INTERACTIVE IMAGE RETRIEVAL BY QUERY FUSION

A. Kushki, P. Androutsos, A.N. Venetsanopoulos

Multimedia Lab, University of Toronto, Canada
{azadeh,oracle,anv@dsp.utoronto.ca}

ABSTRACT

The use of low-level visual features such as colour and shape in content-based image retrieval system leads to several ambiguities. Specifically, due to the many-to-many mapping between the low-level feature space and high-level user concepts, a conceptual user query may not be modelled as a single point in the feature space. Furthermore, conceptually similar images may not be fall close to each other in the low-level feature space. This work addresses these issues by proposing an interactive technique, Query Feedback. This method employs user input to represent a high-level conceptual query as various points in the low-level space. Thus, similarity to a given conceptual query is obtained as the fusion of several low-level representations in the feature space.

1. INTRODUCTION

Content-based image retrieval (CBIR) systems represent and compare images based on visual content such as colour, shape, and texture. Generally, the user initiates a retrieval session by providing an example image to communicate his/her query intentions to the system. The low-level feature representation of this query image is then compared to the low-level representation of the images in the database using some distance measure. The images the fall closest to the query in the low-level feature space are then returned to the user as the best matches to the query. There are, however, two main problems associated with this approach. First, the initial example image provided by the user is merely one of the possibly many realizations of the concept in the user's mind. Therefore, the low-level description of the query available to the system may not capture all aspects of the high-level query concept intended by the user. The second problem is that the many-to-many mapping between low-level feature spaces and high-level semantic user concepts is unknown. This means that images that lie close to each other in the feature space may not actually be similar in terms of their conceptual content.

In order to resolve the above-mentioned ambiguities, several *interactive* systems have been proposed. During a process known as relevance feedback, these systems collect user input on the quality of results. This feedback will be utilized to determine the unknown query point and/or the mapping between the user concepts and low-level feature space. Most of the existing techniques assume the existence of a single query point in the feature space as well as a one-to-one mapping [1, 2], between high-level user concepts and low-level feature space. That is, similar images are assumed to be clustered around the query point. More recent systems have strayed away from this assumptions by allowing similar images to belong to multiple, disjoint clusters. A common approach is to use a Gaussian mixture density to model the likelihood of a particular image belonging to class(es) of similar images [3]. The probability-based approaches suffer from restrictive assumptions regarding the density functions (e.g. Guassianity) and independence of features as well as poor parameter estimates. In other approaches such as [4] and [5], similarity to a conceptual query is obtained by merging the similarity results from various low-level queries. These approaches utilize all images that are deemed relevant by the user as new low-level queries and merge the similarity results. The complexity of the algorithms, thus, increases with the number of positive examples. Furthermore, since the negative or unlabelled images are not used during similarity calculation, these algorithms may get trapped in a sub-optimal solution if additional positive images are not supplied by the user during relevance feedback.

This paper proposes a novel approach, Query Feedback, for interactive content-based image retrieval. This approach, does not impose any of the previously mentioned restrictions in terms of probability models and yet maintains accurate retrieval performance. Query Feedback allows the similar images to come from various classes, each defined by an example image. Specifically, the proposed system intelligently selects a subset of the positive images provided by the user as different realization of the high-level conceptual user query. A novel aggregation scheme is then proposed to combine the results from these different low-level representations and obtain the overall similarity ranking for the database images. Fig. 1 depicts the functionality of Query Feedback.

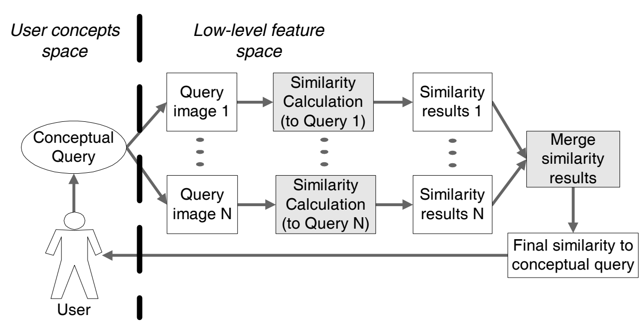


Fig. 1. Overall operation of the Query Feedback system.

The rest of this paper is organized as follows: Section 2 briefly discusses similarity calculations between two images; Section 3 presents the Query Feedback algorithm; Experimental results are reported in Section 4 and the paper is concluded with Section 5.

2. SIMILARITY MEASURE

Query Feedback considers the similarity of an image I_i to a conceptual user query (e.g. "flower") as the *combined* similarity of I_i to multiple realizations of this concept (e.g. "red rose", "white lily", etc.). Thus, the process of relevance feedback consists of three phases: 1) determination of multiple realizations of a given query, 2) similarity calculations between the database images and each of these realization, and 3) fusion of the similarity values from multiple queries. This section addresses similarity calculations between a database image and a given query realization. The remaining components of the system are discussed in Section 3.

In this work, the Unified Framework for Similarity Calculation (UFSC) [6] has been employed for determination of similarity between a query realization and the database images. In this framework, the similarity calculation problem is reformulated as a decision making problem. Specifically, each query image Q is associated with a *conceptual fuzzy set* $\mathcal{S}_Q$ that contains images similar to this query image. Due to the fuzzy nature of $\mathcal{S}_Q$, every database image I_i belongs to this set with to a certain degree or *membership grade* denoted as $\mu_{\mathcal{S}_Q}(I_i)$. This membership grade is essentially the similarity score between I_i and the query image Q. To be precise,

$$\mathcal{S}_Q = \{\mu_{\mathcal{S}_Q}(I_1)/I_1, \ldots, \mu_{\mathcal{S}_Q}(I_N)/I_N\}, \tag{1}$$

where N is the number of images in the database. The membership values $\mu_{\mathcal{S}_Q}(I_i)$ are obtained as a combination of low-level feature descriptor similarities between the two images Q and I_i as depicted in Fig. 2. Specifically, the distance between Q and I_i is first calculated with respect to each feature descriptor p separately and denoted as $D_p(Q, I_i)$. This work employs five of the MPEG-7 colour and texture descriptors, namely, Dominant Colour, Colour Structure, Colour Layout, Edge Histogram, and Homogeneous Texture. Details of these descriptors and distance measures used to obtain the values $D_p(Q, I_i)$ are outlined in [7].

Based on each of the distances $D_p(Q, I_i)$, a decision $d_p(Q, I_i)$ is made regarding the similarity between two images. This decision is obtained by passing the distance through a membership function as detailed in [6]. The membership function used in this work is shown in (2).

$$d_p(Q, I_i) = \frac{1}{1 + \frac{|D_p(Q, I_i)|^\rho}{\lambda}}, \tag{2}$$

where $\lambda = Median_i(D_p(Q, I_i))$ is the median of the distances of all database images to the query, and ρ is a parameter experimentally determined.

The overall similarity is then calculated as a combination or *aggregation* of the feature descriptor decisions, $d_i(Q, I)$. Denoting the aggregation operator as $\odot$, the overall similarity becomes:

$$\mu_{\mathcal{S}_Q}(I_i) = \bigodot_p d_p(Q, I). \tag{3}$$

The aggregation operator, $\odot$, is a compensatory operator defined in (4):

$$x_1 \odot \ldots \odot x_n = \gamma \max(x_1, \ldots, x_n) + (1-\gamma) \min(x_1, \ldots, x_n), \tag{4}$$

where the x_i's are the elements being aggregated and $\gamma \in [0, 1]$. The behaviour of this aggregation operator ranges between the logical AND ($\gamma = 0$) and the logical OR ($\gamma = 1$). Fig. 2 summarizes the UFSC design.

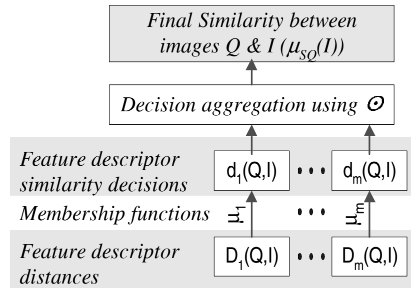


Fig. 2. UFSC design.

3. QUERY FEEDBACK

The previous section described the similarity calculations between a query image and database images. As previously mentioned, however, the query image is merely one realization of a given user concept. In light of this, Query Feedback determines new realizations of the conceptual query at each iteration of relevance feedback. Specifically, after the initial retrieval, the results based on the rankings obtained from (3) are presented to the user. The user then marks the retrieved images that are deemed satisfactory as positive images. At the next iteration, a subset of these positive images are used as new realizations of the conceptual query. Thus, similarity calculations will be based on multiple query images. Once the new query, I_{new}, is chosen by the Query Feedback algorithm, the similarity set, $\mathcal{S}_{new}$, containing the rankings of the database images is generated using (3). The next step is to merge these results with the results based on the previous query image at iteration t of the algorithm, $\mathcal{S}_{Q_t}$. The rest of this section addresses two issues: 1) how the new query images are selected, and 2) how the similarity results from the different query images are combined to provide a single similarity score for a given image. Finally, Section 3.3 provides an algorithmic overview of the Query Feedback method.

3.1. Selection of new query

Query Feedback performs similarity calculations based on multiple example images. Thus, at each iteration of this algorithm one or more example images are chosen from the set of positive images as new instances of the conceptual user query. The objective of new query selection is to ensure that the different realizations of the conceptual query are as different as possible in order for the system to gain the maximum possible information gain regarding the high-level user intentions. For simplicity, we only add one extra query image during each iteration of relevance feedback as opposed to using multiple images. In its simplest form, choosing the new query image translates to choosing the positive image that ranked last during similarity measurement to the query in the previous iteration. This method assumes that the L_1 norm between similarity rank for the query and an image I_i is an indication of the information redundancy between them. That is:

$$\begin{aligned} Q_{new} &= \arg\min_{I_i \in \mathcal{P}} [\mu_{\mathcal{S}_{Q_{old}}}(Q_{old}) - \mu_{\mathcal{S}_{Q_{old}}}(I_i)] \\ &= \arg\max_{I_i \in \mathcal{P}} \mu_{\mathcal{S}_{Q_{old}}}(I_i) \end{aligned} \tag{5}$$

.

There is, however, no indication that this distance is a meaningful measure in the decision space. Thus, a novel *set distance* measure is introduced here for choosing the next query image based on set distances.

By considering an image I_i as the query and using (3), each image in the database, I_p, can be associated with a fuzzy subset, $\mathcal{S}_{I_p}$. This fuzzy subset represents the grade of similarity of each of the images in the database to image I_p. If S_{I_p} and S_{I_r} are *similar* for two given images, the database images will have similar rankings using either I_p or I_r as the query. Therefore, the two images more or less convey the same information. On the other hand, if S_{I_p} and S_{I_r} are dissimilar, it can be concluded that the two images carry different information and as a result, new information may be obtained by combining the two. Various distance measures can be used to find the difference between two fuzzy sets. This paper employs the absolute distance for simplicity. The amount of overlap between two images, $\mathcal{OV}$, is therefore defined in (6).

$$\mathcal{OV}(I_p, I_r) \equiv \sum_i |\mu_{S_{I_p}}(I_i) - \mu_{S_{I_r}}(I_i)|. \tag{6}$$

The new query image is then one that minimizes the overlap with the old query:

$$Q_{new} = \arg\min_i \mathcal{OV}(Q_{old}, I_i). \tag{7}$$

3.2. Query Fusion

After choosing a new query image Q_{new} (as in Section 3.1), a similarity set $\mathcal{S}_{new}$ is generated (Section 2). The membership degree of each database image I_i to this set is denoted as $\mu_{\mathcal{S}_{Q_{new}}}(I_i)$ and represents the similarity score between the images Q_{new} and I_i. The next step is to merge this similarity set from the one from the previous queries to determine an overall similarity. In order to guarantee the convergence of the Query Feedback algorithm, the following modifications are made to the previous set of rankings before fusion of the results. First, the ranks of the images marked as negative by the user are changed to zero. As a direct consequence of the proposed merging scheme, these negative examples will not show up in the top results in subsequent iterations of the algorithm. The second step is to modify the ranks of the images that are unlabelled or not visited by the user by multiplying them by a forgetting factor $\alpha \in (0, 1)$. Thus, if an image is not labelled in many consecutive iterations, its rank will approach zero and it will be treated a negative example. The ranks of the images that are not deemed relevant (i.e. are not ranked within the top N relevant images to a query) are multiplied by this forgetting factor. Thus, at iteration k, the rank of an image that it is not similar to any of the previous k queries, is decreased by α^k. If an image is similar to several positive examples, it has a higher likelihood of being retrieved during initial iterations since it will receive a high rank using any of such positive examples as the query. In Section 4 the effects of this parameter on the accuracy of retrievals are further investigated.

Finally, the modified set of results from the previous iteration, $\mathcal{S}'_{Q_t}$, and the set of results from the current iteration, $\mathcal{S}_{new}$, are merged to generate the new rankings for the images based on user feedback. As shown below, a compensatory operator is used to perform the merging:

$$\mathcal{S}_{Q_{t+1}} = \min(\mathcal{S}'_{Q_t}, \mathcal{S}_{new})^{\beta} \max(\mathcal{S}'_{Q_t}, \mathcal{S}_{new})^{(1-\beta)}. \tag{8}$$

This operator's logical behaviour ranges between the logical AND and OR depending on the parameter β. For $\beta = 1$, a logical AND (or intersection) of the two result sets is obtained and $\beta = 0$ results in a logical OR (or union) of the two resulting sets. For all other values of β a compromise between the AND and OR (or a soft intersection or union) is obtained. The effects of the compensation parameter on the retrieval results will be investigated in Section 4.

Another advantage of the operator of (8) is that it provides a null element that can be used to completely determine the decision regardless of the other elements. If an image is chosen as negative even once, it will receive the lowest possible similarity score (i.e., zero) in all subsequent retrievals regardless of its similarity to any of the query images.

3.3. Algorithmic View

The Query Feedback algorithm is shown below.

Input:
$0 < \alpha, \beta < 1$,
set of positive examples $\mathcal{P}$ & negative examples $\mathcal{N}$,
set of rankings from previous iteration, $\mathcal{S}_{Q_t}$.
Output:
A set of rankings for the images in the database, $S_{Q_{t+1}}$.

Query Feedback algorithm at iteration $t + 1$:
Step 1. Modify similarity set from previous iteration:
$\forall I_i \in \mathcal{P}, \mathcal{S}'_{Q_t}(I_i) = \mathcal{S}_{Q_t}(I_i)$,
$\forall I_i \in \mathcal{N}, \mathcal{S}'_{Q_t}(I_i) = 0$,
$\forall I_i \notin \mathcal{P} \vee \mathcal{N}, \mathcal{S}'_{Q_t}(I_i) = \alpha \mathcal{S}_{Q_t}(I_i)$,
Step 2. Choose the new query Q_{new} from $\mathcal{P}$ from (7).
Step 3. Obtain $\mathcal{S}_{Q_{new}}$ using the UFSC (Section 2).
Step 4. Calculate final ranks $S_{Q_{t+1}}$ from (8).

---

4. EXPERIMENTS

The experiments were carried out on a database of 2850 general images containing a wide variety of subjects such as flowers, sunsets, people, and animals. This type of database is chosen because large variations exist between the low-level representation of images in the same semantic class. Our experiments show that despite this semantic gap, the proposed algorithm performs well in learning the mapping between the low-level features and high-level user concepts.

The performance indicator employed in this work is the *Recall* measure defined in 9.

$$Recall(i) = \frac{N_{relevant}(i)}{N_{total}}, \tag{9}$$

where $N_{relevant}(i)$ represents the number of retrieved images relevant to the user query after i retrievals and N_{total} is the total number relevant images to the query as determined a priori by a human subject. The $Recall$ measure was averaged over 20 different classes to obtain $\overline{Recall}$.

Performance of the proposed system is compared to the Bayesian Estimation [2] and Falcon [4] methods. A plot of $\overline{Recall}$versus iterations of relevance feedback is depicted in Fig. 3 for the three methods. These results show that Query Feedback outperforms the

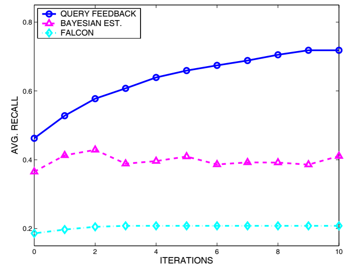
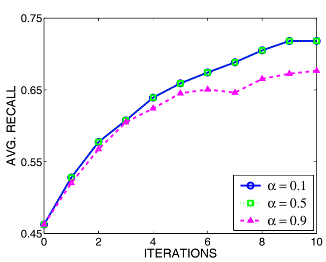


Fig. 3. Comparison of $\overline{Recall}$ for the proposed method, Integrated Probability Function, Bayesian estimation, and Falcon.

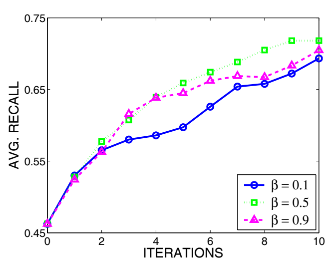
others by providing a peak improvement of 26% compared to 3% and 6% by Falcon, and Bayesian Estimation, respectively.

Fig. 4(a) shows the $\overline{Recall}$ plots for various values of α versus the number of iterations of the relevance feedback algorithm. It can be seen that values of this parameter close to unity cause instability in the algorithm as a drop in $\overline{Recall}$. The higher this value, the less strict the filtering of the outliers. This effect causes a performance degradation as a result of the false positives.

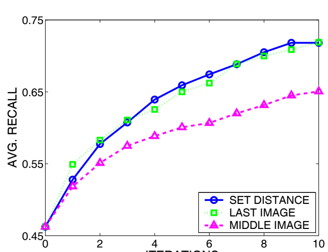
The compensation parameter β is used to merge the results of the various queries to obtain a final set of rankings. The choice of this parameter is dependent on the system, the database, and on the user needs. The $\overline{Recall}$ graph for various values of β are shown in Fig. 4(b) indicating that a value of 0.5 for this parameter results in the highest accuracy in terms of $\overline{Recall}$.



(a) $\overline{Recall}$ (various α values).



(b) $\overline{Recall}$ (various β values).



(c) $\overline{Recall}$ (various query selection methods).

Fig. 4. Effect of system parameters on $\overline{Recall}$.

Finally, the query selection algorithm based on set distance method of (7) is compared to the heuristic method of choosing the last positive example as indicated by (5). The results for a second heuristic method that chooses as the middle positive image as the query are also presented in Fig. 4(c). It can be seen that the proposed method performs slightly better than the heuristic ones. From these results it can be concluded that the complexity of the system can be greatly reduced by employing the heuristic method based on choosing the last positive image as the new query.

5. CONCLUSIONS

Most content-based image retrieval systems assume that the high-level user concepts can be modelled as single points in the low-level feature space. This presumption, however, is an incorrect one due to the semantic gap that exists between high-level user concepts and low-level features. Query Feedback, an interactive, user-centered retrieval technique, has been proposed in this paper to address these shortcomings by performing retrievals based on multiple query examples rather than just one. The overall results are then obtained as a fusion of these low-level concepts. The use of multiple query points allows the system to identify regions of the low-level space that correspond to different manifestations of the concepts of interest after a few iterations of relevance feedback. In contrast to the existing methods, Query Feedback determines the multiple query points by selecting only a subset of the user-provided positive images. These query points are chosen to maximize the information gain from the user feedback.

In addition to its retrieval effectiveness, the Query Feedback system does not require any assumptions regarding probability density functions, user models, or nature of the feature vectors. The design of the Query Feedback algorithm is general enough to accommodate various databases, users, and environments. Parameters such as the query selection algorithm, forgetting factor, and compensation parameter can be adjusted independently of the actual algorithm in order to satisfy specific system requirements in terms of performance and complexity.

6. REFERENCES

[1] Yong Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a powerful tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, 1998.

[2] Z. Su and H. Zhang and S. Li and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 924–937, 2003.

[3] M.R. Naphade, I.V. Kozintsev, and T.S. Huang, "Factor graph framework for semantic video indexing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 40–52, 2002.

[4] L. Wu, C. Faloutsos, K. P. Sycara, and T. R. Payne, "FALCON: Feedback adaptive loop for content-based retrieval," in *Proc. 26th VLDB Conference*, 2000, pp. 297–306.

[5] X. Jin and J. French, "Improving image retrieval effectiveness via multiple queries," in *Proc. of the first ACM international workshop on Multimedia databases*, 2003, pp. 86 – 93.

[6] A. Kushki, "An interactive framework for content-based image retrieval," Master's thesis, University of Toronto, 2003.

[7] Manjunath, B., Salembier, P., and Sikora, T., *Introduction to MPEG-7: multimedia content description interface*, Wiley, 2002.