

Invariant content-based image retrieval using a complete set of Fourier-Mellin descriptors

Stéphane Derrode, Mohamed Daoudi

Faouzi Ghorbel

Groupe de Recherche Images et Formes
ENIC / INT - Cité scientifique de Lille I - rue Marconi,
59658 Villeneuve d'Ascq Cedex, France
Derrode@enic.fr, Daoudi@enic.fr

Ecole Nationale des Sciences de l'Informatique
CERT / DER - 42, rue Asdrubal,
1002 Tunis, Tunisia
Faouzi.Ghorbel@cert.mincom.tn

Abstract

Images retrieval from a large database is an important and emerging search area. This retrieval requires the choice of a suitable set of image features, a method to extract them correctly, and a measure of the similarity between features that can be computed in real-time. This paper presents a complete set of Fourier-Mellin descriptors for object storage and retrieval. Our approach is translation, rotation and scale invariant. Several retrieval examples in large image database are presented.

Keywords-- Content-based retrieval, Pattern recognition, Complete invariant description, Fourier-Mellin transform.

I. INTRODUCTION

Recent advances in computing and communication technology are taking the actual information processing tools to their limits. The last years have seen an overwhelming accumulation of digital data such as images, video, and audio. Internet is an excellent example of distributed databases containing several millions of images. Other cases of large image databases include satellite and medical imagery, where it is often difficult to describe or to annotate the image content.

Techniques dealing with traditional information systems have been adequate for many applications involving alphanumeric records. They can be ordered, indexed and searched for matching patterns in a straightforward manner. However, in many scientific database applications, the information content of images is not explicit, and it is not easily suitable for direct indexing, classification and retrieval. In particular, the large-scale image databases emerge as the most challenging problem in the field of scientific databases.

The Visual Information Retrieval (VIR) systems are concerned with efficient storage and record retrieval. In general, a VIR system is useful only if it can retrieve acceptable matches in real-time. In addition to human-assigned keywords, VIR systems can use the visual content of the images as indexes, e.g. color, texture and shape features. Recently, several systems combine heterogeneous attributes to improve discrimination and classification

results: QBIC [1], Photobook [2], Virage [3]. One point is to determine the appropriate weight given to each attribute. The features determine the type of queries that can be expressed, and VIR systems tend to be application specific because each application will have different retrieval needs.

Most of the features are local since they can be applied to more query situations (shape overlapping, partial retrieval, etc). However they generally need pre-processing stages that can significantly reduce the overall system robustness. Furthermore they rarely take into account the texture information. Thus, this paper focuses on the use of global features for the retrieval of isolated gray-level objects present on an uniform background.

It has been well-known that features should verify a number of properties :

- *Invariance* for a given set of geometric transformations,
- *Stability* under small shape distortions and numerical approximations,
- *Simplicity* and real-time computation.

We base our study on these properties together with another criterion called *completeness*. The latter ensures that two objects will have the same shape if and only if all their descriptors are equal [10]. In section II, we present a complete set of global gray-level image descriptors that satisfies all the criteria cited above. This set is extracted from the *analytical Fourier-Mellin transform*, and it is invariant to translation, rotation and scale changes of the images. From the solid theoretical background, we are able to define a real distance between invariant descriptors that can be computed in almost real-time. Section III introduces the application of the invariant distance in a retrieval scheme. Several classification and retrieval results from two real gray-level object databases are also discussed. Finally, we present the main conclusions and future work from the study.

II. GRAY LEVEL SHAPE INVARIANT REPRESENTATION

In this section, we describe a method to compute a complete and invariant representation of gray-level images that is suitable for content-based retrieval from image databases. The representation is performed using the analytical Fourier-Mellin transform and it is invariant to image translation,

rotation and scaling, i.e. plane similarity transformations. Based on the properties of the invariant set, we use a true mathematical distance between shapes as a similarity measure to compare two images in a database.

A. The analytical Fourier-Mellin transform

Late on the 70's, the optical research community introduced the Fourier-Mellin Transform (FMT) for pattern recognition purposes [6]. In the meanwhile, the Mellin transform was studied for target identification in relation to signal translation and scaling [7]. Few years ago, it was pointed out that the crucial numerical difficulties in computing the Fourier-Mellin transform of an image might be solved by using the Analytical Fourier-Mellin Transform [5]. Let us recall the main bases:

Let $f(r, \mathbf{q})$ be the irradiance function representing a gray-level image defined over a compact set of \mathbf{R}^2 . The origin of the polar coordinates is located on the image centroid in order to offset translation. The analytical Fourier-Mellin transform (AFMT) of f is given by [4]:

$$M_{f_s}(k, v) = \frac{1}{2^p} \int_0^1 \int_0^{2^p} f(r, \mathbf{q}) r^{s-iv} e^{-ik\mathbf{q}} \frac{dr}{r} d\mathbf{q} \quad (1)$$

for all k in \mathbf{Z} , v in \mathbf{R} , and $s > 0$. f is assumed to be square summable under the measure $d\mathbf{q}dr/r$.

The AFMT of an object f can be seen as the usual FMT of the distorted object $f_s(r, \mathbf{q}) = r^{2s} f(r, \mathbf{q})$. The AFMT gives a complete description of gray-level objects since f can be retrieved by its inverse transform given by:

$$f(r, \mathbf{q}) = \int_{-\infty}^{\infty} \sum_{k \in \mathbf{Z}} M_{f_s}(k, v) \cdot r^{-s+iv} e^{ik\mathbf{q}} dv \quad (2)$$

for $r \in \mathbf{R}_+^*$, $\mathbf{q} \in \mathbf{S}^1$ and $s > 0$.

Since no discrete transform exists, three approximations of the AFMT have been designed: the direct, the Cartesian and the fast algorithms [8]. For image content-based retrieval applications, real-time is crucial, and thus the fast algorithm can be performed as follows. With a variable change on the integral ($q = \ln(r)$ instead of r), the equation (1) can be rewritten into Fourier transforms as follows:

$$M_{f_s}(k, v) = \frac{1}{2^p} \int_{-\infty}^{+\infty} \int_0^{2^p} e^{qs} f(e^q, \mathbf{q}) e^{-i(k\mathbf{q}+qv)} d\mathbf{q} dq \quad (3)$$

A fast algorithm is obtained by computing a two dimensional Fast Fourier Transform on the log-polar distorted object $e^{qs} f(e^q, \mathbf{q})$. The log-polar sampling is built from the points obtained by the intersection between N beams originating from the image centroid and M concentric circles with exponentially increasing radii. In our tests, we have chosen $N = 128$ and $M = 128$.

The AFMT of an object is theoretically infinite in extent. When dealing with computers, a finite set of coefficients is only available so that part of the original image content is lost. However, the FMT goes to 0 as $|v|$ and $|k|$ go to ∞ and the information lost by numerical truncation can be as weak as required. Let K and V be the boundaries of the finite extent AFMT so that $M(k, v)$ is available for $k \in [-K..K]$ and $v \in [-V..V]$ with a sampling step value, over axis v , set to 1 and $s = 0.5$. Due to the symmetry property of the Fourier-Mellin transform, the effective size of this representation is $S_{K,V} = [(2K+1)(2V+1)+1]/2$. Figure 1 illustrates the fast AFMT approximation algorithm of a real gray-level image. The computation time for such algorithm is provided in the next section.

Similarly, the original gray-level image can be retrieved from its numerical AFMT by using a fast approximation of the inverse transform (Eq. (2)) and a log-polar to Cartesian coordinates conversion as well. Figure 2 shows the reconstructed Cartesian image of Fig. 1 obtained by using $K = 64$ and $V = 64$. Increasing the values of K and V , the quality of the reconstructed image can be improved. Extensive experiments on AFMT approximation and reconstruction can be found in [8].

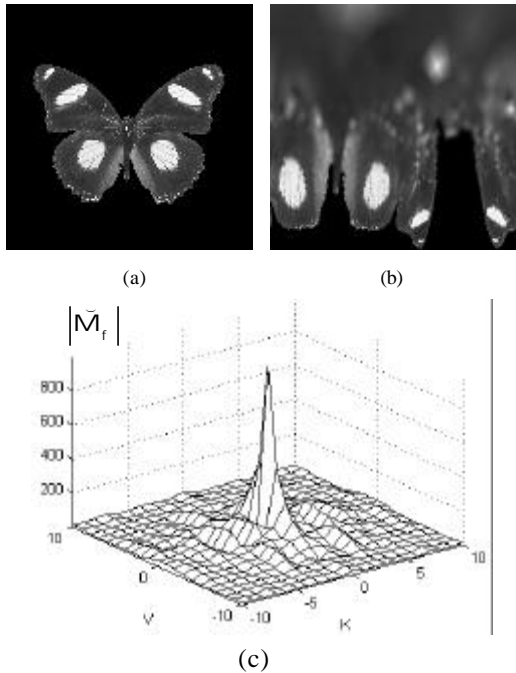


Fig. 1. (a) 128x128 butterfly image and (b) its log-polar re-sampling (128x128). (c) Magnitude of the central Fourier-Mellin coefficients obtained from the fast AFMT algorithm ($S_{10,10}=221$).

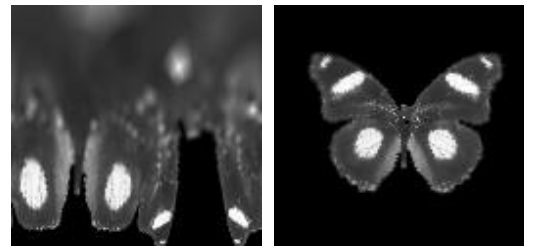


Fig. 2. Log-polar and Cartesian reconstruction of the butterfly in Fig. 1.

Finally, let us recall the transformation law of the analytical Fourier-Mellin transform for planar similarities. Let g be the orientation and size change of an object f by the angle $\mathbf{b} \in [0; 2\pi]$ and the scale factor $\mathbf{a} \in \mathbb{R}_+^*$, i.e. $g(\mathbf{r}, \mathbf{q}) = f(\mathbf{a}\mathbf{r}, \mathbf{q} + \mathbf{b})$. These two objects have the same shape and denoted similar objects. One can easily show that the AFMT of g and f are related by:

$$M_{g_s}(k, v) = \mathbf{a}^{-s+iv} e^{ikb} M_{f_s}(k, v) \quad (4)$$

for all k in \mathbb{Z} , v in \mathbb{R} and $s > 0$.

Equation (4) is called the shift theorem and suggests that the AFMT is well suited for the computation of global shape features which are invariant to the object position, orientation and size.

B. A complete set of Fourier-Mellin features

Since the usual Fourier-Mellin transforms of two similar objects only differ by a phase factor (Eq. (4) without the \mathbf{a}^{-s} term), a set of global invariant descriptors regardless of the object position, orientation and size, is generally extracted by computing the modulus of some Fourier-Mellin coefficients [9]. A set like this is not complete since the phase information is lost and it only represents a signature of the shape. Due to the lack of completeness, one can find distinct objects with identical descriptor values and a classification process may mix up objects, which is critical for content-based retrieval from image database (both false positive and true negative matches).

Recently, a complete family of similarity invariant descriptors based on the AFMT has been suggested [4]. This family can be easily written and applied to any strictly positive s value as follows:

$$I_{f_s}(k, v) = M_{f_s}(0, 0)^{-s+iv} \cdot e^{ik \text{Arg}(M_{f_s}(1, 0))} \cdot M_{f_s}(k, v) \quad (5)$$

for all k in \mathbb{Z} , v in \mathbb{R} .

Each feature $I(k, v)$ is constructed in order to compensate the $\mathbf{a}^{-s+iv} e^{ikb}$ term that appears in the shift theorem (4). The compensation is achieved via the two Fourier-Mellin

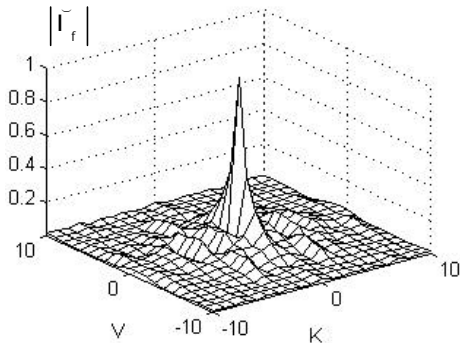


Fig. 3. Magnitude of some central invariant descriptors extracted from the butterfly in Fig. 1.

coefficients, $M(0,0)$ and $M(1,0)$, which are the normalization parameters. The set in Eq. (5) is complete since it is possible (i) to recover the FMT of an object from all of their invariant descriptors and the two normalization parameters by inverting Eq. (5); (ii) to reconstruct the original gray-level image by the inverse AFMT (see Fig. 2). Figure 3 shows the magnitude of the central invariant descriptors obtained from the butterfly in Fig. 1. We obtained as many invariant descriptors as Fourier-Mellin coefficients ($S_{10,10} = 231$ invariants).

C. A true distance between shapes

For pattern recognition purposes, the classification of an unknown object into a set of reference patterns is achieved by several comparison methods. Besides them, the direct comparison of a couple of features, neural networks; or statistical classifiers, by means of intra- and inter-class similarity measures.

Since the invariant set (5) is also convergent for square summable functions, it can be shown that the following function defines a true mathematical distance between shapes [4]:

$$d_2(I_{f_s}, I_{g_s}) = \left(\int_{-1}^{+1} \sum_{k \in \mathbb{Z}} |I_{f_s}(k, v) - I_{g_s}(k, v)|^2 dv \right)^{1/2} \quad (6)$$

This distance is an Euclidean distance expressed in the invariant domain. Theoretically, it is zero if and only if the objects are identical up to a similarity transformation. Due to numerical sampling and approximation, we never have exactly zero and the value of the distance is used for the quantification for the similarity between objects, regardless of their pose, orientation and size in the image.



Fig. 4: Samples of the butterflies database (original size 128x128) and simple Html interface for friendly user access.

III. APPLICATION TO IMAGE ARCHIVAL AND RETRIEVAL

In this section, we use the results obtained from the previous section for content-based image retrieval. Firstly, the two databases used for experiments are presented. Then, the main algorithm, for retrieval and classification, is sketched. Finally, we show different retrieval results and examine the main properties of the invariant set.

A. The test database content

The first database includes a collection of 91 images of butterflies (Fig. 4). 69 images represent distinct butterflies and 22 images show translation, rotation and scale changes of one of the 69 images. This database has been built to test similarity transformation invariance for our descriptor set according to the number of invariants used in experiments. It must be pointed out that searching a set such this is quite difficult, since the global aspect of butterflies shows prevailing shape characteristics and the main difference between butterflies essentially comes from the texture of the wings. Thus, contour-based retrieval scheme might be confused.

The second model collection is the well-known Columbia database which contains 1440 images of 20 different 3D objects : 72 images per object taken at 5 degrees in pose. In this collection, camera and object motion clearly violate the similarity transformation model underlying the image representation. This database is used to test the robustness and stability of our descriptor set to shape distortions.

B. Algorithm and computation time

According to section II, the retrieval of an unknown image from a set of P models can be split into the following two stages:

1. At archival time (off-line), each model f_i , $i \in [1..P]$ is represented by its bi-dimensional matrix of invariant features $\{I_{in}(k,v)\}$, with $k \in [-K..K]$ and $v \in [-V..V]$ by using algorithms derived from Eq. (3) and (5).
2. At query time (on-line), the classification of an unknown

object g is achieved as follows:

- computing their invariant features (if the query does not belong to the database),
- estimating the distance between the input object g and the P models f_i , by using Eq. (6),
- sorting and selecting the models which give the smallest distance to g .

In this way, all images from the database can be compared with any query image, and ranked by the value of d_2 . A fixed number of top-ranked images can then be displayed to the user, enabling browsing through the database. Since invariants are used as indexes, it is noted that a new model can be added to the database without modifying the models already stored. Furthermore, no voting algorithm for selecting the best model is required.

Concerning to the experiments, we compute $S_{15,15}=481$ invariant descriptors for each 128×128 gray-level images. The computation time for deriving the invariant representation is about 4s on a 200 MHz PC (including file I/O, log-polar re-sampling, 2D FFT computation and invariant extraction). It only depends on the size of the image, not on its complexity. At query time, the search through the database is exhaustive and performed linearly. The distance estimation and sorting is almost in real-time.

C. The butterflies database

Figure 5 shows three sample sets of the top 12 ranked images which have been retrieved from the same query image with different amounts of invariants.

For the right query ($S_{15,15}=481$), the method retrieves in priority all the 4 images similar to the query images. Taking less descriptors than 481 for the computation of the distance can produced confusion between shapes. The mix-up can be seen in the middle ($S_{7,7}=113$) and left ($S_{3,3}=25$) retrieves, where the rank of the 4 similar butterfly increases as the amount of invariants decreases. Other experiments have shown that this problem occurs when the scale factor between the query and the target is large. The position of the

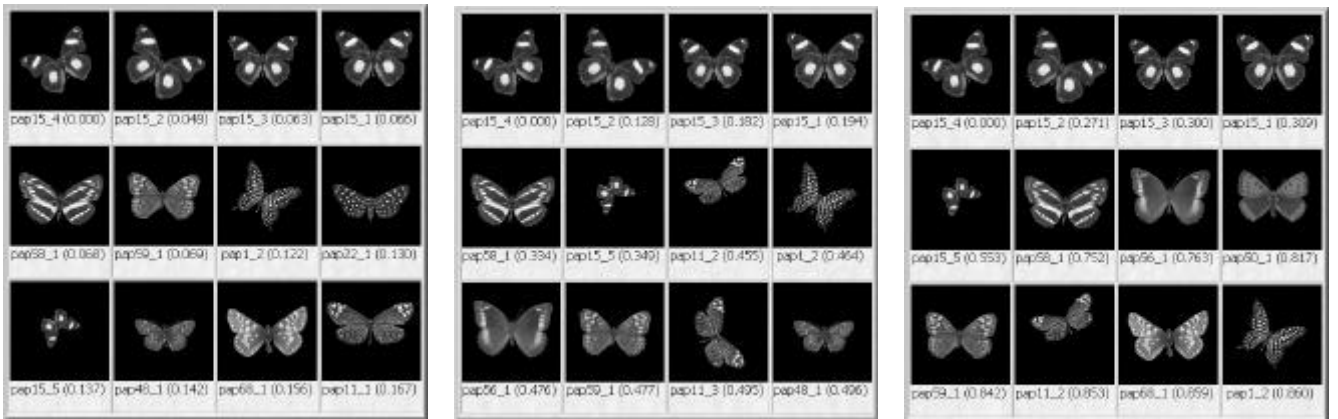


Fig. 5. Results of database retrieval using the query image shown in the top-left corner. Images are ranked from left to right, top to bottom, by decreasing similarity (increasing distance measure d_2). Left: $S_{3,3}=25$, Middle: $S_{7,7}=113$, Right: $S_{15,15}=481$.

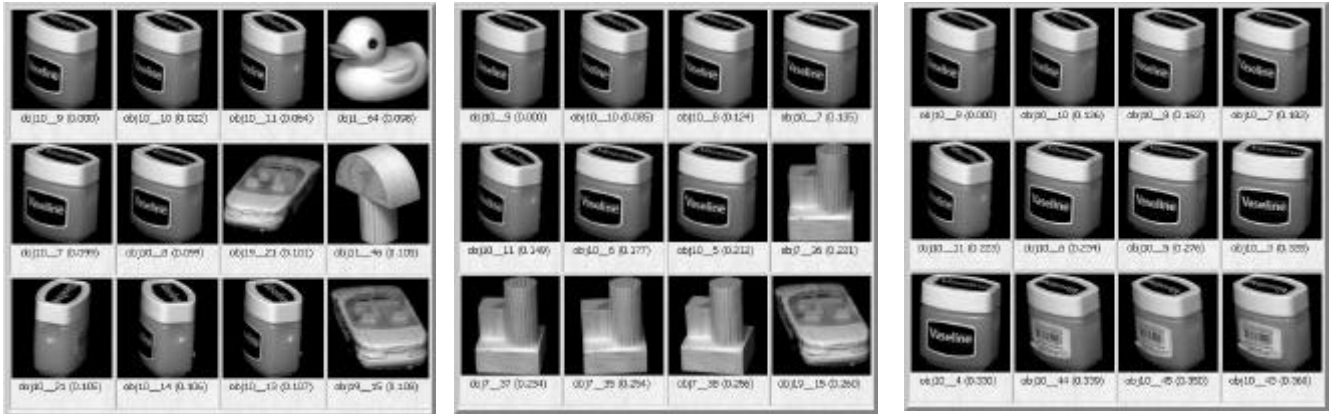


Fig. 6. Results of database retrieval using the query image shown in the top-left corner. Images are ranked from left to right, top to bottom, by decreasing similarity (increasing distance measure d_2). Left: $S_{3,3}$, Middle: $S_{7,7}=113$, Right: $S_{15,15}=481$.

other ranked images also changes according to S , however, most of them are present in the 3 retrievals and show significant resemblance with the query.

D. The Columbia database

Figure 6 shows the top 12 ranked images retrieved from the same query image by larger set of invariants (from left to right). Retrieval results become better as the dimension of the invariant features enlarges, since more and more images of the query object are retrieved.

The top 12 ranked of the right query (481 features) represent images of the query object. The last row of matches is interesting since it presents the back side of the Vaseline bottle, thus we demonstrate the robustness of the method to small distortions (non-similarity transformations). The first non similar object is ranked at the 18th position.

IV. CONCLUSION

We have proposed the use of a complete set of gray-level invariant features regardless of the object pose, orientation and size for the retrieval of an unknown object into a set of models. An efficient algorithm was presented for automatic extraction of a large set of invariant features, and a true invariant distance was then tested for content-based retrieval with two databases. Experimental results have confirmed that classification results can be improved when increasing the dimension of the feature space (with a small additional computing time at query time). The feature set have showed high-quality numerical invariance and good retrieval results when the camera and the object motion clearly violate the similarity transformation model underlying the image representation.

Since our set works better for database containing isolated objects on a uniform background, it seems well-suited for professional database such as : medical, biology, telecommunication (MPEG 7), etc. When the object essentially presents local deformations, such as faces, gray-level local invariants should provide better retrieval results [10].

Our complete invariant set can be used to store and encode complex gray-level shapes uniquely since we have showed

in section II that it is possible to reconstruct an object from its invariant descriptors. Future work will also include the improve of the efficiency for indexing and searching by means of feature quantization and hierarchical search strategies to avoid the exhaustive search through the database as it has been proposed here. Multiple image queries can also be studied to refine retrievals by adding the distances obtained from each individual query.

V. REFERENCES

- [1] M. Flickner et al., 'Query by image and video content: the QBIC system', *IEEE Computer*, pp 23-32, 1995.
- [1] A. Pentland, R. W. Picard and S. Sclaroff, 'Photobook: tools for content based manipulation of image database', in Proc. of SPIE, *Storage and Retrieval for Image and video Database II*, San Jose, CA, pp 34-47, 1994.
- [2] J.R. Bach et al, 'Virage image search engine: an open framework for image management', in Proc. of SPIE *Storage and Retrieval for Image and video Database IV*, San Jose, CA, pp 76-87, 1994.
- [3] S. Matusiak, M. Daoudi, T. Blu and O. Avaro, 'Sketch-based images database retrieval', Sushil Jajodia, M. Tamer Özsu, Asuman Dogac editors, Proceedings Workshop on *Multimedia Information systems*, Lecture Notes in Computer Science Vol. 1508, pp ??-??, 19??.
- [4] F. Ghorbel, 'A complete invariant description for grey-Level images by the harmonic analysis approach', *Pattern Recognition Lett.* 15, pp 1043-1051, 1994.
- [5] D. Casasent and D. Psaltis, 'Scale invariant optical transform', *Opt. Eng.* 15(3), pp 258-261, 1976.
- [6] P.E. Zwicke and Z. Kiss, 'A new implementation of the Mellin transform and its application to ndar classification', *IEEE Trans. Pattern Analysis Mach. Intell.* 5(2), pp 191-199, 1983.
- [7] S. Derrode and F. Ghorbel, 'Robust and efficient Fourier-Mellin transform approximations for invariant gray-level image description and reconstruction', Submitted to *Pattern Recognition*.
- [8] Y. Sheng and H.H. Arsenault, 'Experiments on pattern recognition using invariant Fourier-Mellin descriptors', *J. Opt. Soc. Am. A* 3(6), pp 771-776, 1986.
- [9] T.R. Crimmins, 'A complete set of Fourier descriptors for two dimensional shapes', *IEEE Trans. Syst., Man, Cybern.* 12(6), pp 848-855, 1982.
- [10] C. Schmid and R. Mohr, 'Local gray-value invariants for image retrieval', *IEEE Trans. Pattern Analysis Mach. Intell.* 19(5), pp 530-535, 1997.