

IKONA: INTERACTIVE SPECIFIC AND GENERIC IMAGE RETRIEVAL

N. Boujemaa, J.Fauqueur, M.Ferecatu, F.Fleuret, V.Gouet, B. LeSaux, H.Sahbi

IMEDIA – INRIA Rocquencourt, Domaine de Voluceau, BP. 105, Le Chesnay Cedex 78153 France
<http://www-rocq.inria.fr/imedia/>

ABSTRACT

This paper presents an overview of research activities at the IMEDIA group at INRIA-Rocquencourt research unit. IKONA search engine represents the prototype software for integration and validation of these activities.

1. INTRODUCTION

The main goal of IMEDIA research group is to develop content-based image indexing techniques and interactive search and retrieval methods for browsing large multimedia databases by content. We will describe in the following sections our main research topics:

- Image signatures with fine visual appearance modeling for generic and specific image databases including face detection and recognition,
- Flexible sub-image querying with different image complexity context,
- Smart browsing with database categorization,
- Cross-media indexing and retrieval.

More generally, the IMEDIA team achieves research, collaborations, and technology transfer on the complex issue of intelligent access to multimedia data streams. A prototype software IKONA illustrates the research that is lead at IMEDIA.

2. SPECIFIC AND GENERIC IMAGE SIGNATURES

Visual appearance is automatically measured by numerical signature of image features such as color, texture, shape, or most often a combination of them. More specific image signatures have to be developed for special content and situations. For designing an effective image retrieval system, we find it convenient to divide image databases in two categories.

- The first category concerns specific image databases for which a groundtruth is available. When indexing the database, the designer will consider these groundtruths and tune the models or range of parameters accordingly, maximizing the system efficiency. We have developed specific signature for face recognition and detection, fingerprint identification [7].
- The second category includes databases with heterogeneous images where no groundtruth is available or obvious. Examples include stock photography and the World Wide Web. The user should be assumed to be an average user (not an expert). In this context, generic image signatures are

computed in order to describe general visual appearance such as color and texture.

2.1 Specific databases: face signatures

In this section, we will focus on specific image databases that concern human faces which contain highly semantic information. IKONA handles databases containing faces in complex backgrounds. A first step of face extraction is performed using fast and efficient face detectors [1].

Face localization

Face detection is performed using a hierarchical algorithm based on coarse-to-fine support vector classifiers. We model the face appearance hierarchically by a recursive split of the pose in terms of position, rotation and scale. For all cells with their associated poses in the hierarchy, we build face-background classifiers with an increasing complexity (respectively decreasing invariance and false alarm rates) top-down in the hierarchy. The complexity of each SVM cell detector, in terms of the number of support vectors, is reduced by clustering. We introduce the bias variation technique which allows each simplified SVM classifier to satisfy the face conservation hypothesis as a criterion to get a consistent classifier in terms of detection rate, false alarms and background rejection efficiency. Face detection is performed using a depth-first search and cancel strategy which finds for a given “face pattern”, a root-leaf path with a sequence of positive answers. This “coarse-to-fine” approach allows efficient face-localization and background rejection (cf. Fig.1).

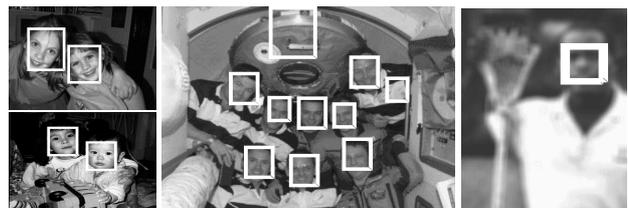


Figure 1. Face localization: some results

Face representation and indexing

Our system assigns for each extracted face a set of dedicated indexes: **DSW** (Dynamic Space Warping). It is a specific face signature [2] that ensures the matching of faces even with partial occlusions and performs registration to compensate for small 2D, 3D pose variations. Faces are pre-processed using the **entropy map** that assigns for each pixel in the face its saliency. This saliency is expressed as the entropy of a local grey level distribution on a region R_x around each pixel x . Given

c_1, \dots, c_r as the gray level quantization, the entropy map is estimated for each pixel x as:

$$H(x) = -\sum_1^r P(c_i) \log(P(c_i)) \quad (1)$$

Using the entropy map for each face k , we construct an index I_k given by:

$$I_k[l] = d_{warp}(k, l) \quad l = 1, n \quad (2)$$

Here n is the number of faces in the database, and d_{warp} is the warping distance between two faces. This warping function allows to transform an entropy map k into l using three kinds of basic operations (substitution, dilation and suppression) and the dynamic programming principle. This mapping finds for each feature in the face k its corresponding feature in l . This signature has proven to be as efficient as existing face signatures for easy databases and outperforms many existing face descriptions for challenging databases. It handles variability compensation and partial occlusions (presence of glasses, scarves, weak face extraction, etc.) successfully.



Figure 2. Some face queries on the ARF database illustration invariance to lighting, partial occlusion and facial expression

2.2 Generic signatures

Global visual appearance is described with a combination of color, shape and texture features. This leads to a high dimensional feature space and time consuming feature matching procedure. We suggest an integrated color-texture signature for more efficient visual appearance description.

Thus, our goal is to revisit the use of color histograms from the perspective of embedding some local information about the statistical and visual relevance or importance of each pixel. We introduce a modified color histogram—the weighted color histogram [4]—and the various measures that describe the local behavior of the colors:

$$\tilde{h}(c) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} w(i, j) \delta(f(i, j) - c), \forall c \in U$$

The weighting $w(i, j)$ is related to a local measure of color non-uniformity (or color activity), computed within a neighborhood of the pixel (i, j) with the color c . The proposed non-uniformity measures are based on the evaluation of perceptual cues (corners and isolated colors, by the use of the Laplacian), statistical color area distribution (by the use of local probability of occurrence

and informational entropy), local color relevance (by a fuzzy typicality and fuzzy entropy) and outlier test based measures (derived from color image filtering). The magnitude of all these measures increases with the local color variability, being minimal for uniform regions.



Figure 3. Retrieval with basic color histogram.



Figure 4. Retrieval with weighted color histogram.

The new, extended histograms have the same size as the usual, first-order color distribution and can be compared by the same metrics as any distribution. Thus, their computational complexity is not excessive. We claim that the weighted histograms can be indeed a valuable upgrade to the traditionally color distribution based image retrieval, since they embed local, textural information into a color only description.

3. PARTIAL QUERIES

Region based queries are being integrated into IKONA. In this mode, the user can select a part of an image and the system will search images (or parts of images) that are visually similar to the selected part. This interaction allows to the user to specify to the system what part or particular object is interesting in the image. In this case, since the query is focused, the system response is enhanced with regards to the user target since the background image signature is not considered. We have developed segmentation based methods as well as point of interest methods to perform partial queries.

3.1 Region-based queries

Region-based queries allow to select and retrieve similar regions of interest allowing more specific search in images than global queries. This approach raises two major problems: how to detect automatically in thousands of images regions of interest for the user, and how to provide a visual description specific to each region?

To address the first problem existing methods propose manual region outline, systematic image subdivision or image segmentation. In Ikona, we have developed a new image segmentation technique [5] to perform coarse region detection. It is based on the classification of local distributions of quantized colors. Colors and color distributions are classified with the Competitive Agglomeration (CA) algorithm which has the advantage to automatically determine the number of classes.

Then a Region Adjacency Graph (RAG) provides information to merge small adjacent regions. Detected regions are coarse and coherent. They encompass a characteristic color variability which makes them visually specific from one another.

The second problem is the description of regions. For images as well as for regions, existing descriptors represent colors from a color codebook predefined for a whole database which contains generally around 200 colors. This coarse representation results in a low color resolution which is sufficient for images but not for regions. Regions are by definition more homogeneous than images and require a finer description to be compared. We propose a region descriptor of fine color variability: the Adaptive Distribution of Color Shades (ADCS) [5]. For each region, its color shades are determined from a color classification of its original pixels at a high classification granularity. Compact and adaptive, the fine descriptor consists of the list of these shades and their population. Color shades are determined for each region among the million of colors of the full color space instead of the 200 colors in existing color descriptions. Region similarity is measured accurately using the color quadratic distance between ADCS descriptors.



Figure 5. Retrieval from top-left lavender region

A retrieval screenshot is shown in Figure 5. The key idea is to detect coarse and visually specific regions of interest and match them with the fine descriptor to improve retrieval results. Further work will be dedicated to speed up the region query process.

3.2 Sub-image retrieval using points of interest

Approaches involving points of interest have been developed for image matching and come to light with object/sub-image retrieval tasks which require more local descriptors. However the solutions met are not optimal, since they are limited to gray value images. We have proposed in [3] a local image descriptor based on color points of interest, i.e. points extracted and characterized from color signal at once. The points are extracted using the Harris color extractor which got the best repeatability compared to the classical gray value operators. Color information is exploited to characterize them too, by generalizing the differential invariants of Hilbert to color images. The points are described in an 8-dimensional space involving only first order derivatives:

$$\vec{v}_{col} = \begin{pmatrix} R \\ \|\nabla R\|^2 \\ G \\ \|\nabla G\|^2 \\ B \\ \|\nabla B\|^2 \\ \nabla R \cdot \nabla G \\ \nabla R \cdot \nabla B \end{pmatrix}$$

This point characterization is invariant to Euclidean transformations and can be made robust to scale, viewpoint and illumination changes. Point comparison is made by using the Euclidean distance on normalized points. The resulting image characterization is more compact than the gray value ones, insofar as it contains a richer photometric information - the color one, while having comparable storage cost. Semi-local constraints of neighborhood are added to the description in order to expand the image characterization with geometrical information. For example, it is possible to consider the spatial distribution of the points.

The indexing step consists in describing each image of the database by a set of color points. The retrieval one consists then first in selecting a part or an object in the image and second in comparing the color points of the selected area to the ones of the indexed image database. The query result is retrieved by using a voting algorithm. The returned image is the one which contains the bigger set of most similar points to the query points.

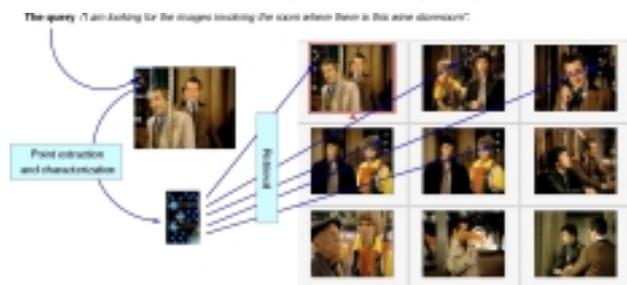


Figure 6. Retrieval example presents a partial query on an image database belonging to a television series

In this example, one wants to retrieve the images which contain a particular background. Indeed, we focused on the upper left part of the query image, which shows partially something like a wine storeroom. The retrieval was attempted on this particular region, described by about thirty points of interest. The best query results obtained under the Ikona platform are presented on the figure 6. The query area has been retrieved in five images, concerning the same room but with different characters. The resulting images differ from the query image in global shape and color, in viewpoint and present some occultations. Global indexing approaches naturally would not have given interesting results for this class of query. Approaches based on region segmentation would not allow the user to make the query on this part, since it represents small regions not easily detectable.

This approach presented is used by the French Judiciary Police as an investigation aids with image similarity retrieval.

4. IMAGE COLLECTION OVERVIEW BY DATABASE CATEGORISATION

The purpose of browsing is to help the user finding a picture by presenting him an overview of the database. We provide a summary of the database, given by a categorization of the database. Images are represented by various signatures: color (weighted histogram), texture (Fourier spectrum), shape and

structure (edge orientation histogram). The categorization in the signature space is made difficult because data are high-dimensional, natural categories are overlapping and have various shapes and the number of categories is unknown.

Since natural categories have various densities and compactness the function does not allow to retrieve low density categories. We have developed an adaptive competition process: the second term in (3) tends to reduce the number of clusters. It is weighted by a factor α_k depending on densities of each cluster [6].

$$J = \sum_{k=1}^c \sum_{i=1}^n u_{ik}^2 d^2(x_i, \beta_k) - \alpha_k \sum_{k=1}^c \left[\sum_{i=1}^n u_{ik} \right]^2 \quad (3)$$

We introduce a noise cluster to collect ambiguous points and outliers, defined as equidistant from all the points. Since natural categories have various shapes Mahalanobis distance is used to distinguish clusters.

The performance of the algorithm is tested on the Columbia Object Image Library. It contains 1440 gray scale images representing 20 objects. For each cluster, the average value of each feature is computed over images, then the average of all images defines a virtual prototype. The real prototype is the nearest image to the virtual one. A discussion with performance comparison is provided in [6].



Figure 7. Overview of the Columbia database with the most representative cluster images.

5. IKONA SYSTEM ARCHITECTURE

Our CBIR software, IKONA, is based on a client-server architecture and aims to be flexible, easily extensible, easy to use, intuitive, and does not enforce special knowledge or training.

The server needs to be fast and is written in C++. It includes image feature extraction algorithms (signatures computation), user interaction modes (retrieval by visual similarity, relevance feedback, region-based query, points-of-interest-based queries, etc...) and a network module to communicate with the clients.

The client needs to be portable and is written in Java; it normally should run on every computer architecture that supports Java Runtime Environment (JRE). It presents the user with an easy to use Graphical User Interface (GUI), sets the query mode for the server and displays the search results. Ikona has these particular properties:

- Native image/text joint description, Multi-users queries support,

- Images (and thumbnails) location handled as URLs: only signatures are stored on the server, images could be distributed,
- Single and Multi-processor machines support.

By default, IKONA performs “retrieval by visual similarity” in response to a query, which means that it searches all images in all databases and returns a list of the most visually similar images to the query image.

6. SUMMARY AND CURRENT WORK

While text indexing is ubiquitous, it is often limited, tedious and subjective for describing image content. Visual content image signatures are objective but have no semantic range. Combining both text and image features for indexing and retrieval is a very promising area of interest of IMEDIA team. We first work on a way to do keyword propagation based on visual similarity. For example, if an image database has been partially annotated with keywords, IKONA can use these keywords for very fast retrieval. Based on the indexed visual features and the keywords index, IKONA can suggest a number of keywords for a non annotated image and their weights. Further research on keyword propagation, semantic concept search and hybrid text-image retrieval mode with feedback mechanism are being carried on.

7. REFERENCES

- [1] Sahbi H. and Boujemaa N. *From Coarse To Fine Skin and Face Detection*, The 8th ACM International Multimedia Conference 2000.
- [2] Sahbi H. and Boujemaa N., *Robust Matching By Dynamic Space Warping For Accurate Face Recognition*, IEEE International Conference On Image Processing, ICIP 2001.
- [3] Gouet V. & Boujemaa N., *Object-based queries using color points of interest*, accepted IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL/CVPR 2001). Hawaii - december 2001, USA.
- [4] C. Vertan and N. Boujemaa, *Upgrading Color Distributions for Image Retrieval: can we do better?*, International Conference on Visual Information Systems Visual2000 Lyon, 2-4 Nov. 2000.
- [5] J. Fauqueur and N. Boujemaa, *Image Retrieval by Regions: Coarse Segmentation and Fine Color Description*, accepted International Conference on Visual Information System (VISUAL'2002), Hsin-Chu, Taiwan, March 2002.
- [6] B. Le Saux and N. Boujemaa, *Unsupervised Categorization for Image Database Overview*, accepted International Conference on Visual Information System (VISUAL'2002), Hsin-Chu, Taiwan, March 2002.
- [7] S. Bernard, N. Boujemaa, D. Vitale ITALE and C. Bricot, *Fingerprint Classification using Kohonen Topologic Map*, IEEE International Conference On Image Processing, ICIP 2001.