Visual browsing in image collections using wavelets

Jérôme Landré (a), Frédéric Truchetet (b) and Olivier Laligant (b)

(a) Université de Reims-Champagne-Ardenne, I.U.T., CReSTIC,
9 rue de Québec, 10000 Troyes, France,
(b) Université de Bourgogne, I.U.T., Le2i,
12 rue de la Fonderie, 71200 Le Creusot, France

ABSTRACT

Visual browsing is an important way of searching for images in large databases. In image retrieval, a lot of problems have to be solved to get a good system: dimensionality curse, users' search context, size of the database, visual features. In this article, a method trying to attenuate these problems is proposed. Each features vector is organized into four signature vectors used in the classification process while building a fuzzy search tree that is proposed to users for visual browsing. Our system gives good results in terms of speed and accuracy by solving several problems of classical image retrieval methods.

Keywords: Image retrieval, clustering, visual browsing

1. INTRODUCTION

Because of the rapid development of imaging technologies, multimedia retrieval is an important challenge for vision researchers. Content-based image retrieval[1] by image request is the classical approach for searching large image databases. Every image is associated to a feature vector whose coefficients are extracted from the image. Distances are computed between image feature vectors and request feature vector. However, this approach suffers limitations when the database size increases. To avoid this kind of problems a new approach was proposed: visual browsing.[2,3]

In visual browsing methods, image collections are organized during an offline stage into families proposed to users for online browsing. In this paper, a visual browsing method based on hierarchical organization of signature vectors into families of visual similar images is proposed.

Our paper is composed of different sections. Section 2 gives a list of related publications. Section 3 recalls several visual psychology results applied to our method. Section 4 shows the principle of our system. Implementation is detailed in section 5. Section 6 describes experimental tests and results. Section 7 is a conclusion of our work and future directions are proposed in section 8.

2. RELATED WORK

Chen et al.[2] gave a method called similarity pyramid in which a quadtree of image families is built offline and proposed to online users. Clustering of images into similar families is performed using K-means algorithm. This method allows both classical image retrieval by example matching and visual browsing. It gives good results in terms of accuracy and speed.

There are several possible improvements for this method:

- Donoho[4] pointed out the problem of dimensionality curse which is the phenomenon observed when working with high-dimension spaces. In these spaces, computing a distance between two points is not significant because of space emptiness. As a conclusion, to be valid, a distance must be computed in low-dimension spaces ($n < 16$).
- Users search context is important. In classical methods, feature vector x_i contains normalized features extracted from shape, texture and color of image i. Even normalized, these features are still very difficult to compare. So feature vectors have to be composed of only one type of features. For instance if a user looks for red cars in a database, features to use are color and shape into two different vectors, texture is not important in this case.

- Multiresolution analysis,[5] proposed by Mallat, can be used to add progressive levels of detail while building the tree and while browsing it.
- In order to decrease error rate, images can belong to more than one family leading to a fuzzy tree.
- The quadtree is restrictive, a tree whose arity is variable can be used depending on the database.

Taking in consideration the abovementionned points, a hierarchical multiresolution image browsing system was developed respecting several psycho-visual rules. This work is the result of my Ph.D. thesis and follows several previous publications.[6,7]

3. PSYCHO-VISUAL CONSIDERATIONS

As our research system is based on visual browsing, several psycho-visual principles[8] must be taken in consideration. Our visual system works by grouping visual atoms called gestalts[9] leading to a coarse-to-fine approach of vision. This coarse-to-fine approach is simulated by multiresolution analysis of images.

Our short time visual memory is limited to 7 ± 2 objects. It means our brain is not able to work with more than 7 ± 2 notions at the same time. This rule leads to the limitation of the number of images proposed to users during browsing (from 2 to 7).

In order to represent a family of images, a model image has to be designed. In our system, a fuzzy image built from the three closest images to the center of the family is our solution because users need to find their desired image by refinements.

4. DESCRIPTION OF OUR SYSTEM

Our system is presented in figure 1. Each image is decomposed using integer to integer lifting scheme transform from Calderbank et al.[10] at three levels of resolution. Several features (color, texture and shape) are extracted from each level, these features are grouped into a feature vector. Several features are extracted from this feature vector and organized into a hierarchy of signature vectors with the help of an expert of the database domain. These signatures are classified automatically into K families by the fuzzy K-means algorithm. At the end of the classification process, a test is performed to decide whether the current signature vector has to be modified or not. A visual research tree is built offline to be browsed online by users.

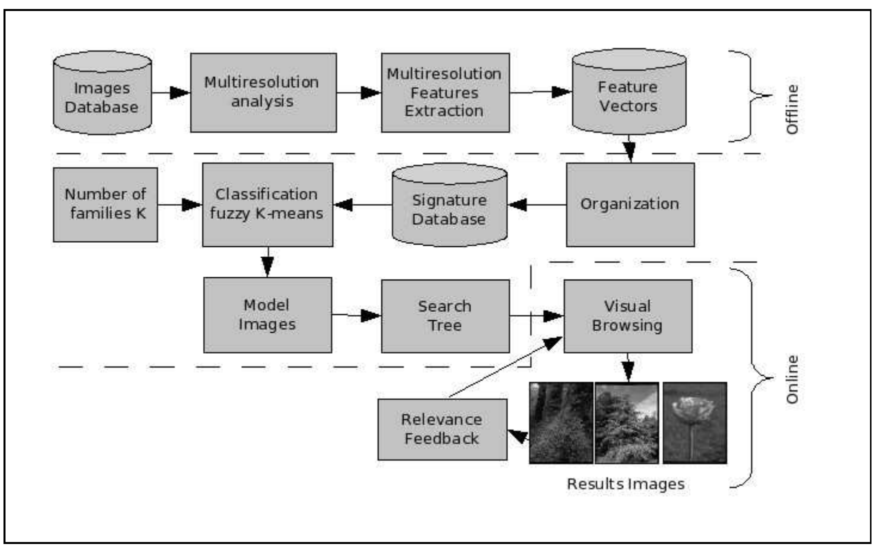

Figure 1. Principle of our system.

Our system is composed of six different steps:

A.	Transformation	B.	Extraction	C.	Organisation
D.	Classification	E.	Representation	F.	Navigation

Each step is detailed in the following description.

4.1. Transformation

Images are first transformed from RGB to Lab and HSV color spaces. Then, a multiresolution analysis is performed at three levels of resolution using integer to integer lifting scheme transform.[10] At each level, an approximation image and three details images are obtained.

4.2. Extraction

In order to characterize images, several features where extracted from approximation and detail images. Shape features are geometric moments from different order and surrounding ellipse of the contour. Texture features are computed from wavelets details energy. Color features are computed from wavelet detail energy of "ab" and "HS" color components.

4.3. Organization

The organization process is described in figure 2. Features are organized into four signature vectors of small but increasing sizes with the help of an expert of the database. It shows many advantages:

- There is no dimentionality curse at all because vectors' sizes are chosen small ($n < 16$).
- Distance computing is more rapid because small size vectors are used first (when many images to classify) and bigger vectors are used last (when small families of images have to be classified).
- Features are organized into vectors regarding to their type. It means signature vectors contain only one kind of features: color or texture or shape. So distances are computed between comparable features.
- Features are organized into pre-defined signature hierarchies taking in account users' search context.

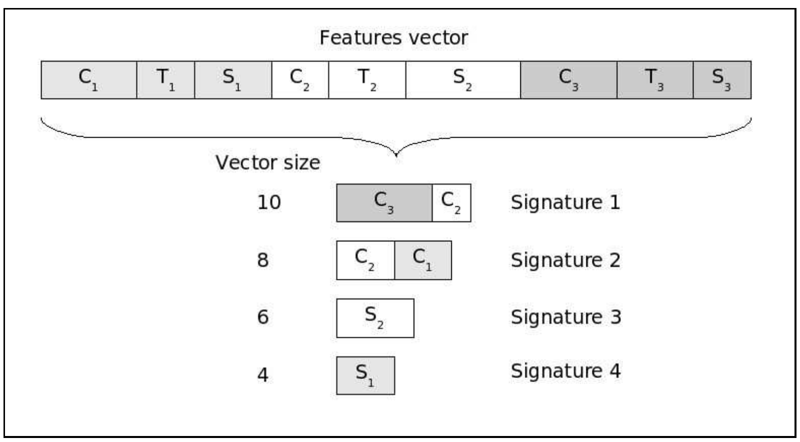

Figure 2. From features vector to signature vectors: organization.

4.4. Classification

Classification is based on K-means algorithm. Signature vectors are classified automatically into K families. K is fixed at the beginning of the algorithm. To take in account psychological principles of early vision, the number of families i.e. the number of images proposed to users while browsing is limited from 2 to 7.

Classification is a recursive process for the building of a search tree. This process ends when the number of images to classify in a family is lower than a given threshold chosen by the expert of the database domain.

4.5. Representation

A model image is necessary to represent a family of images. This model image will be proposed to users during visual browsing. To avoid users to focus their attention on an image while browsing, fuzzy images were used. They are built as the mean of the three closest images to the center of the family.

4.6. Navigation

Visual browsing occurs using a web interface to ensure users can browse the database over Internet easily.

5. IMPLEMENTATION

Our system is built on several librairies. Intel Integrated Performance Primitives (IPP, low-level library) and Intel Open Computer Vision (OpenCV, high-level library)[11] are used in C programs to realize image processing tasks. Visual browsing and management interface are built on PHP dynamic web pages calling C programs. Data are stored in a MySQL database.

Intel optimized libraries in combination with MySQL indexed tables give very good performances in terms of computing time. Figure 3 describes our architecture.

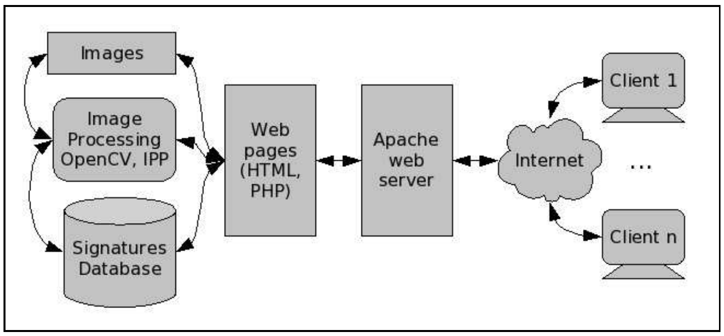

Figure 3. Our client/server web architecture.

In order to visualize families given by fuzzy classification, a 3D viewing tool was developped in Java with the Java3D library. To reduce dimensionality of vectors for 3D representation purpose only, a Principal Components Analysis is computed before 3D scene rendering. An example of interactive 3D visualisation is given in figure 4.

Figure 4. 3D viewing of families.

6. EXPERIMENTAL RESULTS

Our method was tested on two different databases:

- The Trans'Tyfipal paleontological database of the "université de Bourgogne" in France. Two experiments were realized, one with a small subset of 60 images (figure 5) to prove the efficiency of our approach and then another with the whole database set (727 images).
- The Columbia image database which is a free well-known images database of small objects (7200 images, 100 objects under 72 different views).

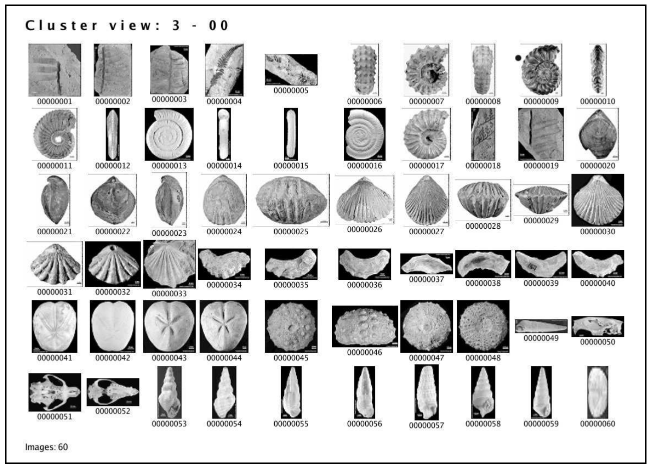

Figure 5. Subset of Trans'tyfipal database: 60 images.

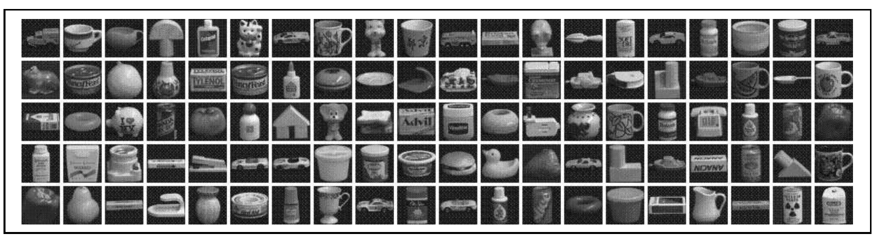

Figure 6. The hundred objects of the Columbia image database.

Users were proposed an easy-to-use interface with choices between several model images with mouse clicks. An example of model images proposed to users for browsing is described in figure 7.

In these tests, one feature vector is computed for each image. This vector is composed of 189 simple features from the three wavelet decomposition levels (63 features per wavelet level). These features are extracted from approximation and horizontal, vertical and diagonal images. Features vector are called fv0, fv1 and fv2 where fv0 is more detailed than fv2 which contains the coarser features. Features are divided into three types: color, texture and shape.

Color features are extracted from the "ab" and "HS" color component of transformed images. They are composed of statistical information about wavelet subbands in the color components of images.

Texture features are extracted from "L" and "V" components. These features mainly consists in energy of wavelet coefficients of each subband.

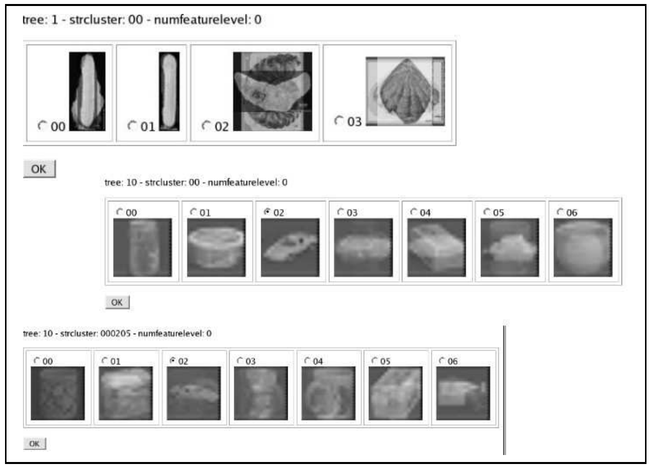

Figure 7. Model images proposed to users during visual searching.

Shape features come from geometrical moments of images. These features are normalized geometrical moments of approximation images, size of surrounding ellipse, geometrical center of images, ...

These features are organized with the help of an expert of the image collection to build a hierarchy of four signatures of increasing size. Only a subset of each feature vector is used in signature vectors leading to small sized signatures of size $n < 16$.

In these tests, the size of signature vectors sv0, sv1, sv2, sv3 are respectively 6, 6, 8 and 8. This increasing size limits dimensionality curse and speed-up the computing time during retrieval and indexing tasks.

6.1. Computing time

Computing time is important in images retrieval because even if computing is made offline, real time classification for small databases (< 10000 images) should be a good challenge. Tests were realized on a Pentium-M 1.1 GHz computer with 512 MB RAM and a 60 GB hard drive running Linux Fedora Core 4 (Apache 2.0.54, PHP 5.1.0, MySQL 5.0.9, Intel OpenCV 0.9.6, Intel IPP 4.1). The following table shows the computing time for reading images, steps A, B and E and writing features and model images.

Base	Imgs	Time/Img	Total time	Img size
Tyfipal60	60	0.3167 s.	19 s.	396x424x24
Columbia	7200	0.0431 s.	5 min. 10 s.	128x128x24

Next table shows computing time for organizing features vector into signature vectors.

Base	Time
Tyfipal60	2 s.
Columbia	43 s.

The following table shows computing time for step D.

Base	K	Read time	Iter.	Classif. time	PCA time	Write time
Tyfipal60	4	0.01 s.	6	0 s.	0 s.	0.03 s.
Columbia	7	0.3 s.	10	0.3 s.	0.17 s.	0.6 s.

Therefore the total classification time is 1.4 seconds for Columbia (7200 images) with seven families (K-Means needed 10 iterations to get the result). It proves our method works near real-time (t<2 s.) with databases whose size is less than 10000 images.

6.2. Accuracy

Accuracy has been tested with users browsing the fuzzy visual search tree. This measure is the result of users' choices while browsing. It is not an absolute measure like precision and recall, the two classical evaluation parameters of content-based image retrieval systems. So results presented below are quite good even if they seem low (about 60% success). Success occurs when users find exactly the same image as the request image they were proposed to retrieve. Even if they find an image close to the request image, it is not considered as success.

Next table gives user success rate as searching six images given to them for Tyfipal60.

Type of tree	2-tree	3-tree	4-tree
Hard	0.40	**0.65**	0.37
Fuzzy	**0.47**	0.60	**0.47**

The following table gives user success rate as searching six images given to them for Columbia.

Type of tree	2-tree	3-tree	4-tree	5-tree	6-tree	7-tree
Hard	0.15	0.15	**0.47**	0.40	0.55	0.50
Fuzzy	**0.35**	0.15	0.30	**0.50**	0.55	**0.60**

These tables prove that fuzzy tree is better than hard tree. It allows images to belong to more than one family allowing users to decrease their error rate during browsing.

6.3. Precision

For any content-based image retrieval system, precision for a given search is defined as the number of accurate images retrieved divided by the number of retrieved images. Precision is an absolute measure of quality for content-based image retrieval systems because it does not rely on users.

Figure 8 shows the precision of the retrieval process for the tyfipal database. The best precision is obtained by using the full size vector with 189 features extracted at the first level of the wavelet decomposition (fv0). The lowest precision is obtained from the full vector with 189 features extracted at level 2 of the wavelet decomposition (fv2).

It is important to remark that signatures have all better precision than fv2. It means our signatures work better than using the full 189 features vectors extracted from the coarser wavelet level. Even if the best result is obtained by fv0, working with sv3 decreases the computing time of about twenty times.

On figure 9, the best precision is for the full feature vector at the first level of details fv0. Again our signatures are better than the full size feature vector from the coarser resolution fv2. And our best signature is again sv3. Precision is not very good in the case of Columbia (due to the classification method) but the speed stays very good.

Figures 10a and 10b show results for full size feature vector fv0 and small size signature vector sv3 during a simple request of similarity. Every image comes with its number in the database and its distance to the request image. Vector sv3 is a slightly less precise than fv0 in this special case.

Figures 11a and 11b show results for full size feature vector fv0 compared to small size signature vector sv3 during a simple request of similarity. In this case, sv3 is not as precise as fv0 but the computing time is twenty times faster.

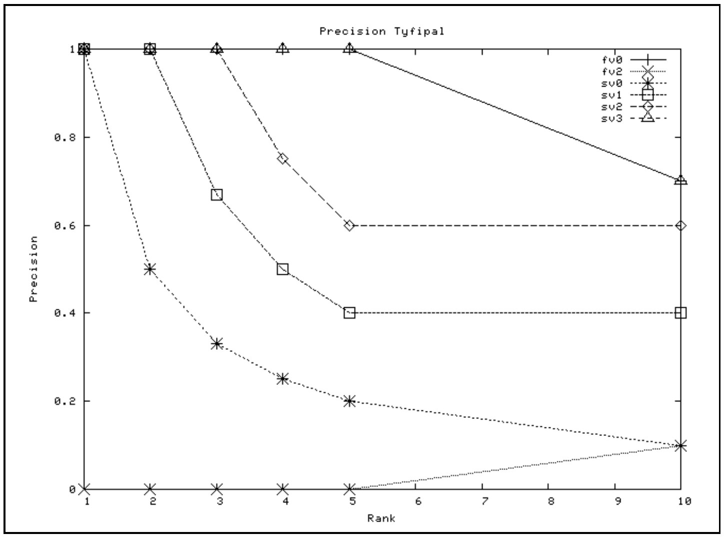

Figure 8. Precision for Tyfipal.

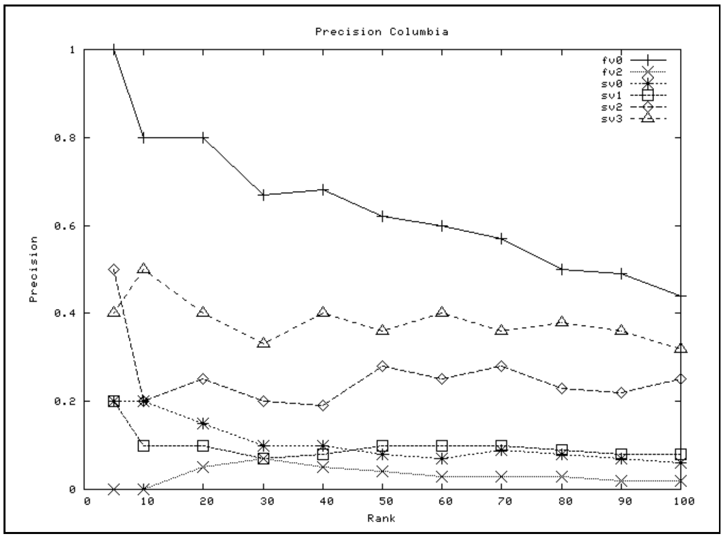

Figure 9. Precision for Columbia.

7. CONCLUSION

In this work, a system for image databases browsing was developed and tested on two bases: Tyfipal and Columbia. Both tests were very good in terms of speed because of the use of Intel optimized image processing libraries. The obtained results show the validity of our method for homogeneous databases. The maximum accuracy was 65% for Tyfipal and 60% for Columbia.

Our method allows users (experts or not) to visually browse image collections. Visual research tree building is computed offline but can be computed online for small size databases (< 10000 images). With the varying size of signature vector, users can manage speed and accuracy of the tree building process depending on their needs. Signature features do not mix different types of features, distance computing is performed on one type of features (colour or texture or shape). Hierarchy of signatures and wavelet transform gives a coarse-to-fine approach during retrieval.

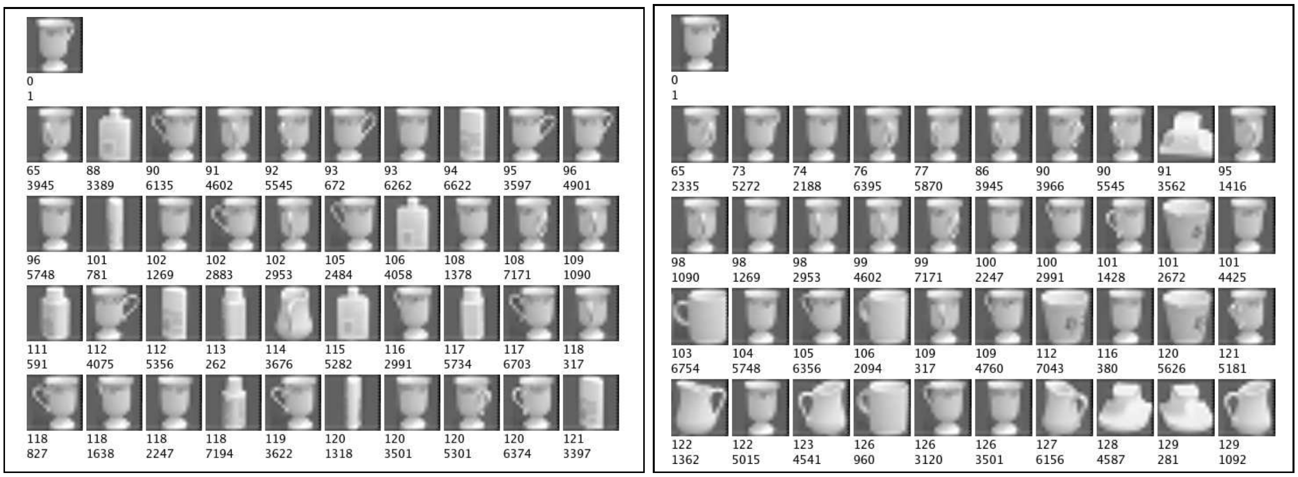

Figure 10. a) Retrieval with fv0. - b) Retrieval with sv3.

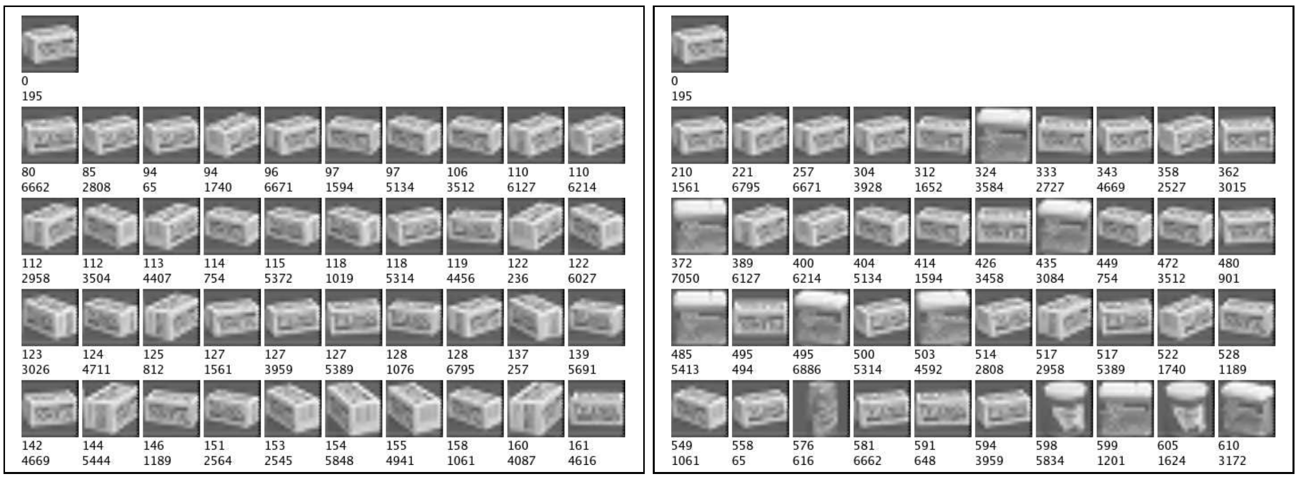

Figure 11. a) Retrieval with fv0. - b) Retrieval with sv3.

Our method gives quite good results and it solves several problems of classical image retrieval systems. Dimensionality curse, psychovisual techniques and users's context are treated in our approach. Our method works well with homogeneous images collections. It is easily extensible (by choosing relevant features to extract depending on the collection) to any image collection. A compromise between speed and accuracy must be chosen depending on the application, our method is adaptable.

8. FUTURE WORK

Due to the number of parameters involved in our method during its six steps, there are many possible improvements. However the most important part of the future work will be to focus on features selection and classification.

A complete study of features aided by an expert of image collections could improve results significantly. Better features will increase distance between different families of images leading to a better classification. Users's search context depends on features so features organization into a hierarchy of signatures is very important for search tree generation.

A classification method based on learning would give better results. Training a classifier with a subset of the collection will give better classification results. Clustering with K-means suffers limitations because K must be fixed during the offline phase of the method.

Working on model images is another way of improvement because model images proposed to users are too fuzzy in certain cases. Proposing a group of three to seven images located near the center of each class is an interesting direction to test.

There are many improvements to try to get a very good visual browsing system in terms of speed and accuracy. Next generation processors give promising results for speed, so accuracy stays the big problem in visual browsing…

REFERENCES

1. O. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions and open issues," *Journal of Visual Communication and Image Representation* **10**, pp. 39–62, March 1999.
2. J. Y. Chen and J. C. Bouman, C. A.and Dalton, "Hierarchical browsing and search of large image databases," *IEEETIP: IEEE Transactions on Image Processing* **9**(3), pp. 442–455, 2000.
3. M. Cord, S. Philipp-Foliguet, P.-H. Gosselin, and J. Fournier, "Interactive exploration for image retrieval," *EURASIP Journal on Applied Signal Processing* **2005**(14), pp. 2173–2186, 2005.
4. D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," in *AMS Mathematical Challenges of the 21st Century*, (University of Los Angeles, California, USA), August 2000.
5. S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 674–693, July 1989.
6. J. Landré, F. Truchetet, S. Montuire, and B. David, "Automatic building of a visual interface for content-based multiresolution indexing and retrieval of paleontology images," *Journal of Electronic Imaging - special issue* **10**, pp. 957–965, october 2001.
7. J. Landré and F. Truchetet, "Multiresolution hierarchical content-based image retrieval of paleontology images," in *Proceedings of SPIE Photonics East 2003*, *Wavelet Applications in Industrial Processing* **5266**, pp. 75–83, (Providence - Rhode Island, USA), December 2003.
8. M. J. Tarr and H. H. Bülthoff, *Object Recognition in Man, Monkey, and Machine*, MIT/ELSEVIER, 1998.
9. A. D. A., L. Moisan, and J.-M. Morel, "Computational gestalts and perception thresholds," *Journal of Physiology - Paris* **97**, pp. 311–322, may-march 2003. Special issue on "Neurogeometry and visual perception".
10. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Applied and Computational Harmonic Analysis (ACHA)* **5**(3), pp. 332–369, 1998.
11. J. Landré, "Programming with intel ipp (integrated performance primitives) and intel opencv (open computer vision) under gnu linux," http://jlandre.ifrance.com, july 2003. version 0.4.