

CONTENT-FREE IMAGE RETRIEVAL USING BAYESIAN PRODUCT RULE

David Liu, Tsuhan Chen

dliu@cmu.edu, tsuhan@cmu.edu
Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, U.S.A.

ABSTRACT

Content-free image retrieval uses accumulated user feedback records to retrieve images without analyzing image pixels. We present a Bayesian-based algorithm to analyze user feedback and show that it outperforms a recent maximum entropy content-free algorithm, according to extensive experiments on trademark logo and 3D model datasets. The proposed algorithm also has the advantage of being applicable to both content-free and traditional content-based image retrieval, thus providing a common framework for these two paradigms.

1. INTRODUCTION

Image retrieval systems aim at providing a user the images in database that are similar to the query she has in mind. Relevance feedback is a technique to let the user interact with the system by giving examples so that the system has more information of what the user needs. The key is how to make the best out of the feedback information. More formally, we define our problem as follows: Given images that the user considers similar to the query she has in mind, rank all database images from most to least similar to that query. It is worth emphasizing that the query does not have to physically exist; all the system needs are the images that the user identifies as similar or relevant to the query.

Content-based image retrieval (CBIR) systems [2] have been researched for over ten years. In CBIR, features of the query image and database images are extracted and compared to facilitate retrieval. On the other hand, content-free image retrieval (CFIR) [1][3] is a rather new topic, which operates in a similar way to online purchase recommendation systems, such as Amazon.com. It accumulates records of user purchase histories, and exploits the fact that two people share a certain degree of purchase preference if their previous purchase records are similar. CFIR opens a new avenue of research direction: it bypasses the difficult problem of image representation faced by computer vision and pattern recognition.

Here we draw an analogy between CFIR and purchase recommendation systems of internet bookstores. It is in general difficult to formulate the query for an image by text description. Hence, the system first asks the user to provide

some positive feedback, i.e., image(s)/book(s) that she considers relevant to her query in mind. If there were previous users who considered the same or similar set of images/books, the system then uses the most popular images/books among those users as retrieval result. As an easy example, if all previous users who bought book B1 also bought book B2, then if a new user picks B1 as a query, the system will automatically recommend B2.

There are a few underlying assumptions for CFIR systems to work well [3]. Most importantly, previous usage history must be more than a certain amount. The sufficient amount of previous usage history was analyzed in [3].

Such kind of user feedback should not be confused with the traditional relevance feedback strategy in CBIR [4][5][6][7]. Traditional relevance feedback operates in the feature space; it facilitates retrieval by alternating the similarity measure, or by adjusting the image representation. However, representing images by features is a difficult problem. The reason that current CBIR systems still have much room for improvement should mostly be attributed to the status of image understanding and image representation, which are still far away from human perception capabilities.

In this paper, we use two image databases to compare several CFIR algorithms. We demonstrate that our Bayesian framework in [8] can be extended and applied to CFIR, and we compare it with the maximum entropy method in [1][3], which is the pioneering work on CFIR. We show by experiments that our method outperforms the maximum entropy method.

This paper is organized as follows. Section 2 describes the Bayesian framework of the proposed CFIR method. Section 3 provides a brief review of the baseline method used in the experiments. In Section 4 we will present experimental results. Finally, we present concluding remarks in Section 5.

2. BAYESIAN CFIR

The system operates by requesting the user to provide one or more images that are relevant to the query she has in mind. This set of images constitutes the positive feedback. The user interface is designed such that the user can browse through the

database arbitrarily and provide positive feedback by checking check boxes under each image.

Consider a database of N images, $\{I_1, \dots, I_N\}$. We denote $X_i = 1$ if the user selects image I_i as relevant to the query in mind; otherwise $X_i = 0$. The set of variables assigned value of one is denoted by $X_E = \{X_{j_1}, \dots, X_{j_E}\}$ and is called the evidence set; the rest of the variables form the hidden set X_H . X_E is in practice much smaller than X_H . The task of content-free image retrieval is formulated as computing the conditional probabilities $P(X_i = 1 | X_{j_1} = 1, \dots, X_{j_E} = 1)$ for all $X_i \in X_H$. The images can then be ranked according to their corresponding conditional probabilities. This can be interpreted in the following way: suppose database images I_{j_1}, \dots, I_{j_E} are relevant to the query the user has in mind, how likely is image I_i the one that the user is looking for?

2.1. CFIR based on the Bayesian product rule

To compute the conditional probability, we assume the following conditional independency holds:

$$\begin{aligned} &P(X_{j_1} = 1, \dots, X_{j_E} = 1 | X_i = 1) \\ &= P(X_{j_1} = 1 | X_i = 1) \cdot \dots \cdot P(X_{j_E} = 1 | X_i = 1) \end{aligned} \quad (1)$$

Applying Bayes' rule, we have

$$\begin{aligned} &P(X_i = 1 | X_{j_1} = 1, \dots, X_{j_E} = 1) \\ &= \frac{P(X_i = 1) \prod_{k=1}^E P(X_{j_k} = 1 | X_i = 1)}{P(X_{j_1} = 1, \dots, X_{j_E} = 1)} \end{aligned} \quad (2)$$

Applying Bayes' rule on Eq.(2) again, we obtain

$$\begin{aligned} &P(X_i = 1 | X_{j_1} = 1, \dots, X_{j_E} = 1) \\ &= \frac{\prod_{k=1}^E \left(P(X_i = 1 | X_{j_k} = 1) P(X_{j_k} = 1) \right)}{P(X_{j_1} = 1, \dots, X_{j_E} = 1) P(X_i = 1)^{E-1}} \\ &\propto \frac{\prod_{k=1}^E P(X_i = 1 | X_{j_k} = 1)}{P(X_i = 1)^{E-1}} \end{aligned} \quad (3)$$

The computation of the pairwise conditional probabilities $P(X_i = 1 | X_j = 1)$ and prior probabilities $P(X_i)$ follow [1]. In later sections we will refer to Eq.(3) as the *Bayesian product rule*.

In [8], we used the Bayesian product rule in the context of a content-based image retrieval system. Hence, this framework is general and applicable to both CBIR and CFIR. The difference lies in the definition of the variables X_i and consequently in the computation of the probabilities $P(X_i = 1 | X_j = 1)$ and $P(X_i)$.

2.2. CFIR based on the Bayesian sum rule

Following Eq.(3), the Bayesian product rule can be approximated by

$$\begin{aligned} &P(X_i = 1 | X_{j_1} = 1, \dots, X_{j_E} = 1) \\ &\propto \left(\sum_{k=1}^E P(X_i = 1 | X_{j_k} = 1) \right) + (1 - E) \cdot P(X_i = 1) \end{aligned} \quad (4)$$

when $P(X_i = 1 | X_{j_k} = 1) \approx P(X_i = 1) \cdot (1 + \delta_{ij_k})$, $\delta_{ij_k} \ll 1$. Kittler et. al. [9] showed that the above sum rule often outperforms the product rule because of its robustness to noise.

3. ZITNICK'S MAXIMUM ENTROPY METHOD

Zitnick's method is used as the baseline method in this paper. This method has recently been proposed by Zitnick in [1] and used in [3]. Zitnick showed that, by maximizing Rényi's quadratic entropy, we have

$$P(X_i = 1 | X_{j_1} = 1, \dots, X_{j_E} = 1) = \lambda_{i,j_1} + \dots + \lambda_{i,j_E} \quad (5)$$

where

$$\begin{aligned} \Lambda_i &\equiv [\lambda_{i,j_1}, \dots, \lambda_{i,j_E}] = \mathbf{p}_i^T \mathbf{P}^{-1} \\ \mathbf{p}_i &= \begin{bmatrix} P(X_i = 1 | X_{j_1} = 1) \\ \vdots \\ P(X_i = 1 | X_{j_E} = 1) \end{bmatrix} \\ \mathbf{P} &= \begin{bmatrix} P(X_{j_1} = 1 | X_{j_1} = 1) & \dots & P(X_{j_E} = 1 | X_{j_1} = 1) \\ \vdots & \ddots & \vdots \\ P(X_{j_1} = 1 | X_{j_E} = 1) & \dots & P(X_{j_E} = 1 | X_{j_E} = 1) \end{bmatrix} \end{aligned} \quad (6)$$

In the case when the matrix \mathbf{P} is singular, a pseudo-inverse should be used [1]. All the vector and matrix elements in \mathbf{p}_i and \mathbf{P} can be looked up from the pairwise conditional probabilities $P(X_i = 1 | X_j = 1)$, $\forall i, j$.

4. EXPERIMENTS

We used two datasets to compare the CFIR methods.

The first dataset (Fig. 1) consists of 1039 real-world trademark logo images. Fourteen trademark examiners in the Intellectual Property Office of Taiwan (rather than people trained for this project) provided the ground-truth data. The ground-truth data collection system operates in the following way. The system exhaustively displays two images from the 1039 images at one time, and asks the human expert whether they are similar (potentially infringing each other) or not, based on their expertise. We hence obtain a fully labelled 1039 by 1039 similarity table \mathbf{S} ; $\mathbf{S}(i, j) = 1$ if image I_i and I_j are potentially infringing each other; $\mathbf{S}(i, j) = 0$ otherwise. Since

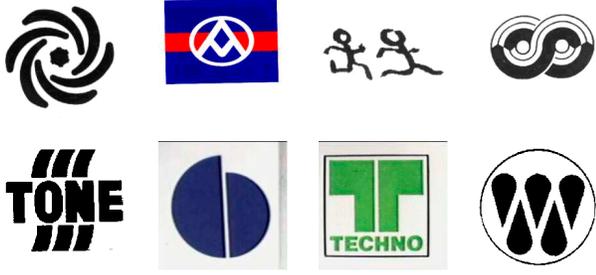


Fig. 1. Sample images from the trademark logo database.

all (i, j) pairs are fully labelled, the user feedback history is complete and hence more than suitable for applying CFIR.

The ground truth similarity table \mathbf{S} provide us a reliable way to compare and evaluate the three CFIR algorithms. We split the ground truth table \mathbf{S} into an evidence table \mathbf{S}_E and an answer table \mathbf{S}_A (which can be interpreted as the training data and test data, respectively) in the following way: Let \mathbf{S}_E and \mathbf{S}_A be tables of the same size as \mathbf{S} and initialized to all zero. For each row of \mathbf{S} , we pick a random subset of the ones and put them in the evidence table \mathbf{S}_E ; the rest of the ones are put in the answer table \mathbf{S}_A . Diagonal terms are always set to one, since an image is surely similar to itself. An example is shown in Table 1,2 and 3: the 1's in $\mathbf{S}(1, 3)$ and $\mathbf{S}(1, 7)$ are put in \mathbf{S}_E ; the remaining 1 in $\mathbf{S}(1, 5)$ is put in \mathbf{S}_A .

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1	0	1	0	1	0	1
X_2	0	1	0	1	0	1	1
X_3	1	0	1	1	1	0	0
X_4	0	1	1	1	0	0	0
X_5	1	0	1	0	1	1	1
X_6	0	1	0	0	1	1	0
X_7	1	1	0	0	1	0	1

Table 1. Ground truth table \mathbf{S}

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1	0	1	0	0	0	1
X_2	0	1	0	0	0	0	1
X_3	1	0	1	0	1	0	0
X_4	0	0	0	1	0	0	0
X_5	0	0	1	0	1	1	0
X_6	0	0	0	0	1	1	0
X_7	1	1	0	0	0	0	1

Table 2. Evidence table \mathbf{S}_E

To evaluate retrieval performance, for each image I_i , we use the relevant images in the evidence table, $\{I_j | \mathbf{S}_E(i, j) = 1\}$, as the evidence set, and compute precision-recall based on

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1	0	0	0	1	0	0
X_2	0	0	0	1	0	1	0
X_3	0	0	1	1	0	0	0
X_4	0	1	1	1	0	0	0
X_5	1	0	0	0	1	0	1
X_6	0	1	0	0	0	1	0
X_7	0	0	0	0	1	0	1

Table 3. Answer table \mathbf{S}_A

the answer table \mathbf{S}_A . More specifically, $P(X_i = 1 | X_j = 1)$ and $P(X_i)$ can be computed as in [1], from which $P(X_i = 1 | X_{j_1} = 1, \dots, X_{j_E} = 1)$ can be obtained for all three CFIR algorithms. Evidence sets smaller than two are discarded. The performance of the three CFIR methods using the trademark image dataset is shown in Fig. 2. The Bayesian product rule significantly outperforms the Bayesian sum rule and the maximum entropy method at recall 0 ~ 60%. After that, the confidence intervals of the three methods overlap.

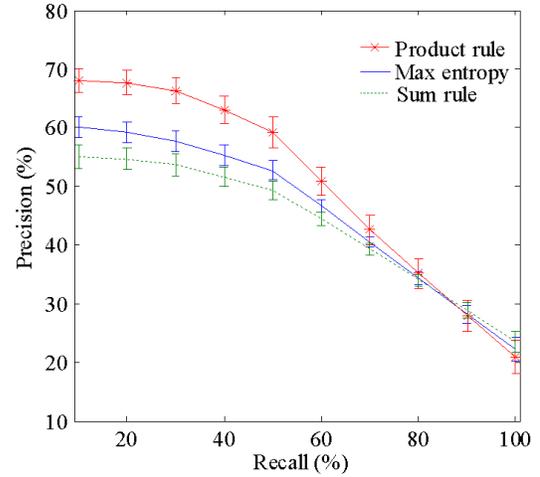


Fig. 2. Precision-recall of CFIR methods. From top to bottom (the higher the better): Bayesian product rule, Zitnick's maximum entropy method, and Bayesian sum rule. The bars are plus-minus one standard deviation.

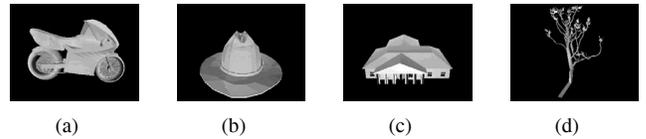


Fig. 3. Sample images from the 3D model database.

The second dataset is a VRML dataset, consisted of 1750 3D images (Fig. 3). The images are classified by graduate stu-

dents into 52 semantic categories (such as airplane, car, building; see [10]) based on their visual appearance; one image can belong to multiple categories. Based on this classification, we build a 1750 by 1750 similarity table \mathbf{S} ; $\mathbf{S}(i, j) = 1$ if image I_i and I_j belong to a common category; $\mathbf{S}(i, j) = 0$ otherwise. Since all (i, j) pairs are fully labelled, the user feedback history is complete and hence more than suitable for applying CFIR.

As noted in [1], sometimes during evaluation the precision-recall curves intersect and it becomes hard to judge which algorithm is better. At this time a single number performance measure is preferred. The accuracy of one trial is defined as in [1]:

$$Accuracy = \frac{\sum_{i=1}^{|\mathbf{D}|-k} \delta(i, \mathbf{R}) h(i)}{\sum_{i=1}^{|\mathbf{R}|-k} h(i)} \quad (7)$$

where $h(i) = 2^{\frac{-i-1}{b-1}}$. Here we use $b = 5$ as in [1]. \mathbf{D} is the whole dataset, \mathbf{R} is the set of relevant images in the training set, $\delta(i, \mathbf{R}) = 1$ if the i^{th} ranked image is in \mathbf{R} . The accuracy will be equal to one when all relevant images are ranked on top. The idea is to assign a different weighting score for each retrieved relevant image; the higher the rank, the higher the weighting score. The accuracy of an algorithm is the average of the accuracy values for a large number of trials.

In Fig. 4 we compare the proposed Bayesian product rule with the maximum entropy method on the VRML 3D model dataset using accuracy as the performance index. The Bayesian product rule CFIR has higher accuracy.

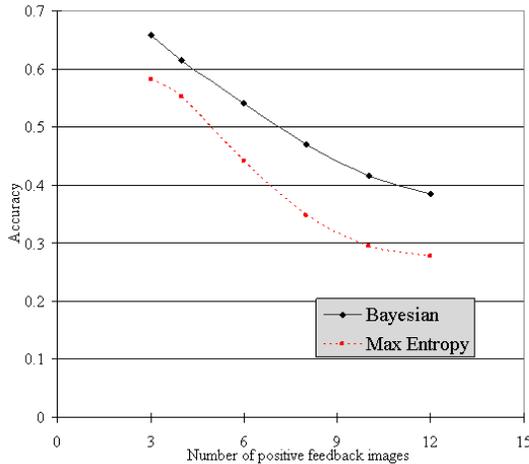


Fig. 4. Plot of accuracy vs. number of positive feedback images. The number of positive feedback images is the number of variables in the evidence set X_E . The Bayesian product rule (top curve) yields higher accuracy than the maximum entropy method.

5. CONCLUSIONS

Content-free image retrieval uses accumulated user feedback records to retrieve images without analyzing image pixels. In this paper we presented a content-free image retrieval algorithm and show it outperforms the maximum entropy algorithm based on extensive experiments on a trademark logo dataset and a 3D model dataset. Our proposed algorithm also has the advantage of being applicable to both content-free and traditional content-based image retrieval [8], thus providing a common framework for these two paradigms.

We believe that a system combining CFIR and CBIR is of largest practical value, since both CFIR and CBIR have its own shortcomings; CFIR requires sufficient amount of usage history, while CBIR suffers from low discriminative image features. The Bayesian framework in this paper should provide a good starting point for a hybrid system.

6. ACKNOWLEDGEMENT

The first author was supported in part by the Taiwan Merit Scholarship TMS-094-1-A-049.

7. REFERENCES

- [1] C. Zitnick, *Computing Conditional Probabilities in Large Domains by Maximizing Renyi's Quadratic Entropy*, Ph.D. thesis, Robotics Institute, Carnegie Mellon University, 2003.
- [2] A.W.M. Smeulders, S. Worning, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349–1380, 2000.
- [3] S. Uchihashi and T. Kanade, "Content-free image retrieval based on relations exploited from user feedbacks," in *IEEE Intl. Conf. Multimedia and Expo*, 2005.
- [4] J.J. Rocchio, "Relevance feedback in information retrieval," in *The SMART Retrieval System*, G. Salton Ed., Prentice-Hall, Englewood NJ, 1971.
- [5] X.S. Zhou and T.S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, pp. 536–544, 2003.
- [6] Y. Rui, T.S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in mars," in *IEEE Intl. Conf. on Image Processing*, 1997.
- [7] I.J. Cox, M.L. Miller, T. Minka, T.V. Papathomas, and P.N. Yianilos, "The bayesian image retrieval system, pichunter: Theory, implementation and psychological experiments," *IEEE Trans. Image Processing*, vol. 9, pp. 20–37, 2000.
- [8] D. Liu and T. Chen, "Probabilistic relevance feedback for image retrieval with binary feature vectors," in *IEEE Conf. Acoustics, Speech, and Signal Processing*, 2005.
- [9] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226–239, 1998.
- [10] C. Zhang and T. Chen, "Annotating retrieval database with active learning," in *IEEE Intl. Conf. Image Processing*, 2003.