement in our method. Conventionally, our method expects users to browse at least 20% images in each iteration, however, which is not a mandatory requirement. We test our method assuming that users reduce their participations in search sessions. Fig. 3 exhibits that our method is very robust despite negative user engagement, though more browsing operations can improve the retrieval performance.

V. CONCLUSION

To deal with the problem of heavy dependence on active user engagement in CBIR, we propose a new approach named PERF by utilizing four implicit feedbacks derived from user's browsing behaviors during search sessions to build a user preference model. PERF can adaptively estimate preference degree of browsed images by a proposed adaptive mechanism. Then the preference values are used to train a preference classifier for each user. The proposed method is evaluated on six data sets consisting of Corel image database and web images. The experimental results demonstrate that our method achieve good retrieval performance in very few iterations, providing the potential to be further developed into a post-processing tool for image retrieval application. Future work includes the improvement of the classifier and testing on large scale image set.
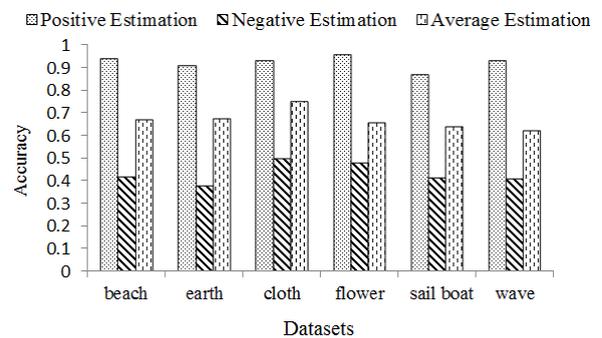

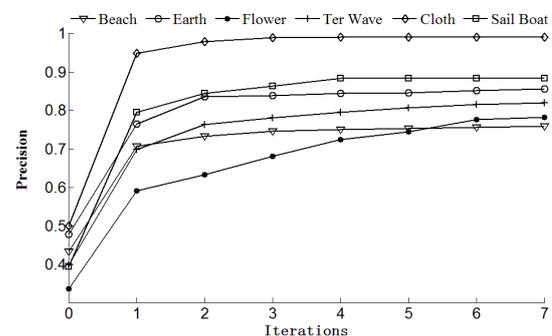Fig. 1. Preference estimation accuracies of six data sets.

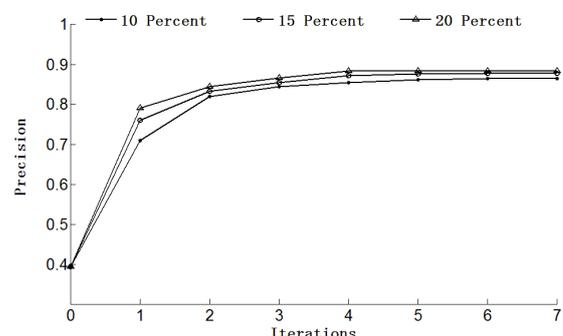Fig. 2. The precision and converge curves by PERF.

Fig. 3. Performance under different user engagement: browsing 10% images, browsing 15% images and browsing 20% images.

REFERENCES

[1] J. H. Su, W. J. Huang, P. S. Yu, and V. S. Tseng, "Efficient relevance feedback for content-based image retrieval by mining user navigation patterns," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, pp. 360-372, 2011.

[2] V. S. Tseng, J. H. Su, J. H. Huang, and C. J. Chen, "Integrated mining of visual features, speech features, and frequent patterns for semantic video annotation," *IEEE Transactions on Multimedia*, vol. 10, pp. 260-267, 2008.

[3] V. S. Tseng, J. H. Su, B. W. Wang, and Y. M. Lin, "Web image annotation by fusing visual features and textual information," in *Proc. the 2007 ACM symposium on Applied computing*, ACM, 2007, pp. 1056-1060.

[4] D. Ritendra, J. Li, and J. Z. Wang, "Content-based image retrieval: approaches and trends of the new age," in *Proc. the 7th ACM SIGMM international workshop on Multimedia information retrieval*, ACM, 2005, pp. 253-262.

[5] V. Khanh, K. A. Hua, and N. Jiang, "Improving image retrieval effectiveness in query-by-example environment," in *Proc. the 2003 ACM symposium on Applied computing*, ACM, 2003, pp. 774-781.

[6] K. Takio and T. Kato, "Learning of personal visual impression for image database systems," in *Proc. the Second International Conference on Document Analysis and Recognition,* IEEE, 1993, pp. 547-552.

[7] T. P. Minka and R. W. Picard, "Interactive learning with a 'Society of models'," *Pattern Recognition*, Elsevier Science Publishing Company, Inc., vol. 30, 1997, pp. 565-581.

[8] R. W. Picard, T. P. Minka, and M. Szummer, "Modeling user subjectivity in image libraries," in *Proc. the 1996 International Conference on Image Processing*, IEEE, vol. 1, 1996, pp. 777-780.

[9] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 644-655, 1998.

[10] D. Kelly and N. J. Belkin, "Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback," in *Proc. the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2001, pp. 408-409.

[11] P. Auer, Z. Hussain, S. Kaski, A. Klami, J. Kujala, J. Laaksonen, A. P. Leung, K. Pasupa, and J. Shawe-Taylor, "Pinview: Implicit feedback in content-based image retrieval," *Journal of Machine Learning Research-Proceedings Track*, vol. 11, pp. 51-57, 2010.

[12] A. Faro, D. Giordano, C. Pino, and C. Spampinato, "Visual attention for implicit relevance feedback in a content based image retrieval," in *Proc. the 2010 Symposium on Eye-Tracking Research and Applications*, ACM, 2010, pp. 73-76.

[13] A. Grigorova, F. G. B. De Natale, C. Dagli, and T. S. Huang, "Content-based image retrieval by feature adaptation and relevance feedback," *IEEE Transactions on Multimedia*, vol. 9, pp. 1183-1192, 2007.

[14] D. Harman, "Relevance feedback revisited," in *Proc. the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 1992, pp. 1-10.

[15] G. Salton and C. Buckley, "Improving retrieval performance by relevance feedback," *Journal of American Society for Information Science*, vol. 41, pp. 288-294, 1990.

[16] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia systems*, vol. 8, pp. 536-544, 2003.

[17] J. Liu, Z. W. Li, M. J. Li, H. Q. Lu, and S. D. Ma, "Human behaviour consistent relevance feedback model for image retrieval," in *Proc. the 15th international conference on Multimedia*, ACM, 2007, pp. 269-272.
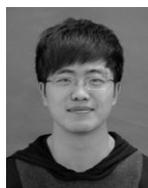
Wei Dai was born in Shandong province, CHN on September 19th. 1991. He is an undergraduate student of Department of Computer Software Engineering at Tianjin University in Tianjin, CHN. His major field of study is data mining.

He is the author of 2 pending patents. His research interests include data mining and image processing.

Wenbo Li was born in Shanxi province, CHN on May 8th. 1991. He is an undergraduate student of Department of Computer Software Engineering at Tianjin University in Tianjin, CHN. His major field of study is image processing, computer vision and data mining.

He is now an Intern in Institute of Automation, Chinese Academy of Sciences in No.95, Zhongguancun Road, Haidian District, Beijing, CHN. His works include Yi Zhang Zhipeng Mo, Wenbo Li, Tianhao Zhao, "A User-driven Model for Content-based Image Retrieval," presented at the APSIPA ASC 2012, Hollywood, CA, USA, 2012. And he is the author of 3 pending patents. His research interests include image processing, computer vision and pattern recognition. Mr Li is a student member of APSIPA.

Zhipeng Mo was born in Hunan province, CHN on July 7th. 1991. He is an undergraduate student of Department of Computer Software Engineering at Tianjin University in Tianjin, CHN. His major field of study is image processing and data mining.

His works include Yi Zhang Zhipeng Mo, Wenbo Li, Tianhao Zhao, "A User-driven Model for Content-based Image Retrieval," presented at the APSIPA ASC 2012, Hollywood, CA, USA, 2012. And he is the author of 3 pending patents. His research interests include image processing and computer vision.

Tianhao Zhao was born in Heibei province, CHN on June 27th. 1992. He is an undergraduate student of Department of Computer Software Engineering at Tianjin University in Tianjin, CHN. His major field of study is image processing and data mining.

His works include Yi Zhang Zhipeng Mo, Wenbo Li, Tianhao Zhao, "A User-driven Model for Content-based Image Retrieval," presented at the APSIPA ASC 2012, Hollywood, CA, USA, 2012. And he is the author of 3 pending patents. His research interests include image processing and data mining. Mr Zhao is a student member of APSIPA.