

# Reconnaissance de Formes en Multi-Orientation et Multi-Echelle sur Documents Techniques

Claude CARIOU<sup>1</sup>, Jean-Marc OGIER<sup>2</sup>, Sébastien ADAM<sup>2</sup>, Joël GARDES<sup>3</sup>, Rémy MULLOT<sup>2</sup>, Yves LECOURTIER<sup>2</sup>

<sup>1</sup>LASTI - Groupe Image – ENSSAT  
3, rue de Kerampont, BP 447, 22305 Lannion Cedex, France

<sup>2</sup>PSI-La3I – Université de Rouen  
76821 Mont Saint Aignan, France

<sup>3</sup>France Télécom-DVSI DES/MBL  
6, avenue des Usines, BP 383, 90007 Belfort Cedex, France

Claude.Cariou@enssat.fr, Jean-Marc.Ogier@univ-rouen.fr

**Résumé** – Nous considérons le problème de la reconnaissance de formes en imagerie dans le cas où les structures à reconnaître sont disponibles dans une orientation et à un facteur d'échelle quelconques. Plus précisément, nous décrivons un outil mis au point pour la rétroconversion automatique de documents techniques, archivés sous forme « papier », du réseau téléphonique géré par France Télécom. La technique employée, qui s'appuie sur la transformée de Fourier-Mellin (TFM), est mise en œuvre sous la forme d'un banc de filtres 2-D. Deux phases de reconnaissance sont proposées. La première s'adresse à des symboles et caractères préalablement identifiés comme isolés. La seconde phase permet la reconnaissance de formes superposées. Dans les deux modes de fonctionnement, un jeu de descripteurs invariants par similitude est calculé et appliqué à un classifieur  $K$ -PPV. Nous présentons des résultats de classification sur une base importante de symboles et de caractères, qui mettent en évidence la supériorité de notre approche sur celle des moments de Zernike.

**Abstract** – The problem of rotation- and scale-invariant pattern recognition is addressed. Specifically, we describe a tool which is designed for the electronic archiving of the French telephone network held by France Télécom, and available on technical paper drawings. The adopted technique uses the Fourier-Mellin transform (FMT) in a filter bank scheme within two steps : the first one tries to recognize isolated characters and symbols, while the second allows the recognition of overlapping patterns. A similitude-invariant features vector is computed in both steps, and is applied to a  $K$ -nearest neighbour classifier. From the analysis of an important image data base, classification rates are given and compared with those provided by Zernike moments, showing the superiority of our approach.

## 1. Introduction

Le traitement automatique de documents techniques pour en produire un archivage numérique est une tâche considérable, non seulement de par la difficulté de passer d'un support papier à un support vectorisé et interprété ; mais également du fait de l'énorme quantité de plans techniques archivés et mis à jour régulièrement. Par exemple, le CNET - France Télécom estime son patrimoine à plus de 2.5 millions de plans relatifs au réseau téléphonique français.

Une interprétation correcte d'un document, quelle que soit la stratégie et la représentation des connaissances employée, ne saurait être efficace sans l'apport de techniques de bas niveau qui soient robustes et de bonne qualité. L'objectif visé est alors de traiter des images numériques (obtenues par scannérisation des plans) en vue de fournir à un plus haut niveau d'abstraction des informations pertinentes. Parmi les techniques de bas niveau, on trouve, dans un ordre croissant de complexité de traitement :

- le filtrage (souvent la suppression de bruit impulsif)
- la segmentation (souvent la binarisation)

- la reconnaissance de symboles caractéristiques et/ou de caractères alphanumériques (parfois sous contrainte d'invariance géométrique).

Dans cette communication, nous présentons une approche de ce dernier point, dans le contexte de l'interprétation de plans du réseau téléphonique. Les spécifications en sont les suivantes :

- *La reconnaissance doit être invariante par rotation et changement d'échelle des symboles et caractères (i.e. par similitude).* En effet, de nombreuses chaînes de caractères et de symboles sont présentes en de multiples orientations et en plusieurs tailles.
- *La reconnaissance doit être robuste vis à vis du placement des formes.* Le recouvrement ou l'interconnexion de symboles doit pouvoir être analysé.

La Figure 1 montre un extrait de document technique où des symboles multi-orientés et des formes interconnectées apparaissent.

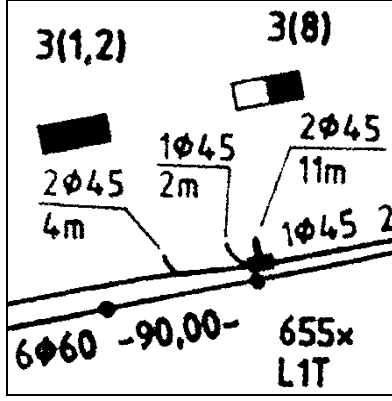


FIG. 1 : extrait d'un plan du réseau

Nous décrivons dans la suite respectivement l'outil développé pour la reconnaissance de symboles et de caractères invariante par similitude, puis les résultats obtenus dans l'analyse des documents techniques de France Télécom.

## 2. TFM, invariance, et mise en oeuvre

L'outil de reconnaissance développé pour cette application repose sur la transformée de Fourier-Mellin (TFM), grâce à laquelle un jeu de descripteurs invariants par similitude peut être dérivé. Plus précisément, on utilise le prolongement analytique de la TFM, défini dans [1], dont la mise en oeuvre, inspirée de [2], requiert pour l'analyse d'images l'emploi d'une structure de banc de filtres RIF.

### 2.1 TFM et prolongement analytique

Soit  $f(\rho, \theta)$  une fonction (la forme) à valeurs dans  $\mathbf{R}$ , exprimée en coordonnées polaires. La TFM est définie comme la transformée de Fourier sur le groupe des similitudes positives  $\mathbf{R}_+^* \times S^1$  [1] :

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

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

Cette intégrale ne converge pas en général, et c'est pourquoi il a été proposé dans [1], l'utilisation de la transformée de Fourier-Mellin analytique (TFMA) définie par :

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

avec  $q \in \mathbf{Z}$ ,  $\theta \in \mathbf{R}$ , et  $\sigma_0 \in \mathbf{R}_+^*$

Le théorème du retard, classique en analyse de Fourier, peut être appliqué à la TFMA : si une forme  $f(r, \theta)$  subit une similitude de paramètres  $(\alpha, \varphi)$ , c'est-à-dire si  $g(\rho, \theta) = f(\alpha\rho, \theta + \varphi)$ , alors on a la relation suivante :

$$\tilde{M}_g(v, q) = \alpha^{-\sigma_0 + iv} e^{iq\varphi} \tilde{M}_f(v, q) \quad (3)$$

### 2.2 Descripteurs invariants

Des descripteurs invariants par rotation de la forme peuvent être obtenus en prenant le module des deux termes dans (2). Toutefois, ces descripteurs ne sont pas invariants par homothétie. Afin de garantir d'une part l'invariance par similitude, la complétude, et la convergence de la description, Ghorbel [1] a proposé le jeu de descripteurs suivants :

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

Avec cette définition, on peut montrer que si  $g(\rho, \theta) = f(\alpha\rho, \theta + \varphi)$ , alors  $I_f(v, q) = I_g(v, q)$ . Les deux propriétés importantes de ces invariants sont les suivantes :

- (i) *complétude* : étant données les quantités  $\tilde{M}_f(0, 0)$ ,  $\tilde{M}_f(0, 1)$  et les fonctions  $\{I_f(v, q)\}$ , il est possible de retrouver l'ensemble des fonctions  $\{\tilde{M}_f(v, q)\}$ , et donc de reconstruire la forme  $f(\rho, \theta)$ , en utilisant la TFMA inverse, définie par :

$$f(\rho, \theta) = \sum_q \int_{\mathbf{R}} \tilde{M}_f(v, q) \rho^{\sigma_0 + iv} e^{iq\theta} dv \quad (5)$$

- (ii) *convergence* : il a été montré que, sous la condition que  $\{\tilde{M}_f(v, q)\}$  soit un ensemble convergent, il existe  $p \in \mathbf{R}$  t.q. la norme  $N_p$  sur les invariants soit convergente.

Une conséquence importante de la complétude et de la convergence de ce jeu d'invariants est l'existence d'une métrique dans l'espace des formes.

### 2.3 Mise en oeuvre sur les images numériques

Plusieurs problèmes sont soulevés lors de la mise en oeuvre de la TFMA sur des images numériques. Le premier provient de la nécessité d'échantillonner la quantité  $\tilde{M}_f(v, q)$ . Une solution simple consiste à échantillonner la variable  $v$  et à calculer  $\{\tilde{M}_f(p, q)\}$  et  $\{I_f(p, q)\}$  pour  $p \in \mathbf{Z}$ .

Une autre difficulté majeure est liée à la structure en réseau carré des images numériques usuelles, et qui ne permet pas la mise en oeuvre directe de la TFMA. Deux options sont alors disponibles pour le traitement des images : l'interpolation [1] ou l'approximation de la TFMA en coordonnées cartésiennes [2]. Avec cette dernière option, il n'est pas nécessaire d'interpoler la forme sur le réseau 2-D ; en effet, on peut remarquer que la TFMA, calculée au centre de gravité  $(k_0, l_0)$  de la forme peut être approchée comme suit :

$$\tilde{M}_f(p,q) \approx \sum_{1 \leq (k^2 + l^2) \leq \rho_{\max}^2} h_{p,q}(k,l) f(k - k_0, l - l_0) \quad (6)$$

où

$$h_{p,q}(k,l) = \frac{e^{-i \left[ \frac{p}{2} \ln(k^2 + l^2) + q \tan^{-1} \left( \frac{l}{k} \right) \right]}}{(k^2 + l^2)^{1 - \frac{\sigma_0}{2}}} \quad (7)$$

Cette formulation montre que la convolution, à son centre de gravité, de la forme avec le filtre de réponse impulsionnelle  $h_{p,q}(\cdot, \cdot)$  donne une approximation du coefficient d'ordre  $(p, q)$  de la TFMA. On peut remarquer que tous les filtres, sauf  $h_{0,0}(\cdot, \cdot)$ , sont à coefficients complexes. D'autre part, ces filtres sont orthogonaux par rapport à la variable circulaire  $q$ , mais pas par rapport à la variable radiale  $p$ . Toutefois, une orthogonalisation de la base complète peut être menée [4]. Enfin, notera la possibilité de paralléliser le calcul des coefficients de la TFMA.

Une fois calculée la TFMA d'une forme donnée par le banc de filtre appliqué à son centre de gravité, on extrait des descripteurs invariants grâce à la relation (4). On peut montrer que ces descripteurs sont tous à valeurs complexes, sauf  $I_f(0,0)$  qui vaut 1 et  $I_f(0,1)$  qui est réel pur. De plus, les invariants possèdent la propriété de symétrie hermitienne.

### 3. Résultats expérimentaux

Nous présentons ici quelques résultats de l'application de la technique exposée plus haut sur des images issues de documents techniques. En particulier, nous mettons tout d'abord en évidence les propriétés discriminantes des invariants sur une base de caractères calibrés, puis nous décrivons la technique adoptée et les résultats de reconnaissance pour les formes isolées (bien conditionnées), puis pour les formes interconnectées.

#### 3.1 Base de caractères calibrés

La Figure 2 est une image numérique représentant 5 caractères, chaque caractère étant donné dans 13 orientations différentes. Ces formes ont été analysées par le banc de filtres  $\{h_{p,q}(\cdot, \cdot)\}$ , avec  $p \in \{-2, \dots, 2\}$  et  $q \in \{0, \dots, 3\}$  et  $\sigma_0 = 1$ . Ce faisant, 33 paramètres non redondants ont été extraits pour chaque forme. Les mêmes formes ont également été analysées à un facteur d'échelle de 0.5, simplement en sous-échantillonnant l'image originale (1 pixel sur deux dans chaque dimension). La Figure 3 correspond aux paramètres invariants calculés pour chaque forme. Par exemple, le premier diagramme représente la superposition des 33 paramètres obtenus pour chacune des 26 formes « 0 » analysées.

Cette première expérimentation montre clairement la pertinence des paramètres extraits à l'aide de la TFMA : la dispersion des paramètres est relativement faible en moyenne, conditionnellement à chaque classe de caractères.

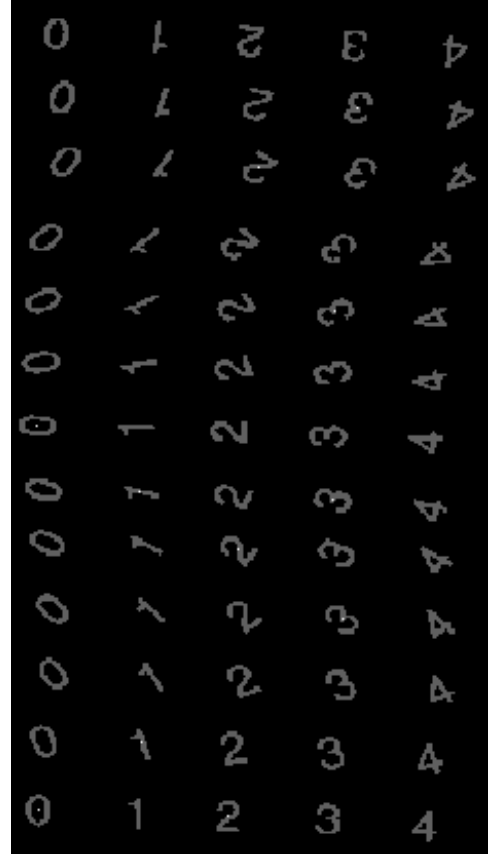


FIG. 2 : caractères multi-orientés

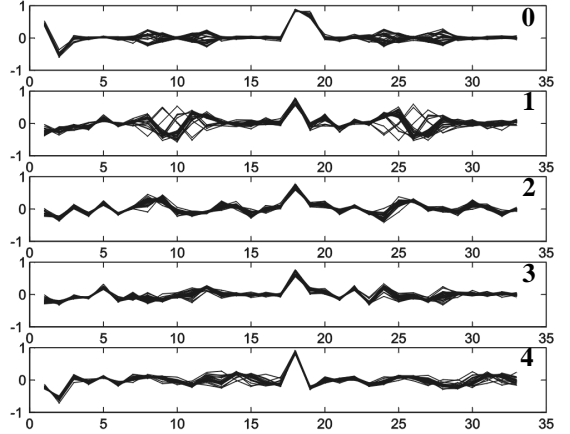


FIG. 3 : paramètres invariants par similitude

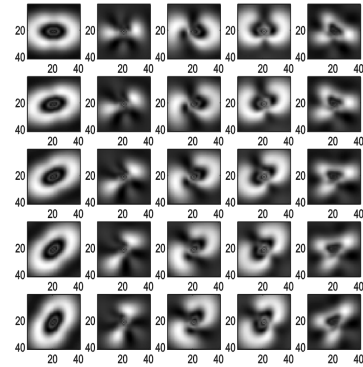


FIG. 4 : reconstruction des caractères

La Figure 4 montre la reconstruction des formes, à partir d'un jeu de 119 paramètres invariants, ce qui correspond à la gamme de variation  $p \in \{-5, \dots, 5\}$  et  $q \in \{0, \dots, 5\}$ .

À partir d'une petite base de caractères calibrés, nous avons pu mesurer le taux de bonne classification, dans les conditions d'analyse énoncées au début de cette section, et le comparer à celui obtenu par une description des formes par les moments de Zernike, autre technique bien connue pour la description de formes à niveaux de gris. La stratégie de classification présentée ici pour les caractères calibrés comme pour les caractères réels s'appuie sur l'algorithme des  $K$  plus proches voisins ( $K$ -PPV). Dans cette étude, nous avons pris  $K = 1$ . D'autres tests basés sur une approche connexionniste de type LVQ (*Learning Vector Quantization*) ont donné des résultats très similaires en termes de taux de classification. Dans le cas présent, la base d'apprentissage comprenait 394 échantillons, et la base de test en comprenait 412. Les résultats, présentés dans la Table 1, montrent la nette supériorité de l'approche proposée.

### 3.2 Base de caractères réels

Une importante série de tests sur des cas réels, issus des documents de France Télécom, a été mise en place de façon à évaluer la qualité de notre approche. Dans cette étude, le nombre de classes considérées était de 51. Nous avons intégré le fait que certaines classes peuvent regrouper des caractères sémantiquement différents, mais identiques à une similitude près : par exemple, avec les polices de caractères utilisées sur les documents techniques, les caractères « b » et « q » appartiennent à la même classe d'équivalence.

#### 3.2.1 Formes isolées

L'objectif de ce premier test était de déterminer le taux de bonne reconnaissance sur des formes « claires » issues des documents techniques, c'est-à-dire non connectées, ni entre elles, ni avec d'autres structures. Pour cela, une base comprenant au total 10000 caractères différents en multi-orientation et multi-échelle a été étiquetée. 4000 de ces caractères ont servi à constituer la base d'apprentissage, les 6000 autres étant réservés au test.

Les résultats de classification sont donnés dans la Table 1. Notre approche a été à nouveau comparée avec celle des moments de Zernike. Les taux de classification obtenus montrent le bon comportement des invariants issus descripteurs de Fourier-Mellin, à la fois par rapport aux moments de Zernike et par rapport aux contraintes industrielles imposées. Les confusions sont presque toujours dues à la mauvaise classification de formes similaires (« B » et « 8 », « S » et « 5 »).

TAB. 1 : Taux de reconnaissance sur base calibrée et réelle

Base	moments de Zernike	Invariants $I_f(p, q)$
calibrée	81	98
réelle	78	88

#### 3.2.2 Formes interconnectées

Le traitement des formes interconnectées peut être mené sur la même base que pour les formes isolées. La seule différence réside dans la stratégie de reconnaissance. En effet, si pour les formes isolées le calcul des invariants est effectué au centre de gravité de la forme, cette dernière notion n'a plus de sens pour les caractères ou symboles se chevauchant. C'est pourquoi il faut envisager une reconnaissance sur la base d'une *détection* de formes connues, autrement dit un *filtrage adapté*. Pour cela, on procède dans un premier temps au calcul des invariants en chaque pixel de l'image. Cette étape est facilitée par l'utilisation, dans le domaine fréquentiel, des filtres définis précédemment, via la FFT-2D. Ensuite, le vecteur d'invariants en chaque pixel est envoyé au classifieur ( $K$ -PPV par exemple), qui propose une classe de forme en ce pixel. La Figure 5 montre un exemple de détection d'une forme connue qui est initialement « collée » à une structure linéaire.

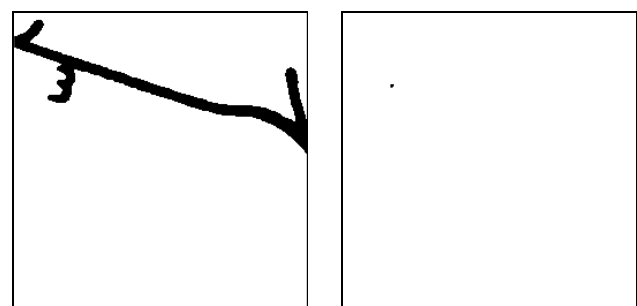


FIG. 5 : reconnaissance de caractères connectés  
(a) image originale  
(b) détection du centre de gravité de la forme « 3 »

## 4. Conclusion

Dans cette communication, nous présentons une procédure de reconnaissance de caractères et de symboles s'appuyant sur les propriétés de la transformée de Fourier-Mellin (TFM) appliquée aux formes en multi-orientation et multi-échelle. Dans le cadre d'une application en rétroconversion de documents techniques, nous démontrons qu'une mise en œuvre sous la forme d'un banc de filtres RIF, ne nécessitant ni reparamétrisation en coordonnées polaires, ni interpolation de l'image, conduit à des résultats qui, en termes de taux de reconnaissance, sont meilleurs que ceux obtenus grâce aux moments de Zernike, sur des bases de caractères calibrés, mais également sur des cas réels. De plus, nous montrons que cette approche permet de reconnaître avec succès des formes interconnectées.

## Références

- [1] F. Ghorbel, « A complete invariant description for gray level images by the harmonic analysis approach », *Pattern Recognition Letters* **15**, pp.1043-1051 (1994).
- [2] G. Ravichandran, M. Trivedi, « Circular-Mellin features for texture segmentation », *IEEE Trans. Image Proc.* **4**(12), pp. 1629-1640 (1995).