

Deep Exploration for Experiential Image Retrieval

Bart Thomee

Mark J. Huiskes

Erwin Bakker

Michael S. Lew

LIACS Media Lab, Leiden University
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
{bthomee, markh, erwin, mlew}@liacs.nl

ABSTRACT

Experiential image retrieval systems aim to provide the user with a natural and intuitive search experience. The goal is to empower the user to navigate large collections based on his own needs and preferences, while simultaneously providing him with an accurate sense of what the database has to offer. In this paper we integrate a new browsing mechanism called deep exploration with the proven technique of retrieval by relevance feedback. In our approach, relevance feedback focuses the search on relevant regions, while deep exploration facilitates transparent navigation to promising regions of feature space that would normally remain unreachable. Optimal feature weights are determined automatically based on the evidential support for the relevance of each single feature. To achieve efficient refinement of the search space, images are ranked and presented to the user based on their likelihood of being useful for further exploration.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Relevance feedback, Search process.*

General Terms

Algorithms, Human Factors.

Keywords

Content-based image retrieval, Relevance feedback, Feature space exploration, Feature selection and weighting, Deep exploration.

1. INTRODUCTION

Over the past years we have seen image collections grow tremendously, both in a personal sense (e.g. home photo collections) and in a public sense (e.g. image databases on the internet). As a result, finding images of interest has become more and more like finding a needle in a haystack. Recent advances in retrieval techniques have progressed the state of the art significantly, but the problem is not solved yet. One popular direction at the moment is long-term learning (e.g. [4]), where from the accumulated feedback of users common denominators between similar images are distilled. Another approach is to take a high-level view by focusing on core image concepts and semantics (e.g. [7]). Finding subspaces or manifolds in feature space where similar

images reside (e.g. [2]) also has received much attention lately. This diversity in techniques notwithstanding, the general consensus is that incorporating relevance feedback leads to improved search results and therefore has been applied in the majority of research from the moment the concept was introduced by Rocchio [1] in 1971.

Despite its potentially great impact on user satisfaction and retrieval performance, the interface is often still a largely ignored component. Most work only focuses on how to present the search results [3], whereas research on how to assist the user to search more efficiently is rather limited (e.g. [7-9]). One of the grand challenges in our field is considered to be the need for experiential exploration systems that allow the user to gain insight into and support exploration of media collections [6]. For users, exploration is the predominant mode of interaction, rather than querying, and therefore interfaces that accommodate for this behavior are needed [5]. In this paper we propose such a system, where the user can visually explore the feature space around relevant images and effortlessly navigate from one area in feature space to another to discover more relevant images using a technique we call *deep exploration*.

Unlike in semantic spaces, where ideally all images of interest are clustered together in a certain area, in low-level feature space these images might be scattered over multiple areas. Feature selection and weighting is one way to transform the feature space so that images that are perceptually close to each other are also close to each other in the resulting space. Many retrieval systems (e.g. [11-13]) analyze the feedback given by the user to figure out which image features are important and also how important they are. In our system, the user focuses the search on the regions in feature space that are relevant, where each of these regions centers on a relevant example image and is bounded by its relevant nearest neighbors. The images contained within these regions are used to automatically determine the optimal set of feature weights for an image when it is explored, based on the evidential support for the relevance of each single feature.

In Section 2 we will look at the proposed exploration technique and the feature weighting approach is discussed in Section 3. Section 4 describes the experiments we performed and we conclude in Section 5.

2. EXPLORING FEATURE SPACE

In low-level feature space images of interest can be spread out over multiple areas. For example in a feature space built up on color features, it is likely that images of differently colored tulips can be found in several parts of the space. Such a search can be performed using multiple query points (e.g. [10]), but exploring the feature space around each query point is often a slow process. Most interfaces only present a limited number of images to the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'09, October 19–24, 2009, Beijing, China.

Copyright 2009 ACM 978-1-60558-608-3/09/10...\$10.00.

user, putting a heavy burden on the user as navigating the space around each query point requires many iterations of feedback. To reduce user effort and allow efficient refinement of the relevant search space, we propose a novel technique whereby the user can visually and interactively explore the feature space around an image and transparently navigate from one area in feature space to another to discover more relevant images.

2.1 Interactive Visualization

In retrieval systems based on relevance feedback, the user indicates his preferences regarding the presented results by selecting images as positive and negative examples. Subsequently, these feedback samples determine a relevance ranking on the image collection, and new images are presented to the user for the next round of feedback.

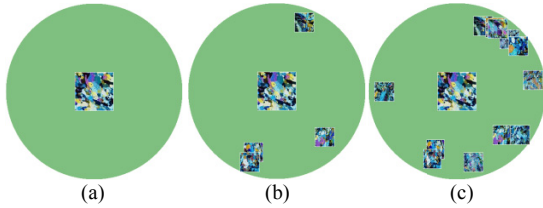


Figure 1. Exploring feature space: a) Initially only the selected image is shown in the center. b-c) The user expands the exploration range several times by scrolling the mouse wheel. The distance of an image to the center increases linearly with their distance in feature space. The feature weights used for calculating the distance are discussed in Section 3.

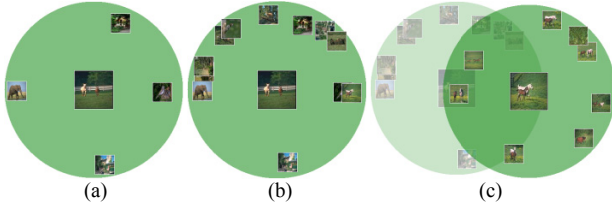


Figure 2. Deep exploration: a) The user is looking for images of horses and explores a relevant image to discover more. However, its nearest neighbors in feature space are not relevant at all. b) The exploration range is increased and a few relevant images are found. c) The user moves the focus of exploration to one of them and this time many of its nearest neighbors are relevant.

In this paper, we propose to integrate feedback selection with a visualization mechanism that allows us to quickly explore the local feature space surrounding an example image. The interaction provides a better sense of the local structure of the database, and allows us to center on examples that best capture the qualities desired by the user. This approach, which we call *deep exploration*, entails the following process.

At the start of an exploration interaction only the selected image is displayed. Then, by adjusting the *exploration front*, more and more of its nearest neighbors are shown, see Figure 1. When the number of nearest neighbors becomes too large to be displayed in a comprehensible manner, a random selection is displayed. The actual deep exploration occurs when the user encounters an image of interest in the exploration range, and decides to transfer the focus to this image to continue the exploration with the new image. This provides the user the opportunity to easily reach other

areas in feature space, see Figure 2 for an example. This can be done as many times as the user wants, jumping from one area in feature space to another. This technique is particularly useful when the search seems to be 'stuck' and cannot improve with the current collection of relevant images. Also, using deep exploration to move from an isolated relevant image to a more densely populated relevant area has direct benefits for the feedback analysis, e.g. the feature selection and weighting approach will perform better with the additional data.

At any stage, the user may decide to treat a centered image as a positive example. In that case all images within the exploration front will also be treated as positive examples, so the exploration range should ideally encompass a high fraction of images considered as relevant. When desired, non-relevant images can be removed by the user at a later stage. Similarly, a selected example can be treated as a negative example. In the end, the retrieval system collects the positive and negative example images corresponding to the selected exploration ranges, and performs the relevance feedback analysis of the next section to determine both (i) the images estimated to be most relevant, and (ii) the images estimated to be most informative, i.e. optimal for display in the next iteration of exploration and feedback.

2.2 Feedback Sets

Let the positive feedback example set S_t^+ at iteration t consist of all selected relevant images gathered thus far

$$S_t^+ = \{(s_1^+, w_1^+, r_1^+), \dots, (s_{n_t}^+, w_{n_t}^+, r_{n_t}^+)\}, \quad (1)$$

where, for each example image s_i^+ , w_i^+ are the corresponding feature weights at the time of exploration and r_i^+ is the exploration range as selected by the user. The negative feedback example set S_t^- at iteration t is defined similarly. When a user re-explores an image, its previous feature weights and exploration range are updated with their new counterparts.

Let A_{ti}^+ be the set of images at iteration t within the exploration range of a positive feedback example

$$A_{ti}^+ = \{x \in D \mid d_{w_i^+}(s_i^+, x) < r_i^+\}, \quad (2)$$

where x is an image from the image database and (s_i^+, w_i^+, r_i^+) are from the i -th tuple of S_t^+ . Let A_{ti}^- at iteration t be defined similarly. We now define the active set A_t as the set of images at iteration t that are in at least one of the positive sets and not in the negative sets

$$A_t = \bigcup_{i=1}^{n_t^+} A_{ti}^+ \setminus \bigcup_{i=1}^{n_t^-} A_{ti}^-. \quad (3)$$

To determine the most informative images, we first calculate the information score T_I of each active image a at iteration t as the minimum distance to their associated feedback examples

$$T_I(a) = \min_{(s,w,r) \in S_t^+} d_w(s, a). \quad (4)$$

Note that an active image can be in the exploration range of several feedback images. Next, we pick the images with the highest information scores, thus maximizing the minimum distances. As a result, images on the border of our search space will obtain the highest information score.

Besides the most informative images, which allow the user to continue exploring the feature space, we keep an image set that contains the best images thus far. For each active image a at

iteration t we calculate a relevance score T_R that is dependent on the distance to its feedback point(s)

$$T_R(a) = \sum_{i=1}^{n_t^+} \frac{\mathbf{1}_{\{x|d_{w_i^+}(s_i^+, x) < r_i^+\}}(a)}{(1 + \gamma d_{w_i^+}(s_i^+, a))}, \quad (5)$$

where $\mathbf{1}_A(x)$ is an indicator function, indicating the membership of x in set A and γ a constant that quantifies the rate of relevance decrease as an active image approaches the border of the exploration range.

3. FEATURE WEIGHTING

The collection of explored feedback images provides us with a convenient setup for *local* feature selection. In particular, it allows us to take into account prior feature density by giving higher weight to feature regions where images cluster unexpectedly. This is desirable given that for features to which the user is indifferent, clustering will naturally occur at the feature regions of high prior density. In our method, the influence of the latter kind of features is suppressed. The resulting local feature weights are used to measure image similarity, (i) to new images to be explored, and (ii) to example images s , through the distance functions $d_w(s, x)$ of equations (4) and (5). For (ii), feature weights w_{ij} are computed for each sample image s_i . In the following we suppress image index i from the notation. Our approach is to estimate the prior feature value density corresponding to the local clustering of examples at the image under study and set the distance function weights accordingly.

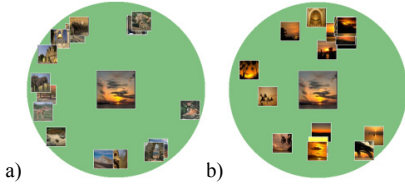


Figure 3. Default feature weights (a) vs. optimized weights (b)

We use the distribution of the relevant examples to establish the feature interval that we should look at for estimating the prior feature value density. Consider the absolute deviations in feature j between image s and each of the active images

$$a \in A_t: |s_j - a_j|. \quad (6)$$

Taking the median of this sequence gives us the median absolute deviation, mad_j , which offers a measure of the spread of the active images around s for each of the features j . We will now let the local feature weight depend on the prior feature value density of the estimated interval $[s_j - \text{mad}_j, s_j + \text{mad}_j]$.

The density p can be estimated using standard non-parametric methods based on quantization. As we have normalized our data, we found that a reasonable and fast approximation of this density can be obtained by means of the standard normal cumulative distribution function Φ

$$\begin{aligned} p_j &= p([s_j - \text{mad}_j, s_j + \text{mad}_j]) \\ &= \Phi(s_j + \text{mad}_j) - \Phi(s_j - \text{mad}_j). \end{aligned} \quad (7)$$

We then transform this density into a weight w_j using

$$w_j = \frac{1}{1 + \beta p_j}, \quad (8)$$

where β is a constant that controls how fast the weight decreases for increasing density. High feature weights are achieved for

intervals of small prior probability and low weights for intervals with high prior probability. Stronger selection can be enforced by first thresholding p_j , i.e. setting p_j to zero when p_j is larger than the threshold. After normalizing the weights, the resulting distance function to image s is

$$d_w(s, x) = \sqrt{\sum_{j=1}^M w_j (s_j - x_j)^2}. \quad (9)$$

As an illustration, in Figure 3a we can see an image of a sunset that is explored with default feature weights. After several iterations of feedback the feature weights have changed and when the image is re-explored, as is shown in Figure 3b, its nearest neighbors are more relevant than they were before.

4. EXPERIMENTS

We used two test databases for the experiments, the Ponce Texture Database [14] and the Corel5k Database [15], where the images are categorized into 25 and 50 classes, respectively. All images are represented by the MPEG-7 Homogeneous Texture Descriptor [16] and by a color histogram, based on a uniform quantization in 152 bins.

We have performed our experiments on three systems, i) one using our proposed exploration interface and deep exploration ('Deep Explore'), ii) one using the exploration interface without deep exploration ('Standard Explore'), and iii) one using a standard interface and the query point movement technique as proposed by Rocchio ('Rocchio'). As one of the strengths of our approach is the interface that provides easy access to additional relevant and non-relevant images, we acknowledge that the Rocchio system cannot be fairly compared with both Explore systems. However, as the systems try to achieve the same goal and are given the same data to work with, we believe that in this sense all systems are comparable.

For each of the image classes we have set up experiments, where the goal is to find all images belonging to that class within at most 20 iterations. Every iteration the user is presented with 40 images. For both Explore systems, these images are composed of the most informative images as calculated by (4). For the Rocchio system, these images consist of the resulting images after performing query point movement. Besides the images on which feedback can be given, a separate result set is kept that contains the best ranking images. For the Explore systems, these images are composed of the top images as calculated by (5). For the Rocchio system, the best ranking set is the same set as the informative set, containing the resulting images after query point movement. In our results we define *precision* as the number of relevant images found over the top 40 best ranking images.

As we did not have access to a large group of users to participate in the experiments, we have simulated realistic user behavior to the best of our ability. As real users generally do not want to give much feedback, in our simulations per iteration a maximum of 5 images are marked as relevant and a maximum of 5 as non-relevant. For the Deep Explore system, the simulated user will attempt to shift the focus of exploration when the number of relevant nearest neighbors of an explored image is small. In this situation, this user will increase the exploration range to only the 100 nearest neighbors and, when one or more relevant images are

contained within the exploration range, the one that is furthest away will be explored. Up to 4 shifts of exploration can be chained together and the image with the largest exploration range is used as a positive feedback image. Each experiment we fill the initial screen with random images. Because retrieval performance depends on which images appear within this initial screen, specifically on how many of these random images belong to the class of interest, we perform the experiment for each class 100 times and average the results. If the initial screen does not contain any relevant image, we generate a new set of random images until at least one relevant image is shown.

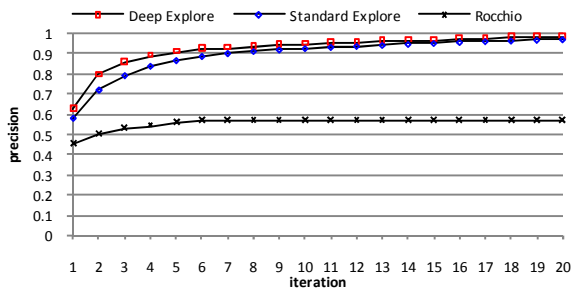


Figure 4. Average precision results on the Ponce database.

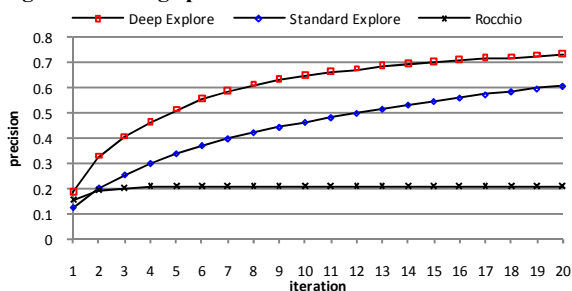


Figure 5. Average precision results on the Corel5k database.

As we can see in Figures 4 and 5, the average precision for both Explore systems is much higher than for the Rocchio system, where the latter system only slightly improves during the first few iterations. As the classes in the Ponce database are quite uniform, both Explore systems quickly improve and are eventually able to find almost all images, with Deep Explore having a small performance benefit due to the advantages of the deep exploration technique. The classes in the Corel5k database are more diverse, and especially in this situation the deep exploration technique helps to boost the performance of the Deep Explore system significantly over the Standard Explore system.

5. CONCLUSIONS

In this paper, we have proposed an exploration-based interface that allows the user to visually and interactively explore the feature space around images. Using the deep exploration technique, relevant areas in feature space can be discovered that would otherwise remain unnoticed. In addition, when an image is explored, its optimal set of feature weights is automatically determined using all images contained within the relevant regions, based on the evidential support for the relevance of each single feature. We performed user experiments on two well-known image databases and the results indicate that the new deep exploration approach leads to an improvement of the amount of relevant images collected.

6. ACKNOWLEDGMENTS

Leiden University and NWO BSIK/BRICKS supported this research under grant #642.066.603

7. REFERENCES

- [1] Rocchio, J.J. 1971. Relevance Feedback in Information Retrieval. *The Smart Retrieval System: Experiments in Automatic Document Processing*, G. Salton, Ed. Prentice Hall, Englewoods Cliffs, 313-323.
- [2] He, X., Cai, D., and Han, J. 2008. Learning a Maximum Margin Subspace for Image Retrieval. *IEEE Transactions on Knowledge and Engineering* 20(2), 189-201.
- [3] Datta, R., Joshi, D., Li, J., and Wang, J.Z. 2008. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys* 40(2), article 5.
- [4] Hoi, S.C.H., Lyu, M.R., and Jin, R. 2006. A Unified Log-Based Relevance Feedback Scheme for Image Retrieval. *IEEE Transactions on Knowledge and Data Engineering* 18(4), 509-524.
- [5] Jain, R. 2003. Experiential Computing. *Communications of the ACM* 46(7), 48-55.
- [6] Lew, M.S., Sebe, N., Djeraba, C., and Jain, R. 2006. Content-Based Multimedia Information Retrieval: State of the Art and Challenges. *ACM Multimedia Systems* 2(1), 1-19.
- [7] Fan, J., Gao, Y., Luo, H., and Jain, R. 2008. Mining Multi-level Image Semantics via Hierarchical Classification. *IEEE Transactions on Multimedia* 10(2), 167-181.
- [8] Nguyen, G.P., and Worring, M. 2006. Similarity Learning via Dissimilarity Space in CBIR. In *Proceedings of the ACM Workshop on Multimedia Information Retrieval*, 107-116.
- [9] Zavesky, E., Chang, S.-F., and Yang, C.-C. 2008. Visual Islands: Intuitive Browsing of Visual Search Results. In *Proceedings of the ACM Conference on Image and Video Retrieval*, 617-626.
- [10] Ortega-Binderberger, M., and Mehrotra, S. 2004. Relevance Feedback Techniques in the MARS Image Retrieval System. *ACM Multimedia Systems* 9(6), 535-547.
- [11] Grigorova, A., De Natale, F.G.B., Dagli, C., and Huang, T.S. 2007. Content-Based Image Retrieval by Feature Adaptation and Relevance Feedback. *IEEE Transactions on Multimedia* 9(6), 1183-1192.
- [12] Tieu, K., and Viola, P. 2004. Boosting Image Retrieval. *International Journal of Computer Vision* 56(1), 17-36.
- [13] Wu, Y., and Zhang, A. 2004. Interactive Pattern Analysis for Relevance Feedback in Multimedia Information Retrieval. *ACM Multimedia Systems* 10(1), 41-55.
- [14] Lazebnik, S., Schmid, C., and Ponce, J. 2005. A Sparse Texture Representation using Local Affine Regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(8), 1265-1278.
- [15] Duygulu, P., Barnard, K., De Freitas, N., and Forsyth, D. 2002. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the 7th European Conference on Computer Vision*, IV:97-112.
- [16] Ro, Y.M., Kim, M., Kang, H.K., Manjunath, B.S., and Kim, J. 2001. MPEG-7 Homogeneous Texture Descriptor. *ETRI Journal* 23(23), 41-5