An Alternative Image Retrieval System Based on Visual and Thematic Corpus Organisation

Gérald Duffing

An Alternative Image Retrieval System Based on Visual and Thematic Corpus Organisation

Gérald Duffing
UMR Loria
BP 239 — F-54506 Vandoeuvre-les-Nancy cedex, France.
e-mail: duffing@loria.fr

Abstract

Retrieving images in large databases involves either thematic or visual querying capabilities. We think that both approaches can be efficiently combined to provide a thematically and visually relevant retrieval. This paper deals with bi-modal access in image retrieval systems. Our approach relies on a corpus organisation based on both visual and thematic clustering, allowing access to non-indexed images. An adapted retrieval process has been designed to reformulate queries, taking into account users' judgements that deal with thematic and visual aspects of their information needs.

1. Introduction

It becomes very easy to gather large amounts of images into a corpus. When dealing with thousands of images, it remains however difficult to examine each picture and to index it. We think that *partially indexed* corpora can be organised in a way that facilitates retrieval, provided the retrieval process is adapted. Current existing image retrieval systems [1] split into two main categories. Systems allowing only *keyword querying* consider the indexed part of the corpus, whereas *content-based retrieval* systems [20, 10, 18] allow retrieval by means of visual similarity comparison. Hybrid methods try to integrate both approaches [14, 13]. We believe that existing content-based systems allowing visual retrieval give little attention to the thematic part of the query: keywords are used for pre-filtering purposes, or as labels embedding static visual properties. To bridge the gap between text- and visual-based queries without loss of high semantic search capabilities, we propose a new retrieval strategy based on a prior corpus organisation that takes into account both thematic and visual aspects.

2. Corpus analysis and organisation

2.1. Classification basics

For corpus organisation purposes we focus on clustering techniques, which aim at grouping similar objects. An improvement of retrieval effectiveness may be expected, as stated in the clustering hypothesis [17]: *"closely associated documents tend to be relevant to the same requests."*. Retrieval efficiency is also expected, as query is only matched against cluster representatives [9]. A corpus can be clustered into a partition, or into a hierarchical structure [5]. In this paper, we intend to use it as a first organisation tool in order to build two different classifications: a thematic and a visual one. We think that links can be established between these two structures. We use an agglomerative hierarchical clustering process to compute a hierarchical structure called "dendrogram".

2.2. Describing images

We use a partially-indexed corpus. A simple thematic similarity measure is defined based on WordNet [6], which is a lexical reference system where words are organised into synonym sets called "synsets", being connected to each other by means of semantic relationships. We will here focus on nouns and make use of *specific, generic, meronym, and holonym* relationships.

Each image can be characterised by a set of visual features computed by image analysis techniques, describing colour, texture and shape [16, 8, 19, 3, 11], even if these feature are not always robust [12]. When dealing with heterogeneous corpus, as we cannot assume domain-dependent knowledge that could help selecting relevant features, we use "generic" ones. **Texture** is often computed on gray-scale images. Let's cite stochastic attributes; co-occurrence matrices; multi-scale analysis, etc. **Colour** is represented in a given colour space. Colour histograms have been extensively used to represent the colour distribution of images. **Shape** can be described by many different features. The main problem is to extract the relevant shape from the image, with no additional noise.

2.3. Text and visual integration for image description

Our goal is to determine how some concept is represented in the corpus, in terms of visual features, in order to allow both thematic and visual retrieval. To achieve this, we propose the concept of **realisation**, derived from the *schema* [4] and defined here as *a set of visual features appearing simultaneously in an image or part of an image, possibly associated with a keyword.* A realisation stores different visual features computed from an image clip. The set of all the realisations associated to a given concept (keyword) provides the system with knowledge about the visual properties of this concept in the corpus. Associating a keyword with its possible realisations is an indexing task manually carried out by the corpus administrator. The realisation is used for **image comparison**, based on visual features matching. The task here is to determine in what extent visual features stored in the realisation are present in both images. Each realisation can consist in more than one feature, and each stored feature has its own way of computing similarity. Thus it is necessary to elaborate a global similarity measure, that can be a weighted sum of involved features. This scheme allows to give more importance to "perceptually relevant" features — such as colour — against others.

2.4. Initial corpus organisation

Given a set of image properties — thematic and visual — and associated similarity measures, image clusters can be constructed. As visually similar images are clustered, and provided one image in the cluster is also thematically indexed, then every image in the same cluster gets a chance to be retrieved and displayed to the user, even if the query contains no visual constraint. The corpus organisation runs in two steps: in **thematic clustering**, only indexed images are affected. This yields a first document hierarchy. In **visual clustering**, images are classified based on visual similarity. Thematically indexed images appear in both dendrograms which are then connected via these images. We call this connection a "tunnel".

3. An adapted retrieval process

All the visual features we have computed remain uncertain. To alleviate this lack of accuracy, an adapted retrieval process is defined that will take advantage of all information given by users. The vocabulary that can be used as query items is limited to nouns from WordNet. Realisations may have been associated to keywords, allowing users to browse through visual examples. If he selects realisations, visual conditions are added to the query. Otherwise, this thematic query is submitted, and will however be automatically enriched with visual conditions, due to relevance feedback.

3.1. Image database search

The retrieval process outlined below consists in distinct phases: thematic extraction and ranking, and visual retrieval and ranking.

Thematic retrieval. — The thematic hierarchy is searched using the thematic part of the query (1). This yields cluster $C1$, containing image A among others. Depending on cluster size, though, not all images may be equally relevant to the query: this motivates an additional thematic ranking process (2) that will improve precision.

Visual retrieval using tunnels. — The visual hierarchy is searched using images retrieved in the previous step that lead, by means of tunnels, to clusters in the visual hierarchy: in our example, A also belongs to cluster $C2$ (3). All images in $C2$ are visually similar, but not necessarily thematically similar. Some of them are indexed (and can be ranked according to the thematic part of the query (4)), whereas others are not.

Visual retrieval using realisations. — The previous phase retrieved images that may not correspond to the visual part of the query. Here we are searching for images visually similar to the query only. This process (7) yields cluster $C3$. Note that cluster $C3$ and $C2$ are *distinct*, although only one cluster has been drawn on figure 1, for sake of simplicity.

All retrieved images from (2), (4) and (5) are ranked visually (6), according to the visual part of the query, constituting an ordered set of proposed images. The following formula is used to rank images: $S = w_t S_t + w_v S_v$, with $w_t + w_v = 1$. Note that visual score (S_v) and thematic score (S_t) may be weighted by users so that it is possible to give more or less importance to visual search against thematic search.

The overall advantage of this method is that non-indexed images have a chance to be retrieved. As a consequence, images retrieved in that way may be thematically irrelevant. However, when a non-indexed image has been retrieved and *judged by user*, then in some cases we can make assumptions about its theme, if the image has been judged as "thematically relevant". Thus, the system incrementally improves its knowledge about the corpus and noise should be reduced in future retrieval sessions.

3.2. User feedback and query reformulation

Content-based image retrieval systems barely address relevance feedback issues based on visual similarity. Often, this is achieved by allowing users to select a relevant image among those proposed by the system, and to use this image for subsequent retrieval. More realistic approaches have been proposed, that can combine several image judgements on images to build a new query [2]. Here, we want to get more precise information from users. Actually, we may have retrieved images for which one point of view only (theme or visual aspect) is relevant: the system needs to know in what extent a retrieved image is relevant or not to the user. To achieve this, we propose a threefold judgement on each image: **theme**, **colour**, and **layout**. In this work, "layout" refers to spatial localisation of visually homogeneous zones. User can "accept", "reject" or have "no idea"; he may also weight his decision with a coefficient ranging from 1 to 10.

From this judgement data, the system derives a representation of its understanding of user's needs. We are looking for discriminating information between accepted and rejected images, in terms of colours, and layout. To model this information, we introduce the **virtual image** concept as a collection of weighted realisations. From the user feedback and from the features associated with displayed images, we build two sets of realisations (wanted and unwanted image features), in order to reformulate the query.

For *thematic reformulation* purposes, the keywords' weights are smoothly updated according to user feedback: weights are either increased or decreased if user keeps on wanting or refusing associated themes, or they are initialised to a default value if he changes his mind.

Visual reformulation is more complex, as each type of feature has its own strategy to alter its component depending on user feedback. For colour feature, a global colour histogram has been computed over the whole corpus, and 128 representative colours have been extracted by a quantization process. Thus wanted and unwanted colours can be modelled as a vector admitting a positive or negative normalised value. This vector is updated according to user feedback.

4. Experimentations

We give here some details on our current experimentations dealing with visual similarity evaluation. We consider two retrieval sessions $s1$ and $s2$: images shown on table 1 have been retrieved using single-word queries and we try to find visually similar images.

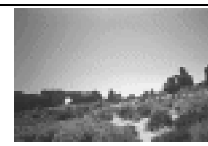
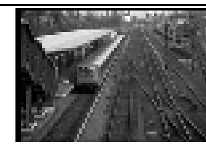
Session	$s1$	$s2$
Query	"rock"	"train"
Image		

Table 1: Test images for feature-based retrieval tests.

Useful visual features are close to human perception, in order to take easily into account user judgements [15]. We chose basic colour, texture and shape features. To improve localisation and to allow "layout" comparison, features are computed over the entire image, *and* on 32×32 image tiles.

4.1. Texture features

We compute Haralick's features [7] as texture indices. Fourteen values are computed at four different orientations in every image tile. Only local characterisation is used with Haralick feature. Our experiments show that those indices are powerful enough to provide a rough characterisation when images are tiled into 32×32 squares, and when image layout is quite identical: in table 2, it is clear that the large homogeneous sky permitted good retrieval results. For very complex scenes, however, results are less satisfying (table 3), suggesting that our image division strategy is too rough.

4.2. Colour features

For colour features computations, 128 representative colours are extracted from the entire corpus. Each image is then quantised according to this reduced colour map. A global frequency histogram is computed, as well as local histograms on image tiles. To evaluate homogeneous region detection in images, we have considered three different colour spaces: RGB, L*u*v*, HVC. Not surprisingly, L*u*v* achieved the best results, as it has been designed for colour comparison.

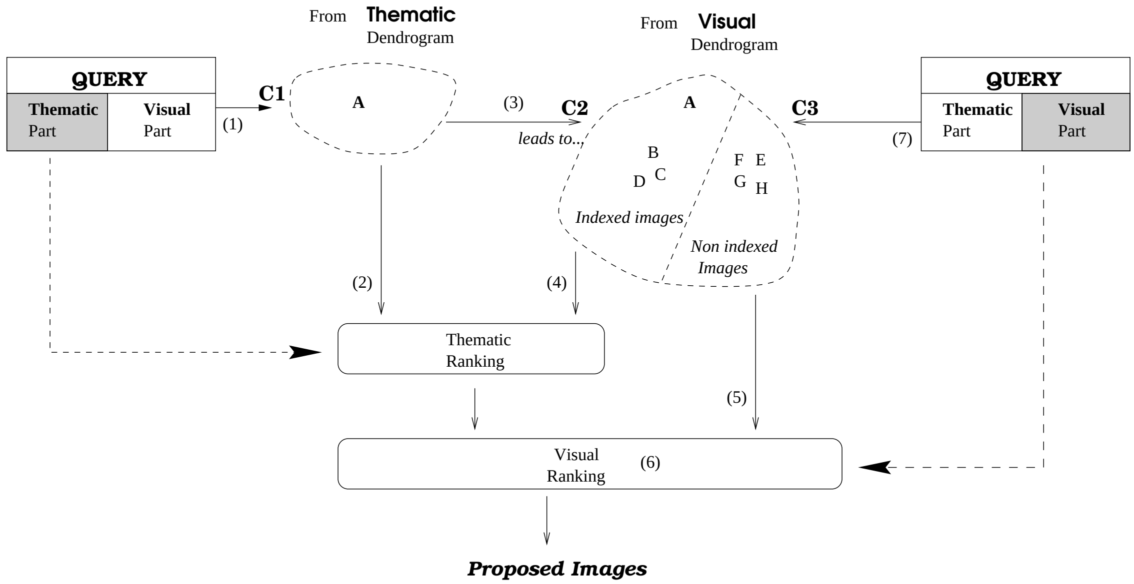
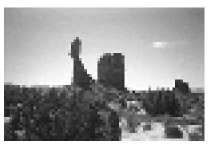
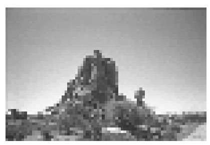
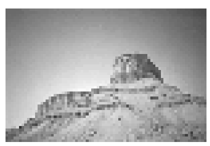


Figure 1: The Retrieval Process

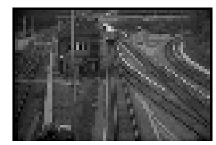
 

0.137 0.148

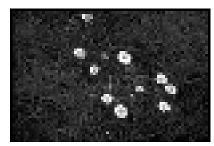
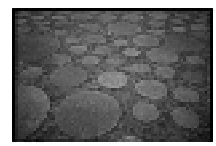
 

0.153 0.161

Table 2: Texture-based retrieval test 1.

0.131 0.136

0.139 0.140

Table 3: Texture-based retrieval test 2.

Table 4 illustrates a test involving global colour vectors, whereas table 5 displays results for local colour comparisons. Our experiments show that global colour vectors achieve best results: computing image layout based on local colour features has to be further refined.

We observe that colour features are well suited for global similarity computation, whereas texture features work well locally. This suggests a combination of these two features.

5. Conclusion

We have presented an integrated method that allows thematic and visual image retrieval in a partially indexed corpus, relying on a double – visual and thematic – image characterisation. The corpus is then organised into two hierarchical structures by a clustering process. This integration has led to the adaptation of the retrieval process. To give particular attention to user feedback, we introduced a specific image judgement scheme, that will facilitate query reformulation. Our approach aims at allowing retrieval of non-indexed images, and to produce some interesting indexing hypothesis.

References

[1] P. Aigrain, H. Zhang, and D. Petkovic. Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review. *Multimedia Tools*

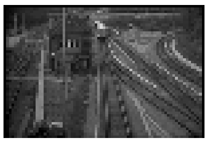
0.122

0.126

0.157

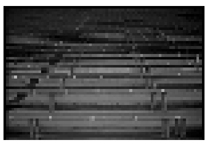
0.163

Table 4: Global colour-based retrieval test.

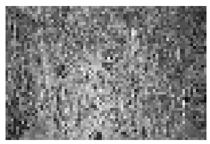


0.404

0.416



0.426



0.427

Table 5: Local colour-based retrieval test.

and Applications, 3:179–202, 1996.

[2] A. Celentano and E. Di Sciascio. Feature integration and relevance feedback analysis in image similarity evaluation. *Journal of Electronic Imaging*, 7(2):308–317, 1998.

[3] A. Del Bimbo and P. Pala. Image indexing using shape based visual features. In *ICPR'96 13th int. IAPR conf. on Pattern Recognition*, volume 3, Vienna, Austria, August 1996. IEEE.

[4] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman. The Schema System. *International Journal of Computer Vision*, 2:209–250, 1989.

[5] A. El-Hamdouchi and P. Willett. Comparison of hierarchic agglomerative clustering methods for document retrieval. *The Computer Journal*, 32(3):220–227, 1989.

[6] C. Fellbaum, editor. *WORDNET: An Electronic Lexical Database.* MIT Press, 1998.

[7] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC–3(6):610–621, November 1973.

[8] R.M. Haralick and L.G. Shapiro. *Computer and robot vision.* Addison-Wesley, 1992.

[9] N. Jardine and C.J. van Rijsbergen. The use of hierarchical clustering in information retrieval. *Information Storage and Retrieval*, 7:217–240, 1971.

[10] T. Kato. Database architecture for content-based image retrieval. In *Storage and Retrieval for Image and Video Databases*, pages 112–123, 1992.

[11] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1998.

[12] M. De Marsicoi, L. Cinque, and S. Levialdi. Indexing pictorial documents by their content: a survey of current techniques. *Image and vision computing*, 15(2):119–141, february 1997.

[13] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project: querying images by content using color, texture and shape. In Wayne Niblack, editor, *Storage and Retrieval for Image and Video Databases*, pages 173–181, San Jose, CA, 1993. SPIE.

[14] V.E. Ogle and M. Stonebraker. CHABOT: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, September 1995.

[15] R.W. Picard and T.P. Minka. Vision Texture for Annotation. *Multimedia Systems*, 3:3–14, 1995.

[16] W.K. Pratt. *Digital Image Processing.* John Wiley & Sons, New York, second edition, 1991.

[17] C.J. van Rijsbergen. *Information Retrieval.* Butterworths, second edition edition, 1979.

[18] J. R. Smith and S.-F. Chang. Querying by color regions using the VisualSEEk content-based visual query system. In Mark T. Maybury, editor, *Intelligent Multimedia Information Retrieval*, pages 23–41. AAAI Press, Menlo Park, 1997.

[19] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[20] T. Whalen, E.S. Lee, and F. Safayeni. The Retrieval of Images from Image Databases. *Behaviour & Information Technology*, 14(1):3–13, 1995.