A Bayesian Approach to Hybrid Image Retrieval

Pradhee Tandon and C. V. Jawahar

Center for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500032, INDIA
{pradhee@research.,jawahar@}iiit.ac.in

Abstract. Content based image retrieval (CBIR) has been well studied in the computer vision and multimedia community. Content free image retrieval (CFIR) methods, and their complementary characteristics to CBIR has not received enough attention in the literature. Performance of CBIR is constrained by the semantic gap between the feature representations and user expectations, while CFIR suffers with sparse logs and cold starts. We fuse both of them in a Bayesian framework to design a hybrid image retrieval system by overcoming their shortcomings. We validate our ideas and report experimental results, both qualitatively and quantitatively. We use our indexing scheme to efficiently represent both features and logs, thereby enabling scalability to millions of images.

1 Introduction

Retrieval of similar images and videos from large databases, has received significant attention in recent years [1]. There are two prominent approaches to solve this problem: (i) Content-based image retrieval (CBIR), popular in the computer vision community (ii) Content-free image retrieval (CFIR), which has received some amount of attention in the database community. A CBIR method typically converts an image into a feature vector representation, and matches with the images in the database to find out the most similar images. On contrary, CFIR methods exploit the co-occurrence information (for example in a collaborative filtering framework) in the logs of image-access to model the similarity across images [3, 5]. If a user accesses/accepts two images together, then these images are treated as semantically related. There are also attempts which tried to combine both these approaches [6, 8, 10].

We are interested in designing a practical image retrieval system which (i) naturally scales to large number of images (ii) allows the simultaneous use of ideas from visual similarities as well as user behavior patterns (iii) allows overcoming the limitations (Section 3) of CBIR and CFIR by exploiting their complementary nature. We meet these objectives by reasoning in a Bayesian framework, where the *a priori* information comes from the logs, and visual similarity acts as the evidence. We conduct extensive experiments and report results to validate the superiority of the hybrid solution both qualitatively and quantitatively. We also demonstrate the scalability and efficiency of the solution. We can

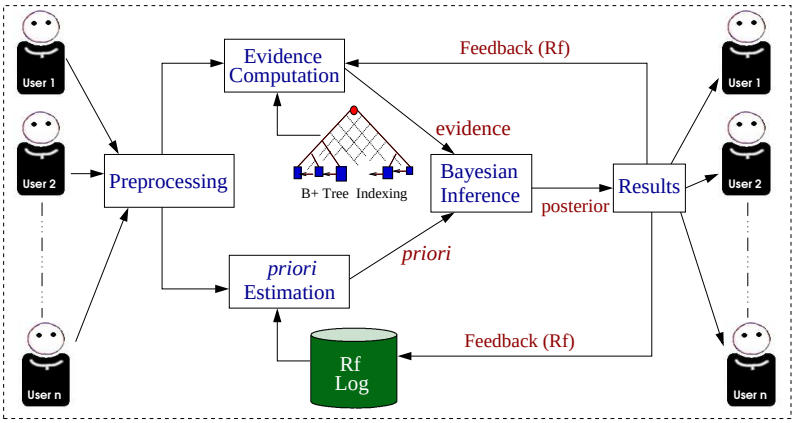

Fig. 1: Architecture of our scalable Bayesian Image Retrieval system

successfully retrieve from millions of images in interactive (sub-second) time as demonstrated in Section 4.

2 A Scalable Indexing Scheme

Our indexing scheme is an extension of [7] and [2], which were primarily derived for CBIR in presence of a changing similarity metric. [7] demonstrated the utility of a B+-tree based indexing scheme for efficient approximate nearest neighbor (ANN) computation in high dimensional vector spaces. Later on, [2] exploited it based on the fact that most concepts get clustered in the feature spaces.

The indexing scheme used in this work (briefly sketched in Figure 1) allows simultaneous indexing of both visual clues (raw features) and user interaction patterns in the form of logs (unlike in [2, 7]). Logs are represented as relationships across images in a MySQL database. Images are represented as fixed-length feature vectors in a B+ tree index structure as in [7]. A computationally efficient reasoning (Section 3), which combines these two factors with the help of a set of *learned* weights, is carried out for the interactive search. We use the logs of retrieval process to provide co-occurrence information for pairs of images. When processing the query, we use these co-occurrence relations, as explained in the next section. At the end of the retrieval, the database of logs is refined based on user acceptability of images. Note that this log contains significant amount of subjectivity (and therefore uncertainty).

Efficiency of our indexing scheme can be attributed to: (a) indexing the image feature-by-feature makes the indexing scheme consistent even when relative importance of features changes. (b) feature vectors are bulky (due to large number of dimensions) and they are represented in B+ trees. Logs are compact relationships and they are represented in MySQL. This makes our indexing scheme space efficient. (c) Since ANN (feature-based retrieval) and log based retrieval (from a

standard database) are individually fast and our fusion scheme is based on very few multiplications, our scheme is overall efficient. This meets our requirement of interactive retrieval from large databases.

3 Bayesian Image Retrieval

CBIR Vs CFIR*:* CBIR methods use low level features for representing and retrieving images. They, typically, are unable to represent human perception of visual content. The semantic gap is the primary bottleneck of CBIR methods. Many of the previous methods extensively explored the use of feedback based learning for improved performance [4, 9]. However, they are either critically dependent on features or computationally infeasible. CFIR on the other hand uses only feedback based co-occurrence among images. Therefore they are able to capture semantic relations among images and predict accurately their relevance to other seen images. User feedback and logs are difficult to obtain. Unless the system is functional, users do not provide any feedback. This creates a deadlock and cold start. In addition, CFIR has no clue about previously unseen images. Prediction accuracy is also critically dependent on the availability of logs.

Bayesian Integration*:* CBIR and CFIR thus provide two complimentary estimates of similarity among images. Their effective integration can overcome the critical dependencies of both of them and provide improved accuracy. We use a Bayesian framework for fusing the two approaches. Bayes theorem provides a method to calculate the probability of a hypothesis based on its prior probability, the probabilities of observing various data given the hypothesis, and the observed data itself. We formulate the image retrieval problem as one of estimating the probability of retrieving an image, as a posterior estimation problem. We model the *a priori* on the co-occurrence information from the history logs. The visual similarity, between the query and the database images, is used as the evidence in favor of the match. The two are combined in the Bayesian inference to estimate the *a posteriori* of the image being relevant to the query. If $R(\mathbf{q}, \mathbf{a})$ denotes the event of retrieving the image $\mathbf{a}$ given $\mathbf{q}$ as the query. The *a priori* probability of this event $P(R)$ can be computed from the co-occurrence as

$$P(R) = \frac{n(\mathbf{a}, \mathbf{q})}{n(\mathbf{q})} \tag{1}$$

where $\mathbf{q}$ has been found relevant by users, $n(\mathbf{q})$ times and it has been relevant with $\mathbf{a}$, $n(\mathbf{a}, \mathbf{q})$ times. It is 1 when $\mathbf{a}$ and $\mathbf{q}$ were always retrieved together. It is zero, when they never co-occur as acceptable images together. $n(\mathbf{a}, \mathbf{q})$ and $n(\mathbf{q})$ are initially assumed to be 1 to avoid inconsistencies.

Let $S(\mathbf{q}, \mathbf{a})$ be the feature-level similarity of the images $\mathbf{q}$ and $\mathbf{a}$. We learn the query concept as weights for the features $\mathbf{w}$ as discussed later. Using these weights we also learn feature weights $\mathbf{c}$, for the popular concept in individual images. $S(\mathbf{q}, \mathbf{a})$ can then be estimated as

$$p(S|R) = f(\mathbf{c}, \mathbf{w}, \mathbf{q}, \mathbf{a}) \tag{2}$$

We can now estimate the posterior using Bayes Rule as in

$$p(R|S) = p(S|R)P(R) \tag{3}$$

We do not consider the denominator of the Bayes rule, since it does not modify the relative ranking of the database images, given the query. In practice, one could use alternate 'definitions' of the probabilities, as long as they satisfy the basic axioms of probability. The top N images with the maximum *a posteriori* probability can be returned to the user.

Bayesian Image Retrieval Process: In a query-by-example framework like ours, each image in the system is represented as a vector of numeric feature values $[X_1, \ldots, X_d]^T$, constituting a multi-dimensional space where each image is a point. The database is pre-processed while the query is processed online to extract the set of features. The query is then compared for similarity with a subset of the dataset, using the Bayesian integration scheme discussed above and top N results are returned for interaction.

Given a query vector, $\mathbf{x_q}$, we retrieve the p ($p >> N$) nearest data points from each B+-tree. The trees are enumerated in order of decreasing relevance ($\mathbf{w}$) making it likely to retrieve the closest points earlier. The scheme is analyzed in detail in [2, 7]. Both relevant and irrelevant images, as marked by the user (relevance feedback), are used for incrementally learning $\mathbf{w}$. $\mathbf{w}$ is learnt by iteratively estimating the relevance of a feature, s_j, based on the dispersion of the feature over relevant and irrelevant sets. At the end of the query session the relevance of features, $\mathbf{c}$, for every relevant image, is updated using $\mathbf{w}$. The expressions for estimating and updating relevances have been discussed in detail in [2]. Visual similarity S_i between x_q and image x_i is computed using a weighted Mahalanobi's metric as in

$$S_i = \left[\left(\mathbf{W}^T \left[\mathbf{x}_i - \mathbf{x_q} \right] \right)^T \mathbf{M} \left(\mathbf{W}^T \left[\mathbf{x}_i - \mathbf{x_q} \right] \right) \right]^{\frac{1}{2}}$$

where $\mathbf{W}$ is the *diagonal* matrix of $\mathbf{w}$. Co-variance matrix M is computed initially.
Co-occurrence information is summarized into $\mathbf{V}$. In the first iteration retrieval happens only on visual similarity but next one onwards the feedback pattern is used for estimating the closest concept and only these samples are used for posterior computations.

Co-occurrence information is updated either at the end of every query session or deferred by a few. We try to discover the implicit concepts in the co-occurrence information using an incremental k-means clustering on the vectors for images. This results in $\mathbf{V}$ concepts, $\mathbf{V}_i, ..., \mathbf{V}_k$ which represent the relationships existing in the database as of now. This helps us learn higher level concepts and also prunes the search space. Incremental clustering, though repetitive, allows self discovery of concepts as logs improve. Being modeled as an off-line process it does not effect retrieval. Thus together the index and the off-line summarizing allows sub-second retrieval times.

4 Experiments and Discussions

Datasets: For our accuracy experiments, we have used two datasets with different characteristics. The first, $\mathbf{D_1}$, is a completely annotated, 58 category set of $12,000$ real natural images, collected from Flickr, COREL and cartoon videos. We represent it using MPEG-7's Color Structure and Edge Histogram descriptors. The second set, $\mathbf{D_2}$, is also completely annotated and comprises of the Caltech-256 dataset. We use the state-of-art *bag of words* approach to represent these images using a $2,000$ word *SIFT* vocabulary. For the scalability study we use a dataset, $\mathbf{D_3}$ of *1 million* points generated from a uniform distribution. For collecting logs, users were asked to select a query and provide feedback to a randomly selected set of 20 images from the dataset, supposedly similar to the query. For $\mathbf{D_2}$ we used the available annotations, instead of users, for automatically creating logs. Note that logs do not provide the complete similarity distribution (users do not see all the examples in the database). Typically only 10% of the valid co-occurrences are available while testing the system.

Precision improvement: In this experiment we use human logs to show how our approach achieves better accuracy compared to a pure CBIR. We used 5 random queries from each of the 58 categories in $\mathbf{D_1}$. We computed average precision for both the approaches and compare it for a few categories in Table 1(a). A similar comparison for $\mathbf{D_2}$ using annotation based logs can be found in Table 1(b).

Learning in BSIR: Our Bayesian systems learns across users and is able to retrieve with better accuracy by using improved co-relevance information or the *a priori* and feature relevance in images, $\mathbf{c}$.

Qualitative Comparison: Next we visually compare with CBIR using the top few results for some queries in Figure 2. As can be seen, our Bayesian retrieves with better accuracy in both the examples. The leftmost image is also the query.

Efficiency and Scalability: In Figure 2 we show how our approach retrieves in sub-second times both with increase in database size and the number of dimensions using $\mathbf{D_3}$ and 5 randomly selected queries. We have optimized on the retrieval time by designing *a priori* updates as an off-line process and storing the co-occurrence matrix in MySQL.

Approach	**Category**		
	1	2	3
BAYES	59.91%	75.69%	63.94%
CBIR	31.38%	39.38%	41.38%

(a)

Approach	**Category**		
	1	2	3
BAYES	72.08%	67.22%	59.35%
CBIR	42.00%	47.08%	28.84%

(b)

Table 1: Tables shows the improved precision for 3 categories with our approach over CBIR. (a) uses real user logs while (b) uses annotation based logs.

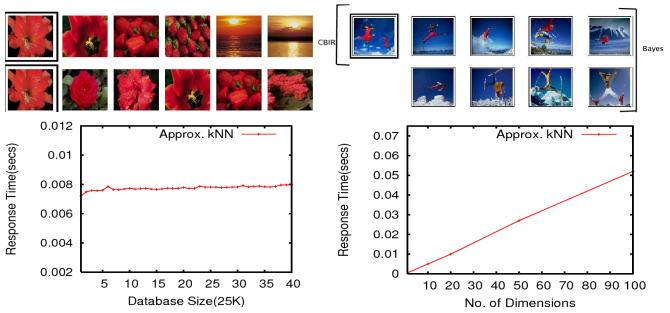

Fig. 2: (Clockwise from top-left) Top 6 results from CBIR (first row) and our Bayesian (second row); More semantically similar images got added to the CBIR result(first row) set with our Bayesian(first and second rows); Avg. retrieval time with increasing number of dimensions; Avg. retrieval time with increasing database size

5 Conclusions

We have proposed a Bayesian inference based hybrid image retrieval system which fuses complimentary techniques of CBIR and CFIR and overcomes many of their shortcomings. We have presented extensive experiments to validate the advantage in terms of accuracy, interactive retrieval times and efficient learning. We would like to further extend concept discovery in our future work.

References

1. Ritendra Datta, Dhiraj Joshi, Jia Li and James Z. Wang: Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Survey (2008) 1–60
2. Pradhee Tandon, P. Nigam, V. Pudi and C. V. Jawahar: FISH: a practical system for fast interactive image search in huge databases. ACM CIVR (2008) 369–378
3. Zhong Su, HongJiang Zhang, Stan Z. Li and Shaoping Ma: Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning. IEEE Transactions on Image Processing (2003) 924-937
4. Xiang S. Zhou, Thomas S. Huang: Relevance feedback in image retrieval: A comprehensive review. Multimedia Systems (2003) 6, 536–544
5. Xiangdong Zhou, Qi Zhang, Liang Zhang, Li Liu, Baile Shi: An Image Retrieval Method Based on Collaborative Filtering. IDEAL (2003)
6. Prem Melville, Raymod J. Mooney and Ramadass Nagarajan: Content-boosted collaborative filtering for improved recommendations. NCAI (2002) 187–192
7. Nataraj Jammalamadaka, V. Pudi and C. V. Jawahar: Efficient Search with Changing Similarity Measures on Large Multimedia Datasets. MMM (2007) 206-215
8. Kai Yu, Anton Schwaighofer, Volker Tresp, Wei-Ying Ma and HongJiang Zhang: Collaborative Ensemble Learning: Combining Collaborative and Content-Based Information Filtering via Hierarchical Bayes. UAI (2003) 616-623
9. D.R. Heisterkamp: Building a latent semantic index of an image database from patterns of relevance feedback. ICPR (2002) 134–137
10. J. Han, K. Ngan, M. Li, and H. Zhang: A Memory Learning Framework for Effective Image Retrieval. IEEE Trans. on Image Processing (2005) 511–524