

Bases de données multimédia III – Espaces de représentations

ENSIMAG
2014-2015

Matthijs Douze & Karteek Alahari



Espaces de représentation : Plan

- Du monde réel à la description numérique
- Représenter : pour quelle application ?
- Difficultés
- Indices visuels
- Représentation globale vs représentation locale

Du monde réel à la représentation numérique

- À notre échelle, les signaux que nous désirons acquérir sont continus
- Leur représentation sur des supports numériques nécessite une opération de discrétisation destructrice
→ par opposition aux supports analogiques (bruit)
- On distingue deux types d'opérations de discrétisation
 - ▶ L'échantillonnage
 - ▶ La quantification

Inria



L'échantillonnage

- C'est l'opération qui consiste à choisir un support discret pour le signal
 - ▶ support temporel (son, vidéo, ...)
 - ▶ support spatial (image, vidéo)
 - ▶ support 3D (scanner, ...)
- Il est généralement régulier, mais pas nécessairement

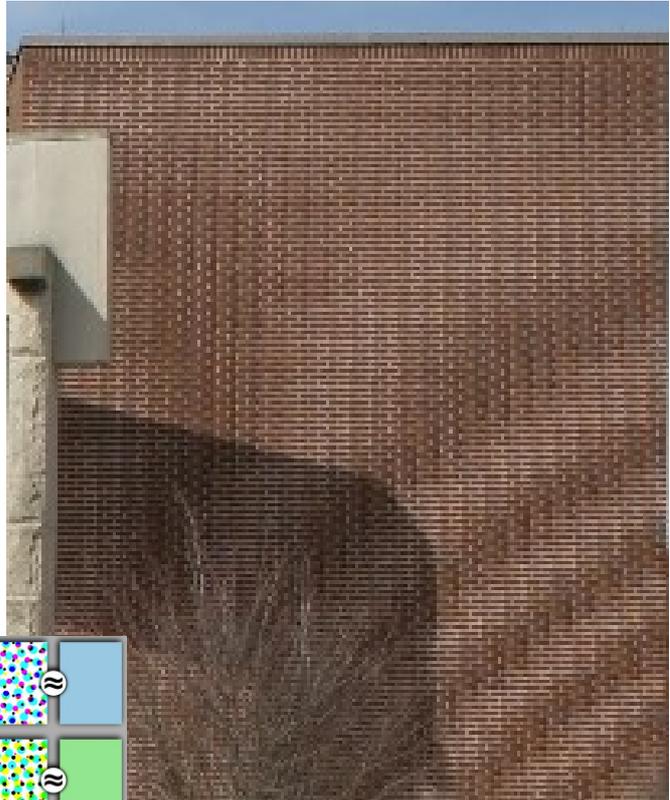
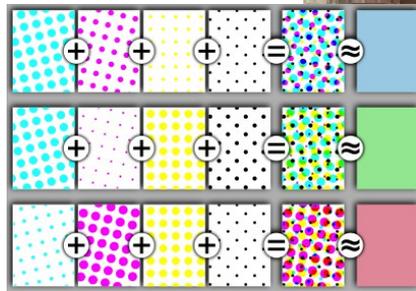


Inria



L'échantillonnage

- Le niveau de discrétisation requis est fonction de la plage de fréquence que l'on veut pouvoir restituer
- Théorème d'échantillonnage de Nyquist-Shannon: stipule que la fréquence d'échantillonnage doit être au moins 2^* la fréquence maximale contenue dans le signal (sinon aliasing)
- Effet de moiré – fréquence vidéo
- halftoning pour éviter le moiré :



Inria

Grenoble INP
ensimag

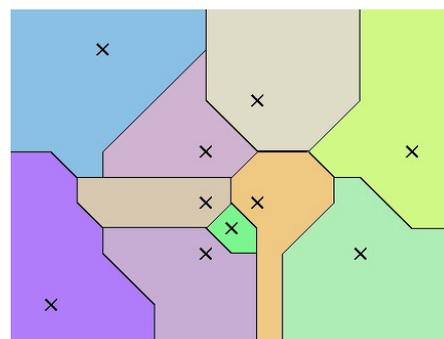
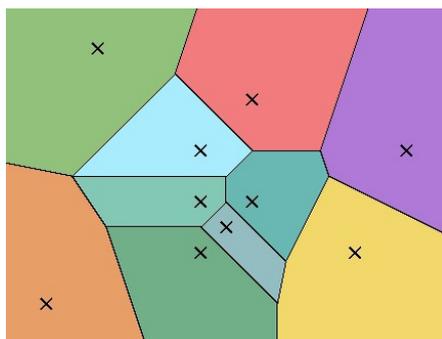
Quantification (début)

- C'est l'opération de discrétisation des amplitudes du signal
- Exemple au tableau
- Dans un sens plus général, la quantification peut être vectorielle

$$f: \mathbb{R}^d \rightarrow F, \text{ avec } F = \{1, 2, \dots, k\}$$

ensemble fini d'entiers

- La quantification définit un diagramme de Voronoï
→ une décomposition d'un espace métrique déterminée par les distances à un ensemble discret de points



Inria

Grenoble INP
ensimag

Quantification (suite)

- Quantification uniforme
 - ▶ Calcul de l'erreur de reconstruction si le signal est uniforme
 - ▶ Impact de l'augmentation de la précision sur le rapport signal à bruit
 - au tableau
- Quantifications usuelles (la plupart sont *structurés*)
 - ▶ Quantification uniforme
 - ▶ Quantification par lattice (lattice hexagonale, lattice E8, lattice de Leech) – sous-groupe discret
 - ▶ JPEG2000: quantification dead-zone ou codée en treillis (JPEG2000)
- Pour plus d'infos, voir [5]

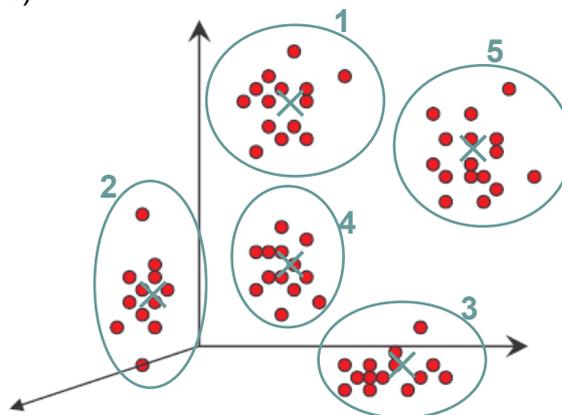
[5] R. M. Gray et D. L. Neuhoff, Quantization, IEEE Transactions on Information Theory, vol 44, pages 2325-2384, Oct. 1998

Inria



Quantification (fin)

- Quantificateur non structurés : appris sur les données
 - ▶ regroupe les données similaires en leur associant le même index
 - ▶ k-means (Lloyd's)
 - au tableau



- Exercice :
 - ▶ appliquer le k-means aux points (0,0) (2,0) (0,1) (3,4) (4,2) (5,0), $k=2$
 - ▶ prouver la convergence de l'algorithme

Inria



Exemple de discrétisations de signaux

SON

- Échantillonnage : Régulier
 - ▶ 8 kHz pour une conversation téléphonique (fréquence de la voix < 4kHz)
 - ▶ 44.1 kHz (CD) ou 48 kHz (professionnel), parfois 96 kHz
- Quantification :
 - ▶ 8 bits (premières cartes son, qualité pauvre)
 - ▶ 16 bits (CD), le PSNR est de 98 dB
 - ▶ 20 bits ⇒ bruit de quantification de 122 dB

IMAGE (couleur)

- 3 grilles de valeurs, 1 grille par composante de couleur
 - ▶ Bayer pattern
 - RGB : 8 bits de quantification pour chaque couleur
- ⇒ 24 bits par pixels (pixel=élément du support du signal)

Inria



La couleur

- Espace de représentation RGB (Red, Green, Blue)
 - ▶ représentation additive des couleurs sur la base de la lumière émise
 - ▶ le plus souvent, chaque canal est représenté sur 8 bits
 - ⇒ 16,7 millions de couleurs possibles
- YUV (Y: luminosité, U et V: chrominance)
 - ▶ utilisé dans PAL
- YCbCr, Y'CbCr
 - ▶ Mise à l'échelle de YUV pour les signaux numériques
 - ▶ utilisé dans JPEG, MPEG2
 - ▶ permet une quantification plus grossière des valeurs de chrominance
- HSV (Hue, Saturation, Value) = HSB
 - ▶ Bonne représentation perceptuelle des couleurs, utilisé dans les logiciels
- CMYK : Espace de couleur soustractive (application des encres)
- CIELAB: espace de couleurs "perceptuel"

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.13983 \\ 1 & -0.39465 & -0.58060 \\ 1 & 2.03211 & 0 \end{bmatrix} \begin{bmatrix} Y \\ U \\ V \end{bmatrix}$$

Inria



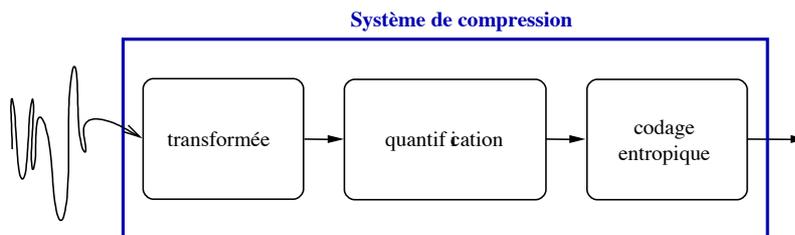
Espaces de représentation : Plan

- Du monde réel à la description numérique
- Représenter : pour quelle application ?
- Difficultés
- Indices visuels
- Représentation globale vs représentation locale

Inria



Représenter pour reconstruire : compression



- **Transformée** : étape qui consiste à représenter le signal dans un espace où l'information pertinente se concentre sur quelques coefficients
- **Quantification**
- **Codage entropique** : étape qui consiste à minimiser l'espérance de la longueur de représentation d'un ensemble de valeurs à support fini
- Transformée et codage entropique sont deux étapes réversibles (respectivement en continu et en discret). Seule la quantification introduit une perte sur le signal.

Inria



Représenter pour décrire

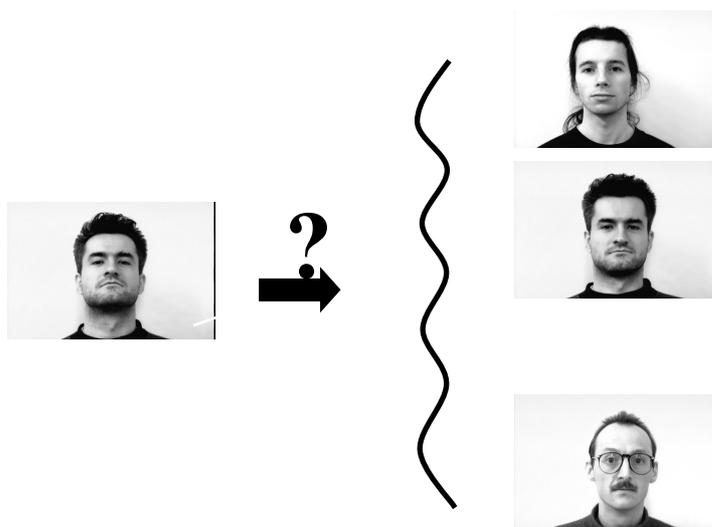
- Application de recherche par le contenu
 - ▶ focus sur les images
 - ▶ description des images pour les retrouver par leur contenu
- Problème très différent de celui de la compression
 - ▶ on ne cherche pas à reconstruire (la plupart des informations n'ont pas d'utilité directe pour la recherche, alors qu'en reconstruction elles contribuent au confort visuel)
 - ▶ on cherche à extraire ce qui sera pertinent pour l'application visée
- Applications très variées

Inria



Reconnaissance de visage : identification d'une personne

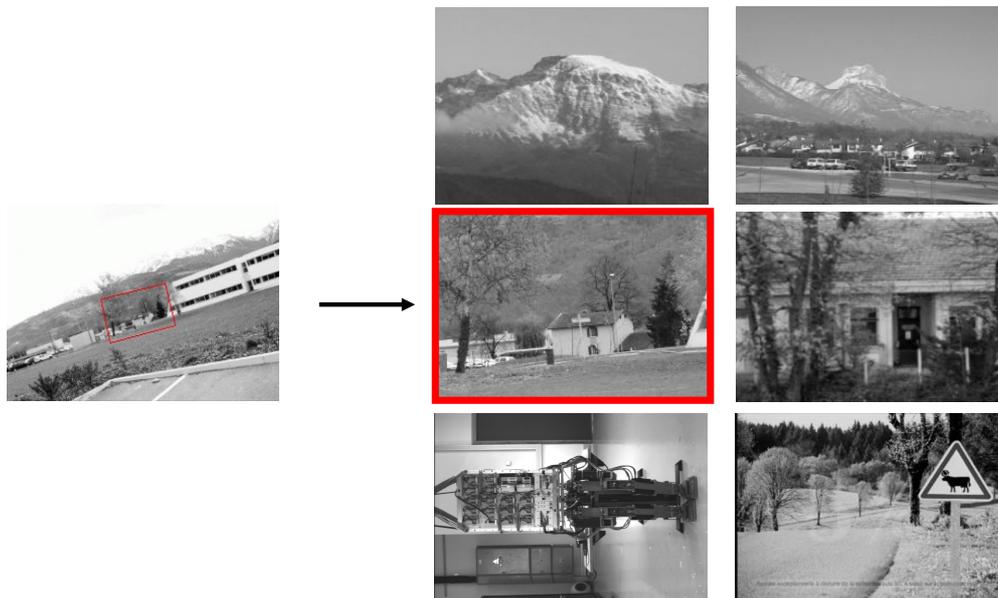
- Au sein d'une base de visages...



Inria



Reconnaissance d'un lieu



Inria

Grenoble INP
ensimag

Problèmes rencontrés

- Pour l'image, par ordre de difficulté croissante
- Contenu à reconnaître
 - ▶ une même image que l'image requête
 - ▶ une image contenant un objet donné
 - ▶ une instance d'une classe d'objet
- Tâche à effectuer
 - ▶ reconnaissance (image similaire, l'image contient l'objet ou classe d'objet)
 - ▶ détection (sous-partie de l'image similaire, bounding box de l'objet)
 - ▶ segmentation (détection des bords de l'objet)

Inria

Grenoble INP
ensimag

Espaces de représentation : Plan

- Du monde réel à la description numérique
- Représenter : pour quelle application ?
- **Difficultés**
- Indices visuels
- Représentation globale vs représentation locale

Inria



Difficultés

- Transformations image (rotation, échelle)

rotation image



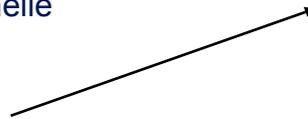
Inria



Difficultés

- Transformations image (rotation, échelle)

changement d'échelle



Difficultés

- Transformations image (rotation, échelle)
- Changement de luminosité



Difficultés

- Transformations image (rotation, échelle)
- Changement de luminosité
- Visibilité partielle / occultation



Inria

Grenoble INP
ensimag

Difficultés

- Transformations image (rotation, échelle)
- Changement de luminosité
- Visibilité partielle / occultation
- Clutter (objets supplémentaires)

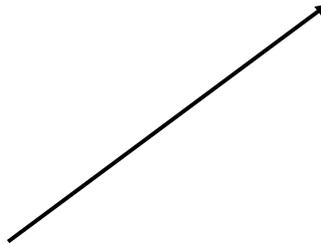


Inria

Grenoble INP
ensimag

Difficultés

- Transformations image (rotation, échelle)
- Changement de luminosité
- Visibilité partielle / occultation
- Clutter (objets supplémentaires)
- Objets 3D



Inria

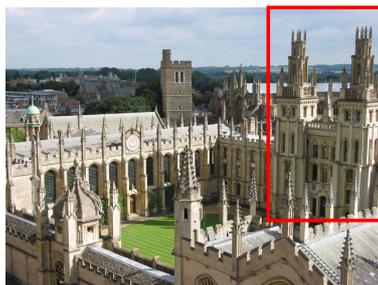
Grenoble INP
ensimag

Exemple du « Oxford Building Dataset »

Requêtes (objets)



Exemples d'images à retourner

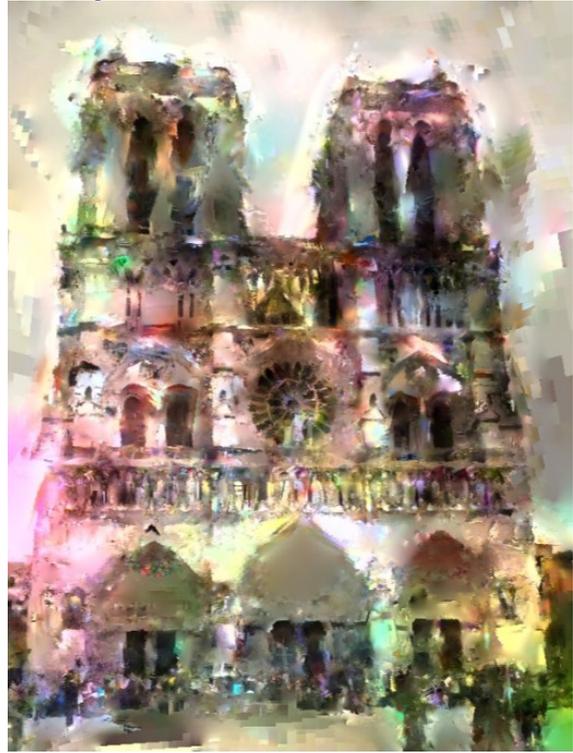


Inria

<http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/index.html>

Grenoble INP
ensimag

Que peut-on reconstruire à partir de ce qui sert à indexer?

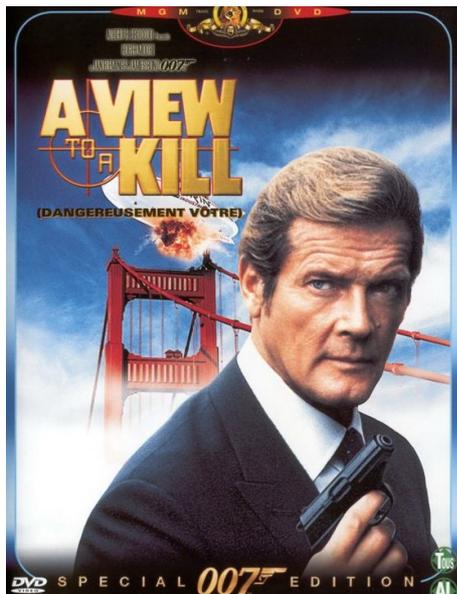


Inria

P. Weinzaepfel, H. Jégou, P. Pérez, "reconstructing an image from its local descriptors", CVPR 2011



Que faut-il reconnaître ?



Inria



Premier aperçu : segmentation

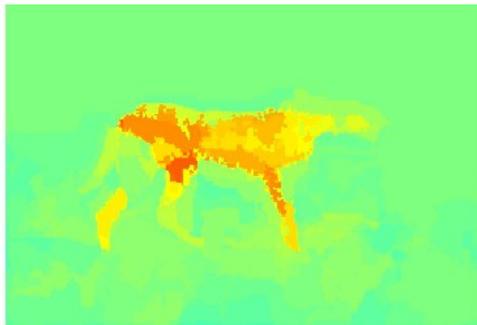
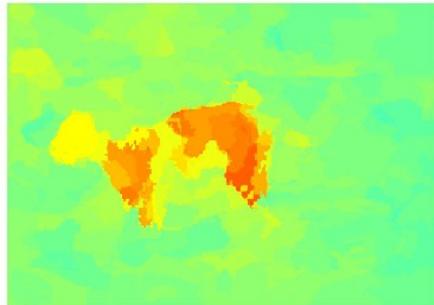
Ce qu'on aimerait



Inria

Premier aperçu : segmentation

Ce qu'on a



Inria

Difficultés

- Transformations image (rotation, échelle)
 - Changement de luminosité
 - Visibilité partielle / occultation
 - Clutter (objets supplémentaires)
 - Objets 3D
 - Nombre important d'images dans la base
-
- Nécessité d'apprendre les variations d'apparence → un modèle visuel

Inria



Espaces de représentation : Plan

- Du monde réel à la description numérique
- Représenter : pour quelle application ?
- Difficultés
- **Indices visuels**
- Représentation globale vs représentation locale

Inria



Indices visuels

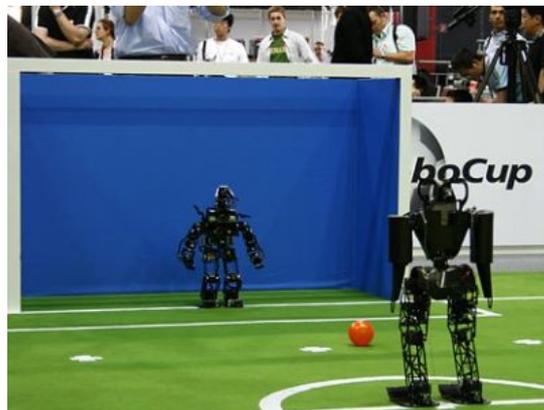
- Ici, focus sur l'image, mais s'applique à la vidéo
- **Indice visuels** : caractéristiques de l'image, au sens de notre perception d'humain, que l'on cherche à utiliser pour la tâche considérée
- Principaux indices visuels
 - ▶ couleur
 - ▶ forme
 - ▶ texture
 - ▶ régions
 - ▶ mouvement (pour la vidéo)

Inria



Indice visuel : couleur (début)

- Significatif pour la catégorisation (= trouver la classe d'objet) ou le tracking (suivre des objets)
- Indice visuel ayant reçu beaucoup d'attention
 - ▶ de nombreux descripteurs (globaux ou locaux)
- En indexation : utilisé depuis longtemps pour faire de la recherche d'image par similarité (description le plus souvent indépendante de la position et de l'orientation des objets dans l'image)
- Robocup 2006...



Inria



Indice visuel : couleur



<http://labs.ideeinc.com/multicolour/>

Inria



Indice visuel : couleur (suite)

- Problèmes
 - ▶ choix de la représentation de la couleur
 - ▶ distance entre couleurs
 - ▶ Invariance ?
- Indice visuel insuffisant si utilisé seul, car peu discriminant :
 - ▶ pomme rouge vs Ferrari
- *Plus tard dans ce cours* : exemple de description (globale) utilisant la couleur
- Remarque : une image en niveaux de gris ou même noir et blanc suffit à un humain pour effectuer la tâche demandée !

Inria



Couleur

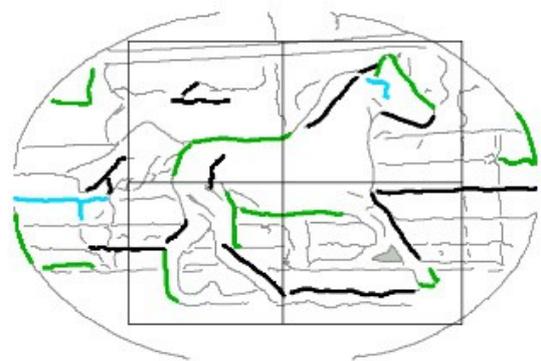


Inria

Grenoble INP
ensimag

Indice visuel : forme (début)

- Nombreux algorithmes d'extraction de formes
- Bestiaire (non exhaustif) :
 - ▶ extraction des contours
 - ▶ chaîne de points de contours ou approximation polygonale
 - ▶ mise en correspondance de graphes
- indice visuel différent des régions (1D vs. 2D)



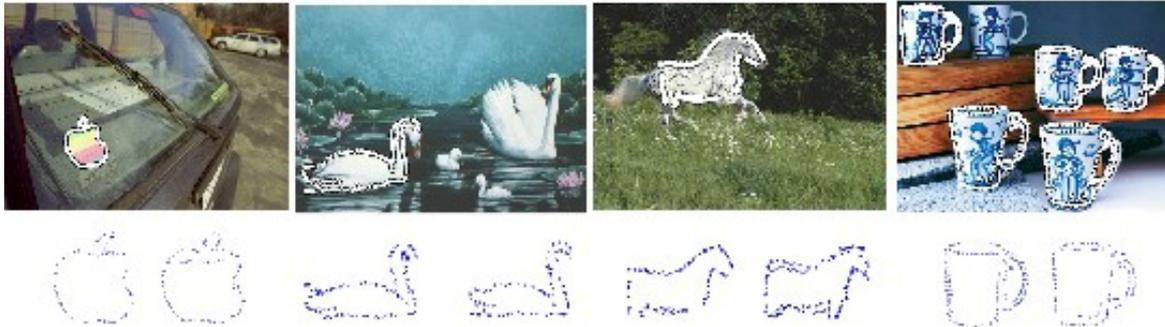
(a)

Inria

|||

Indice visuel : forme (fin)

- Difficultés
 - ▶ l'étape d'extraction des contours est souvent instable
 - ▶ séparer la forme du fond est difficile (problème de segmentation)
- En général, ces algorithmes mènent à une description qui n'est pas invariante à des gros changement d'échelle, ni aux rotations (car invariance en rotation introduit alors forte perte de discriminance pour cet indice visuel)



V. Ferrari, L. Février, F. Jurie, C. Schmid, *Groups of Adjacent Contour Segments for Object Detection*, PAMI 2008

Inria



Indice visuel : texture

- Variations répétitives de l'intensité
- Elles résultent de propriétés physiques des surfaces 3D (herbe, nuages), ou de propriétés photométriques de surface (zèbre, tissu à fleurs)
- De nombreuses parties des images/vidéos peuvent être caractérisées par des textures
- L'humain reconnaît facilement les textures, mais difficile à définir (et donc représenter): transition micro/macro, textures irrégulières...
- Deux types de méthodes
 - ▶ statistiques : distribution spatiales des intensités de l'image
→ co-occurrences, auto-corrélation
 - ▶ géométriques : segmentation de motifs qui se répètent

Inria



Régions texturées

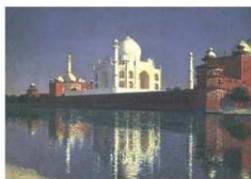


Inria



Indice visuel : régions

- Prend en compte la répartition spatiale de zones « homogènes »
+ attributs intrinsèques aux régions (taille, couleur, etc)
- Relations spatiales entre ces régions (ciel en haut/mer en bas)
- Sensible au bruit et aux occultations



Inria



Comparaisons des indices visuels

- Pour une utilisation "brute" des différents indices visuels
- Notion de descripteur associé à un indice visuel
- Du niveau sémantique de bas en haut

Indice visuel	Descripteur Utilisé	Avantages	Limitations
Couleur	Histogramme, moments	Extraction très facile, propriétés globales de l'image	Pas d'information spatiale, peu de lien avec la sémantique
Forme	Polygones, approximations, invariants moments	contient presque toute l'information sémantique	sémantique difficile à extraire de la forme
Texture	Matrices de co-occurrences, filtre de Gabor, Fourier	caractéristique facile à reconnaître	difficile à définir : pb de poule et d'oeuf
Région	Surface, position, moments, histogramme de couleurs	Intègre les relations spatiales (ciel en haut de l'image)	

Inria



Espaces de représentation : Plan

- Du monde réel à la description numérique
- Représenter : pour quelle application ?
- Difficultés
- Indices visuels
- **Représentation globale vs représentation locale**

Inria



Description globale

- **Une description globale** est une représentation de l'image dans son ensemble, sous la forme d'un vecteur de taille fixe
- Elle se contente le plus souvent d'exploiter un indice visuel unique
- mesure de (dis-)similarité définie sur l'espace de ces descripteurs
- exemple : histogramme de couleurs

Inria



Description locale

- Une description locale est une représentation de l'entité à décrire par un *ensemble* de descripteurs (vecteurs)
 - ▶ Vecteurs calculés localement pour différentes parties de l'entité
ex: plusieurs sous-parties d'une image
 - ▶ Ainsi, plusieurs descripteurs par entité
 - ▶ Mesure de similarité nécessite la mise en correspondance (appariement) des descripteurs
- Étude détaillée de ce type de description dans la suite de ce cours

Inria



Représentation globale ou locale ?

- Avantages de la représentation globale
 - ▶ certains invariances sont faciles à obtenir
 - ▶ extraction pour un coût de calcul plus faible
 - ▶ comparaison plus simple : comparaison des entités avec un vecteur unique, donc pas de mise en correspondance géométrique
 - ▶ en description locale, chaque information individuelle n'est pas très discriminante

- Avantages de la représentation locale
 - ▶ + de résistance à la plupart des transformations
 - ▶ Robuste: statistiques dépendantes de plusieurs descripteurs
 - ▶ description plus précise d'une image dans son ensemble (souvent la description est aussi volumineuse que l'image d'origine)
 - ▶ on peut les agréger en représentation globale...

Inria

