LONG-TERM SIMILARITY LEARNING IN CONTENT-BASED IMAGE RETRIEVAL

J. Fournier *M. Cord*

Equipe Traitement des Images et du Signal (CNRS UPRES-A 8051)
University of Cergy-Pontoise / E.N.S.E.A, FRANCE
E-mail: {fournier,cord}@ensea.fr

ABSTRACT

This paper presents a new learning technique for the similarity model refinement in CBIR systems. We propose a whole retrieval strategy based on a new relevance feedback scheme and on a long-term similarity learning algorithm which uses feedback information of previous sessions. We introduce this technique as the simple evolution of the short-term relevance feedback approach into a long-term similarity learning, without additional need of user interaction. Our algorithm is validated via a quality assessment realized on a heterogeneous database of 1,200 color images.

1. INTRODUCTION

The semantic gap between the low-level similarity and the high-level user's query is the main problem in Content-Based Image Retrieval. It leads classical search algorithms to erroneous results. CBIR researchers have studied various relevance feedback methods in order to bridge this gap by allowing systems to interactively improve their discriminatory capabilities [1] [2] [3].

Content-based image retrieval methods are divided into two approaches. The first one is the *probabilistic method*. A probability distribution representing the uncertainty about the user's aim is build for the whole database. Relevance feedback helps to update this distribution thanks to user interaction [1]. The second approach is the *search-by-similarity* or search-by-example. Images are ranked according to a similarity (or dissimilarity) measure processed with reference to the user's query (usually an example image) [3] [4]. Relevance feedback allows here to refine the query and/or the similarity function.

Even if PicHunter [1] uses off-line measurement trials in order to build the *user model* of its Bayesian similarity measure, and even if FourEyes [5] tries to build across session groupings, the information acquired thanks to relevance feedback is often lost at the end of the search. Long-term learning techniques which try to propagate this information across retrieval sessions still appears as an open issue [6].

This paper presents a new learning technique based on short-term and long-term relevance feedback processes. The search strategy uses the similarity function refinement algorithm presented in our previous work [4], as well as a new compound query technique. We also propose a long-term learning technique which builds and refines a "semantic" similarity thanks to information collected during previous experiments (previous search sessions).

Our short-term retrieval strategy is presented in section 2, section 3 introduces the long-term similarity learning and section 4 is dedicated to the quality assessment.

2. RELEVANCE FEEDBACK AND SHORT-TERM RETRIEVAL STRATEGY

Our retrieval strategy belongs to the search-by-similarity approach. It is designed for target search as well as category search [1]. The image is indexed by $M = 2$ statistical distributions (one for the color model and one for the texture model) of N bins (The reader may refer to [7] for details).

The goal of relevance feedback is to refine the similarity function and the query, in order to improve the system's similarity model. The process is iterative and interactive: the user annotates (relevant/irrelevant) the displayed images and the feedback information is used to update the similarity model.

2.1. Similarity function and competition between models

The similarity between the request R and the target image T (respectively indexed by $\boldsymbol{R}$ and $\boldsymbol{T}$) is processed via a weighted sum of the similarities provided by each of the models:

$$S(\boldsymbol{R},\boldsymbol{T}) = \sum_{k=1}^{M} \beta^{(k)} S^{(k)}(\boldsymbol{R},\boldsymbol{T}) \qquad (1)$$

where $\beta^{(k)} \in \mathbb{R}^+$ and $S^{(k)} = 1 - d^{(k)}$ and $d^{(k)}$ is the city-block distance between two histograms.

Since each model leads to different classifications of the target images [8], it is important to detect and reinforce (via $\beta^{(k)}$) the most discriminating models for the current search. This process is called *competition between models*.

For a given target image, the aim is to minimize the quadratic error between the content-based similarity (also called visual similarity) and the desired similarity assessed via the user's annotations. For a target T, the desired similarity S_D^T is equal to 1 if T is relevant and 0 otherwise.

Our competition of models belongs to the *optimization-based* methods [9] since it aims at minimizing an objective criterion. The optimization is performed thanks to the LMS rule [10] ($1 \leq k \leq M$, $\mu^T \in \mathbb{R}^+$):

$$\beta^{(k)*} = \beta^{(k)} + \mu^T (S_D^T - S(\boldsymbol{R}, \boldsymbol{T})) S^{(k)}(\boldsymbol{R}, \boldsymbol{T}) \beta^{(k)} \quad (2)$$

2.2. Compound query and similarity merging

The relevant images are not clustered in the feature spaces [8]. In order to cope with multi-modal and scattered distributions, all examples collected via user interaction are gathered in the *compound query*. If L images have been annotated as relevant since the beginning of the search, the query becomes: $\boldsymbol{R} = \{\boldsymbol{R}_l, 1 \leq l \leq L\}$

Each example provides a similarity measure to the target. The similarity calculation between the compound query and the target is based on a fusion operator called *barycenter*:

$$S^{(k)}(\boldsymbol{R}, \boldsymbol{T}) = \frac{\sum_{l=1}^{L} a_{k,l} \; S^{(k)}(\boldsymbol{R}_l, \boldsymbol{T})}{\sum_{l=1}^{L} a_{k,l}} \quad (3)$$

where the weights are the similarities: $a_{k,l} = S^{(k)}(\boldsymbol{R}_l, \boldsymbol{T})$

On the one hand, if the relevant image distribution is unimodal, the behavior of this operator is close to the mean of similarities. On the other hand, maximum similarities are reinforced if relevant images have a multi-modal or scattered distribution. This merging scheme is robust to isolated examples and appears as an alternative to Kernel based methods [11] [12].

2.3. Asymmetrical learning

As written by Huang and Zhou [9]: *"the positive examples are all good in the same way, but bad examples are bad in their own ways"*. If irrelevant images may be informative for the competition between models, this information may also be inconsistent. The main problem encountered in our optimization process is the small size of the learning set which may contain more counter-examples than examples. The stability of the optimization process is not ensured because of counter-examples, that is why we reduce the influence of the irrelevant images in the LMS rule by decreasing the learning rate for the counter-examples. This parameter is larger for relevant (μ_{rel}^T) than for irrelevant images (μ_{irrel}^T): $\mu_{rel}^T = 10 \cdot \mu_{irrel}^T$

Otherwise, the learning set is made of all relevant and irrelevant images accumulated since the beginning of the search session. The asymmetrical learning rule combined with the example accumulation helps to make the optimization process more stable.

The competition of models and the compound query rely on varied sets of relevant and irrelevant images. Whereas classical search-by-similarity algorithms simply sort the images according to the similarity measure, we have developed an original browsing process making the system able to retrieve "remote" images [8]. Actually, browsing is necessary in order to cope with multi-modal and scattered distributions of the relevant images.

3. LONG-TERM SIMILARITY LEARNING

In CBIR systems, all information collected during the search is lost at the end of the session. We propose to build a new similarity measure which exhibits semantic properties thanks to the information gathered during previous retrieval sessions.

3.1. Similarity learning rule

For a given query, the user's annotations reflect the similarity between images from a semantic point of view. The aim of long-term similarity learning is to build, via experiments, a *semantic* similarity measure between two images.

Given a request image (R), a target image (T) and the desired similarity provided by user's annotations (S_D^T), the long-term similarity between R and T is learned thanks to the following rule:

$$S_{LT}(\boldsymbol{R}, \boldsymbol{T}) = S_{LT}(\boldsymbol{R}, \boldsymbol{T}) + \mu(S_D^T - S_{LT}(\boldsymbol{R}, \boldsymbol{T})) \quad (4)$$

This long-term similarity only depends on the user's annotations. This new information source is used in parallel of the content-based similarity (eq. (1)). Both similarities S and S_{LT} are combined in a new measure S_{comb}:

$$S_{comb}(\boldsymbol{R}, \boldsymbol{T}) = S(\boldsymbol{R}, \boldsymbol{T}) \cdot S_{LT}(\boldsymbol{R}, \boldsymbol{T}) \quad (5)$$

The long-term similarities between all image pairs are initialized to 0.5. It corresponds to a prior non-informative value (no learned similarity available). For a compound query $\boldsymbol{R} = \{\boldsymbol{R}_l, 1 \leq l \leq L\}$, the long-term similarity is processed as the mean over all examples: $S_{LT}(\boldsymbol{R}, \boldsymbol{T}) = \frac{1}{L} \sum_{l=1}^{L} S_{LT}(\boldsymbol{R}_l, \boldsymbol{T})$. All examples and counter-examples are stored during the search and the long-term similarity updating rule (eq. (4)) is applied at the end of every retrieval session.

3.2. Simulation

In order to illustrate the long-term learning process, we have simulated searches in a small image database containing 319 images arranged in 5 semantic categories. The learned long-term similarities are stored in a squared matrix $\boldsymbol{M}$ of size 319×319 where $M_{i,j}$ stands for the similarity between a query image i and a target image j. Figure 1 shows this similarity matrix.

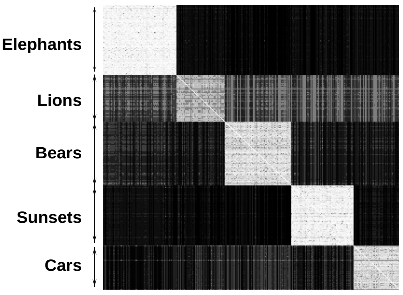


Fig. 1. Example of a long-term similarity matrix and semantic categories obtained after learning.

The learned values are close to 1 (white) for the image pairs of the same semantic category whereas they are close to 0 (black) for images of different classes. The ideal matrix (or semantic similarity matrix) is a block diagonal matrix with binary entries (1 in the blocks). This simple example clearly shows that the long-term similarity reflects the semantic similarity between images.

3.3. Discussion

PicHunter [1] proposes off-line trials to tune the user model at the heart of its Bayesian similarity measure. In our approach, the similarity between the image pairs is directly learned independently of the image features.

Our long-term learning algorithm is close to reinforcement learning [13] and S_D^T can be interpreted as the reinforcement signal. The search strategy manages the *exploration/exploitation* dilemma. On the one hand, exploration aims at building a reliable long-term information for each of the image pairs. On the other hand, the retrieval strategy exploits its knowledge (long-term and content-based similarities) to satisfy the user's request.

The combination function that we use to merge long-term and visual similarities may be changed in order to tune the influence of the short-term or long-term model. In other words, the system works even if no long-term information is available. Actually, the retrieval effectiveness increases with the number of uses.

The number of uses the system needs to build a reliable long-term similarity for all image pairs depends on the exploration effectiveness and on the size of the database. Nevertheless, it is possible to speed up learning thanks to refined algorithms as Dyna-Q in Q-learning [13], for instance.

4. EXPERIMENTS

Experiments have been carried out on a general database made of 1,200 color images found on Internet. For the automatic evaluation, these images are gathered in 14 semantic categories like elephants, lions, cars, aerial images, *etc.*

Figure 2 presents the performances of our short-term relevance feedback strategy compared to a random search. The recall criterion is averaged for 50 queries of the "lion" category.

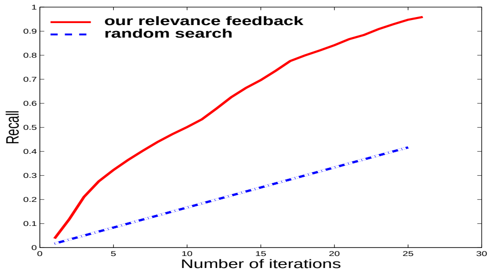


Fig. 2. Short-term relevance feedback performances.

These curves show that our relevance feedback is effective in order to improve the search performances of the short-term approach ([8] provides the comparison to reference methods).

Figure 3 draws the performances (averaged precision-recall curves without relevance feedback, for 100 different queries of the "lion" category) without long-term learning (only visual similarity), with long-term learning over 60 queries and with long-term learning over 120 queries (10 feedback steps for each search).

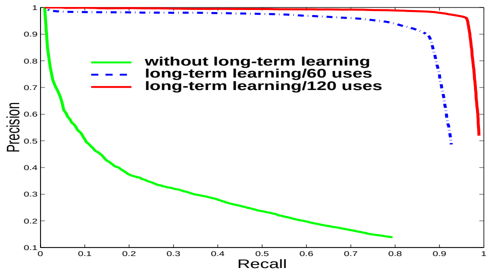
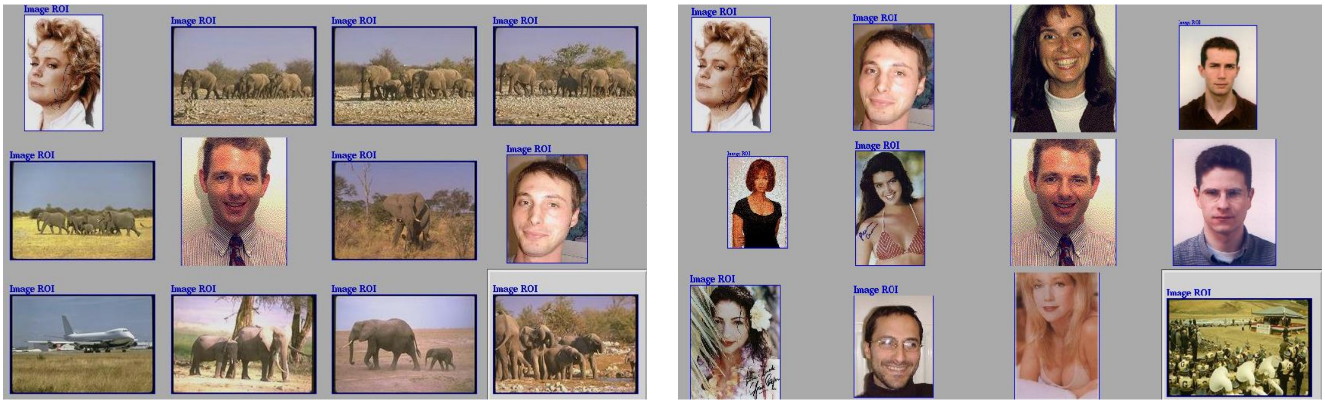


Fig. 3. Evaluation of the retrieval method with and without long-term learning.

The performances are significantly greater using the long-term similarity. Moreover, performances increase with the number of system uses. In this example, the lion category is perfectly learned after 120 uses. In spite of the simplicity of the proposed method, this test validates the contribution



(a) Without long-term similarity

(b) With long-term similarity (120 system uses)

Fig. 4. Portrait retrieval. The query is the top-left hand image.

of long-term learning. This new information source complements the visual similarity since it exhibits some kinds of semantic properties.

As a last illustration, figure 4 presents the result of a search in our general database of 1,200 images. The query image is a portrait (top left-hand image). These results show the lack of consistency between the content-based similarity (fig. 4 (a)) and the semantics of the relevant images. The long-term learning helps to reduce the semantic gap (fig. 4 (b)) and helps to retrieve a wide set of relevant images.

Even if our strategy implies a great number of uses in order to build a reliable information, the system works with coarse long-term similarity values. Moreover, refined versions of the algorithm could be employed in order to speed up the learning process.

5. CONCLUSION AND FUTURE WORK

We have presented a new retrieval strategy based on relevance feedback and long-term similarity learning. User interaction allows the refinement of the current search results. Moreover, the interaction information is used in order to build a “semantic” similarity between images. This measure is updated across sessions and combined with the content-based similarity. The experiments and the quality assessment bear out the efficacy of the proposed method.

We are currently working on extensions of the long-term learning approach as well as on the validation of the method for large databases in multi-user context.

6. REFERENCES

[1] I.J. Cox, M.L. Miller, T.P. Minka, T.V. Papathomas, and P.N. Yianilos, “The bayesian image retrieval system, PicHunter: Theory, implementation and psychophysical experiments,” *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 20–37, 2000.

[2] J. Peng, B. Bhanu, and S. Qing, “Probabilistic feature relevance learning for content-based image retrieval,” *Computer Vision and Image Understanding*, vol. 75, no. 1-2, pp. 150–164, July-August 1999.

[3] Y. Rui and T.S. Huang, “Optimizing learning in image retrieval,” in *Conf on Computer Vision and Pattern Recognition (CVPR)*, Hilton Head, SC, June 2000.

[4] J. Fournier, M. Cord, and S. Philipp-Foliguet, “Back-propagation algorithm for relevance feedback in image retrieval,” in *International Conference in Image Processing (ICIP’01)*, Thessaloniki, Greece, October 2001, vol. 1, pp. 686–689.

[5] T.P. Minka and R.W. Picard, “Interactive learning with a “society of models”,” *Pattern Recognition*, vol. 30, pp. 565–581, 1997.

[6] N. Vasconcelos, “Content-based image retreival from image databases: current solutions and future directions,” in *International Conference in Image Processing (ICIP’01)*, Thessaloniki, Greece, October 2001.

[7] J. Fournier, M. Cord, and S. Philipp-Foliguet, “Retin: A content-based image indexing and retrieval system,” *Pattern Analysis and Applications Journal, Special issue on image indexation*, vol. 4, no. 2/3, pp. 153–173, 2001.

[8] J. Fournier and M. Cord, “A flexible search-by-similarity algorithm for content-based image retrieval,” in *International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP’02)*, Duram, North Carolina, USA, March 2002.

[9] T.S. Huang and X.S. Zhou, “Image retrieval with relevance feedback: From heuristic weight adjustment to optimal learning methods,” in *International Conference in Image Processing (ICIP’01)*, Thessaloniki, Greece, October 2001.

[10] B. Widrow and M.E. Hoff, “Adaptive switching circuits,” in *IRE WESCON*, New-York, 1960, pp. 96–104.

[11] C. Meilhac and C. Nastar, “Relevance feedback and category search in image databases,” in *IEEE International Conference on Multimedia Computing and Systems (ICMCS’99)*, Florence, Italy, June 1999, pp. 512–517.

[12] Y. Ishikawa, R. Subramanya, and C. Faloutsos, “MindReader: Querying databases through multiple examples,” in *Proc. 24th Int. Conf. Very Large Data Bases, VLDB*, 1998, pp. 218–227.

[13] R. Sutton and A. Barto, “Reinforcement learning: An introduction,” MIT Press, Cambridge, MA., 1998.