A MULTI-SCALE AND MULTI-ORIENTATION RECOGNITION TECHNIQUE APPLIED TO DOCUMENT INTERPRETATION : APPLICATION TO FRENCH TELEPHONE NETWORK MAPS

S. Adam[1&2], F. Rousseau[1] , J.M. Ogier [1], C. Cariou[3], R. Mullot[4], J. Labiche[1] and J. Gardes[2]

[1] Laboratory PSI, University of Rouen, 76 821 Mont Saint Aignan, France
Sebastien.Adam@univ-rouen.fr
[2] France Telecom R&D, DMI/GRI, 2, avenue Pierre Marzin - 22307 Lannion Cedex - France
Joel.Gardes@rd.francetelecom.fr
[3] LASTI, ENSSAT Lannion, 6, rue Kerampont, BP 447, 22305 Lannion, France
Claude.Cariou@enssat.fr
[4] Laboratory L3i, University of la Rochelle, Avenue Michel Crépeau, 17 042 La Rochelle – France
remy.mullot@univ-lr.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 and the use of a new set of features allowing the classification of multi-oriented and multi-scaled patterns, as well as the rotation and scale parameters estimation. This set of invariants is based on the Fourier-Mellin transform (FMT). This transformation has two interesting properties in the context of our study. The first relies on the excellent classification rate which is obtained with this method. Secondly, it provides a robust estimator of orientation and scale of the processed patterns through the use of the shift theorem of the Fourier transform. A comparison with parameter estimation issued from the Zernike moments is also given.

1. INTRODUCTION

The current improvements of intranet structures allow large companies to develop internal communications between services. The representation of the heritage of such 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 digital 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 recognizing characters and symbols, which often appear on technical documents in several orientations and sizes. Such a constrained pattern recognition problem has been the object of intense and thorough studies. An excellent state of the art describing different possible approaches can be found in [5]. In this communication, 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 herein is the automatic analysis of documents issued from the prime French telephonic operator, *France Telecom*.
The paper will be organized as follows. In the second section, we briefly describe our pattern description tool, based on a set of invariants derived from the Fourier-Mellin transform. In the third section, we explain how these properties can be exploited in order to obtain an estimation of rotation and scale parameters w.r.t. a reference pattern. Indeed, the orientation estimation is an important factor in order to reconstruct strings in a further step of recognition. Then, in the fourth section, we present some results obtained by using our tool, and compare them with another known approach, i.e. Zernike moments. Finally, in the fifth section, we conclude and define some potential perspectives to this work.

2. THE FOURIER-MELLIN TRANSFORM : PROPERTIES AND DERIVED INVARIANTS

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 respect to any combination of rotation and scaling of a pattern, i.e. any geometric positive similitude. 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 a graphic part of the document, leaving any classical pattern recognition technique useless. Finally, the last constraint is the ability to estimate the orientation and scale of a pattern in order to group characters with similar orientations into strings.
The strategy that we propose here covers the three above 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 Fourier-Mellin transform (FMT) properties, which are very interesting for our application. Basically, the technique developed herein is a combined use of the works of Ghorbel [6], Ravichandran and Trivedi [7]. In the following, we first recall the definition of the FMT, its analytic prolongation (AFMT) and a set of complete and stable similitude invariant features, first proposed in [6]. The properties of these invariants will then be recalled.

2.1. 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, as:

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

with $q \in \mathbb{Z}, \quad v \in \mathbb{R}$.

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 $f(r,\theta)$.

2.2. Analytic prolongation of the Fourier-Mellin Transform (AFMT)

In order to alleviate the above difficulty, Ghorbel [6] has proposed the use of the AFMT, defined as:

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

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

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

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

Taking the modulus of both terms in Eq. (3) yields features which are invariant under any rotation of the pattern but not under scaling. Moreover, invariants of this type do not have the completeness property, i.e. there is no bijection between the dual representations of a single pattern, since the phase information is dropped.

In [6], the following set of rotation and scale invariant features was proposed :

$$I_f(v,q) = \tilde{M}_f(v,q)\left[\tilde{M}_f(0,0)\right]^{-1+i\frac{v}{\sigma_0}}\left[\tilde{M}_f(0,1)\right]^{-q}\left|\tilde{M}_f(0,1)\right|^{q} \qquad (4)$$

Then, if $g(r,\theta)=f(\alpha r,\theta+\beta)$, it can be easily shown that $I_f(v,q)=I_g(v,q)$, thus showing the invariance of the set of FMT descriptors under change of scaling or rotation. Other important properties of these features are their completeness and their convergence (see [6] for details). For implementation, a value of $\sigma_0=1$ was taken in the above equations, and the computation of the AFMT was performed without Cartesian to polar conversion by using a filtering scheme [8].

3. EXPLOITATION OF PROPERTIES FOR INTERPRETATION

In this part, we describe the use of the AFMT for technical document interpretation. First, we recall some pre-processing steps that are necessary to recognize shapes on documents. Then we expose recognition results obtained through the use of AFMT, combined with classification tools. Finally, we introduce some theoretical aspects and give results concerning the estimation of rotation and scale parameters with the AFMT.

3.1. Pre-processing steps

Generally, in order to apply the (A)FMT, a unique pattern-invariant and pattern-representative point must be chosen as the center of development (CoD) of the transform onto the pattern. In the particular case of document interpretation, this requires to isolate each pattern on the document image before applying the AFMT, typically at its center of inertia. The process used to extract sole patterns is essentially based on the combination of a local adaptive segmentation and a connected components extractor. These tools are precisely described in [8]. We show on Fig. 1 results obtained by the two steps of binarization and classification of connected components.

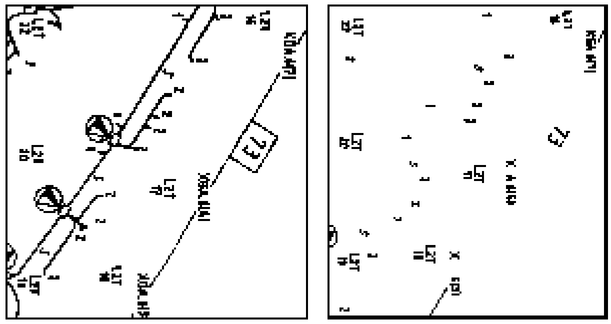

Fig. 1. An original gray scaled image (left part) and the corresponding character layer (right part)

One can note that after these two steps, some characters have been left out. In [8], our approach concerning such inconsistencies is precisely described.

3.2. Recognition results

The strategy that we have adopted for our tests concerning the classification of isolated patterns can be split in two parts.

First, in order to validate the methodology, we applied our technique on a set of "clean" characters, issued from a database especially provided for this application, as can be seen on Fig. 3. The database consisted in 394 training samples and 412 test samples).

In the second part, we considered a set of characters issued from real *France Telecom* documents. The training database was composed of 4890 samples and the testing database was composed of 14878 samples. The classification rate is quite good with regard to our industrial constraints since the correct classification rate reaches 95,74% with the K – nearest neighbors (KNN) approach and 95,11% with the learning vector quantization (LVQ) classifier. A study of the confusion matrix shows that confusions encountered are still due to similar shapes, for example like "B" and "8", "5" and "S". These patterns are sometimes difficult to distinguish because of the noise on the image, which may come from an inadequate binarization. Finally, in order to have an objective point of view about the proposed invariant features, we have compared them with other techniques that are generally considered in the literature as being excellent. Thus we compared the Fourier-Mellin features with the Zernike moments [9] and the circular primitives [10]. The results of these tests, which highlight the superiority of the Fourier Mellin invariants, are presented in [8].

3.3. Scale and rotation parameters estimator

Another important feature of the FMT is the possibility to determine orientation and scale of shapes from the set of descriptors extracted, through a comparison with extracted descriptors from reference shapes. Indeed, from Eq. 3 and through the knowledge of a reference shape, it is possible to obtain values of α and β (α being the scale factor and β the angle between the unknown shape and a corresponding reference shape). In order to compute these similitude parameters (also called *movement parameters*), we compute the Euclidean distance between the two objects in the Fourier Mellin space. If f and g are belong to $L^1(G,d\mu_G)\cap L^2(G,d\mu_G)$, the distance between them in the FM space can be defined as :

$$d_2(f,g) = \sqrt{\sum_{k\in\mathbb{Z}}\int_{-\infty}^{+\infty}\left|\tilde{M}_g(v,q)-\tilde{M}_f(v,q)\right|^2 dv} \qquad (5)$$

Squaring Eq. (5) yields the square error between f and g in the FM space, which for identical patterns is a function of movement parameters :

$$E_{fg}(\alpha,\beta) = \sqrt{\sum_{k\in\mathbb{Z}} \int_{-\infty}^{+\infty} \left| \tilde{M}_g(v,q) - \alpha^{-\sigma_0+iv} \exp(iq\beta)\, \tilde{M}_f(v,q) \right|^2 dv} \quad (6)$$

Thus movement parameters can be extracted by minimizing E_{fg}, by computing :

$$\left(\alpha^*,\beta^*\right) = Arg \min_{(\alpha,\beta)} E_{fg} \; ; \alpha \in \mathbb{R}_+^*, \beta \in [0,2\pi[\quad (7)$$

Nevertheless, the calculation of the AFMT for a finite set of (v,q) values renders the above criterion non convex : several *local minima* exist. So an optimization method where gradient information are not necessary is to be used. We have chosen Nelder-Mead's algorithm [11] in order to perform this optimization, and we have also compared its performance with a genetic algorithm (GA) approach [12]. Note that the number of local minima of the criterion generally equals $q_{\max}$, the maximum order of q which is chosen in the computation of the AFMT.

In Fig. 2, the global minimum of the function E_{fg} is obtained for correct values of (α, β). In this example, the criterion is computed between a reference shape (letter "A") and another "A" with the same size, but rotated by a 100 degrees angle. Indeed, the estimation of movement parameters using 33 invariant features yields $\alpha^* \approx 1$ and $\beta^* \approx 100°$ which are correct results. Also, in this example, we have chosen $q_{\max} = 3$, and yet 3 local minima can be observed for E_{fg} in Fig. 2.

In Fig. 4, we show the results obtained with this technique for the estimation of the orientation on the different patterns presented in Fig. 3, by using the patterns in the leftmost column as the reference ones. 12 clusters of points can be visually identified, each of them corresponding to the common orientations of the 5 patterns, while actually each one is present in 13 different orientations.

Table 1 gives a comparison between the two tested optimization methods on a sample of analyzed patterns. As can be seen, results are quite the same in both cases. However, the choice of the Nelder-Mead algorithm is to be preferred, because the GA approach implies a high computational burden.

It is important to note that identical tests have been performed for scale factor estimation and that the results are essentially the same.

4. COMPARISON OF TWO ROTATION ESTIMATORS

Kim and Kim have proposed, in [13], a method providing an orientation estimator based on Zernike moments (combination of regular moments). In order to validate our approach, we have compared this method with our approach based on the AFMT.

4.1. Brief description of the method

Zernike polynomials are orthogonal polynomials which are often used for multi-oriented and multi-scaled pattern recognition. They are given by :

$$Z_{nm} = \frac{(n+1)}{\pi} \sum_{\text{unit}} \sum_{\text{circle}} V_{nm}^*(r,\theta) f(r,\theta) \quad (8)$$

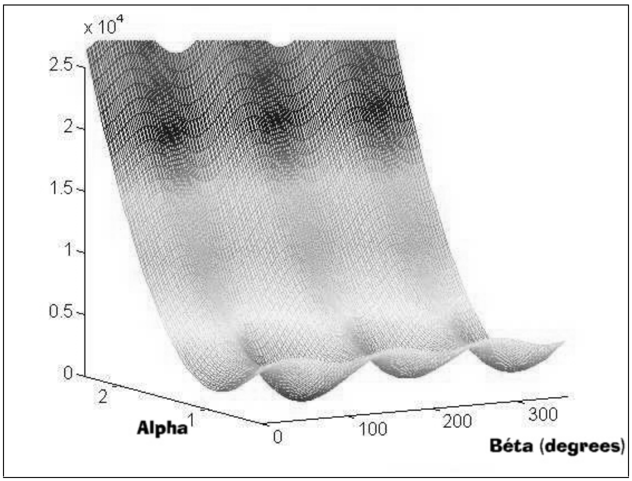

Fig. 2. Square error function calculated between a reference pattern (letter "A") and another "A" with the same size, and rotated by a 100 degrees angle.

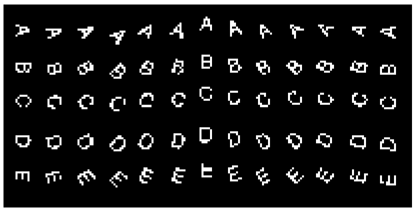

Fig. 3. Set of shapes used for the orientation estimation. Characters of the first column are used as reference shapes

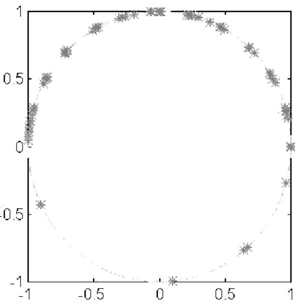

Fig. 4. Estimation of the rotation angle in radians of the shapes from column 2 to 13 in comparison with the reference shapes of the first column.

TABLE 1: COMPARISON OF TWO OPTIMIZATION PROCEDURES FOR SCALE AND ORIENTATION ESTIMATION (α / β).

Character	True orientation	Nelder-Mead	GA
E	0	1 / 9.8 E-6	0.99 / 359
A	0	1 / -1.26 E-5	0.95 / 1.47
C	0	1 / -2.77 E-7	1.00 / 2.72
E	15	0.49 / 12.15	0.49 / 12.72
A	45	1.01 / 46.56	1.01 / 47.23
A	60	0.99 / 61.40	1.02 / 63.83
C	90	0.98 / 93.74	0.98 / 93.32
A	90	1.01 / 90.74	1.01 / 90.78

With

$$V_{nm}(r,\theta) = R_{nm}(r)\exp(im\theta) \quad (9)$$

$$R_{nm}(r)=\sum_{s=0}^{(n-|m|)/2}(-1)^s\frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}+s\right)!}r^{n-2s} \quad (10)$$

In order to obtain rotation invariance, it is necessary to use only the magnitude of these values. In such a case, if

$$g(r,\theta)=f(r,\theta+\beta) \quad (11)$$

then

$$Z_{nm}^{(g)}=Z_{nm}^{(f)}\exp(im\beta), \quad (12)$$

Indicating that β can be estimated.
For more details, we strongly encourage the reader to read [13].

4.2. Experimental results

The results obtained with this method are also quite good since error on the estimation of the orientation never exceeds 10%. In Table 2 are given compared estimates of orientation between the two different approaches for the analysis of a character "A". Actually, results are quite similar in most cases, although a slightly better mean error is obtained with the FM approach. Our final objective with this estimation being to reconstruct strings of characters, errors (which are always less than 10 %) are small enough regarding our constraints.
Obviously, further experiments must absolutely be led in order to validate the approach. The validation of this kind of technique on a consequent data base is a very heavy operation since it requires to constitute a data base for which each sample is labeled with its orientation. Two alternatives are possible when dealing with this kind of problem : the first one consists in constituting a synthetic data base, for which the data are rotated automatically. This approach is being led in our laboratory. The second consists in working on a real data set and in labeling each sample manually. This approach is the most reliable technique for the validation of the approach, but raises the problem of the assignment of a precise orientation to each shape in the labeling stage. As far as we know, on a real data set, there is no reference dealing with this kind of problem. In this domain, our current works deal with an estimation based on the fusion of different contextual information.

5. CONCLUSION

In this communication, we have proposed an original methodology, allowing the detection and recognition of multi-oriented and multi-scaled patterns. This methodology is based on the Fourier-Mellin transform and the supports on which it was applied are technical documents representing the network of the French Telephone operator (*France Telecom*). Since recognition of isolated patterns using this methodology have already been presented in a recent paper [8], we have focused here on a particular innovating point concerning the estimation of character orientation. Results obtained on this point are very. 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. From another point of view, we are currently working at the integration of contextual information within the system approach in order to validate the obtained estimation of orientation. Indeed, as an example, a set of aligned characters provides some information concerning a string of characters, and as a consequence, can validate the final result. 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 a better interpretation of technical documents should be achieved. Furthermore, all these aspects concerning string and symbol recognition constitute an interesting alternative to the growing needs concerning document indexing and content based information retrieval. Indeed, textual and symbolic information are among the most important cues for document indexing.

TABLE 2: COMPARISON OF ORIENTATION ESTIMATORS

True orientation	FM estimate	Zernike estimate
0	0	0
15	16	16
30	30	23
45	47	48
60	61	61
75	78	76
90	91	88
105	105	107
120	118	122
135	136	137
150	148	153
165	164	167
180	175	175
Mean error	**1,46**	**2,38**

6. BIBLIOGRAPHY

[1] Boatto L. *et 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) (1992), pp. 928-940.

[3] J.-.M. Ogier, R. Mullot, J. Labiche, 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] O.D. Trier, A. K. Jain, T. Taxt, Features extraction methods for character recognition – a survey, *Pattern Recognition* **29** (1996), pp. 641-662.

[6] F. Ghorbel, A complete invariant description for gray level images by the harmonic analysis approach, *Pattern Recognition Letters* **15** (1994), pp. 1043-1051.

[7] G. Ravichandran, M. Trivedi, Circular-Mellin features for texture segmentation, *IEEE Trans. on Image Processing* **4** (1995), pp. 1629-1640.

[8] C. Cariou *et al.*, A multi-scale and multi-orientation recognition technique applied to document interpretation : application to French telephone network maps, *Int. J. of Pattern Recognition and Artificial Intelligence (IJPRAI)* **13**(8) (1999), pp. 1201-1218.

[9] M. Teague, *Image analysis via the general theory of moments,* Journal of Optical Society of America, **70** (1980), pp 920-930.

[10] M. P. Deseilligny, H. Le Men, G. Stamon, «Character string recognition on maps, a rotation-invariant recognition method », Pattern Recognition Letters, **16**, (1995), pp. 1297-1310.

[11] J. A. Nelder and R. Mead, A simplex method for function minimization, *Computer Journal* **7** (1965), pp. 308-313.

[12] D. Zongker and A. Jain, *Algorithms for Feature Selection : An Evaluation*, Proceedings of ICPR 96, **Vol II (**1996), pp 18-22.

[13] W.-Y. Kim and Y.-S. Kim, *Robust rotation angle estimator, IEEE Trans. on PAMI* **21**(8) (1999) pp. 768-773.