

# Character Recognition, Orientation, and Scale Estimation Thanks to the Fourier Mellin Transform

Sébastien Adam<sup>1,3</sup>, Jean Marc. Ogier<sup>1</sup>, Claude Cariou<sup>2</sup>, Rémy Mullot<sup>1</sup>,  
Joël Gardes<sup>3</sup>, and Jacques Labiche<sup>1</sup>

<sup>1</sup> Laboratory PSI, University of Rouen, 76 821 Mont Saint Aignan, France  
Jean-Marc.Ogier@univ-rouen.fr

<sup>2</sup> LASTI, ENSSAT Lannion, 6, rue Kerampont, BP 447, 22305 Lannion, France  
Claude.Cariou@enssat.fr

<sup>3</sup> France Télécom, DVSI DES/MBL, 6, Avenue des Usines, 90 007 Belfort, France  
Joel.gardes@cnet.francetelecom.fr

**Abstract.** In this paper, we consider the general problem of technical document interpretation, applied to the documents of the French Telephonic Operator, France Telecom. More precisely, we focus the content of this paper on the computation of a new set of features allowing the classification of multi-oriented and multi-scaled patterns. This set of Invariant is based on the Fourier Mellin Transform. The interests of this computation rely on the possibility to use this Fourier Mellin transform within a “filtering mode”, that permits to solve the well known difficult problem of connected character recognition. In this paper, we also present an original technique allowing to compute an estimation of the orientation of each shape to be recognized.

## 1 Introduction

The current improvements of intranet structures allow large companies to develop internal communications between services. The representation of the heritage of huge companies like network managers firms is often represented through paper documents, which can be either graphic or textual. As a consequence, the sharing of these kind of information will stay very difficult as long as the storage format will not be digital. This explains the current development of studies concerning the automatic analysis of cartographic or engineering documents which comes as a result of the growing needs of industries and local groups in the development and use of maps and charts. The aim of the interpretation of technical maps is to make the production of documents easier by proposing a set of stages to transform the paper map into interpreted numerical storage [1][2][3][4]. An important step of this conversion process consists in the recognition of characters and symbols, which often appear, on technical documents, in different orientation and size. In this paper, we focus our attention on an original technique which allows the recognition of multi-oriented and multi-scaled characters and

strings. The application which is considered in this paper is the automatic analysis of a French telephonic operator documents, France Telecom.

The paper will be organized as follows. In the second section, we will give a synthesis of the bibliographic sources dealing with multi-oriented and multi-scaled pattern recognition problem. In the third section, we will then propose our new pattern description tool, based on a set of invariant issued from the Mellin Fourier Transform. This part will also include the description of the interesting properties of this tool and it's different modes of utilization that permits to solve the difficult problems of connected patterns recognition. The orientation estimation, which is an important factor in order to reconstruct strings will also be presented in this section. In the fourth section, we will present how this multi-oriented and multi scaled OCR (Optical Character Recognition), based on two utilization modes, is implemented in the particular case of connected patterns recognition .Preliminary results will be given in this part. Then, as usually, in the fifth section, we will try to have a critical point of view of our approach and to define some potential perspectives to this work.

## 2 Characters and Symbols Recognition : Classical Approaches

The problem dealing with character and symbol Recognition constitutes a difficult point of the current CAD conversion process. From a definition point of view, we will consider as “symbols”, the shapes the size of which is similar to the characters’ one. This definition takes into account the “alphanumeric characters” as well as the graphic symbols identified by A.Chhabra in [5]. The works integrating particular constraints like orientation or scale changes are less numerous in the literature. Let’s note the works of Deseiligny [6] in the particular case of maps, the one’s of Dori [7], based on NETS system, and Trier [8]. In this field of research, three main approaches may be distinguished.

- The first suggests a preliminary computation of the shape’s orientation, and tries, through a normalisation and a rotation step, to obtain a pattern in a reference position which can be introduced into a classical OCR system. However, methods based on such a strategy are not frequently used because of the lack of methods that enable a reliable computation of the orientation to be obtained. Moreover, distortions due to sampling errors appear during the geometric transformations that are needed to normalise the pattern.
- A second approach consists in using a multi-layer feed forward classifier, fed by the original image of the pattern. In this kind of context, the classifier renders the problem invariant with respect to the desired transformations [8].
- The last approach, which is probably the most frequently used, consists in extracting from the shape a set of descriptors, which are invariant to the desired transformations. An excellent description of the state of the art in this domain can be found in Trier [8]. Generally, it is possible to note that the features used to describe patterns independently from their position, size, and rotation, can be split up into two groups, as shown in the following.

## 2.1 Descriptors Based on the Global Aspect of the Pattern

Many features can be used to describe the global aspect of a shape. Since the works of Hu in 1961, invariant moments [9], which are based on combinations of regular moments, have been very often used. One can thus cite Zernike or pseudo-Zernike moments [10], Bamieh moments [11], and Legendre moments. These invariant moments, which can be extracted from a binary or a grey-scaled image, generally offer properties of reconstructability, thus ensuring that extracted features contain all the information about the shape under study. Good comparative studies about moment invariants can be found in [12], both showing the superiority of Zernike moments in terms of recognition accuracy. Nevertheless, these studies also proved that moment-based approaches are sensitive to noise and that they are time-consuming, even if complexity optimisation methods can be found in the literature [13].

## 2.2 Descriptors Based on a Local Approach

Besides the approaches presented above, a geometric invariant description can also be accomplished using features which are supposed to contain most of the pattern information. For example, contours are commonly used in order to obtain invariant descriptions of patterns through Fourier descriptors [14] or elliptic Fourier descriptors [15]. In terms of simplicity and robustness, the potential interest of these descriptors was shown by Taxt [16] through a comparative study between these descriptors. Structural invariant features can also be extracted from thinned characters. One can thus cite the number of loops, the number of T-joints or X-joints, the number of bend points. However, it has been shown that such features used alone do not lead to robust recognition systems [17]. Circular primitives, which are, by definition, well adapted to rotation invariant recognition, have been used in [18]. These are based on the analysis of the shape through a set of circles. A comparative study, available in [19], shows that they yield better results than Hu's moments. In the field of invariant pattern recognition, there is a consensus [12][8] about the fact that the feature extraction method is probably the single most important factor. A large number of feature types is reported in the literature. As we have seen above, some of them are based on the global aspect of the pattern (e.g. moments) whereas others try to select particular points from the shape (e.g. circular primitives). Given this large number of existing methods, one could argue that it is not necessary to develop new invariant features. Nevertheless, comparative evaluation studies [12] have shown that the features presented above are not perfect in terms of recognition accuracy, especially when images are noisy. Also, one must notice the lack of literature concerning the analysis of connected characters. Indeed, in the document processing application field, exception made of a few papers presenting some recognition techniques based on topographic or topologic criteria [6], as far as we know, there is no work about the direct recognition of connected shapes without prior segmentation or introduction of some *a priori* knowledge.

### 3 The Fourier-Mellin Transform and Its Properties

As we said previously, a strong constraint in the global interpretation of the document comes from the fact that characters and symbols can have any orientation and size. The consequence is that the recognition procedure to be applied must be invariant with regard to any combination of rotation and scaling of a pattern, i.e. any geometric similitude transformation. Another strong constraint relies on the robustness of the recognition procedure. In fact, after the binarization step, many characters are still connected either together, or to the network, leaving any classical pattern recognition technique useless. The strategy that we propose covers both constraints within a uniform framework. It is based on the application of the generalized Fourier analysis to the particular geometric group of positive similitude. More precisely, we make use of the properties of the Fourier-Mellin transform, the properties of which are very interesting for our application. Basically, the technique developed herein is a combined use of the works of Ghorbel [20], Ravichandran and Trivedi [21]. First, we will recall the definition of the Fourier-Mellin transform (FMT). Then, we will recall the analytic prolongation of the FMT (AFMT) and a set of complete and stable similitude invariant features, first proposed in [20]. The properties of this set of invariant will then be exposed.

#### 3.1. The Fourier-Mellin Transform (FMT)

Let  $f(r, \theta)$  be a real-valued function (the pattern) expressed in polar coordinates. The FMT of this function is defined as the Fourier transform on the group of positive similitude:

$$M_f(v, q) = \int_{\rho=0}^{+\infty} \int_{\theta=0}^{2\pi} \rho^{-iv} \exp(-iq\theta) f(\rho, \theta) \frac{d\rho}{\rho} d\theta \tag{1}$$

with  $q \in \mathbf{Z}$ ,  $\theta \in \mathbf{R}$

In this expression,  $i$  is the imaginary unit. It is well known that the Fourier-Mellin integral does not converge in the general case, but only under strong conditions for  $f(r, \theta)$ .

#### 3.2. Analytic Prolongation of the Fourier-Mellin Transform (AFMT)

In order to alleviate the above difficulty, Ghorbel [17] has proposed the use of the AFMT, defined as:

$$\tilde{M}_f(v, q) = \int_{\rho=0}^{+\infty} \int_{\theta=0}^{2\pi} \rho^{-iv+\sigma_0} \exp(-iq\theta) f(\rho, \theta) \frac{d\rho}{\rho} d\theta \tag{2}$$

with  $q \in \mathbf{Z}$ ,  $v \in \mathbf{R}$ , and  $\sigma_0 \in \mathbf{R}_+^*$

An important property of the AFMT (as well as the FMT) relies on the application of the shift theorem for the Fourier transform. Let  $g(\rho, \theta) = f(\alpha\rho, \theta + \beta)$  be a scaled and rotated version of  $f(\rho, \theta)$ , then we have the following :

$$\tilde{M}_g(v, q) = \alpha^{-\sigma_0+iv} \exp(iq\theta) \tilde{M}_f(v, q) \tag{3}$$

Taking the modulus of both terms in Eq. (2) yields features which are invariant under any rotation of the pattern but not under scaling. To obtain scale invariance on this basis, one could use the following set of features :

$$I_f(v, q) = \left| \tilde{M}_f(v, q) \left[ \tilde{M}_f(0, 0) \right]^{-1} \right| \tag{4}$$

This set of invariant features provides a simple representation of shapes. However, it does not respect the completeness property, i.e. there is no bijection between the dual representations of a single pattern, since the phase information is dropped.

In [20], the following set of rotation and scale invariant features was proposed :

$$I_f(v, q) = \tilde{M}_f(v, q) \left[ \tilde{M}_f(0, 0) \right]^{-1+i\frac{v}{\sigma_0}} \left[ \tilde{M}_f(0, 1) \right]^{-q} \left| \tilde{M}_f(0, 1) \right|^q \tag{5}$$

### 3.3. Properties

Taking Eq. 5, if  $g(\rho, \theta)=f(\alpha\rho, \theta+\beta)$ , it can be easily shown that  $I_g(v, q)=I_f(v, q)$ , thus showing the invariance of the set of descriptors under change of scaling or rotation. Other properties of these *features* rely on:

(i) their completeness : given the quantities  $\tilde{M}_f(0, 0), \tilde{M}_f(0, 1)$  and  $\{I_f(v, q)\}$ , it is possible to return to the whole set of FM coefficients  $\{\tilde{M}_f(v, q)\}$ , and thus to reconstruct  $f(\rho, \theta)$ , by using the inverse AFMT in the following way :

$$f(\rho, \theta) = \sum_q \int_{\mathbf{R}} \tilde{M}_f(v, q) \rho^{\sigma_0+iv} \exp(iq\theta) dv \tag{6}$$

(ii) their convergence : it is proven, in [17] that under the assumption that  $\{\tilde{M}_f(v, q)\}$  is a convergent set, there exists  $x \in \mathbf{R}, x > 1$  such that

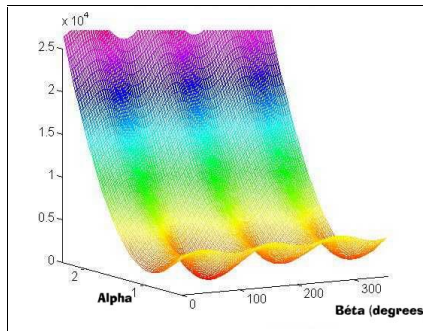
$$\left( \int_{\mathbf{R}} \sum_{\mathbf{Z}} \left| I_f(v, q) \right|^x dv \right)^{1/x} < +\infty \tag{7}$$

One important consequence of the completeness and convergence of this set of invariant features lies in the existence of a metric in the shape representation space, which enables a classification process. Another consequence is the possibility to determine

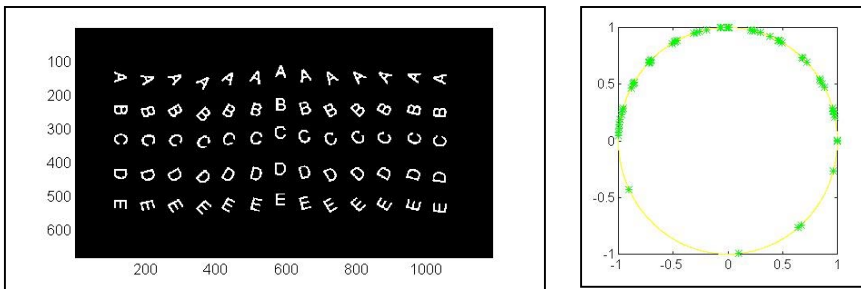
orientation and scale of shapes from the set of descriptors extracted, through a comparison with descriptors extracted from reference shapes. Indeed, from Eq. 3 and through the knowledge of a reference shape, it is possible to obtain values of  $\alpha$  and  $\beta$  ( $\alpha$  being the scale factor and  $\beta$  the angle between the unknown shape and a reference corresponding shape). In order to compute these similitude parameters  $\alpha$  and  $\beta$ , it is for example possible to minimize the quadratic error criteria  $E_{f,g}$ :

$$E_{f,g}(\alpha, \beta) = \int \sum_{v,q} \left| \tilde{M}(v,q) - \alpha^{-\sigma_o+iv} + e^{iq\theta} \tilde{M}_f(v,q) \right|^2 dv \tag{8}$$

On Fig. 1, one can see that the global minimum of the function  $E_{f,g}$  is obtained for correct values of  $(\alpha, \beta)$ . Indeed, in this example, the criteria, calculated with 33 invariant features is minimized for  $\alpha \approx 1$  and  $\beta \approx 100$  which are quite correct results.



**Fig. 1.** Quadratic error calculated between a reference shape and a 100 degrees rotated version.



**Fig. 2.** Left :Set of shapes used for the orientation estimation. Characters of the first column are used as reference shapes. Right: Estimation of the rotation angle in radians of the shapes from column 2 to 13 in comparison with the reference shapes

On the right part of Fig. 2, we present the results which are obtained with this technique for the estimation of the orientation of the different shapes of the left part, by using the first column as reference shapes. One can see 12 “groups” of stars, each of them corresponding to the common orientations of the 5 shapes. These preliminary

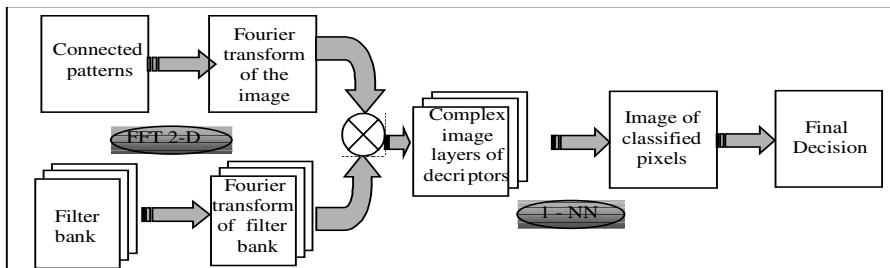
results are very interesting since the estimation of the orientation and the scale of a shape are fundamental information in order to reconstruct strings.

### 3.4. Discussion

In this part, we have recalled the definition of the FMT and its analytic prolongation (AFMT). We have seen that the use of the AFMT should be interesting for multi-orientation and multi-scale character recognition since it is a complete and convergent transformation. In terms of system implementation, this tool can be used within two sequential steps. The first one concerns isolated patterns, which can be easily extracted from a document through the use of a component extractor. The methodology and the obtained results concerning this step are precisely exposed in [22]. Details about application to 2-D square lattice images and obtained results on theoretical and practical data are given in this paper. These results are excellent since classification rates reach 95 % on real data issued from technical documentation and since a comparison with classical approaches shows the superiority of the AFMT methodology [22]. The second step has for objective to detect and recognize characters or symbols which are connected to each other or to graphical parts of a document. This particular point is presented in the next section.

## 4. Application of the AFMT for the Recognition of Connected Shapes on Technical Documents

In this section, we will detail the implementation of the AFMT in the particular case of the recognition of connected patterns. This operation is performed thanks to the possibility of using the AFMT in a filtering mode, as shown in [21] for texture classification. More precisely, the image is first convolved by a set of Fourier Mellin filters and one tries to locate the pixels for which a pre-specified response is obtained and which may be identified as the centroid of a pattern. The scheme for the analysis of images with connected patterns is shown of Fig. 3.



**Fig. 3.** Scheme for the analysis of images containing connected patterns

As one can see on this figure, the first step consists in computing the 2-D Fourier transform of both the image and the set of filters. This technique is used in order to reduce computational burden by changing convolution into multiplication in the Fourier domain. Then, the Fourier transform of the image is multiplied by the Fourier transform of each filter, what enables, after an inverse Fourier Transform applied to the result of this multiplication, to obtain, for each pixel of the initial image, a set of Fourier Mellin descriptors (See Fig. 4: “Complex Image layers of descriptors”). Then, a classification process is applied for each of these vectors in order to assign a class to each pixel by comparison with the responses obtained for isolated characters. Of course, a reject is introduced in this classification process in order to avoid to have too many detection responses. This reject is defined through the use of a confidence value given by the classifier (strategy n°1 on the figure 4). Finally, using responses and confidences given by the classification phase, the system takes a “final decision” for the image. In order to take this final decision, different solutions can be used. Indeed, thanks to the possibility to use response given for each pixel, it is also interesting to take into account spatial information since pixels in the neighborhood of the theoretic centroid of a pattern will also respond with the class of this pattern (strategy n°2 on the figure 4). On the next figure, we give the results obtained on synthetic images with 2 different decision methods. It is important to note that, through the use of the classical isolated patterns approach, 15 images were not recognized. Results shown on this figure are quite logical. Indeed, one can see that the “3” is recognized in most images whereas “8” are logically recognized on some configurations. This is for a major part due to the fast decay of the magnitude of  $h_{p,q}(.,.)$  as the radius from the center point of the filter increases. Significant tests on France Telecom technical documents are currently in progress in order to provide a reliable evaluation of these methods ; final results will be given in the final paper.

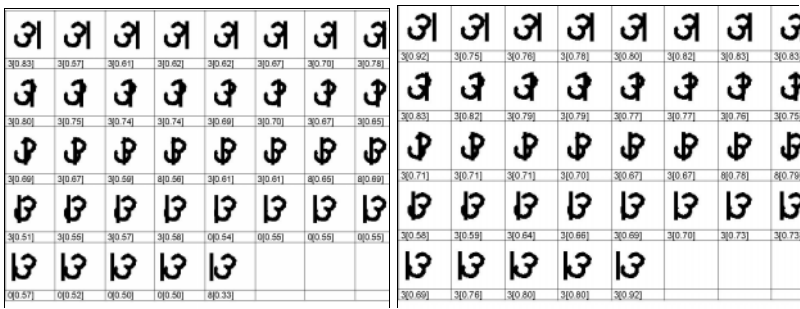


Fig. 4. Classification Result obtained on synthetic images : left part - strategy 1, right part - strategy 2.

### 5. Conclusion and Perspectives

In this paper, we have proposed a original methodology, allowing the detection and recognition of multi-oriented and multi-scaled shapes. Since recognition of isolated



patterns using this methodology have already been proposed in a recent paper [22], we focus our attention in this submission on two particular innovating points. The first one deals with the estimation of characters orientation. From this point, the obtained results are very encouraging and the integration of this tool is currently in progress in order to reconstruct strings of characters. Indeed, on technical documents, “consistent” strings must have the same orientation and the same size. The second point deals with the detection and the recognition of connected patterns. In this submission, the applied process is precisely described and very encouraging preliminary results are shown. The perspectives of this work are numerous and concern different points. First, concerning the orientation estimation approach, a study is currently under way in order to find the optimal optimization methods between classical tools such Genetic Algorithm or “simulated annealing”. Concerning the recognition of connected patterns, we are currently in test in order to choose the optimal decision method. All these considerations and the current results make us very optimistic concerning the future of this project since the results and the possible improvements of the methodology seem to indicate that classification rates should still increase of several points.

## References

1. Boatto L. and al., An interpretation system for land register maps, *IEEE Computer Magazine* **25** (1992), pp 25-33.
2. S.H. Joseph, P. Pridmore, Knowledge-directed Interpretation of Line Drawing Images, *IEEE Trans. on PAMI*, 14(9), pp 928-940, 1992
3. J.M. Ogier, R. Mullot, J. Labiche and Y. Lecourtier, Multilevel approach and distributed consistency for technical map interpretation: application to cadastral maps, *Computer Vision and Image Understanding(CVIU)*, **70**, (1998) pp 438-451.
4. P. Vaxivière, K. Tombre, CELESTIN : CAD conversion of Mechanical Drawings, *IEEE Computer Magazine* **25** (1992), pp 46-54.
5. K. Chhabra. Graphic Symbol Recognition: An Overview, *Lecture Notes in Computer Science*, vol. 1389, pp 68-79, 1998.
6. M.P. Deseilligny, H. Le Men, G. Stamon, Character String Recognition on maps, a rotation-invariant recognition method, *Pattern Recognition Letters*, 16, pp 1297-1310, 1995.
7. D. Dori, L. Wenyin, Automated CAD Conversion with the Machine Drawing Understanding System : Concepts, Algorithms, and Performances, *IEEE Trans. on SMC – Part A*, 29(4), pp 411-416, 1999.
8. O.D. Trier, T. Taxt, A.K. Jain, Features Extraction Methods for Character Recognition – a Survey, *Pattern Recognition*, 29(4), pp 641-662, 1996.
9. M.K. Hu, Visual pattern recognition by moment invariants, *IRE Transactions on Information Theory* **8** (1962), pp 179-187.
10. M. Teague, Image analysis via the general theory of moments, *Journal of Optical Society of America* **70** (1980), pp 920-930.
11. R. Bamieh, R. De Figueiredo, A general moments invariants/attribution graph method for the three dimensional object recognition from a single image, *IEEE Journal of Robotics Automation* **2** (1986), pp 240-242.
12. S.O. Belkasim, M. Shridar, M. Ahmadi, Pattern recognition with moment invariants: A comparative study and new results, *Pattern Recognition* **24** (1991), pp 1117-1138.

13. M. Dai, P. Baylou, M. Najim, A efficient algorithm for computation of shape moments from run-length codes or chain codes, *Pattern Recognition* **25** (1992), pp 1119-1128.
14. S.C. Pei, C. N. Lin, Normalisation of rotationally symmetric shapes for pattern recognition, *Pattern Recognition* **25** (1992), pp 913-920.
15. B.C. Lin, « New forms of shape invariants from elliptic Fourier Descriptors », *Pattern Recognition* **20** (1987), pp 535-545.
16. T. Taxt, J. B. Olafsdotir, M. Daehlen, Recognition of handwritten symbols, *Pattern Recognition* **23** (1990), pp 1155-1166.
17. Mori, C.Y.Suen, K. Yamamoto, Historical review of OCR Research and Development, Proceedings of the IEEE, 80(7), pp 1029-1058, 1992 .
18. N. Kita, Object locating based on concentric circular description, in *Proc. ICPR'92*, Den Hague, Vol. 1 (1992), pp 637-641.
19. S.A. Dudani, K. J. Bredding, R. M. McGhee, Aircraft identification by moment invariants, *IEEE Transactions on Computers* **26**, (1977), pp 39-45.
20. F. Ghorbel, A complete invariant description for gray level images by the harmonic analysis approach, *Pattern Recognition Letters* **15** (1994), pp 1043-1051.
21. G. Ravichandran, M. Trivedi, Circular-Mellin features for texture segmentation, *IEEE Trans. on Image Processing* **4** (1995), pp 1629-1640.
22. S. Adam, J.M. Ogier, C. Cariou, , , R. Mullot, J. Gardes, Y. Lecourtier, “ Symbol and Character Recognition : Application to Engineering Drawings ”, submitted to the International Journal on Document Analysis and Recognition (IJ DAR).