

Modélisation et Reconnaissance des formes

Introduction

Marie-Odile Berger

<http://members.loria.fr/moberger>

September 18, 2018

Objectifs du cours

- ▶ cadre général: la reconnaissance et l'identification des formes
- ▶ **forme**= image nuage de points, signal de parole, des capteurs...
- ▶ Quelques problèmes:
 - ▶ reconnaître la présence d'un objet/des objets dans une image (mais pas seulement)
 - ▶ localiser ces objets dans l'image
 - ▶ connaître précisément sa position géométrique (ex: robotique, conduite automatisée...)
 - ▶ tenir compte de l'incertain sur l'identification pour effectuer une action

But

modéliser des formes, les reconnaître, les localiser...

en tenant compte de l'incertain, des connaissances a priori, des données (exemples) disponibles

Objectif: Doter les machines des capacités de l'homme à reconnaître des caractères, des objets, des sons, des signes des signaux temporels...

Deux grands objets d'étude:

- ▶ Étudier de quelle manière l'être humain effectue cette reconnaissance (touche à des domaines comme psychologie, physiologie, biologie)
- ▶ Viser le développement de théories et de techniques permettant d'effectuer certaines tâches de reconnaissance (domaines: informatique, statistique, mathématiques)

Qu'est ce qu'une forme?

- ▶ Exemples de formes: empreintes digitales, écriture manuscrite, visages, parole, images, des objets temporels...
- ▶ La RF consiste à étudier comment les machines peuvent
 - ▶ apprendre à extraire des structures d'intérêt,
 - ▶ prendre des décisions en observant un environnement
 - ▶ reconnaître, décrire ou classifier des formes

Au départ, la RF est surtout du traitement du signal

- ▶ test de la présence d'un signal
- ▶ identification de sources multiples
- ▶ traitement de la parole

et progressivement, on a envisagé des tâches plus complexes...

L'humain fait beaucoup de choses:

reconnaître

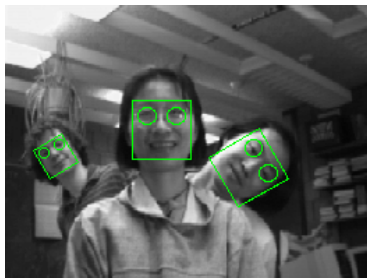
- ▶ des visages
- ▶ des sons
- ▶ des formes

et ceci indépendamment

- ▶ du point de vue sous lesquels on les observe
- ▶ des conditions d'observation
- ▶ de leur variabilité

Mais le domaine est en forte progression depuis 2012.

Quelques exemples en images



exemple: segmentation

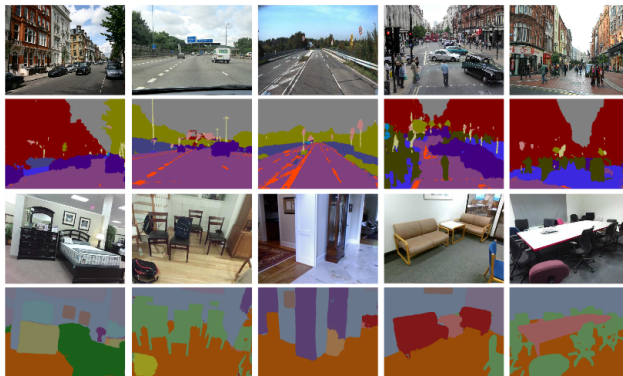
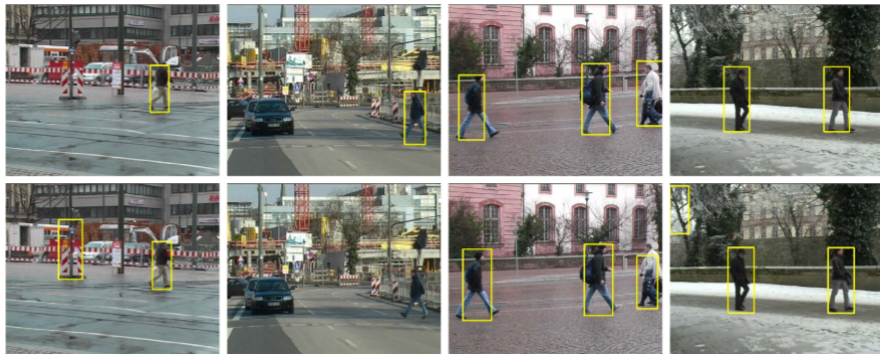


Figure: segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [Badrinarayanan 2016]

demo

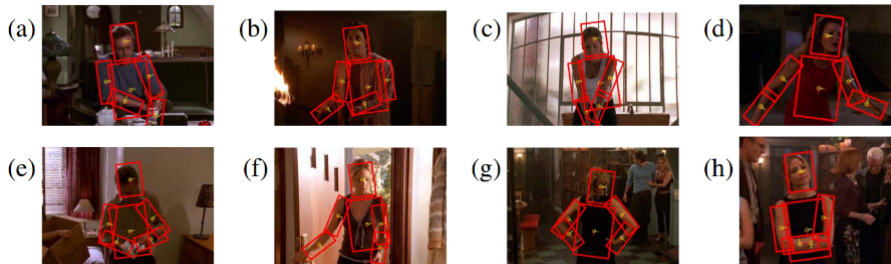
Détection précise ou approximative?

Détection de piétons [Andriluka CVPR 2009]: on peut se contenter de détecter des boîtes contenant des piétons



Détection précise ou approximative?

Où avoir une estimation précise de la position du corps



Joint reconstruction and semantic segmentation [Hane CVPR 2013]

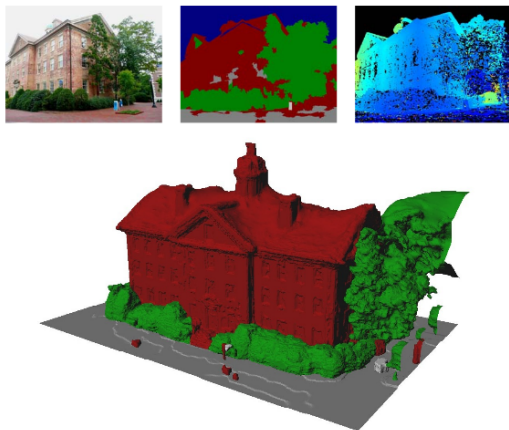


Figure 1: Top: Example of input image, standard image classification result, depthmap. Bottom: Our proposed joint optimization combines class segmentation and geometry resulting in an accurately labeled 3D reconstruction

Quelques exemples classiques de RF

Problème	Entrée	Sortie
Analyse de documents	image du document	mots, graphique
Filtres internet	Emails	classification en SPAM
Analyse du langage naturel	texte	informations sémantiques
Reconnaissance de la parole	spectrogramme	mots
Recherche multimédia	son, images, vidéo	Ident. d'évènements
Reconnaissance biometrique	empreintes digitales, iris	authentification
Identifications de défauts	images	pièces au rebut
Surveillance médicale	signaux (ECG, temp)	émission d'alertes
Identification et suivi	vidéo	trajectoire de la cible

cognitivistes contre comportementalistes

- ▶ faut-il s'inspirer de nos connaissances sur la perception humaine pour concevoir des systèmes d'inspiration cybernétique.
- ▶ le comportementaliste ne cherche pas à analyser le concept mais essaie de collecter un maximum de prototypes différents pour en extraire des régularités et des moyens de classification

Apprentissage / représentation

Étant donné un ensemble de formes caractérisant une classe, doit on

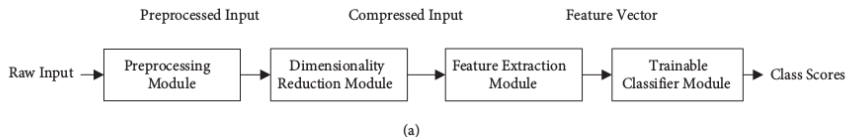
- ▶ utiliser une représentation sophistiquée des formes (ce qui nécessite souvent des connaissances explicites sur le domaine)
- ▶ travailler directement sur les données sans a priori (tout repose sur le processus de classification)

- ▶ RF statistique
- ▶ RF syntaxique ou structurelle
- ▶ Systèmes à base de connaissances

Dans ce cours, on parlera surtout de RF statistique. D'autres modules du master envisagent les autres aspects.

La RF est à la confluence de plusieurs domaines: maths, stats, probas, apprentissage, biologie, informatique, parallélisme

l'approche conventionnelle:



- ▶ Les caractéristiques des données sont extraites (de manière statique) **indépendamment** du processus de classification

Un système de reconnaissance des formes

l'approche par réseaux convolutionnels (CNN): extraction des caractéristiques et entraînement du classifieur **ne sont pas dissociées**:

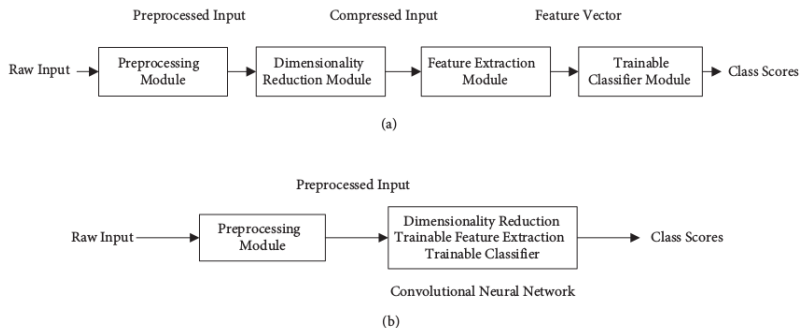


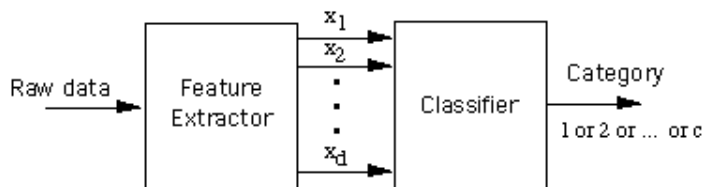
Figure 1. Pattern recognition approaches: (a) conventional, (b) CNN-based.

- ▶ Reconnaissance d'objet **Demo Berkeley**

- ▶ Les CNN et le Deep Learning font l'objet d'un module dans AVR (E. Vincent/J. Fix) Voir aussi les intervention de **Yann Le Cun** au collège de France.
- ▶ Questions communes à tous les systèmes: on cherche une fonction de classification minimisant une métrique qualifiant la distance entre les résultats obtenus et les résultats souhaités
 - ▶ Quelle distance utiliser? comment calculer le minimum? La solution est elle sensible à la présence de données aberrantes?
- ▶ Tout ne peut se résoudre par les CNN qui demandent beaucoup de données d'apprentissage!

Les problèmes de la RF statistique

- ▶ Collecter des données
- ▶ Représenter les données d'entrée (souvent de taille importante): → Extraire les caractéristiques de ces données pour réduire la dimension du problème
- ▶ Modéliser les classe d'objets
- ▶ Choisir une procédure adéquate pour classier un objet d'après son vecteur de caractéristiques. Évaluer la qualité du classifieur

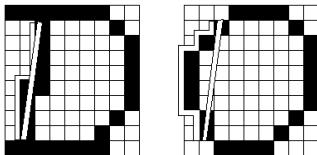


- ▶ But : Amasser des données représentatives d'un phénomène et également des données permettant ultérieurement de tester le système de RF
- ▶ Problème: Comment savoir que l'on a assez de données ? (Sont elles bien représentatives du phénomène?)
- ▶ A ce propos, consulter l'article 2018 de **Sciences et Avenir**: Intelligence artificielle : la reconnaissance faciale est-elle misogyne et raciste ?

Objectifs: réduire la taille du problème en décrivant la donnée par quelques caractéristiques discriminantes (valeurs semblables pour des motifs semblables, valeurs différentes pour des motifs différents)

- ▶ dépendance du domaine et des connaissances a priori sur le domaine (formes paramétrables)
- ▶ **invariance souhaitable** vis à vis de l'espace de mesure (invariances en général vis à vis de rotation, translation et changement d'échelle). Invariance souhaitable vis à vis d'occultation partielle ou de déformations.
- ▶ souhait d'une **représentation compacte** pour accélérer la reconnaissance
- ▶ **Note:** avec les CNN, une forme d'invariance vient de l'introduction dans les données d'apprentissage de diverses données donnant lieu à la même interprétation

Caractériser les données: exemple 1



caractéristiques possibles: aire, périmètre, compacité, histogramme ...

Empreinte:
et les point

urcations

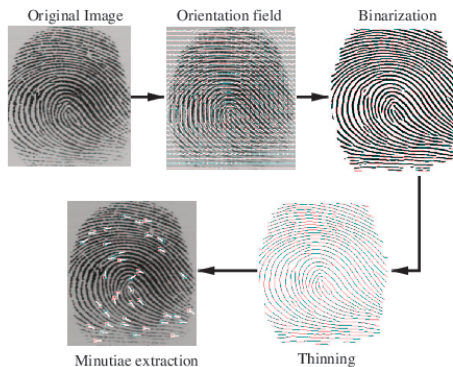


Figure 3: Various stages in a typical minutiae extraction algorithm [1].

Caractériser les données: exemple 3

les modèles d'apparence [cootes95]: plutôt que de caractériser des images de visages par l'extraction de caractéristiques a priori (lèvres, nez, yeux...) extraire directement des caractéristiques communes aux images à partir d'un groupe d'exemples.

Représentation linéaire des variations d'un ensemble de formes par rapport à la forme moyenne (voir le cours sur l'ACP et les modèles linéaires):



Figure: (1): extrait de la base de données. (2): les modes extraits

Caractériser les données: exemple 3



Figure: représentation d'une forme en utilisant 1, 2, 3 ... modes

Il existe de nombreuses façons de représenter des classes. L'objectif est d'avoir une représentation compacte et de faciliter l'identification de la classe d'appartenance d'un nouvel exemple.

Quelques représentations fréquentes:

- ▶ représentation de la classe par la base des exemples: cout du test d'appartenance prohibitif car on passe tous les éléments en revue.
- ▶ représentation d'une classe par une distribution de probabilité (souvent paramétrique) suivie par les exemples (Gaussienne par exemple)
- ▶ connaissance des fonctions de séparations des classes (frontières linéaires, quadratiques ...)

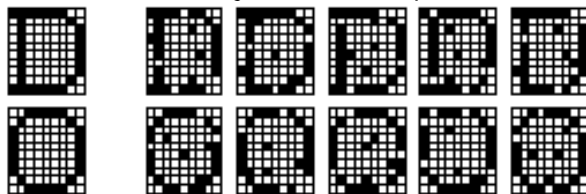
- ▶ Modèles génératifs:
 - ▶ Basé sur l'étude de $P(x|Class)$. On décrit la classe sans chercher à regarder ce qui différencie cette classe des autres classes.
 - ▶ exemple: ACP
- ▶ Modèles discriminants:
 - ▶ on cherche à maximiser la qualité de la classification sur une base d'apprentissage.
 - ▶ exemple: SVM (hyperplan séparateur), perceptron, (erreurs),...

Exemple de représentation des classes

Étant donné un certain nombre d'exemples des objets à reconnaître, étant donné un nouvel objet x , on veut affecter x à l'une des classes.

On peut:

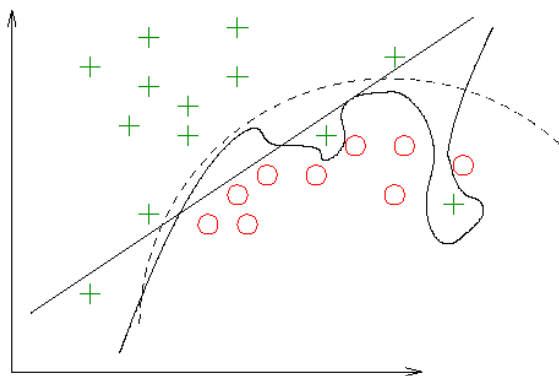
- ▶ ne pas structurer la base des exemples et mesurer la concordance la meilleure entre l'objet et les exemples de la classe (pattern matching)



- ▶ Produire une représentation statistique de la base (moyenne, variance) et l'utiliser pour la reconnaissance

En présence de données aberrantes

Le contrôle d'une forme de régularité des classes ou des distributions, permet d'éviter de créer des frontières inutilement complexes, risquant de refléter uniquement le bruit...



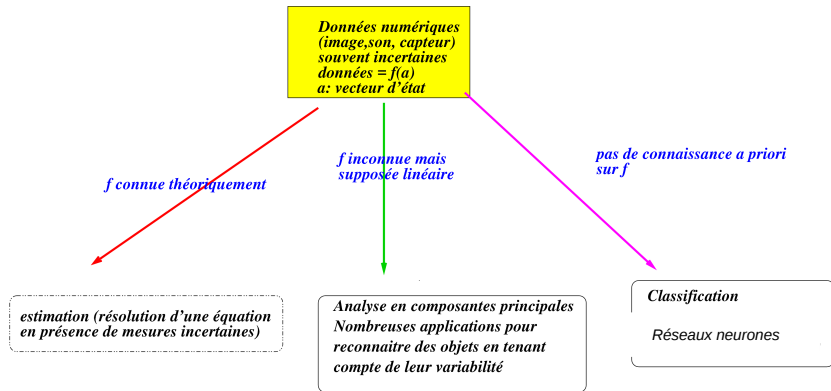
Gérer les données incorrectes, rares ou bruitées est un problème

Quelques problèmes importants

- ▶ Avoir des outils statistiques/probabilistes/numériques permettant
 - ▶ de détecter des mesures invalides/inadaptées: **théorie des tests**
 - ▶ de travailler dans un espace de dimension réduite: **réduction de la dimensionnalité**
 - ▶ des moyens de grouper/modéliser des données similaires (une classe d'objets) avec des critères dépendant de l'application: **plus proche voisin, classification, modélisation linéaire, méthodes probabilistes**
 - ▶ d'estimer les paramètres caractéristiques d'une mesure, sa classe étant connue: **théorie de l'estimation**
- ▶ Des mesures atypiques ou erronées peuvent être présentes dans les données → faire en sorte qu'elles n'influencent pas le processus de reconnaissance (notion de **robustesse**)

La RF en fonction des connaissances disponibles:

objectif: à partir de données numériques (image, parole, capteur, signal), étudier, caractériser ou classifier le phénomène ayant donné lieu à ces observations



Notions abordées dans ce cours:

- ▶ Données et incertitudes
 - ▶ Représentation des données
 - ▶ Modélisation et prise en compte de l'incertain
 - ▶ Tests adéquation mesure/modèle
- ▶ Modélisation linéaire
 - ▶ l'analyse en composante principales (ACP), l'analyse en composantes indépendantes (ACI), l'analyse linéaire discriminante (ALD)
 - ▶ Exemples de systèmes de reconnaissance de visage basés sur des classifieurs linéaires (ADABOOST)
- ▶ Estimation
 - ▶ Les bases de l'estimation paramétrique
 - ▶ Estimation robuste
 - ▶ Algorithme EM
- ▶ Quelques problèmes de reconnaissance des formes dans le domaine de l'image

Part I

Représenter les données

- ▶ **la taille**: Les données sont rarement de taille raisonnable (spectrogramme, images, . . .) → il faut adopter une représentation des données de taille raisonnable avec le moins possible de perte d'information
- ▶ **l'invariance**: les données peuvent être enregistrées dans des repères différents (ex orientation différente). Les données peuvent aussi être des mesures indirectes d'un même phénomène: les mesure ne sont donc pas directement semblables même si elles concernent un même phénomène
 - ▶ Utiliser des mesures invariantes (ou le plus possible) pour caractériser des formes.
 - ▶ dans les cas complexes, il n'y a pas de caractéristiques évidentes résumant au mieux les données et tenant compte des variations d'apparence. Celles ci **doivent être apprises**.

Comment caractériser l'invariance?

Il peut y avoir de simples mouvements de l'objet, des changements de points de vue, des changements d'illumination, des occultations....



Problème (**malédiction (*)**) de la dimension

*: terme inventé par Richard Bellman pour parler de la difficulté de travailler avec des données appartenant à des espaces de grande dimension.

- ▶ Représenter une forme par un vecteur de caractéristiques de **petite taille** permet de limiter la complexité des processus
- ▶ Un grand vecteur de caractéristiques peut avoir tendance à modéliser l'accessoire (le bruit) plutôt que l'essentiel des données.
- ▶ malédiction: il faut énormément de données pour obtenir une bonne estimation. Soient 100 observations d'un phénomène faites dans l'intervalle $[0, 1]$. Pour réaliser dans $[0, 1]^{10}$ une couverture équivalente à celle des 100 points il faudrait $100^{10} = 10^{20}$ observations, ce qui est la plupart du temps inenvisageable.

Exemples de représentation: descripteurs de Fourier

A la différence des moments géométriques qui ne nécessitent pas de segmentation, les descripteurs de Fourier sont calculés à partir du contour de la forme.

si $f(x)$ est une périodique de période T , on définit les coefficients a_n et b_n par

$$a_0 = 2/T \int_0^T f(x)dx, \quad a_n = 2/T \int_0^T \cos(2\pi nx/T)f(x)dx$$
$$b_n = 2/T \int_0^T \sin(2\pi nx/T)f(x)dx$$

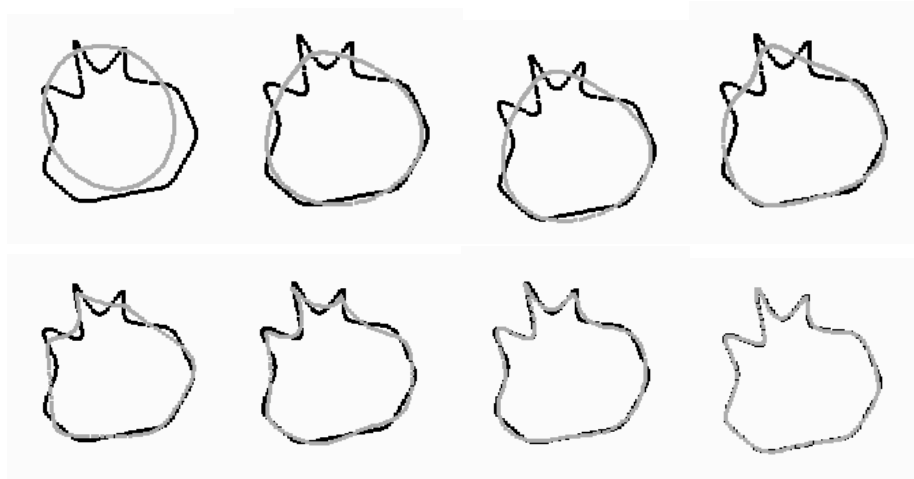
f peut être approximée par

$$a_0 + \sum a_n \cos(2\pi nx/T) + \sum b_n \sin(2\pi nx/T)$$

en se limitant en pratique à un certain nombre de termes.

Représentation: descripteurs de Fourier

Exemples d'approximations successives par série de Fourier:



Représentation: descripteurs de Fourier

Propriétés utiles des descripteurs de Fourier: coeffs de $f(x - \tau)$.

$$a'_n = a_n \cos(2\pi n\tau/T) - b_n \sin(2\pi n\tau/T)$$

$$b'_n = a_n \sin(2\pi n\tau/T) + b_n \cos(2\pi n\tau/T)$$

Il n'y a pas d'invariance par translation, **mais** le module des coefficients est invariant.

Troncature :

La limitation en pratique à un certain nombre de termes n'est pas forcément évidente: une marche nécessite une infinité de coefficients pour être bien représentée.

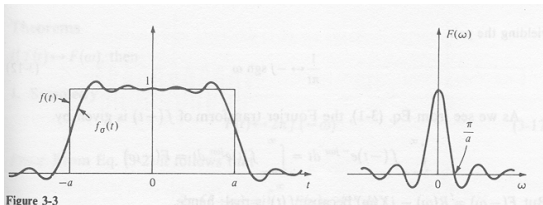


Figure 3-3

intérêt: une modélisation hiérarchique en terme de détails assez naturelle

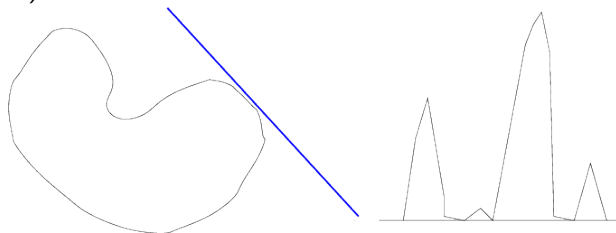
Exemple de représentation: signatures

Utiliser des représentations moins constructives mais effectivement invariantes à des groupes de transformations.

Exemple: **codage de la pente**

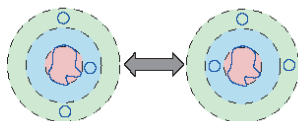
Balayer un contour. Construire la distribution $f(x)$, ou l'histogramme de l'angle polaire de la tangente à la courbe.

En utilisant comme codage le **module de la transformée de Fourier de f** , on a l'invariance par rotation (rotation: addition d'une constante à l'angle polaire)



Autre exemple: utiliser l'angle du contour par rapport au rayon issu du centre de gravité et passant par ce point.

Inconvénient: dans tous les cas, deux formes différentes peuvent partager la même signature



Remarque: l'invariance à des transformations 2D peut être prise en compte à deux niveaux:

- ▶ dans le descripteur: on a en général une description plus pauvre mais qui permet une recherche rapide des éléments ressemblants. Il reste à faire un peu de ménage parmi les candidats.
- ▶ dans le processus de recherche/classification: on recherche les éléments ressemblants modulo un groupe de transformations défini.



S. Agarwal and D. Roth.

Learning a sparse representation for object detection.

In *Proceedings of 7th European Conference on Computer Vision, Copenhagen (Denmark)*, 2002.



Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele.

Pictorial structures revisited: People detection and articulated pose estimation.

In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2009.

Best Paper Award Honorable Mention by IGD.



Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla.

Segnet: A deep convolutional encoder-decoder architecture for image segmentation.

CoRR, abs/1511.00561, 2015.

Bibliographie II

 T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham.

Active shape models -their training and application.

Computer Vision and Image Understanding, 61(1):38–59, 1995.

 R. O. Duda and P. E. Hart.

Pattern Classification and Scene Analysis.

Wiley-InterScience, 1973.

 C. Häne, C. Zach, A. Cohen, R. Angst, and M. Pollefeys.

Joint 3d scene reconstruction and class segmentation.

In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 97–104, June 2013.

 A. K. Jain, R. P. W. Duin, and J. Mao.

Statistical Pattern Recognition: A Review.

IEEE Transactions on PAMI, 22(1):4–37, January 2000.



Yann LeCun, Fu-Jie Huang, and Leon Bottou.

Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting.

In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC (USA), 2004.



L. Rabiner.

A tutorial on hidden markov models and selected applications in speech recognition.

Proc. IEEE, 77:257–286, 1989.



A. Webb, editor.

Statistical Pattern Recognition.

wiley, 2002.