The Image Shape Spectrum for Image Retrieval

Chahab Nastar

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

The Image Shape Spectrum for Image Retrieval

Chahab Nastar

N ° 3206

Juillet 1997

THÈME 3

Rapport de recherche

ISSN 0249-6399

The Image Shape Spectrum for Image Retrieval

Chahab Nastar*

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet syntim

Rapport de recherche n ° 3206 — Juillet 1997 — 22 pages

Abstract: We present an appearance-based technique for image characterization and retrieval. Our method is translation/rotation and scale- invariant and encodes the significant data in the image without using any segmentation. It is also very well suited to small viewpoint changes and is robust to noise and occlusion. We present several retrieval examples in large benchmark databases, including face databases and a database of 3D objects, for which the method reaches an ideal recognition rate.

Key-words: image databases, image indexing, image retrieval, appearance, invariance.

(Résumé : tsvp)

* Chahab.Nastar@inria.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

Le spectre de forme pour l'indexation d'images

Résumé : Nous présentons une technique fondée sur l'apparence pour caractériser et rechercher des images dans une base. Notre méthode est invariante à la translation, à la rotation et au changement d'échelle et capte l'information significative de l'image sans recours à la segmentation. Elle est également très bien adaptée au faible changement de point de vue et elle est également robuste au bruit et aux occultations. Nous présentons des exemples de recherche dans plusieurs grandes bases de données, contenant des images de visages ou des images d'objets tridimensionnels. Pour cette dernière base la méthode atteint un taux de reconnaissance idéal.

Mots-clé : bases d'images, indexation d'images, recherche d'images par le contenu, apparence, invariance.

1 Introduction

Multimedia is dominated by images. Digital movies, videos, photographs, paintings and drawings are widely used by scientists, professionals, and also the average person with a computer. Today, the new technologies related to archiving, retrieving, and editing images have to integrate high-level techniques of computer vision. The fields of application range from smart surveillance and medical diagnosis to photo editing, design, art and education.

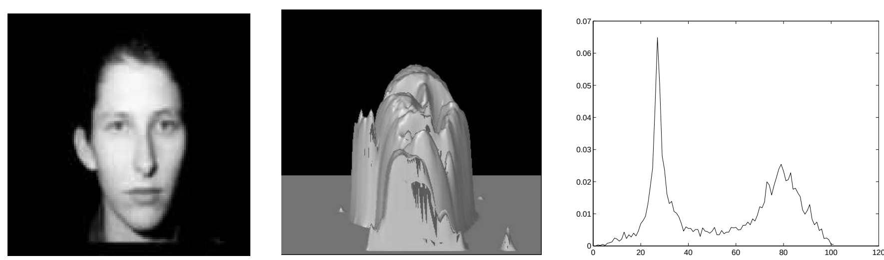

Figure 1: An image, its XYI representation, and its image shape spectrum (ISS)

Image databases typically contain several thousands of images, but the tools for browsing them are still rather primitive. Traditionnally, textual features (such as keywords) have been used for image indexing. This approach is sometimes effective but has several problems, since it models the content of the image quite subjectively. To a computer vision scientist, the image retrieval problem is the following: the user selects a query image and asks the computer to find more images that are perceptually *similar* to the query image. This is a very hard problem that has multiple solutions, depending on the application, the user, and the content of the images in the database. An index (called *image signature*) is usually derived from each image of the database, and a *similarity metric* is computed for grading the images in the database from the most similar image to the less similar image. Finally, a user-friendly *browser* is needed for querying and visualizing the images.

Most of the approches derive image signature from a single image attribute, like color [32, 8, 22, 10], texture [25, 15, 14, 30], or shape [11, 16, 1, 29, 31, 17, 2]. However, a number of recent studies attempt to integrate several image attributes, since a single attribute may simply not be present or lack sufficient discriminatory power for a number of real-world applications [24, 9, 26, 12].

In this framework, we have previously [21, 19, 20] modeled an image as a greylevel intensity surface (XYI representation) (see also [3]), and we have used that for matching and retrieval in a face database. On the other hand, Dorai and Jain [5, 6] use Koenderink's definition of the *shape index* [13] for characterizing a surface, with an application to matching and recognition of free-form surfaces in range images.

In this paper, we unify and extend the two techniques above with an application to image retrieval in large databases. Our approach avoids any preprocessing such as segmentation or edge detection. We use the shape index of the intensity surface of an image as the image attribute, and form its corresponding histogram as the image signature. Our method can be seen as a *geometric* alternative to the *physics*-based modeling of XYI similarity [20], as well as a *generalization* of the shape spectrum [5] to standard video imagery.

2 Image Shape Spectrum (ISS)

Similarly to [20], we use the XYI representation (viewing an image as its greylevel surface). This representation has the advantage of unifying the *texture* and *shape* attributes, as the intensity surface is embedded in the 3D XYI space with variations in the XY (shape) and I (texture) domains. In [20], we have used a *physics*-based analogy: two images are "similar" when their corresponding intensity surfaces can be matched with little strain energy. In this paper, we propose a *geometric* alternative by measuring a global statistic over the whole image: two images are similar when their intensity surfaces have a similar local shape. It should be pointed out that the physics-based representation of [20] preserves the spatial arrangement of the image, tuning the method to the typical variations observed in face images; on the other hand, our geometric method is invariant to translation/rotation and scaling of the image, a desirable property for image retrieval independently of image content.

By *image shape*, we mean the local shape of the intensity surface of the image. This can be efficiently derived by using standard differential geometry

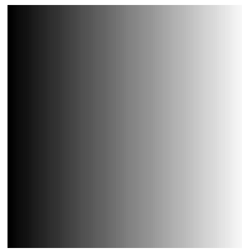

Figure 2: A planar intensity surface has no texture and no ISS

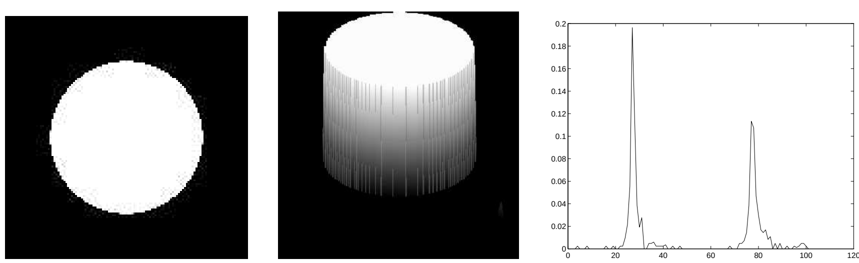

Figure 3: When the data is flat, the ISS only encodes the contours but still characterizes the data

of surfaces [4]. However, Koenderink and van Doorn [13] have shown that classical surface curvature measures such as the Gaussian and mean curvature are not very indicative of local shape. They introduce two more significant measures: the curvedness and *shape index* of a surface. This latter indicator has the important *scale-invariance* property, and captures the intuitive notion of local shape particularly well. Similarly to [13] and [5], we define the *image shape index* as a quantitative measure of the local shape of the intensity surface at a point $p = (x, y, I(x, y))$:

$$S_I(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \tag{1}$$

where k_1 and k_2 are the principal curvatures of the intensity surface, and $k_1 \geq k_2$. The principal curvatures are derived by using Gaussian filtering of width σ for image smoothing (we used the algorithm described in [27] with $\sigma = 2$ pixels). Note that the values of the curvatures at the image borders (of width $4\sigma = 8$ pixels) cannot be confidently estimated and should be ignored.

With this definition of the image shape index, all intensity surface shapes can be mapped into the interval $S_I \in [0, 1]$ and every distinct shape corresponds to a unique value of S_I, except the planar shape whose points have an indeterminate shape index since $k_1 = k_2 = 0$.

We now build the *image shape spectrum* (ISS) as the histogram of the image shape index over the entire image (we use 0.01 as the bin width). We normalize the ISS so that it sums up to unity. An example of an image, its greylevel surface, and its ISS[1] are represented on figure 1.

3 Properties of the ISS

3.1 Significance

The shape index and therefore the ISS are undefined for planar intensity surfaces. *A planar intensity shape represents an untextured image region* (figure

[1]The x axis represents discrete values of the shape index, the bin width being unity (101 bins in total); at each bin, the y axis represents the proportion of surface points that have the same shape index.

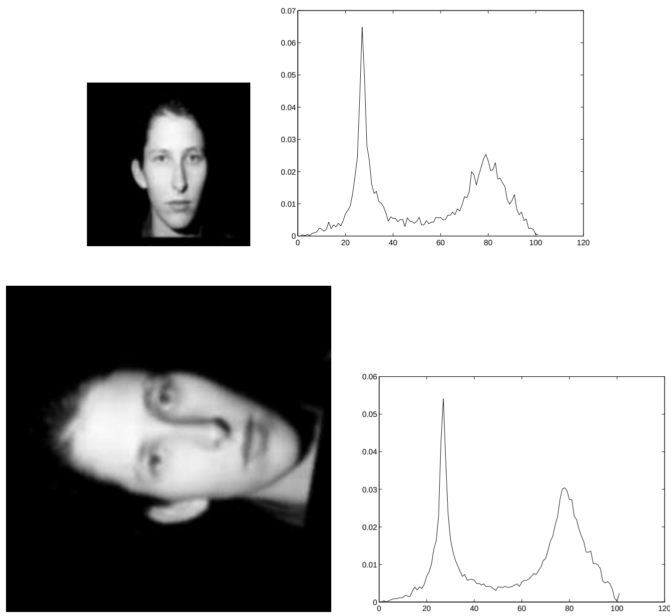

Figure 4: Top row: an image and its ISS. Bottom row: a rotated/2×scaled version of the image and the corresponding ISS. The r.m.s. error between the spectra is less than 2%

2). This is a nice property since by definition, the ISS will automatically ignore uniform and untextured regions (which are generally part of the background) to adapt to the free-form part of the surface which contains the data. On the other hand, if the data contains large uniform regions (flat intensity), the ISS will ignore those regions and encode the *region contours*, thus still characterizing the data (figure 3). Note that the ISS also captures image contrast via surface curvature. *Finally, the ISS encodes the significant information contained in the data particularly well.*

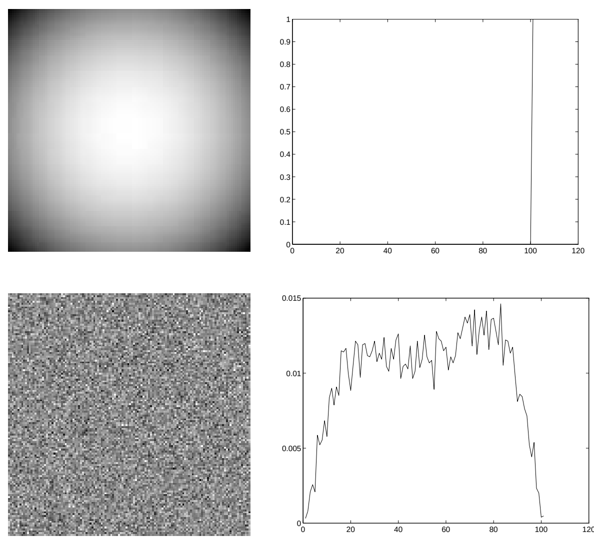

Figure 5: The ISS of a spherical cap has a single peak ($k_1 = k_2$ everywhere); the ISS of white Gaussian noise is irregular

3.2 Invariance

Due to the intrinsic nature of image shape, the principal curvatures and hence the ISS are fully invariant to in-plane translation/rotation of the image. Like the shape index, the ISS specifies *shape* quite indenpendently of *size* [13]; in other words, the ISS is also invariant to changes in scale. *Finally, the ISS is a* *rotation/translation and scale- invariant image signature*, a desirable property for image retrieval. In figure 4, the error between the spectra is less than 2%.

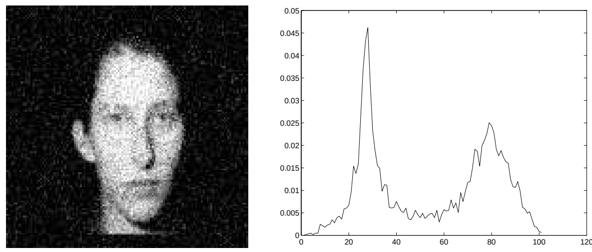

Figure 6: White Gaussian noise of width 3 pixels was added to the face image in figure 1, providing the ISS on the right. The error between the biased and original ISS is less than 2%. Note that for clarity, the amount of noise is 10 times amplified in the left image

3.3 Robustness

The image shape index measures the second order structure of the intensity surface in the neighborhood of any one of its points. The use of a Gaussian filter allows for robust estimation of the principal curvatures (and the shape index). For this same reason, pixel noise has little effect on the ISS. Figure 6 shows that a noisy version of an original image produces a very similar ISS. The error between the spectra in figures 1 and 6 is less than 2%.

3.4 Skewness

It appears from the various figures that the *rut* and the *ridge* are the main intensity shape categories in a number of images, thus giving a comparable skewness to all image shape spectra. In these configurations, the value of the shape index is respectively 0.25 and 0.75, and the ISS has two peaks at these values. This happens when $|k_1| \gg |k_2|$ or $|k_1| \ll |k_2|$ (equation (1)), in other words, at the vicinity of *image edges*. Any image that somehow contains edges (i.e. almost all real-world images) will produce an ISS with two distinctive peaks at $x = 0.25$ (concave edge) and $x = 0.75$ (convex edge). Examples of images not producing the two peaks are: a spherical cap (with borders not visible) and random noise (figure 5).

4 Application to Image Retrieval

4.1 ISS similarity

Now that we have defined the ISS as a significant image signature with nice properties of invariance and robustness, we need to define a similarity metric between two signatures $\mathbf{P} = (P_1, \ldots, P_N)$ and $\mathbf{Q} = (Q_1, \ldots, Q_N)$, where N is the number of bins (in our experiments the bin width is 0.01 thus $N = 101$).

A number of distances can be defined as dissimilarity metrics. We have implemented various L^p distances, as well as distances between moments, the Hellinger distance, and the Kullbach-Leibler divergence.

None of these metrics stands out as the most suitable, and all of them give good results on various experiments, leading to the rather pleasant conclusion that *the metric is not a main issue*, as opposed to the signature. The same observation is made in a number of image retrieval studies (e.g. [12]).

We eventually keep the Euclidean distance (i.e. L^2 error or sum of squared differences), which is the most natural dissimilarity metric:

$$d_{L^2}(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^{N} (P_i - Q_i)^2 \tag{2}$$

4.2 Retrieval in a Database of Elementary Images

We synthetize a set of 12 images representing free-form intensity surfaces (e.g. spherical cap), binary plain shapes (e.g. disk) and binary curvilinear shapes (e.g. rectangle). The retrieval with a dome-like structure (ellipsoid) as the query image is shown in figure 7. Note that the best matches are respectively: a sphere, a disk and a circle. *The retrieval result is thus coherent and perceptually satisfactory.* We can now move to extensive experimentation on real-world images.

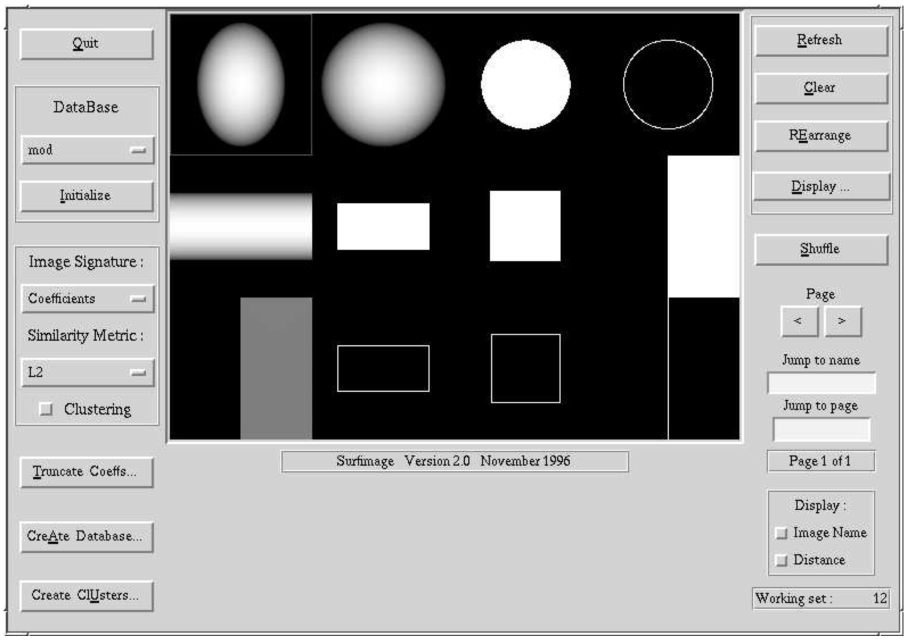

Figure 7: Retrieval of the ellipsoid (top left) in a set of elementary shapes. Similarity decreases from left to right and top to bottom

5 Experimental results

5.1 Computation Time

The computation time for deriving the ISS of an image depends only on the size of the image, not on its complexity. For a typical 128×128 image, the computation time is about 0.3 second per image on a SPARC-20 workstation, thus approximately 5 minutes for a database of 1000 images. After this pre-processing stage, the distance computation and sorting is on-line and *real-time*.

5.2 The ORL Face Database

We use the Olivetti Research Laboratory face database which contains 400 images (10 different images of 40 distinct subjects, although this information was not injected into the system). For some of the subjects, the images were taken at different times, with varying lighting, pose, facial expressions and facial details. The ISS-based retrieval of a face image is presented in figure 8. In this example, the method retrieves in priority *all* images of the query image. Note in the wide changes in image scale, head pose, facial expression, and facial details.

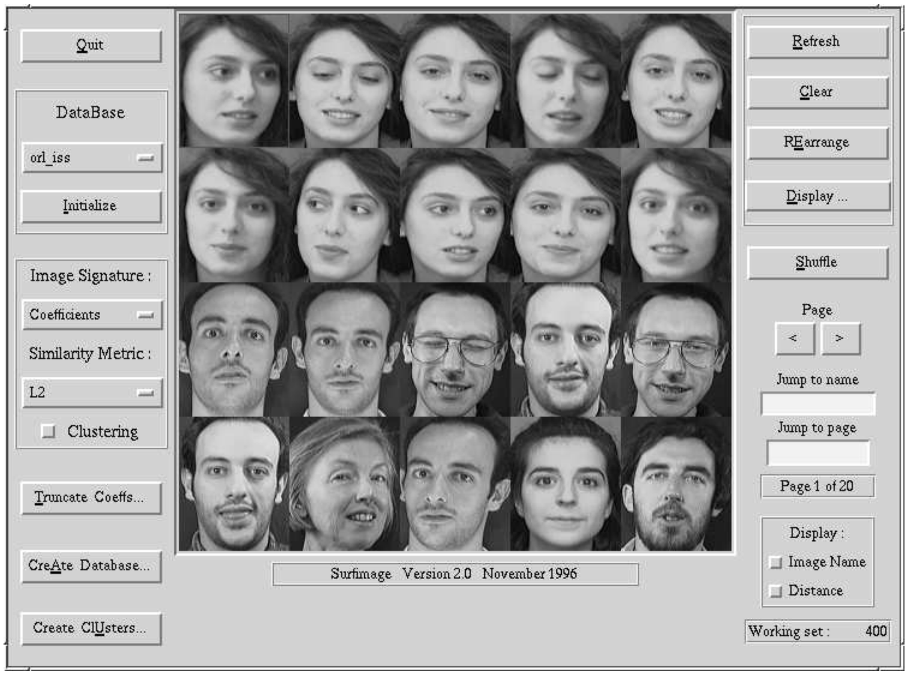

Figure 8: Retrieval of the top left face. Note in particular the viewpoint and scale changes. The system retrieves all other images of the query face among the best matches

5.3 The MIT Face Database

The MIT database contains 7,562 mugshot face images of approximately 3,000 people [33, 23, 20]. Each person has at least 2 images in the database. The faces are roughly aligned in the same position and scale in all images. The example in figure 9 shows the robustness of the ISS approach to occlusion. In particular, the second and third best matches are images of the query person without sunglasses.

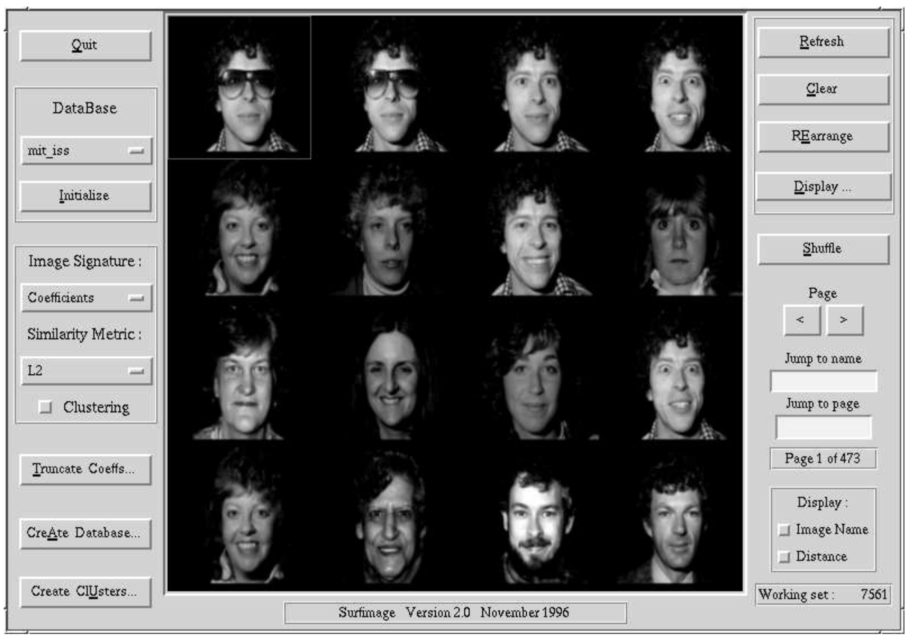

Figure 9: Retrieval of the top left face. Note that in particular the second and third best matches are faces of the query person without sunglasses

5.4 The Columbia Database

The Columbia database contains 1440 images of 20 different 3D objects with a wide variety of properties ranging from uniform reflectance and simple shapes to complex textural properties. The database contains 72 images per object, taken at 5 degrees incremented in pose. Several studies report 100% recognition rate on this database [18, 28].

Figure 10 presents a retrieval example on the Columbia database. Note that in our case the search is performed on the whole database of 1440 images. The last row of matches are interesting since they present the back side of the object, thus demonstrating the robustness of the method to image perturbations.

To assess the recognition rate, we use a *nearest neighbor rule*. If the most similar image was of the same object, then a correct recognition was scored. In our experiment, the system produced a recognition accuracy of 99.7% which corresponds to only 5 mistakes in matching views of 1440 objects. Note that at this point no *a priori* knowledge about the existence of 20 objects in the database was introduced in the system.

Figure 10: Retrieval of the top left image. Note that among the best matches are rear views of the vaseline bottle

5.5 Clustering the Signatures

Several objects of the Columbia database are very similar. In some queries, the system can retrieve a different similar object among the best matches. This is an interesting property if the goal of the user is to find an object from another similar one: figure 11 shows a TYLENOL pack as the query image for which the top fifteen matches are TYLENOL and ANACIN packs. Note again the large viewpoint changes in the retrieved images.

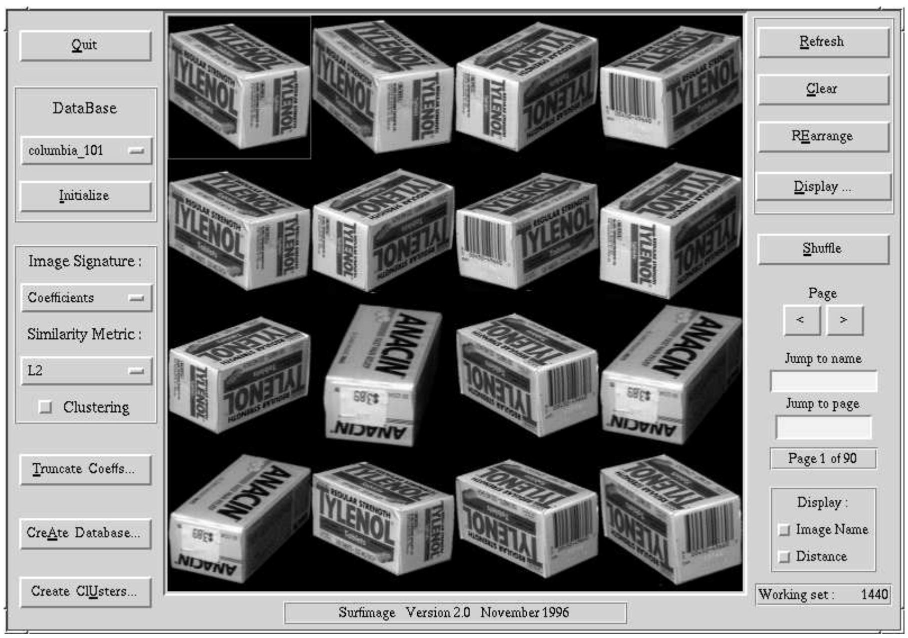

Figure 11: Retrieval of the TYLENOL pack on the top left. Note that among the best matches are TYLENOL and ANACIN packs

If needed, we can inject the *a priori* knowledge of the 20 objects, by clustering the Columbia database into 20 clusters, each cluster containing 36 images of an object (incremented 10 degrees). The other half of the database (720 images) will be used as the test set. We use the classic Mahalanobis distance [7] of the query image $\mathbf{Y}$ to the nearest cluster $\mathcal{C} = \{\ldots, \mathbf{X}_i, \ldots\}$ as the dissi-

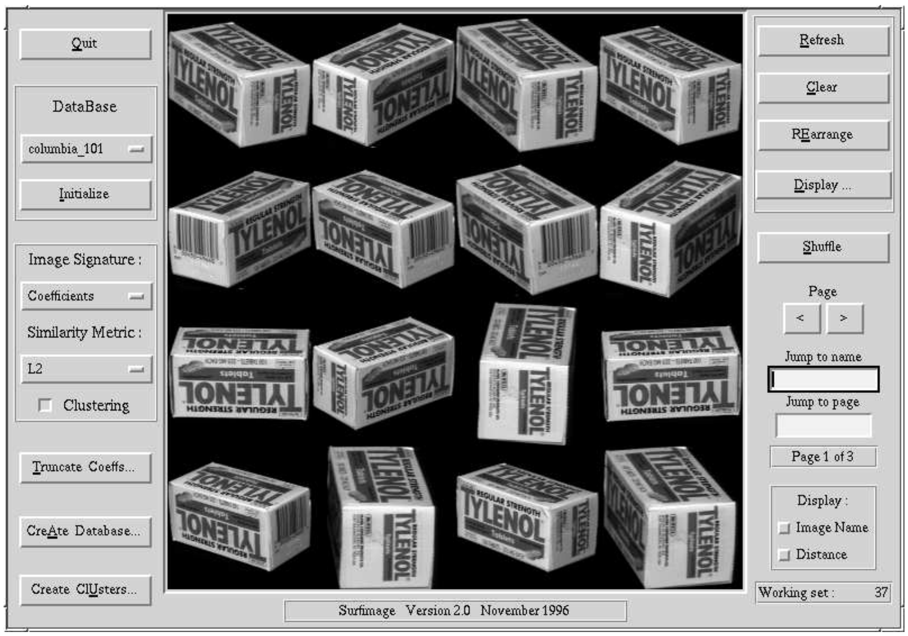

Figure 12: Retrieval of the TYLENOL pack on the top left by using the nearest cluster. All retrieved images are TYLENOL packs

milarity metric (figure 12):

$$d_M(\mathbf{Y},\mathcal{C}) = (\mathbf{Y} - \bar{\mathbf{X}})^T \, \mathbf{W}^{-1} \, (\mathbf{Y} - \bar{\mathbf{X}}) \tag{3}$$

where $\bar{\mathbf{X}}$ is the centroid and $\mathbf{W}$ is the covariance matrix of the cluster. Note that the clustering is performed in feature space (X and Y are images signatures). The addition of the clustering phase solves all ambiguities and the system reaches the ideal recognition rate of 100% with the nearest neighbor rule.

6 Summary and Conclusions

We presented an appearance-based technique for image characterization. Our technique uses all the information contained in the image without any kind of alteration or segmentation. Our experiments show that the image shape spectrum (ISS) is a significant image signature for retrieval for most databases, independently of image content. It has the advantage of being rotation/translation and scale invariant, is robust to noise and occlusion, and encodes the data particularly well. The ISS also deals with viewpoint changes. The technique is particularly easy-to-implement, needs no segmentation, tuning parameters or dedicated algorithm. It was widely validated on large databases (in particular the ORL, the MIT and the Columbia database) and gave excellent recognition rates (in particular 100% recognition on the Columbia database).

Finally, note that the histograming of the shape index as proposed in this paper should only be seen a convenient way of providing invariance properties to the image signature. However, the goal of our research is more generally to introduce image cheracterization via coding of the the intensity (or gray) surface shape (using for example surface curvatures and the shape index).

Acknowledgements

Thanks to Y. Marrec for the implementation of the *Surfimage* browser, and to V. Prinet for providing the curvature extraction algorithm.

References

[1] J. Bigun, S. Bhattacharjee, and S. Michel. Orientation radiograms for image retrieval: An alternative to segmentation. In *Proceedings of 13th International Conference on Pattern Recognition*, Vienna, Austria, 1996.

[2] A. Del Bimbo and P. Pala. Image indexing using shape-based visual features. In *Proceedings of 13th International Conference on Pattern Recognition*, Vienna, Austria, 1996.

[3] T.F. Cootes and C.J.Taylor. Modelling object appearance using the grey-level surface. In *Proceedings of the British Machine Vision Conference*, 1994.

[4] M. P. do Carmo. *Differential Geometry of Curves and Surfaces.* Prentice-Hall, Englewood Cliffs, 1976.

[5] C. Dorai and A. K. Jain. COSMOS: a representation scheme for free-form surfaces. In *Proceedings of the International Conference on Computer Vision 1995 (ICCV'95)*, Cambridge, MA, June 1995.

[6] C. Dorai and A. K. Jain. View organization and matching of free-form objets. In *Proceedings of the IEEE International Symposium on Computer Vision*, Coral Gables, FL, November 1995.

[7] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, 1973.

[8] B. M. Mehtre et al. Color matching for image retrieval. *Pattern Recognition Letters*, 16, March 1996.

[9] M. Flickner et al. Query by image and video content: the qbic system. *IEEE Computer*, 28(9), 1995.

[10] T. Gevers and A.W.M. Smeulders. A comparative study of several color models for color image invariant retrieval. In *First International Workshop on Image Databases and Multi-Media Search*, Amsterdam, August 1996.

[11] P.W. Huang and Y.R. Jean. Using 2D C^+ strings as spatial knowledge representation for image database systems. *Pattern Recognition*, 27(9), 1994.

[12] A. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8), 1996.

[13] J.J. Koenderink and A.J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10 (8), 1992.

[14] F. Liu and R.W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(7), 1996.

[15] W.Y. Ma and B.S. Manjunath. Texture-based pattern retrieval from image databases. *Mutlimedia Tools and Applications*, 1(2), 1996.

[16] R. Mehrotra and J. Gary. Shape-similar retrieval in shape data management. *IEEE Computer*, 28(9), 1995.

[17] F. Mokhtarian, S. Abbasi, and J. Kittler. Efficient and robust retrieval by shape content through curvature scale space. In *First International Workshop on Image Databases and Multi-Media Search*, Amsterdam, August 1996.

[18] H. Murase and S. Nayar. Visual learning and recognition of 3D objects from appearance. *Int. Journal of Computer Vision*, 14(1), 1995.

[19] C. Nastar, B. Moghaddam, and A. Pentland. Generalized image matching : Statistical learning of physically-based deformations. In *Proceedings of the Fourth European Conference on Computer Vision (ECCV '96)*, Cambridge, England, April 1996.

[20] C. Nastar, B. Moghaddam, and A. Pentland. Flexible images: Matching and recognition using learned deformations. *Computer Vision and Image Understanding*, 35(2), February 1997.

[21] C. Nastar and A. Pentland. Matching and recognition using deformable intensity surfaces. In *IEEE International Symposium on Computer Vision*, Coral Gables, USA, November 1995.

[22] G. Pass and R. Zabih. Histogram refinement for content-based image retrieval. In *Proceedings of the Workshop on Applications of Computer Vision*, Sarasota, Florida, December 1996.

[23] A. Pentland, B. Moghaddam, T. Starner, and M. Turk. View based and modular eigenspaces for face recognition. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, 1994.

[24] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. *Int. Journal of Comp. Vision*, 18(3), 1996.

[25] R. Picard and T. Minka. Vision texture for annotation. *ACM/Springer Journal of Multimedia Systems*, 3, 1995.

[26] R.W. Picard. A society of models for video and image libraries. Technical Report TR-360, MIT Media Lab, 1996.

[27] V. Prinet and O. Monga. Adaptative filtering and geometrical invariants in face depth maps. In *Workshop on Geometrical Modeling and Invariants in Computer Vision*, December Xi'an (China) 1995.

[28] R.P.N. Rao and D.H. Ballard. Object indexing using an iconic sparse distributed memory. In *Proceedings of the International Conference on Computer Vision 1995 (ICCV'95)*, Cambridge, MA, June 1995.

[29] Y. Rui, A. C. She, and T. S. Huang. Modified fourier descriptors for shape representation - a practical approach. In *First International Workshop on Image Databases and Multi-Media Search*, Amsterdam, August 1996.

[30] C. Schmid and R. Mohr. Combining greyvalue invariants with local constraints for object recognition. In *IEEE Proceedings of Computer Vision and Pattern Recognition (CVPR '96)*, San Francisco, June 1996.

[31] S. Sclaroff. Encoding deformable shape categories for efficient content-based search. In *First International Workshop on Image Databases and Multi-Media Search*, Amsterdam, August 1996.

[32] M. Swain and D. Ballard. Color indexing. *Int. Journal of Computer Vision*, 7 (1), 1991.

[33] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399