Image retrieval : a first step for a human centered approach

Anne Guérin-Dugué, Stéphane Ayache, Catherine Berrut

Image retrieval : a first step for a human centered approach[1]

Anne Guérin-Dugué; Stéphane Ayache; Catherine Berrut

Laboratoire CLIPS-IMAG
BP-53, 38041 Grenoble cedex 9, France
{anne.guerin; stephane.ayache; catherine.berrut; }@imag.fr

Abstract

Image indexing using content analysis is known as a difficult task, involving the vision research domain. Using these tools in the context of a retrieval system is generally frustrating for users, due to a lack of interfaces development, and to the difficulty for users to understand the low-level features managed by the system. We propose in this paper a general point of view for introducing a link between such systems and potential users. This includes image features based on visual perception models, a relevance feedback model, and a graphical interface to express the information need through user-system interactions.

Keywords

Content analysis, Image retrieval, Perception, User design, Relevance feedback

1. Introduction

The goal of content-based image retrieval (CBIR) systems is to help users in finding images among wide image collections. Actually, such systems have a double goal: building a system able to help users, but also characterizing the content of images such that they could be retrieved by the system. This double goal is known to be antagonistic because on one hand, the system computes the similarities between images from low level features, and on the other hand, the user looks for images with his own interpretation and motivation, which are mostly and unconsciously semantically driven.

The domain of image characterization aims at automatically extracting features (color, texture, shape, appearance, etc.). This is a very challenging task in many Computer Science and Vision laboratories. During the past decade, a number of systems have been developed, showing an intensive research activity : [1,2] give firstly a good review of the domain and of existing systems, and secondly future trends.

However, the qualitative performance of systems is bounded by a semantic gap because these systems are mainly based on low-level features. To bridge this gap, some systems have been provided with "relevance feedback" [3, for example] derived from the information retrieval domain [4]. Upstream, there exist experimental studies on perceptual similarities, in order to capture the semantic meaning of the images, and to understand the main visual categories driving visual perception [5]. In fact, these studies imply perceptive metrics to be integrated in such systems [6, 7]. Statistical models are proposed to get benefit of low level features distribution in natural images, for visual categories [8]. Downstream, protocols already established in the Information Retrieval domain are used in experimental evaluations [9]. This work is developed in the framework of the SCOPIE project whose objectives are the design and conception of a user centered CBIR system, according to three axis, (i) image characterization through visual perception models, (ii) validation by psychological experimentations, and (iii) framework for user interactions. This article is focused on this last issue, developed into two points, firstly on the interface for interaction and secondly on a non parametric model of relevance.

2. Our analysis

As far as the system understands low-level features, and the user generally makes a correlation between relevant documents and a semantic need, the misunderstanding between the user and the system exists and generally emphasized with the usage of relevance feedback.

Our analysis of existing systems leads to several criticisms :

[1] This work is a part of the SCOPIE project founded by the federations ELESA and IMAG, and realized in collaboration with the LIS and the LPNC laboratory.

Indexing with features, initially developed for another context in computer vision, to a user front-end system, such as an image retrieval system ;

Querying : Most systems interact from a QBE (query by example) interface with a query-answer-query-answer-… dialog, with a ranked list of retrieved images. In such a context, the user has to query and query without any help in the global session he is having.

Relevance feedback : Generally systems introducing feedback use the idea issued from the information retrieval domain. The drawback of this calculus stands in its uniform view of the relevant and not-relevant documents, and does not take into account subsets which could emphasize individual needs. Even if there exist systems which provide feedback through each indexing modality, the individual way of choosing an image should be more taken into account.

3. Our point of view

Our global position related to the state of the art consists in emphasizing the human needs and expression possibilities. Such approaches are rare, and very limited in existing systems We underline that efforts should be made in order to make the user having priority in the development and analysis of such systems. This could be introduced at different levels:

Indexing : The use of features based on visual perception models in the indexing process. Managing the images with these algorithms provides at least a way of interpreting images such as a user would perceive. The challenge here consists in providing a more intuitive images ranking according to perceptual dimension giving a global semantic interpretation : for example, artificial/natural scenes, indoor/outdoor scenes, … In [10], the authors show how perceptual properties of natural images (naturaless, openness, roughness, ruggedness and expansion) are correlated with global and local spectral information in the scene. From a projection on these perceptual dimensions, the semantic of the scene can be inferred without segmentation and object identification.

Querying : The development of a user dedicated workspace: the gap system/user exists. Since this is our reality, we must provide to the user a double level of expression: querying a system, AND/OR a workspace for the management of his retrieval. In such a context, there is an evident need for the development of tools dedicating to the management of the user session. Very interesting approaches has been developed in [11] where the interactions to extract semantic are developed through a similarity based model. Here, we use the "ostensive model" developed in [12] for Information Retrieval and our aim is to generalize this model in the context of an CBIR system. In this model, the user doesn't explicitly formulate a query : he just expresses his information need through a relevance feedback. The progressive development of the information need is integrated inside a session.

Relevance feedback : Adaptation of the system to every user by creating retro-analysis of the user's interaction is needed in order to individually adapt the system. We first propose to study the feedback inside each indexing modality : this should emphasize each user points of interest within each modality. For example, let us suppose that a user is interested by a special landscape in autumn and winter. Within the chrominance modality, there will roughly exist two separate sets of relevant documents : the white ones, and the orange-brown ones. Classical feedback would minimize the weight of this modality due to the large within-distance between all the relevant documents. We propose to model the relevance and the non-relevance of the documents with a non parametric model inspired from [13] in which no assumption is considered about the distribution of the relevance and non relevance documents. We also propose to analyze feedback within the global session, in order to study it as the real expression of the user's need. Going on with evaluations by down (classical system evaluation like in information retrieval) and upstream (user data on perceptual similarities) experimentations, which guarantee a good system-user interaction. Going back to the ostensive model where the temporal dimension is introduced, we develop the relevance feedback in this model which is entirely based on the reformulation.

4. Low level perceptual features

This CBIR system uses classical low level features, color and texture, through a visual perception point of view.

4.1 Color representation

For image retrieval systems, color information is one of the most important features. We propose a perceptive color space based on a biological model of the retina inducing a kind of color-constancy behavior [14]. In the retina of man and primates, the three kinds of photoreceptors are known to present a non-linear and adaptive transduction function. If $C_i(x,y)$ is the image of the i-th colour component activating the photoreceptors of class i, the response of this class is a compressed image :

$$c_i(x,y) = \frac{C_i(x,y)}{C_i(x,y) + h(x,y) * C_i(x,y)}, \quad (1)$$

where the term $h(x,y) * C_i(x,y)$ is a kind of local mean of the input image $C_i(x,y)$, obtained by the convolution with a normalised low-pass kernel: $h(x,y)$. The result is that (i) each photoreceptor produces an output image which is more compressed in highlights, and (ii) the resulting colour of each

pixel is affected by the local surrounding colour of the scene, i.e. mainly the colour of illumination. Then if we name $R(x,y)$, $G(x,y)$ and $B(x,y)$ the "stimulation color space", we can name the compressed signals $r(x,y)$, $g(x,y)$ and $b(x,y)$ the "perceptive color space" from (1). Then we can derive three more orthogonal signals: as $l_c = [r(x,y) + g(x,y) + b(x,y]/3$, $a_c = r(x,y) - g(x,y)$, $b_c = b(x,y) - [r(x,y) + g(x,y)]/2$. These three signals define the new perceptive compressed color space with the corresponding luminance (l_c) and chrominance (a_c, b_c) components. The chrominance modes of the images are well enhanced allowing an efficient parsimonious statistical model by a 2D Gaussian mixture where the number of modes is automatically selected. Visual dissimilarities by comparing two chromatic distributions, are implemented through the Kullback-Leibler divergence. Spatial information is introduced after segmentation and then the spatio-chromatic distribution is derived. The spatio-chromatic dissimilarity is naturally deduced in the same way. [15] summarizes this approach and presents numerical experiments where this method using this new color space is advantageously compared with classical color spaces like Lab and CbCr.

4.2 Luminance information by "Independent Component Analysis"

Low level features are directly extracted from images with Independent Component Analysis (ICA) [16]. When ICA is applied to a set of natural images, it provides band-pass-oriented filters, similar to simple cells of the primary visual cortex [17]. These filters compose a new basis function set in which images are encoded by independent features. Moreover, the property of independence allows us to use a more adapted dissimilarity function than the classic norm based distance : it provides a justified way to estimate the Kullback-Leibler divergence between two images by the sum of the divergences between the marginal densities of both [18]. Figure 1 illustrates the process with one filter belonging to the filters set extracted by ICA.

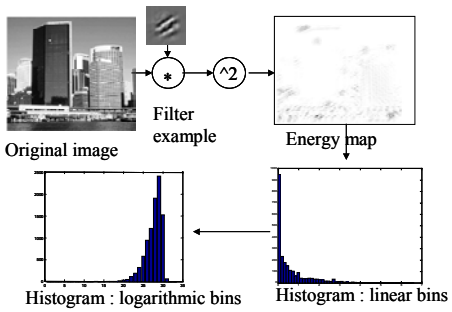

Figure 1 : Example of a distribution according to a filter extracted by ICA.

** means convolution.*

One hundred and twenty filters are extracted from 32×32 patches of a learning images database. For each image and each filter, the histogram of the energy values inside logarithmically distributed bins compose the signature (33 bins) from which the dissimilarity is processed in the same Kullback-Leibler divergence framework. See [18] for details.

4.3 Information for browsing

The choice here is to use the same coherent approach for all the features : model the distribution with a selected method according to the data, a parametric model with a mixture a gaussian functions for chrominance (global and local distributions), and a non-parametric method with log-histogram for the energy distributions. Due to the choice of the ICA filters and the assumption of independence between luminance and the two chromatic dimensions in the proposed color space, the dissimilarity between images can be naturally derived as a sum of simple Kullback-Leibler divergences, and then we justify the fusion process as a weighted sum of divergences :

$$Diss(F,G) = \sum_i \omega_i D_{KL}(f^i, g^i), \qquad (2)$$

with i=1,2,3 for color, spatial and energy distribution, respectively. For this first version of the system, all the ICA filters have the same weight. We use a symmetric version of the KL divergence, then this cumulative measure defines only a topological space. The weights will be adaptively tuned through the relevance feedback process. So for browsing, the system knows or can compute on line, for each couple of images, a triplet of dissimilarity values.

5. Starting with new browsing interfaces

Coming back to the idea of a user's workspace associated with the queries, we think that querying and managing a session should be integrated in a single environment, so that the user may find a flexible and intuitive process when using a image retrieval system. This can be done through a *browsing strategy* applied to organized image databases (images with neighborhood relationships). Because we must avoid the necessity for a user to find ONE query, the system should allow the user to build his sets of solutions without identifying it as the result of a single query. There exist interesting approaches allowing this way of interacting with a retrieval system, in the information retrieval domain [12], in the user interface domain [19], or in groupware technologies. Our first approach with our system is to re-use these ideas in the context of a CBIR system.

5.1 Page by page browsing

This interface implements a "page by page browsing" base strategy. This can be useful to select a "zero page" to initialize a browsing. Each page identifies a cluster of images

(automatically or manually extracted). The interface is based on the principle of a magnifying glass in order to view a great number of images. It also provides all tools allowing an easy browsing : zoom on interesting images, usage of a magnifier which size is also a parameter, …

5.2 Static browsing

Let us consider one unique matching function between documents, i.e. one unique dissimilarity function between images. By example, this function can be set with a choice by default for the weighting parameters in the cumulative dissimilarity. Figure 3 shows one of the interface we propose to query the system, inspired from the ostensive model [12]. The idea is again that the user works in the context of a session to find the set of interesting documents he needs. Instead of being a step by step interface, the system here shows its global interaction : from left to right the user queries using images and results. The results are displayed on a sphere, which allows a good visibility on the set of retrieved and selected documents. The global answer to the user is the set of images kept by the user, whatever the queries which were needed for that. We have also developed the necessary tools for the good usage of our interface : camera motion, zoom on the central documents, cut on not interesting documents, …

Figure 2 : Example of the "fish eye" view of a part of a session.

5.3 Dynamic browsing

This static model is interesting thanks to the global view of the session to express the information need. It can be enhanced by taking into account the sequence of the selected documents : this is the *dynamic model*.

If the sequence of the selected documents is used to adapt the matching function, that means the system answer depends not only on the selected documents but also on the own browsing sequence of each user. Browsing is a sequential process in time. In [12], the authors point out that the user knowledge and his information need vary in time, and the system must take into account this evolution. Then in the ostensive model, temporal profiles are implemented to modulate the system answer. Three profiles are discussed, (i) constant in time, (ii) decreasing profile (forgetting factor), and (iii) increasing (memory factor). The second one is proposed in [12] where the selected documents are more important in a near past than in a far past.

The integration of a temporal profile is one way to individualize the system answers between users, which may have different browsing strategies to reach the same target.

Even if this temporal aspect is important for a centered user approach, it remains limited due to the semantic gap, if it is not associated to a relevance feedback process. Now, we will explain how it is tightly coupled in this CBIR system.

6. Relevance feedback model

6.1 General principle

The relevance feedback model is one part of the interaction model we have defined, as it is illustrated at figure 3.

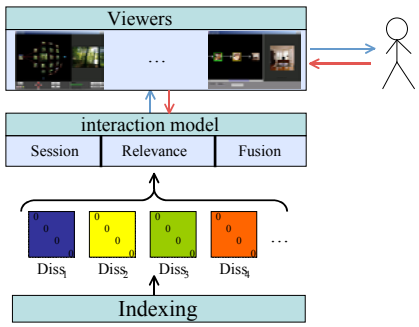

Figure 3 : Three components of an interaction model inside a CBIR system.

At the top, we have a front end interface (here the "fish eye" viewer, cf §5.2). At the middle, the interaction model is composed of (i) a session model (ostensive model, cf §5) with a given temporal profile, (ii) a relevance model and (iii) a fusion model to integrate to different image feature representation (cumulative fusion, cf 4.3). And at the bottom, we have the result of the indexing process : dissimilarity between images for each feature space.

6.2 Relevance model

One of the major difficulty is to model the distribution of the relevant and non relevant documents, where multi modal distributions must be considered. In our system, we have used a non parametric method by kernel gaussian functions (K^+, K^-). The user selects his relevant (R^+ set) and non relevant (R^- set) documents and this relevant information is diffused to the neighborhood in each feature space. For an image G, in a feature space i, we have this Rocchio like equation :

$$R^i(G)=\alpha \sum_{S\in R^+} K^+(Diss(g^i,s^i))-\beta \sum_{S\in R^-} K^-(Diss(g^i,s^i)) \cdot \qquad (3)$$

This relevance is integrated over time according to the given temporal profile. Then the dissimilarity proposed to the user

U when he selects the image F and asks for its neighborhood, is defined by :

$$Diss_U(F,G)=\sum_i \omega_i \Phi\left\{Diss(f^i,g^i);R^i(G)\right\}. \quad (4)$$

This equation extends the equation (2) where the relevance of the image G is integrated to find the neighborhood of F. With the function Φ, the dissimilarity is linearly modulated in order to bring together images if the relevance is high. This measure is no more symmetric : the user relevance is only integrated for the images G which will be retrieved by the system from the selected image F. The relevance of the image F had been integrated in a previous query. The browsing framework is seen here as a chain along which the user relevance is diffused. At last, the weights ω_i are estimated with simple heuristics which are often used in this framework, (i) inertia of the relevant documents, (ii) average rank and (iii) average relevance.

6.3 Experimentations

First experiments had been realized only on a limited database (321 scenes of indoors, landscapes (beach, mountain, forest, …), cityscapes). More extensive experiments on a larger database (several thousands) are works in progress. 16 subjects were invited to solve an information need in a given number of iterations (3, because the number of images these experiments is too small). The quality factor is estimated by the ratio of the number of relevant images at the last iteration versus all the relevant images in the whole session. Due to the limited size of the database, the temporal profile has not be integrated and has been held constant. In order to test one of the model dimensions, the other has been configured on the baseline. The first dimension concerns the fusion (the baseline is the inertia approach) and the second one, the relevance (the baseline is the model without the non relevance feedback) .

	Inertia,α=β	Inertia, α>β	Inertia, β=0
Relevance ratio	33,8%	35,4%	33,2%
	Inertia, β=0	Rank, β=0	Sum, β=0
Relevance ratio	32,8%	35,4%	41,6%

7. Conclusions, Perspectives

These first results confirm in this browsing framework, the role of the non-relevance judgment (α=2β) and shows an advantage for the fusion guided by the relevance ("sum" strategy) which is the more flexible approach, and is coherent with the choice of a non parametric method for relevance modeling. By this way, images are recoded in a new space which is a "relevance space" whose the dimension is the number of features. The underlying idea is to obtain a better clustering in agreement with the user relevance. This centered user approach is very promising, must be still developed to integrate more perceptual features and to improve this relevance model, to be validated by larger psychophysical experiments.

8. References

[1] Del Bimbo A., Visual Information Retrieval, Morgan Kaufmann Pub, 1999

[2] Smeulders A., Worring M., Santini S., Gupta A. and Jain R., Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions on PAMI, vol. 22, n° 12, pp. 1349-1380, 2000

[3] Chakrabarti K., Porkaew K., Mehrotra S., Efficient Query Refinement in Multimedia Databases, CVPR'98, Santa Barbara, CA, June 1998.

[4] Rocchio J.J, Relevance feedback information retrieval. In Gerard Salton, ed., The Smart retrieval system| experiments in automatic document processing, pp. 313-323. Prentice Hall, NJ, 1971

[5] Rogowitz B., Frese T., Smith J.R., Bouman C.A., Kalin E., Perceptual Image Similarity Experiments, Vision and Electronic Imaging III, SPIE, n°3299, San Jose, CA, January 26-29, 1998.

[6] Mojsilovic A., Rogowitz B., "Capturing image semantics with low-level descriptors", Proc of ICIP'01, vol 1, pp 18-21, Greece, 2001.

[7] Vailaya A., Jain A.K., Zhang H.J., On image classification : City images vs Landscapes, Pattern Recognition, vol. 31, n°12, pp. 1921-1936, 1998.

[8] Torralba A., Oliva A., Statistics of natural image categories, Network: Computation in neural systems, vol. 14, pp. 391-412, 2003

[9] Salton G., McGill MJ., Introduction to modern information retrieval, McGraw-Hill, New York, 1983

[10] Oliva A., Torralba A., Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope, IJCV, vol. 42, n°3, pp. 145-175, 2001

[11] Santini S., Jain R., Gupta A., A User Interface for Emergent Semantics in Image Databases. DS-8 1999: 123-143

[12] Campbell I, Interactive evaluation of the Ostensive Model, using a new test-collection of images with multiple relevance assessments, J. of Information Retrieval, vol II, issue 1, 2000

[13] Laaksonen J., Koskela M., Oja E., PicSOM : Self-Organizing Maps for Content-Based Image Retrieval, In Proceedings of IEEE IJCNN'99, Washington, DC. July 1999.

[14] Alleysson D., Hérault J., Variability in color discrimination data explained by a generic model of retina with non linear and adaptative processing. Int. Color Vision, Society Annual Meeting, Göttingen, August 1999.

[15] Guérin-Dugué A., Biernacki C., Hérault J., Statistical modelling of a perception based colour representation for image characterisation and retrieval, ICIP'01, Thessaloniki, Greece, October 2001.

[16] Hyvärinen A., Karhunen J., Oja E., Independent Component Analysis, John Willey & Sons, 2001.

[17] Bell A.J., Sejnowsky T.J., The Independent Components of Natural Scenes are Edge Filter, Vision Research, vol. 36, pp. 287-314, 1997.

[18] Le Borgne H., Guérin-Dugué A., Antoniadis A., Classification of images with independent features, Submitted to Pattern Recognition Letters, december 2002.

[19] Nigay L, Vernier F., A Framework for the Combination and Characterization of Output Modalities. DSV-IS'2000, 5-6 June 2000, Limerick, Ireland.