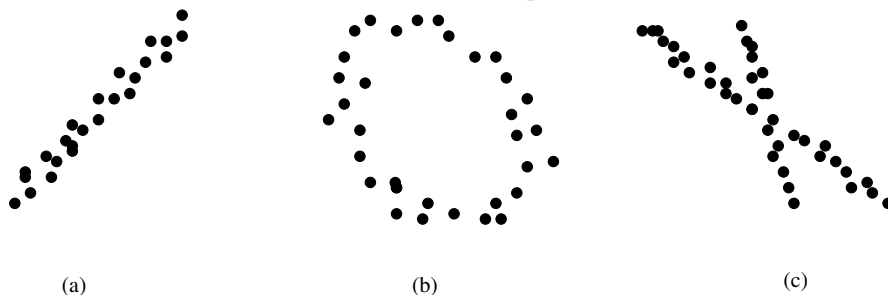


Analyse de formes et estimation

Epreuve correspondant au cours de Marie-Odile Berger
Documents distribués en cours autorisés

1 ACP et l'ICA

1. La figure ci-dessous montre trois exemples de données dans le plan. Expliquer dans chacun des cas si une ACP peut correctement décrire ces données. Justifiez votre réponse.



2. Une analyse en composantes indépendantes fonctionnerait-elle mieux sur ces données ?

2 Covariance d'une estimation aux moindres carrés

On se place dans le cadre classique de l'estimation aux moindres carrés où on cherche à estimer un vecteur $p \in R^n$ à partir d'un vecteur de mesures $Z \in R^m$, avec $m \gg n$, liés par la relation $Z = Ap + u$, u étant un bruit blanc de matrice de covariance la matrice identité $Id_{m,m}$. Je rappelle que l'estimation aux moindres carrés de \hat{p} est $\hat{p} = (A^t A)^{-1} A^t Z$.

Démontrer que la covariance sur l'estimation aux moindres carrés $cov(p - \hat{p})$ est $(A^t A)^{-1}$.

3 Estimation robuste

On souhaite ici étudier d'un peu plus près la mise en oeuvre de l'algorithme RANSAC vu en cours. Rappelons que cet algorithme cherche à construire le plus grand ensemble de mesures cohérentes avec un modèle. Cet algorithme est rappelé ici dans le cas de la détermination d'une droite aux moindres carrés.

Objectif : réaliser l'estimation robuste d'une droite aux moindres carrés à partir d'un ensemble \mathcal{S} de points $(x_i, y_i)_{\{1 \leq i \leq n\}}$.

Algorithme RANSAC :

1. Sélectionner aléatoirement un ensemble de 2 points parmi l'ensemble \mathcal{S} et calculer les coefficients de la droite à partir de ces deux points.
2. Déterminer l'ensemble \mathcal{S}^* des mesures qui sont à une distance inférieure à s de la droite calculée.
3. après N essais, le plus grand ensemble de consensus est conservé et la droite est re-estimée en utilisant les points de cet ensemble.

On rappelle que qu'on appelle *outliers*, les mesures écartées par l'algorithme et *inliers* les mesures acceptées par l'algorithme.

Cet algorithme contient plusieurs paramètres : le seuil s pour constituer l'ensemble de consensus et le nombre N de tirages aléatoires à effectuer. On souhaite ici mettre en oeuvre des méthodes pour choisir au mieux ces paramètres.

1. Expliquer l'intérêt d'un algorithme de type RANSAC par rapport à un algorithme de type M estimateur.
2. Estimation du nombre N d'essais à effectuer : on suppose tout d'abord que la proportion ϵ d'outliers est connue. Indiquer le principe d'une méthode permettant de déterminer N .
3. Si au lieu de prendre en compte le nombre minimal (ici 2 dans le cas d'une droite) de points permettant de définir une droite, on en choisit plus (par exemple 3 ou 4 points), quels sont les avantages et les inconvénients de ce choix sur l'algorithme ? Ce choix nécessite-t'il de modifier les paramètres de l'algorithme ?
4. On se place maintenant dans le cas où la proportion d'outliers n'est pas connue. Expliquer pourquoi l'algorithme suivant permet de calculer une valeur raisonnable de N .

Calcul de N si ϵ est inconnu

- (a) $N = \infty$, $\text{compteur}_{\text{exemples}} = 0$
- (b) tant que $N > \text{compteur}_{\text{exemples}}$, répéter :
 - Sélectionner aléatoirement 2 points et compter le nombre d'inliers
 - $\epsilon = 1 - (\text{nombre d'inliers})/\text{nombre total de points}$
 - étant donnée la valeur courante de ϵ , calculer N en utilisant la méthode proposée en 2.
 - $\text{compteur}_{\text{exemples}} = \text{compteur}_{\text{exemples}} + 1$;

5. Montrer que RANSAC minimise en fait la fonction de coût suivante :

$$C = \sum \rho(r_i^2)$$

où ρ est définie par :

$$\rho(r^2) = \begin{cases} 0 & r^2 < s^2 \\ \text{constant} & r^2 \geq s^2 \end{cases}$$

Quel est l'intérêt de remplacer C par la fonction suivante C_2 :

$$C_2 = \sum \rho_2(r_i^2)$$

où ρ_2 est définie par :

$$\rho_2(r^2) = \begin{cases} r^2 & r^2 < s^2 \\ \text{constant} & r^2 \geq s^2 \end{cases}$$

4 Reconnaissance des formes

Nous nous plaçons ici dans un cadre similaire à l'article [2] *Shape Contexts Enable Efficient Retrieval of Similar Shapes* étudié en cours. Nous disposons d'une base de données d'images $S_{i\{1 \leq i \leq n\}}$ représentant des classes d'objets, chaque objet étant représenté par plusieurs images.

En phase de reconnaissance, étant donnée une nouvelle image inconnue, l'objectif est de déterminer à quelle classe appartient cette image.

On utilise la démarche suivante en deux étapes :

- Dans une première étape, on extrait des points caractéristiques de chaque image de la base de donnée, chaque point étant décrit par un vecteur de caractéristiques de taille m décrivant le voisinage de ce point. Le choix des méthodes permettant d'obtenir ces points et leur vecteur de caractéristique n'est pas l'objet de cet exercice et on supposera ces méthodes disponibles. Il faut cependant garder à l'esprit que la taille m du descripteur est en général très grande ($m=256$ est courant par exemple).
- Ces descripteurs sont ensuite groupés en K groupes, K étant donné, en utilisant un algorithme, par exemple de type de K -means, sur l'ensemble des descripteurs recueillis dans toutes les images de la base de données. On choisit de représenter chaque groupe par son centre de gravité. On note $[c_1, \dots, c_K]$ le centre des K groupes.

Étant donnée une nouvelle image I , on la représente grâce aux groupes déterminés ci-dessus de la façon suivante :

- On extrait les points caractéristiques de I avec la méthode utilisée pour traiter la base de données. On calcule un descripteur d pour chacun de ces points. Chaque descripteur est affecté à un groupe en déterminant le centre du groupe dont il est le plus proche : $\operatorname{argmin}_{i \in 1, \dots, K} \text{distance}(d, c_i)$. On quantifie ainsi l'espace des descripteurs et chaque descripteur est maintenant étiqueté par un entier entre 1 et K selon sa distance aux centres des groupes.
- L'image I est représentée par l'histogramme des étiquettes des points extraits de l'images, noté $\text{Shape}(I)$. Plus précisément, avec $K=5$, si dans une image on a extrait 5 points étiquetés 3, 11 points étiquetés 2 et 1 point étiqueté 5, le vecteur représentant l'image sera $(0,11,5,0,1)$. Pour trouver la classe de l'objet présent dans l'image, on cherche l'image de la base S_i qui minimise la distance entre $\text{shape}(S_i)$ et $\text{shape}(I)$. La classe attribuée à I sera alors la classe de S_i .

Étude de la représentation

- Quel est le gain en taille obtenu par cette représentation ?
- Cette description est elle invariante à la position de l'objet ? aux occultations de l'objet ?
- Un même vecteur de caractéristiques peut-il caractériser des formes (très) différentes ?
- Si une image contient plusieurs objets décrits dans la base, cette méthode permettra-t-elle de les reconnaître ?

Incidence de la valeur de K

On se pose ici la question de l'influence de la valeur de K sur la qualité de la reconnaissance.

Expliquer pourquoi le caractère discriminant d'un descripteur d peut être mesuré par $p(d|C)/p(d|\bar{C})$. La probabilité conditionnelle $p(d|C)$ est ici calculé, pour un descripteur d , à partir du nombre d'occurrences de ce descripteur parmi les images appartenant à la classe C (\bar{C} est le complémentaire de C).

La figure 1, extraite de [1], compare le caractère discriminant du descripteur classique et du descripteur quantifié sur une base de donnée d'images. Comment expliquez vous cette chute du caractère discriminant avec la quantification ? (indication : Expliquer pourquoi les descripteurs les plus informatifs sont ceux présents dans une (ou peu de) classe(s). En déduire pourquoi la quantification affecte la reconnaissance).

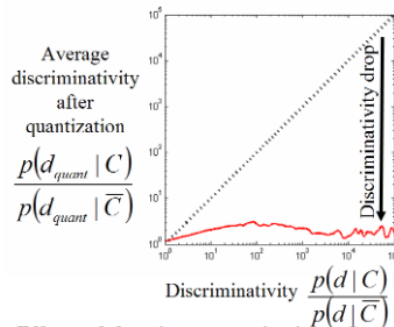


Figure 2. **Effects of descriptor quantization – Severe drop in descriptor discriminative power.** We generated a scatter plot of descriptor discriminative power before and after quantization (for a very large sample set of SIFT descriptors d in Caltech-101, each for its respective class C). We then averaged this scatter plot along the y-axis. This yields the “Average discriminative power after quantization” (the RED graph). The display is in logarithmic scale in both axes. NOTE: The more informative (discriminative) a descriptor d is, the larger the drop in its discriminative power.

FIGURE 1 – Effet de la quantification sur le pouvoir discriminant des descripteurs. Extrait de [1]

5 ACP avec un petit nombre de formes de grande dimension

Etant donné un ensemble de formes représentées par un vecteur de mesures $X_{1 \leq i \leq p}^i \in R^n$, l'objectif de l'ACP est de représenter au mieux l'ensemble de ces formes par $\mathcal{F} = \{\bar{m} + \sum_{k=1}^q \alpha_k u_k\}$. Il arrive fréquemment que la taille n des formes exemples soit très supérieure au nombre p de formes utilisées pour le calcul de l'ACP, induisant des coûts élevés lors du calcul des vecteurs propres. C'est le cas par exemple de la modélisation de la tête parlante vue en cours où la dimension de chaque tête (150 points \times 3 coordonnées) est très élevée alors

qu'on ne dispose que d'un petit nombre $p \ll n$ d'exemples de têtes.

- Quelle est la taille de la matrice de covariance $V = \frac{1}{p}XX^t$ dont on doit extraire les vecteurs propres (X est ici la matrice $n \times p$ dont chaque colonne est constituée des formes initiales)?
- Pourquoi peut on calculer au plus p composantes principales?
- Montrer que le calcul des valeurs propres de $V = \frac{1}{p}XX^t$ peut se déduire du calcul des valeurs propres de la matrice $A = \frac{1}{p}X^tX$ de taille $p \times p$, réduisant ainsi largement le coût du calcul des composantes principales.

Références

- [1] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [2] G Mori, S Belongie, and J Malik. Shape contexts enable efficient retrieval of similar shapes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii (USA)*, 2001.