

Learning Weighted Distances for Relevance Feedback in Image Retrieval

Thomas Deselaers^{1*}, Roberto Paredes^{2†}, Enrique Vidal^{2†}, and Hermann Ney^{1*}

¹ Computer Science Department – RWTH Aachen University, Aachen, Germany

² Instituto Tecnológico de Informática – Universidad Politécnica de Valencia, Valencia, Spain

E-mail: deselaers@cs.rwth-aachen.de

Abstract

We present a new method for relevance feedback in image retrieval and a scheme to learn weighted distances which can be used in combination with different relevance feedback methods. User feedback is a crucial step in image retrieval to maximise retrieval performance as was shown in recent image retrieval evaluations. Machine learning is expected to be able to learn how to rank images according to users needs. Most image retrieval systems incorporate user feedback using rather heuristic means and only few groups have formally investigated how to maximise the benefit from it using machine learning techniques. We incorporate our distance-learning method into our new relevance feedback scheme and into two different approaches from the literature. The methods are compared on two publicly available databases, one which is purely content-based and one which uses additional textual information. It is shown that the new relevance feedback scheme outperforms the other methods and that all methods benefit from weighted distance learning.

1. Introduction

Content-based Image Retrieval (CBIR) deals with finding images based on their content, without using any additional textual information. CBIR has been investigated for quite some time, with main focus on system centred methods [10]. In textual information retrieval, it was shown that relevance feedback, i.e. having a user judge retrieved documents and using these to refine the search can lead to a significant performance improvement [8]. In image retrieval (content-based, as well as using textual information), relevance feedback has been discussed in some papers, but in contrast to the text retrieval domain, where the Rocchio relevance feedback method can be considered as a reasonable and well-established baseline, no standard method is defined.

*Partly sponsored by the German research foundation (DFG) under contract number Ne-572/6

†Work supported by the Spanish Project Consolider Ingenio 2010: MIPRCV (CSD2007-00018)

A survey on relevance feedback techniques for image retrieval until 2002 was presented in [12]. Most approaches use the marked images as individual queries and combine the retrieval results. More recent approaches follow a query-instance-based approach [4] or use support vector machines to learn a two-class classifier [9]. The approach presented here, is similar to the approach presented in [4], because it also follows a nearest neighbour search for each query image, but instead of using only the best matching query/database image combination, we consider all query images (positive and negative) jointly.

The nearest-neighbour approach is appealing, because nowadays, most CBIR systems are based on nearest neighbour searches on arbitrary vectorial features rather than restricting the features to sparse binary feature spaces as they are used in the GNU Image Finding Tool (GIFT) [6] and therefore the technique can easily be integrated into most image retrieval systems.

Here, we present a technique that has both advantages: the individual queries (i.e. positive *and* negative images from relevance feedback) are considered independently and in a nearest neighbour-like manner, but machine-learning techniques are applied to optimise the search criterion for the nearest neighbours.

The learning step consists of tuning of weights for the distance function used in the nearest neighbour search. The learning of weights for nearest-neighbour searches (using L_2 -norm) was investigated in [7], where a weight for each component of the vectors to be compared is determined from labelled training-data. In contrast to this work, here we use the L_1 -norm, which is known to be better suited for histograms. Additionally, here the amount of available training data is much smaller than normally used for distance learning [7].

2. Relevance Feedback

In the following, we present our new methods for relevance feedback and compare it to relevance score [4] and Rocchio's method [8].

The methods presented can use negative feedback, but also work without. The number of feedback images

is not limited, so that a user is free to mark as many images as relevant/non-relevant as he likes, so that a user who marks many images is likely to obtain better results than a user who only marks a few images.

In our setup, given an initial query, the user is presented with the top-ranked 20 images from the database, and can choose to mark images as relevant, non-relevant, or to not mark them at all. Then, these judgements are sent back to the retrieval system which can use the provided information to refine the results and to provide the user with a new set of (hopefully) better results.

In each iteration of relevance feedback, the system can use machine learning techniques to refine its similarity measure and thus to improve the results. In Section 3, we present the new distance learning technique which can be used in combination with all retrieval schemes.

2.1. Combination of Classifiers

Combining classifiers is a well-known way of fusing information from different cues [5]. Here, we consider each marked image as basis for a nearest neighbour classifier with only one training sample and consider each database image to be a test example that has to be classified into the classes *relevant* (“*r*”) or *non-relevant* (“*n*”).

We assume that, given a relevant query image q_+ , i.e. an image which was marked relevant in the feedback process, the probability that an image x from the database is relevant $p_x(c = r|q_+)$ is

$$p_x(r|q) \propto \exp(-d(x, q_+)), \quad (1)$$

where $d(x, q_+)$ is an appropriate distance function comparing the images x and q_+ respectively their descriptors.

Analogously, we assume that the probability for an image being non-relevant $p(n|q_i)$ has the same relationship for negatively marked images q_- :

$$p_x(n|q_-) \propto \exp(-d(x, q_-)). \quad (2)$$

Given that $p(r|q_-) = 1 - p(n|q_-)$, we can use the sum rule [5] to fuse the output of the individual classifiers for the set of positive queries Q^+ and the set of negative queries Q^- :

$$p_x(r|(Q^+, Q^-)) = \frac{\alpha}{|Q^+|} \sum_{q_+ \in Q^+} p_x(r|q_+) + \frac{1-\alpha}{|Q^-|} \sum_{q_- \in Q^-} (1 - p_x(n|q_-)), \quad (3)$$

where we added additional weighting factor α to allow for flexibly changing the impact of the negative and positive queries.

This method can also be considered to be a kernel densities approach, because here not the classification itself but the ranking of the images is important and thus

prior probabilities can be disregarded. In the following sections we will refer to this method as Classifier Combination (CC).

2.2. Relevance Score

Relevance score (RS) has been inspired by the nearest neighbour classification method [4]. Instead of finding the best match for each query image among the database images, for each database image only the best matching query image is considered among the positive and negative query images. The ratio between the nearest relevant and the nearest non-relevant image is considered for ranking the images. RS is computed as

$$RS(x, (Q^+, Q^-)) = \left(1 + \frac{\min_{q_+ \in Q^+} d(x, q_+)}{\min_{q_- \in Q^-} d(x, q_-)} \right)^{-1} \quad (4)$$

and then images are ranked such that the images with highest relevance score are presented first.

Note that using the quotient of the probabilities from Section 2.1 is also possible and can be compared to using likelihoods for two-class discrimination but turned out to perform worse in informal experiments.

2.3. Rocchio Relevance Feedback

Rocchio’s method for relevance feedback [8] can be considered a de facto standard in textual information retrieval. In CBIR, it has been investigated in the context of the GIFT system [6]. In Rocchio relevance feedback, the individual query documents are combined into a single query according to

$$\hat{q} = q + \beta \left(\sum_{q_+ \in Q^+} q_+ \right) - \gamma \left(\sum_{q_- \in Q^-} q_- \right), \quad (5)$$

where \hat{q} is the new query, q is the query from the last feedback iteration and β and γ are weighting factors to determine the influence of relevance feedback, commonly the parameters are chosen $\beta = 1/|Q^+|, \gamma = 1/|Q^-|$.

Once, \hat{q} is determined it is used to query the database and find the most similar images, in a normal nearest-neighbour (or in textual information retrieval *tf/idf*-fashion).

3. Learning Weighted Distances

In the retrieval techniques described above, the distance function d comparing image descriptors is central. We use a weighted version of the L_1 distance, which is known to be a good choice to compare histograms:

$$d(x, q) = \sum_{i=1}^D w_i |x_i - q_i|, \quad (6)$$

where w_i is the weight for the i -th histogram bin of the query q_i and the database image x_i . If all w_i are chosen to be 1, this is the L_1 distance.

To learn the weights w_i for the distance function, we proceed analogously to the procedure proposed for weighted- L_2 [7] and consider the feedback images as training images for the nearest neighbour system. To improve the performance, we learn the weights such that the distances among the positively marked images are minimised whereas the distances between positively marked images and negatively marked images are maximised. In total the following term has to be minimised with respect to the w_i in the distance function d .

$$\sum_{x \in Q^+} \sum_{q_+ \in Q^+ \setminus \{x\}} \sum_{q_- \in Q^- \setminus \{x\}} \frac{d(x, q_+)}{d(x, q_-)} \quad (7)$$

and analogously a term for all negative query images has to be maximised. If only one positive or only one negative query image is given, only the respective other can be optimised.

The optimisation is done using gradient descent and effectively learns weights which simultaneously minimise the distance between relevant images while maximising the distances to non-relevant images and thus is expected to improve retrieval accuracy.

4. Experimental Results & Databases

We evaluate the methods on two publicly available databases, the MSRC database and the ImageCLEF 2007 photo retrieval database. The images from both databases are represented by a 512-dimensional colour histogram and a 512-dimensional Tamura texture feature histogram, which was shown to be a reasonable baseline for CBIR [2]. For the ImageCLEF database, we have an additional 9845-dimensional histogram representing the English textual description of the image. All histograms are compared using L_1 -distance and weighted L_1 . The weights for the descriptors are optimised individually, but in principal it would also be possible to concatenate the histograms, consider them to be one descriptor, and learn the weights jointly.

The experiments were performed fully automatically with ‘*simulated user feedback*’. Each query was performed and then all relevant images retrieved among the top 20 results were added to the set Q^+ and all non-relevant images among these were added to the set Q^- , thus effectively simulating a user who judging each of the twenty top-ranked images regarding its relevance. This procedure is repeated five times, leading to six precision values, the first one after the initial query and then five succeeding ones after each iteration of relevance feedback.

4.1. MSRC

The MSRC database was published by the Machine Learning and Perception Group from Microsoft Research, Cambridge, UK and is available on-



Figure 1. Example images from the databases (a) MSRC database, (b) ImageCLEF 2007 photo retrieval database.

Table 1. Results on the MSRC database using up to 5 iterations of relevance feedback.

method	$P_0(20)$	$P_1(20)$	$P_2(20)$	$P_3(20)$	$P_4(20)$	$P_5(20)$
CC	0.518	0.731	0.810	0.849	0.869	0.880
CC weighted	0.518	0.750	0.830	0.859	0.890	0.892
RS	0.518	0.701	0.787	0.839	0.872	0.897
RS weighted	0.518	0.722	0.807	0.849	0.872	0.895
Rocchio	0.518	0.639	0.684	0.697	0.702	0.704
Rocchio weighted	0.518	0.672	0.702	0.721	0.741	0.742

line¹. It consists of 4320 images from 33 classes such as aeroplanes, bicycles/general, bicycles/sideview, sheep/general, sheep/single and is generally considered a difficult task [11]. Some example images from this database are shown in Figure 1. Experiments on the MSRC database are carried out in a leaving-one-out manner. That is, each image is used as a query to retrieve relevant images (i.e. images from the same class) from the remainder of the database.

Experimental results are in Table 1 and Figure 2. Interestingly, the improvement is larger for the CC method (red) which even without weighted distance (dotted) clearly outperforms RS (green) in particular in the first relevance feedback iterations. This is an interesting and desirable property of an image retrieval systems where the user wants to obtain an acceptable precision with as little effort as possible. Furthermore, it is worth to mention that the distance weight learning (dotted lines) consistently improves all the methods and that both CC and RS clearly outperform Rocchio’s method (blue) although the gain obtained from distance learning in Rocchio’s method is considerable.

4.2. ImageCLEF 2007 Tasks

The ImageCLEF 2007 photo retrieval database was used for the 2007 ImageCLEF image retrieval evaluation [1]. This database consists of a total of 20,000 images with 60 queries where relevant images have been determined manually according to the textual descrip-

¹<http://research.microsoft.com/vision/cambridge/recognition/default.htm>

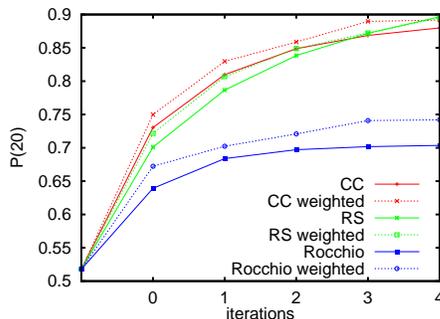


Figure 2. Results on the MSRC database.

Table 2. Results on the ImageCLEF photo dataset using up to 5 iterations of relevance feedback

method	$P_0(20)$	$P_1(20)$	$P_2(20)$	$P_3(20)$	$P_4(20)$	$P_5(20)$
CC	0.222	0.463	0.578	0.643	0.696	0.727
CC weighted	0.222	0.467	0.598	0.693	0.728	0.728
RS	0.256	0.450	0.535	0.587	0.627	0.647
RS weighted	0.256	0.449	0.550	0.591	0.649	0.649
Rocchio	0.179	0.266	0.300	0.316	0.323	0.324
Rocchio Weighted	0.182	0.282	0.351	0.450	0.479	0.482

tion of the intended meaning of the queries. Each of the queries consists of 3 images which are used to initialise the set Q^+ (thus the differences in the initial retrieval result).

Experimental results are in Table 2 and Figure 3. The results obtained for this dataset again show that the CC approach is clearly better than the other methods. The CC approach (red) reaches 12% of improvement over the RS (green) technique and is more than twice as good as the Rocchio method (blue). Here again, the proposed distance weight learning (dotted) consistently leads to an improvement for all methods. Again Rocchio’s method is improved most which shows that the distance weighting is able to compensate for the limited flexibility imposed by the single-prototype query.

For comparison, the best result using user interaction in the 2007 ImageCLEF photo retrieval evaluation was $P(20)=0.459$, obtained by the submission of the University in Chemnitz, Germany using textual and visual information, user feedback (unspecified number of iterations) and automatic query expansion. The CC method is slightly better after one iteration of relevance feedback and clearly outperforms with more iterations.

5. Conclusion

We presented a novel relevance feedback scheme based on classifier combination and a method to automatically tune weights in a distance function for content-based image retrieval which can be incorporated into most distance-based image retrieval system. The weighting scheme is integrated into the novel scheme and two

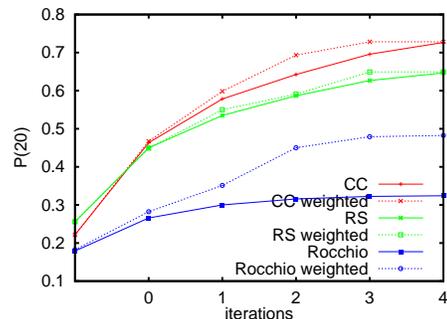


Figure 3. Results on the ImageCLEF database.

schemes from the literature and leads to clear improvements in all. All in all the classifier combination scheme in combination with the weight learning outperforms all other methods which was evaluated on two different databases, one purely content-based task, and another one incorporating textual and visual information.

References

- [1] P. Clough, M. Grubinger, A. Hanbury, and H. Müller. Overview of the imageclef 2007 photographic retrieval task. In *CLEF 2007 Workshop*, LNCS, in press, Budapest, Hungary, 2008.
- [2] T. Deselaers, D. Keysers, and H. Ney. Features for image retrieval: An experimental comparison. *Information Retrieval*, in press, 2008.
- [3] G. Giacinto. A nearest-neighbor approach to relevance feedback in content-based image retrieval. In *CIVR*, Amsterdam, The Netherlands, July 2007.
- [4] G. Giacinto and F. Rolli. Instance-based relevance feedback for image retrieval. In *NIPS*, Vancouver, Canada, Dec. 2004.
- [5] J. Kittler. On combining classifiers. *PAMI*, 20(3):226-239, Mar. 1998.
- [6] H. Müller, W. Müller, S. Marchand-Maillet, and D. M. Squire. Strategies for positive and negative relevance feedback in image retrieval. In *ICPR 2000*, pp. 1043-1046, Barcelona, Spain, Sept. 2000.
- [7] R. Paredes and E. Vidal. Learning weighted metrics to minimize nearest neighbor classification error. *PAMI*, 28(7):1100-1110, 2006.
- [8] J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retrieval System: Experiments in Automatic Document Processing*, pp. 313-323. Prentice-Hall, Englewood Cliffs, NJ, USA, 1971.
- [9] L. Setia, J. Ick, and H. Burkhardt. SVM-based relevance feedback in image retrieval using invariant feature histograms. In *Workshop on Machine Vision Applications*, Tsukuba Science City, Japan, May 2005.
- [10] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *PAMI*, 22(12):1349-1380, Dec. 2000.
- [11] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *ICCV*, volume 2, pp. 1800-1807, Beijing, China, Oct. 2005.
- [12] X. S. Zhou and T. S. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8:536-544, 2003.