Chapitre 1

Introduction

1.1 Présentation de la reconnaissance des formes

La reconnaissance des formes (ou RdF) est issue de différentes disciplines qui sont les mathématiques (probabilités et statistiques), les sciences de l'ingénieur, l'informatique et l'intelligence artificielle. C'est à partir des années 60 que la reconnaissance des formes est devenue une discipline spécifique. L'extraordinaire développement des ordinateurs ces dernières années a donné un élan à la RdF en permettant des applications temps réel, en particulier dans le domaine des applications visuelles et auditives. Les procédés d'acquisition tels que caméra, scanner sont très accessibles, ainsi que des ordinateurs à la fois puissants et bons marchés. Ils permettent le traitement de nombreuses données en un temps raisonnable comme cela est souvent nécessaire en RdF.

L'objectif de la RdF est de réaliser des systèmes informatisés qui simulent les activités humaines de perception, de reconnaissance et de compréhension : reconnaissance de l'écrit, de la parole, interprétation de scènes, robotique, reconnaissance de signaux médicaux EEG (électroencéphalogramme), ECG (électrocardiogramme). Cela implique aussi une certaine pluridisciplinarité pour comprendre l'aspect physique des capteurs, les aspects mathématiques de la classification, ceux relatifs à l'informatique.

Les systèmes de reconnaissance des formes intègrent toute la chaîne perception-reconnaissance depuis l'acquisition des données brutes jusqu'à la compréhension élaborée de ces données. Ces dernières ont entre temps subi de nombreuses transformations. De ce fait, la RdF fait appel à des disciplines connexes telles que le traitement du signal et de l'image, l'intelligence artificielle ou le traitement automatique des langues (TAL). Prenons l'exemple de la lecture d'images de textes imprimés : l'OCR (Optical Character Recognition). Il ne s'agit pas seulement de reconnaître des caractères mais aussi de segmenter l'image en zones, sélectionner les zones de texte, découper celles ci en lignes, mots et caractères. Ces analyses font appel au traitement du signal. Après la reconnaissance proprement dite, celle ci peut être améliorée en utilisant des modèles de langage.

1.2 Domaines d'application

Parmi les domaines les plus populaires de la RdF on trouve l'interprétation d'images aériennes ou satellites conduisant à la surveillance ou aux prévisions agricoles, celles des images et signaux médicaux pour des tâches de comptage ou de repérage de cellules ou d'événements anormaux, la détection de défauts (pièces industrielles, industrie alimentaire, ...). Sans être exhaustifs, citons aussi d'autres domaines récents et très actifs : la reconnaissance des gestes, de l'attitude et même des émotions à partir de séquences vidéos ou audios, la reconnaissance des opinions à partir de tweets.

Parmi les domaines les plus populaires de la RdF on trouve la reconnaissance de l'écriture et de la parole. En ce qui concerne l'écrit, on pense souvent à l'OCR (Optical Character Recognition) qui est la reconnaissance des caractères dans les textes imprimés. On parle de système monofonte, multifonte, omnifonte selon que le système traite une, plusieurs ou n'importe quelle police de caractères On trouve dans le commerce de bons systèmes OCR multifontes qui nécessitent toutefois des documents de bonne qualité pour atteindre les taux de reconnaissance affichés. La reconnaissance des adresses postales, montants de chèques, formulaires et courriers manuscrits sont des applications industrielles importantes qui traitent l'écriture manuscrite non contrainte. Le vocabulaire doit cependant être limité. La variabilité intra-scripteur et inter-scripteurs de l'écriture sont un défi majeur pour la reconnaissance. L'écriture manuscrite en ligne est obtenue à partir d'un stylet adapté et d'une tablette graphique. Des paramètres de vitesse, pression du mouvement de l'écriture sont alors enregistrés pour la reconnaissance des mots ou l'authentification [1] de scripteurs.

Pour le traitement de la parole, on parlera de même de systèmes mono-locuteur si celui est adapté à une personne donnée, multi-locuteurs s'il est adapté à plusieurs personnes connues du système ou omni-locuteurs pour les usages tout public. Ici encore, le degré de difficulté est différent suivant que l'on traite des mots énoncés séparément ou de la parole continue où le défi est justement de segmenter les données en mots. Les systèmes de sécurité utilisent la voix pour identifier ou authentifier des locuteurs. D'autres paramètres biométriques comme les empreintes digitales, la signature, les images (2D, 3D) de la main et du visage, sont également utilisés.

1.3 Schéma général d'un système de reconnaissance

Une chaîne de traitement dans un système de reconnaissance comprend plusieurs modules (Figure 1.1) et plusieurs espaces de travail. L'objectif de la reconnaissance des formes va être de définir une suite d'opérations permettant de passer de l'espace des données ou formes, à l'espace des classes où la catégorie de la forme est estimée. Ces opérations sont en pratique des procédures informatisées.

Avant d'estimer la classe d'une forme, celle-ci doit être numérisée puis représentée.

1. On distingue l'identification de l'authentification. Lors d'un processus d'authentification, l'individu clame son identité et le système vérifie que l'individu n'est pas un imposteur.

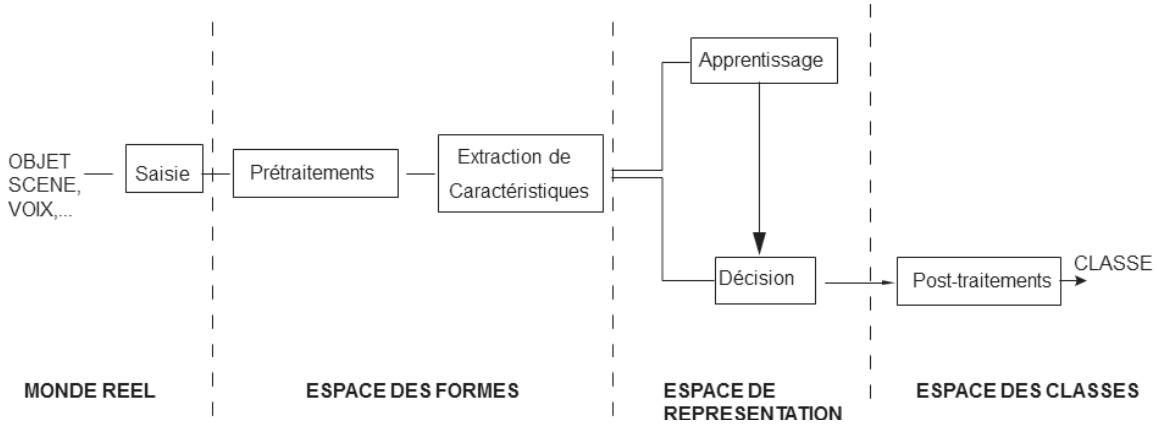

FIGURE 1.1 – Schéma général d'un système de reconnaissance des formes.

Les représentations peuvent être diverses : vecteurs ou séquences de vecteurs, arbres ou graphes. La fonction de décision peut être construite à partir de représentations extraites sur des formes issues d'un ensemble d'apprentissage : c'est la phase dite d'apprentissage. Une décision peut alors être prise pour une forme inconnue, dont on a extrait au préalable sa représentation.

1.4 Système d'acquisition et prétraitements

Suivant la nature du signal, un capteur (caméra, microphone, ...) est nécessaire pour acquérir le signal [2] sous forme numérique ou analogique (il faut alors le convertir en numérique) pour qu'il soit traitable par un système informatisé. On passe ainsi du monde réel au monde des formes ou données.

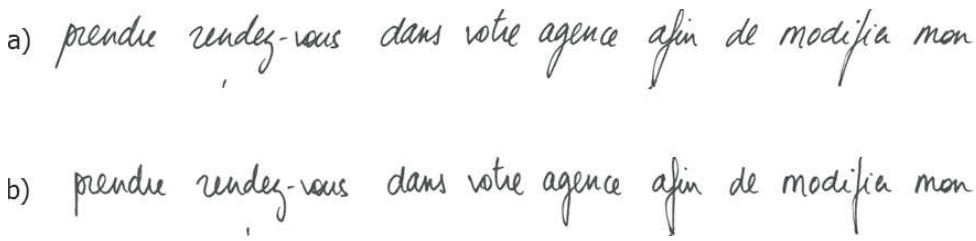

FIGURE 1.2 – Exemple de prétraitement effectué sur une image de ligne de texte manuscrite : correction de l'inclinaison des traits verticaux.

Les prétraitements sont spécifiques à un domaine et ont pour utilité de réduire les bruits de capteurs ou inhérents au signal. Ils peuvent servir aussi de préparation aux phases suivantes en réduisant la variabilité du signal. La Figure 1.2 montre le résultat d'un prétraitement servant à redresser des mots cursifs dans une ligne de texte.

2. Le terme signal désigne ici aussi bien un signal mono-dimensionnel comme la parole, qu'une image ou un signal multi-dimensionnel comme les images multi-spectrales

1.5 Extraction de caractéristiques

La fonction de décision d'un système de reconnaissance des formes repose largement sur les données. Ces données sont des mesures ou observations, symboliques ou numériques, extraites des formes . Par exemple une température, une longueur, la luminance reçue par un capteur, un voltage,etc...Bien choisir ces caractéristiques et en prendre un nombre suffisant est vital pour développer un bon système de reconnaissance.

Le système de reconnaissance va utiliser ces caractéristiques, en nombre d, pour déterminer la classe d'une forme inconnue. Le système considère K classes différentes. Celles ci sont par exemple des types d'animaux, des lettres, la présence ou absence d'une tumeur, etc. Dans le graphique de la Figure 1.3, $K = 3$, $d = 2$ et chaque classe est représentée dans l'espace de représentation (c-à-d l'espace des caractéristiques) par une forme de point différente. Les peuvent être aussi représentées par des couleurs de points différentes.

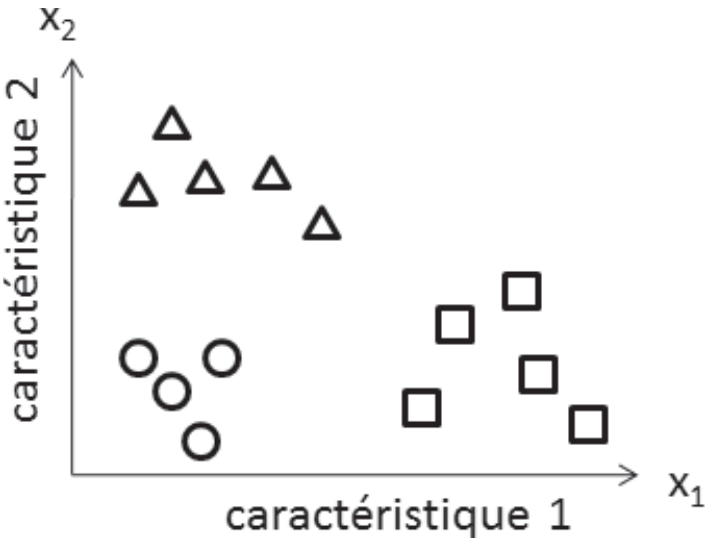

FIGURE 1.3 – Caractéristiques extraites sur des formes appartenant à 3 classes dans un espace de représentation bi-dimensionnel. $d = 2$ et $K = 3$.

Les caractéristiques peuvent être entachées de bruit ou d'erreurs, elles peuvent être aussi continues ou discrètes. Elles sont choisies pour résoudre un problème précis et conduire à une erreur de classement faible. Ceci sera possible si les caractéristiques sont discriminantes : c-à-d si elles varient peu parmi les formes d'une classe donnée et diffèrent autant que possible entre formes de classes distinctes.

Les caractéristiques doivent aussi être choisies afin d'éviter les prétraitements. Le concepteur cherchera à extraire des caractéristiques de type RST [3] , c-à-d invariantes à la rotation, l'échelle et la translation . Il prendra aussi en compte les déformations les plus probables des formes et leurs répercussions dans la chaîne de traitement.

L'objectif de la reconnaissance des formes est de regrouper les formes en catégories à partir de leur caractéristiques communes. On peut par exemple décrire le chiffre 3 par : forme possédant 2 boucles non fermées l'une au dessus de l'autre. Cette description par caractéristiques, ici relative à la structure de la forme, est aussi une réduction

3. RST pour *Rotation, Scale, Translation*

notable d'information par rapport à l'image du chiffre elle même. De nombreuses possibilités sont offertes au concepteur du système pour le choix des caractéristiques et de leur représentation :

– de type statistique, en extrayant un vecteur de caractéristiques x qui consiste en différentes mesures de type numériques extraites de manière systématique sur les formes. Un tel vecteur peut s'écrire :

$$x = [x_1, x_2, ..., x_d]^T. \tag{1.1}$$

où les $x_i, i = 1, \ldots, d$ sont les caractéristiques [4], x_i étant la ième caractéristique de la forme représentée par x. L'indice d correspond à la dimension de l'espace de représentation (ou espace de caractéristiques). $d = 2$ est bien pratique pour l'affichage des données, mais en réalité la plupart des systèmes de reconnaissance utilisent des vecteurs où $d \gg 2$.
Une séquence de tels vecteurs peut aussi être extraite en balayant la forme de de manière ordonnée avec une fenêtre glissante par exemple.

– de type structurel : en recherchant à décomposer la forme en constituants élémentaires appelés primitives. La représentation est alors une ordonnée de primitives. Les primitives peuvent être aussi associées sous forme de graphe ou d'arbre.

Les caractéristiques doivent permettre de distinguer les différentes classes de formes entre elles. Elles sont dépendantes des domaines d'application. L'expérience et l'intuition guident souvent l'analyste dans le choix des caractéristiques dont le nombre est déterminé, en rapport avec la taille de l'ensemble d'apprentissage. On peut citer pour les objets plans :

– les caractéristiques directionnelles : histogrammes de gradients et caractéristiques SIFT (*Scale Invariant Feature Transform*) invariantes à l'échelle comme leur nom l'indique. Les coefficients de Gabor.

– les caractéristiques à contexte de forme (en anglais *shape context*) : à chaque point de contour on calcule les distances et les angles de ce point à tous les autres points du contour. Ces valeurs sont ensuite quantifiées et représentées par un histogramme bi-directionnel (distance, angle).

– les descripteurs de Fourier : calculés à partir des points de contour. Ces caractéristiques sont proches de la théorie du Signal. Elles présentent de bonnes propriétés relatives aux déformations causées par translation, homothétie et rotation.

– des caractéristiques de zones (en anglais *grid features*) (voir Fig. 1.4) : on encadre la forme par un rectangle que l'on découpe en zones et dans chaque zone, on compte le pourcentage de pixels noirs.

4. appelées attributs ou paramètres, et en anglais *features*.

– les moments (voir Section 1.10)

– pour les mots d'un texte appartenant à un corpus de documents : la mesure TF-IDF *Term Frequency-Inverse Document Frequency* et ses variantes.

En traitement de la parole, les caractéristiques sont basées sont les MFCC (*Mel Frequency Cesptral Coefficients*).

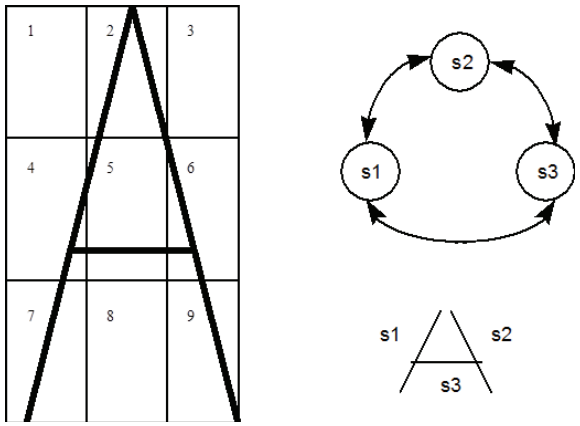

FIGURE 1.4 – Exemple simplifié d'extraction de caractéristiques sur une image de caractère. Gauche : l'image est découpée en 9 zones. Sur chacune d'elles, on extrait le pourcentage de pixels noirs. L'ensemble des 9 valeurs forme le vecteur de caractéristiques. Droite : la forme est découpée en segments s1, s2 et s3. L'ensemble de ces segments est agencé sous forme de graphe : s1 est lié à s2, s3 coupe s1 et s2,.... .

Pour réduire la quantité d'information à traiter, notamment la dimension du vecteur de caractéristiques, on le projette dans un espace où les caractéristiques ne sont pas corrélées. Cet espace est obtenu par la méthode d'analyse en composantes principales ou ACP (cf. aussi chapitre 5). L'ACP est aussi une méthode d'analyse de données qui s'attache à visualiser et décrire les données dans l'espace de représentation.

1.6 Métriques

Quelques distances (métriques) utilisées dans les méthodes statistiques sur les vecteurs X et Y à d composantes sont :

– la distance euclidienne : $d(X,Y) = \sqrt{\sum_{k=1}^{d} |x_k - y_k|^2}$

– la distance L1 : $d(X,Y) = \sqrt{\sum_{k=1}^{d} |x_k - y_k|}$

– la distance de Chebychev ou L_∞ ou convergence uniforme : $d(X,Y) = max_k|x_k - y_k|$

Ces métriques servent à constituer des groupements dans l'espace de représentation, notamment par les méthodes statistiques à apprentissage non supervisé. Reprenons les points de la Figure 1.3 et supposons que les classes ne soient plus connues, nous

obtenons la Figure 1.5. Nous distinguons toujours les 3 classes qui correspondent aux 3 groupements indiqués.

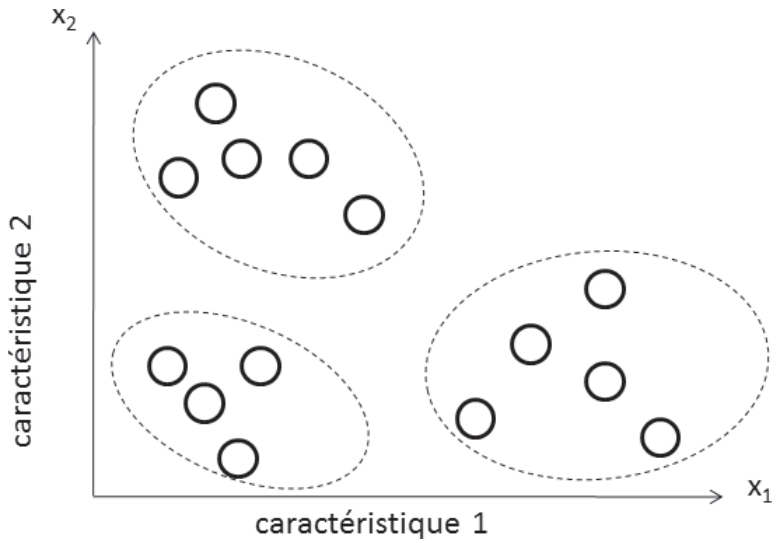

FIGURE 1.5 – Caractéristiques extraites sur des formes dans un espace de représentation bi-dimensionnel. Cas non supervisé.

1.7 Principes de décision

L'étape suivante consiste à assigner une catégorie ou classe à la représentation extraite d'une forme. Il s'agit de trouver, lors d'un apprentissage, une fonction de classement d qui permet de passer de la représentation à l'étiquette.

Soit Ω, l'espace des formes, i.e. l'ensemble de toutes les formes possibles des objets à analyser. Chaque forme $\omega \in \Omega$ constitue un événement élémentaire. Soit R, l'espace de représentation, ensemble des caractéristiques ou représentations extraites sur les formes. Soit J, l'espace des classes, i.e. ensemble des étiquettes. Cet ensemble est supposé fini, de cardinal K.

Exemple : dans un problème de reconnaissance d'images de chiffres isolés, Ω est l'ensemble de toutes les images possibles représentant les chiffres de 0 à 9, J est l'ensemble des étiquettes $J = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$, le nombre de classes est $K = 10$. L'espace de représentation est par exemple $R = \mathbb{R}^{16}$ si les caractéristiques extraites représentent des mesures ou comptages dans 16 zones de l'image des chiffres.

Soit T , la fonction de classement idéale, qui associe à chaque forme sa classe véritable sans jamais se tromper.
$T : \Omega \to J$

Soit X, la fonction qui à une forme associe sa représentation :
$X : \Omega \to R$

La fonction de classement d cherchée opère sur la représentation des formes : $d : R \rightarrow J$

L'espace Ω peut alors s'écrire : $\Omega = \bigcup_{i=1,...K} \omega_i$, les ω_i étant des parties de Ω constituées des formes de la ième classe : $\omega_i = T^{-1}(i)$. Dans l'exemple de reconnaissance de chiffres mentionné plus haut, le nombre de classes est $K = 10$ et $\omega_1, \ldots \omega_{10}$ sont des ensembles où chaque $\omega_i, i = 1, ...9$ correspond à toutes les formes du chiffre i et ω_{10} aux formes du chiffre 0.

Les $\omega_i, i = 1, \ldots K$ forment un système complet d'événements car :

$$\begin{cases} \omega_i \cap \omega_j & = \emptyset \quad i \neq j \\ \bigcup_{j=1,...K} \omega_j & = \Omega \end{cases}$$

Soient $\omega_i, i = 1, K$ les classes possibles. On a également :

$$\sum_{i=1}^{K} P(\omega_i) = 1 \tag{1.2}$$

où $P(\omega_i)$ est la probabilité *a priori* de la classe ω_i, i.e. $P(\omega_i)$ est la fréquence d'apparition de la classe ω_i. Il s'agit d'une probabilité de type discret. Nous reviendrons sur ces notions au Chapitre 2.

Parfois il est nécessaire d'introduire une "classe" de rejet, notée ω_0, à laquelle sont attribués les vecteurs x dont la forme est ambiguë. La classe ω_0 ne correspond pas vraiment à des formes. Il faut la considérer comme une décision supplémentaire dans le processus de décision. Le classifieur traite ainsi des formes appartenant à K classes et produit $K + 1$ décisions.

Quand on décide d'attribuer une forme à la classe de rejet, un processus généralement plus fin et plus coûteux est mis en place : appel à un classifieur plus complexe utilisant plus de caractéristiques, ou décision manuelle faisant appel à l'utilisateur.

La construction de la fonction d va rencontrer les difficultés suivantes : données altérées par le bruit, distorsions de forme, variabilité intra-classe (formes ou tailles différentes dans une même catégorie). Il faut aussi distinguer les classes entre elles, alors qu'elles peuvent être de forme similaire (lettres 'O' et 'Q', chiffre '1' et lettre 'l'). Un apprentissage, à partir d'exemples, est nécessaire pour construire la fonction de décision.

1.8 Apprentissage

L'apprentissage consiste à apprendre les caractéristiques communes aux classes et à distinguer les différentes classes entre elles. On constitue un ensemble d'apprentissage à partir d'exemples issus des différentes classes.