

---

**Examen de Bases de données multimédia**

30 janvier 2013

*durée : 2h00*

Documents autorisés : transparents du cours, notes de cours.

Calculatrice autorisée (mais inutile).

Bonus = points en plus hors barème.

---

**EXERCICE 1. Questions à réponses courtes (10 pts)**

Répondez de manière *concise*. Sauf si cela est explicitement demandé, vous n'êtes pas obligé de faire des phrases ni de justifier vos réponses.

[Question 1.] QCM : Choisissez une réponse aux propositions suivantes.

**Attention : les mauvaises réponses sont pénalisées.**

- Si on compresse un fichier MP3 de 1 Mo avec un utilitaire de compression sans perte (zip ou gzip), produira-t-il un fichier de (a) 900 ko, (b) 400 ko, (c) 150 ko ?
- Un logiciel de lecture multimedia sur ordinateur affiche le nom de la chanson pour chaque piste d'un CD. Lit-il ce nom sur le CD ? (a) oui (b) non
- Amazon mechanical turk est une interface de programmation permettant d'exploiter (a) des très gros clusters de calcul, (b) de la main-d'œuvre bon marché, (c) un grand espace de stockage.
- Pour une chaîne Hi-Fi, quel est celle qui produit un signal de meilleure qualité : (a) PSNR=70, (b) PSNR=50.
- La vérité terrain (*ground truth*) est : (a) un algorithme de base par rapport auquel il faut se comparer, (b) une mesure de pertinence pour l'indexation multimedia, (c) le résultat que donnerait un algorithme parfait.
- Quelle est la requête la plus difficile, pour un système d'indexation multimedia : (a) une image est imprimée puis scannée, retrouver l'original ; (b) trouver les images de voitures ; (c) trouver les images de la tour Eiffel à partir d'une photo de celle-ci.
- Le descripteur SIFT est invariant à l'inversion des niveaux de gris de l'image : (a) vrai, (b) faux.
- Pour vérifier la cohérence temporelle, l'algorithme du Dynamic Time Warping est moins robuste a un changement de framerate (vitesse de lecture des videos) qu'une transformée de Hough 1D : (a) vrai, (b) faux.

**[Question 2.]** On a deux histogrammes  $(a_1, \dots, a_m) \in \mathbb{R}_+^m$  et  $(b_1, \dots, b_n) \in \mathbb{R}_+^n$  tels que  $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$ . On cherche à calculer le Earth mover's distance (EMD) entre les deux histogrammes, sachant que le coût de transfert d'une unité de  $a_i$  vers  $b_j$  est  $q_{ij} \geq 0$ .

1. écrire la définition de EMD sous forme d'un problème d'optimisation.
2. quel algorithme permet de résoudre ce problème d'optimisation ?

**[Question 3.]** Dans un espace de dimension  $d$ , soit l'hypercube  $H$  de côté  $c$ . On considère que la périphérie  $P$  de l'hypercube est l'espace qui se situe à moins de  $c/10$  de l'extérieur de  $H$ .

1. Calculer le volume de  $H$
2. Calculer le volume de  $P$
3. Pour quelle valeur approximative de  $d$  le volume de  $P$  représente-t-il plus de la moitié du volume de  $C$  ?

**[Question 4.]** On veut reconnaître des images scannées avec un scanner de résolution connue. On utilise pour cela une mise en correspondance de descripteurs locaux (Harris + SIFT).

1. quel modèle géométrique est approprié dans ce cas ?
2. combien de paramètres faut-il estimer ?
3. Une homographie est un sur-ensemble de cette transformation. Quel est l'inconvénient d'estimer une homographie ?

**[Question 5.]** Un système d'indexation d'une base de 1000 documents renvoie une liste de 20 documents en réponse à une requête. L'utilisateur sait qu'il y a 15 documents corrects dans la base, et 5 de ces documents sont parmi les réponses du système. Quels sont la précision et le rappel du résultat ?

---

**EXERCICE 2. Descripteurs d'image denses.**

Sur une image de  $400 \times 300$  pixels, on utilise des descripteurs locaux denses. Des imagerie (patches) sont échantillonnés régulièrement sur une grille. Les imagerie font 21 pixels de côté et leurs centres sont sur une grille espacée de 4 pixels.

On veut calculer un descripteur SIFT sur chaque imagerie.

**[Question 6.]** Quelles sont les étapes de calcul d'un descripteur SIFT à partir d'une imagerie de  $21 \times 21$  pixels ?

**[Question 7.]** Quelles étapes peuvent être précalculées sur l'image en entier ?

**[Question 8.]** Quelles sont les quantités qui déterminent la taille  $d$  du vecteur SIFT ?

Le SIFT est constitué d'histogrammes. Cependant, les entrées (bins) des histogrammes ne sont pas incrémentées de 1 pour chaque pixel du patch, mais par un poids issu de plusieurs interpolations (soft-assign).

**[Question 9.]** Quelles sont les deux interpolations qui interviennent lorsqu'un pixel contribue à une entrée du descripteur ?

---

**EXERCICE 3. Indexation de vecteurs**

On décide d'implémenter un système d'indexation d'images utilisant des descripteurs locaux (comme par exemple ceux introduits dans l'exercice précédent). Pour cela, on indexe les descripteurs locaux des images de la base, et on utilise comme requête les descripteurs locaux de l'image requête. La métrique de comparaison entre descripteurs locaux est la distance euclidienne.

Pour accélérer le système, on décide d'utiliser une structure de fichier inversé.

**[Question 10.]** Par rapport à une comparaison exhaustive, comment un fichier inversé permet-il d'accélérer la recherche de descripteurs ? A quel prix cela se fait-il ?

Pour simplifier, on considère que l'image requête ne contient qu'un **seul descripteur local**  $Q \in \mathbb{R}^n$  ( $n = 4$ ), et qu'il y a deux images dans la base. La première a pour descripteurs  $(B_1, B_2, B_3)$ , la seconde  $(B_4, B_5)$ .

$$Q = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 5 \end{bmatrix}, B_1 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ 4 \end{bmatrix}, B_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 4 \end{bmatrix}, B_3 = \begin{bmatrix} 1 \\ 6 \\ 1 \\ 2 \end{bmatrix}, B_4 = \begin{bmatrix} 5 \\ 0 \\ 0 \\ 3 \end{bmatrix}, B_5 = \begin{bmatrix} 5 \\ 2 \\ 3 \\ 0 \end{bmatrix}$$

**[Question 11.]** faites un k-means sur les vecteurs de la base pour trouver 3 centroïdes (pour l'initialisation, vous prendrez les 3 descripteurs de l'image 1).

**[Question 12.]** Calculer le Bag-of-words et le VLAD pour les deux images.

**[Question 13.]** Dessinez une représentation de la structure de données utilisée pour stocker la base avec le fichier inversé.

**[Question 14.]** Lors de la recherche, à quel centroïde le descripteur requête est-il affecté ?

**[Question 15.]** Quelles opérations sont effectuées par le système pour retourner les vecteurs semblables ?

**[Question 16.]** Quels sont les scores des deux images de la base ?

**Les questions suivantes peuvent être traitées indépendamment des précédentes.**

Au lieu d'utiliser un fichier inversé, on décide de comparer directement les descripteurs locaux. Pour chaque descripteur requête on vote pour les images correspondant aux trois plus proches descripteurs de la base.

On décide d'utiliser un quantificateur produit (QP) pour calculer les distances entre descripteurs. Le principe du QP est de découper les vecteurs en  $K$  morceaux (de tailles égales), puis d'appliquer une quantification vectorielle sur chaque morceau.

Pour un vecteur requête  $Q \in \mathbb{R}^n$ , un vecteur de la base  $B \in \mathbb{R}^n$  et un dictionnaire  $D = \{c_\ell \in \mathbb{R}^{\frac{n}{L}}, \ell \in \llbracket 1, L \rrbracket\}$  (qu'on suppose identique pour chaque morceau  $k \in \llbracket 1, K \rrbracket$ , avec  $n$  multiple de  $K$ ), alors le calcul de distance en version symétrique s'exprime comme :

$$d_{\text{QP}}(Q, B)^2 = \sum_{k=1}^K d_2^2(c_{Q_k}, c_{B_k})$$

où  $Q_k$  et  $B_k$  sont les indices dans le dictionnaire des centroïdes correspondant à la quantification respective de leur  $k$ -ème morceau, et  $d_2$  est la distance euclidienne.

[Question 17.] Quel est l'intérêt de QP par rapport à une distance euclidienne ?

[Question 18.] Quel est son inconvénient ?

[Question 19.] Comment pourrait-on l'améliorer à cet égard ?

On considère le dictionnaire suivant avec  $K = 2$ ,  $L = 3$  :

$$D = \left\{ \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix} \right\}$$

[Question 20.] Comment sont stockés les vecteurs de la base avec QP ? Listez les entrées. Cela a-t-il un avantage par rapport aux descripteurs initiaux ?

[Question 21.] Calculez la distance QP de la requête  $Q$  avec tous les descripteurs.

[Question 22.] Quels sont les scores des deux images de la base ?

[Question 23.] [bonus] comparez les résultats des deux méthodes.