



Données multimedia images et vidéo

CES Data Scientist – Mars 2016 Michel Roux



Décrire le contenu des images

Notion de signature visuelle





Regardez attentivement ...















Qu'avez-vous vu ? Décrivez les images à votre voisin



Planète Vénus





Vénus de Willendorf (23000 av JC)





Sabot de Vénus



Nombril de Vénus

La naissance de Vénus de Botticcelli (vers 1485)









Que voyez-vous?







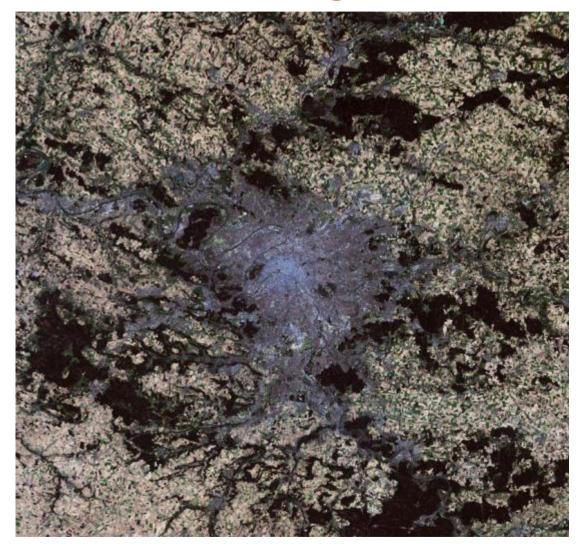


Des niveaux variés de description

- Bas niveau, proche de notre perception : description en terme de couleurs, formes et textures
- Haut niveau, proche de notre interprétation : usage de mots, phrases, sémantique
- Subjectif : dépend de celui qui regarde l'image (sa perception) et de son niveau de connaissance



Décrivez cette image...



Décrivez cette image...







Décrivez cette image...





variété des éléments à repérer

On doit pouvoir retrouver :

Des objets localisés : Bâtiments, rond-points, chars, ... à des échelles variées

Des éléments géométriques répartis sur l'image : Axes routiers, fluviaux

Des zônes d'intérêt (paysages) : forêt, champs, villes, mer, ...

Et dautres éléments, suivant le contexte

Echelle / résolution



Décrire une image permet de l'indexer

Catalogue :

 Exploitation de données globales sur le document visuel, indépendantes du contenu

Classification:

 Exploitation d'une caractéristique globale du document visuel

Indexation

 Exploitation d'une analyse fine du contenu du document visuel



Définition

- Indexer = extraire une information synthétique des images afin de faciliter l'accès à leur contenu
 - domaines scientifiques concernés : fouille de données (data mining), classification, ingénierie des connaissances, vision artificielle, SGBD, ...
 - information = élément susceptible d'être codé pour être conservé, traité, communiqué
 - index = clé d'accès à l'information contenue dans l'image





- "Fossé sémantique" : comment accéder à/coder une information visuelle ?
- Spécificité des images :
 - Photos, médicales, satellitaires (optique, radar), 2D ou 3D
- Diversité des formats des images
 - panchromatiques, spectrales
- Grande taille des bases de données => vers une automatisation des procédés
- Comment exploiter des informations associées (« multimodalité ») ?



TP0: image numérique (15')

Objectifs:

- premier contact avec l'environnement de travail
- Lecture / écriture / visualisation d'une image avec OpenCV (Python)



Comment faire?

- Avec des mots => indexer = annoter
 - 10 000 images ? Des terabytes d'images ?
 - L'INA et les indexeurs...
- Avec des modèles mathématiques => indexation par le contenu
 - Favorise l'automatisation face à la masse de données
 - Exploite des outils d'apprentissages
 - Adéquation à notre « perception » ? le fossé sémantique



Indexation d'images à partir du texte



Restructurations : organiser la résistance économique par des moyens exceptionnels

13/11/2013 - Industrie & innovation

Arnaud Montebourg a présenté un plan de résistance économique doté de 380 millions d'euros pour secourir des entreprises en difficulté et pallier ainsi le désinvestissement des banques.



Le "Made in France" fait salon

07/11/2013 - Industrie & innovation

Le "Made in France" a le vent en poupe. Pour sa 2e édition, MIF Expo, le salon des produits et innovations fabriqués en France se tient Porte de Versailles les 9, 10 et 11 novembre. Cette année, le salon sera particulièrement dédié à l'innovation et présentera le meilleur des nouveaux objets de la France industrielle.



L'innovation, facteur clé de la montée en gamme de notre économie

05/11/2013 -

Un an après le lancement du Pacte national pour la croissance, la compétitivité et l'emploi, le Premier ministre en a présenté un premier bilan à Saint-Etienne (Loire). Il a aussi dévoilé de nouvelles mesures pour dynamiser l'innovation.

Trouvé sur le site du gouvernement

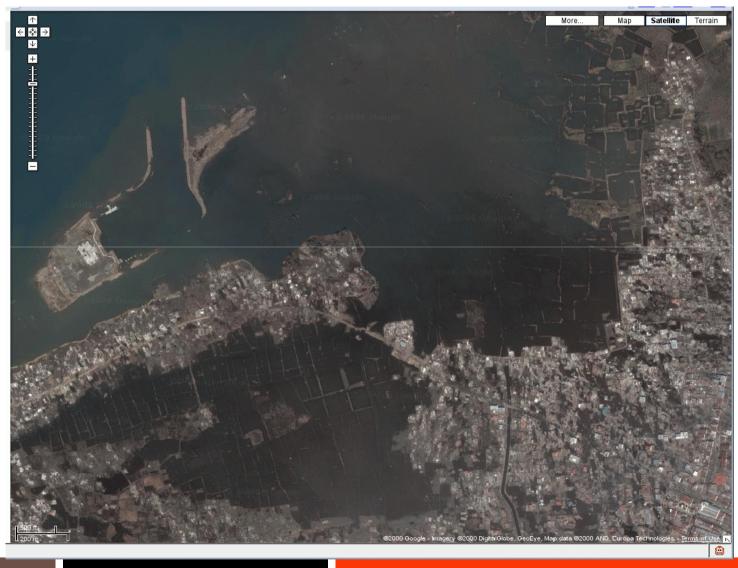


Automatiser l'annotation?





Automatiser l'annotation?



Limite des requêtes textuelles



Planète Vénus





Vénus de Willendorf (23000 av JC)





Sabot de Vénus



Nombril de Vénus

La naissance de Vénus de Botticcelli (vers 1485)



Indexation d'images à partir du texte

- Une image illustre un document mais le document ne décrit pas forcément l'image;
 - Le texte accompagnateur d'une image ne suffit généralement pas
- La richesse du contenu d'une image dépend du domaine de connaissances de celui qui la regarde ;
 - Nécessité de modéliser ce domaine (ontologies)
- Aspects subjectifs du contenu
 - Plus facile de faire une description des objets contenus dans l'image que des émotions qui en résultent
 - vers une indexation automatique du contenu!



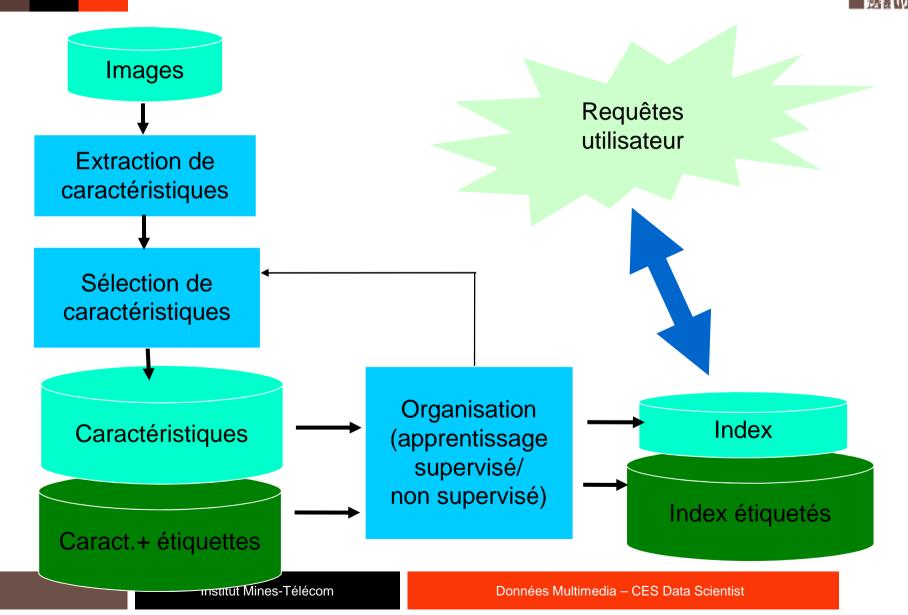
Les signatures classiques

Couleur, forme et texture



Indexation par le contenu : principe





Principe de l'indexation par le contenu



- Off-line : production d'index issus de l'analyse du contenu des images
 - Extraction de caractéristiques pertinentes ("signatures")
 - Réduction de la dimension
 - Organisation par classification
- On-line : gestion des requêtes d'un utilisateur
 - « fossé sémantique »
 - La requête est exprimée soit par du texte (si des étiquettes sont disponibles), soit par des images exemplaires (voire contre-exemplaires)





http://youtu.be/MTDJsbysoqM

Université de Laval, David Gabay

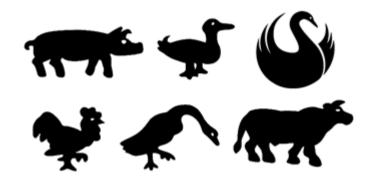
Des espaces de couleur différents :

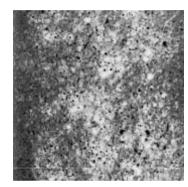
RGB (Red Green Blue) cônes de perception

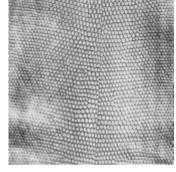
HSV (Hue Saturation Value) « psychologique »



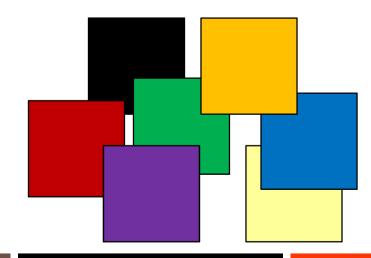
Couleurs – formes - textures













Exemples



Plutôt textures?



Plutôt géométriques ?



Lenna Soderberg (1978, november Playmate): mélange harmonieux de couleurs, formes et textures ?



TP1 (30')

Objectifs:

- Un exemple de caractéristique globale pour les images : moyenne et variance de l'intensité des pixels de l'image
- Visualisation dans un espace 2D
 - Qu'observez-vous?

Institut Mines-Télécom

Comment l'interprétez-vous ?





Caractéristiques :

- Locales, spécifiques : points saillants (Harris, Deriche, ...), minuties, SIFT, ...
- Régionales, générales : couleur, texture, forme
- Combinaison : descripteurs régionaux autour de points caractéristiques

Globale ou locale :

Blocs arbitraires, segmentation, information topologique





Problème de représentation

 Caractéristiques numériques, symboliques, vecteurs vs graphes,...

Invariances

- Translation, rotation, homothétie, colorimétrie
- déformations : taille, cadrage, masquage, perspective, contraste, éclairage, vieillissement
- toujours nécessaires ?



Comparer des signatures d'images

Distances, similarités, ... et notre perception



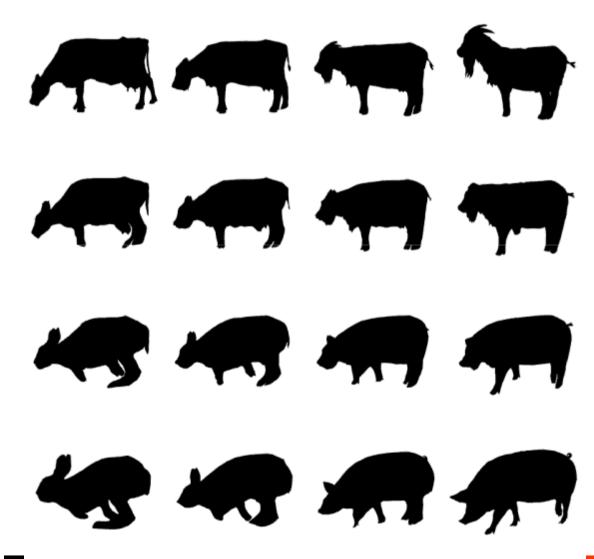


Comparaison d'images

- Comparer deux images == comparer leurs attributs (caractéristiques, signatures)
- Similarité visuelle ?=? similarité des attributs

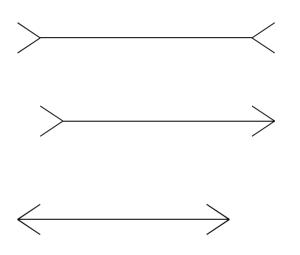


ges et notion de (dis-)similarité





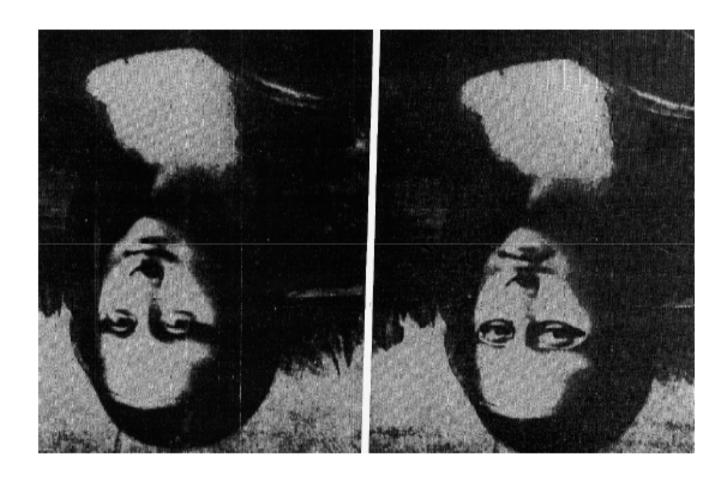




Lequel de ces segments est le plus grand?

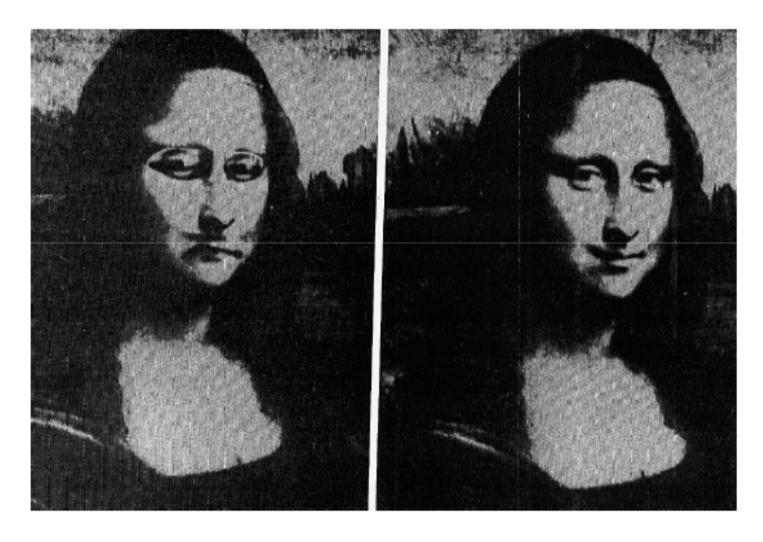


Images similaires ?



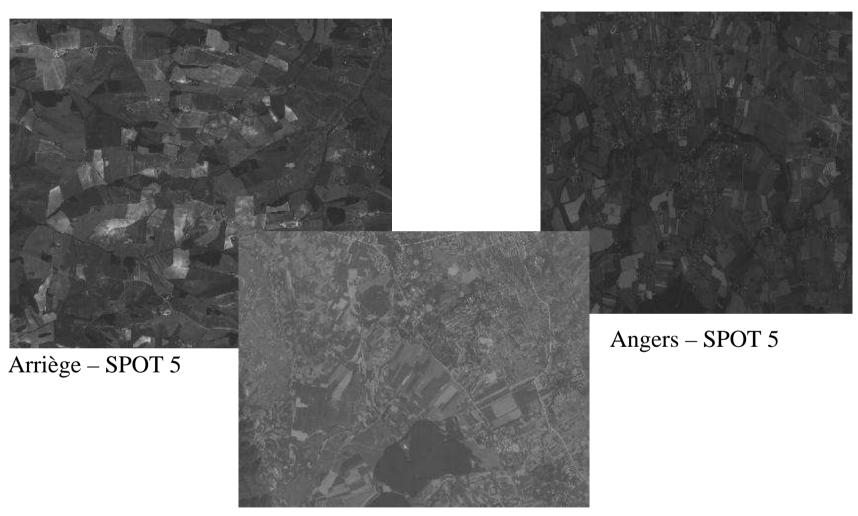


Images similaires ?









Zhongdong – SPOT 5







De l'oeil au cerveau : drogues et hallucinations visuelles LeProfDuWeb.com

http://youtu.be/Ku24ICS_ePU

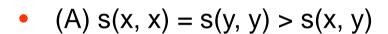
Institut Mines-Télécom

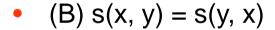
Similarité mathématique



- Deux images seront comparées par l'intermédiaire des caractéristiques extraites
 - Sim(Im1, Im2) = Sim(f(Im1), f(Im2)), avec f la fonction d'extraction des caractéristiques

Mesures de similarités, distances



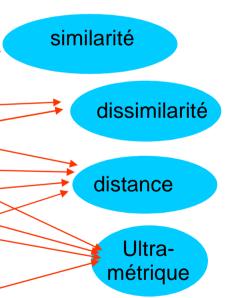


• (C) d(x, x) = 0

• (D) $d(x,y) = 0 \Leftrightarrow x = y = 0$

• (E) $d(x,y) \le d(x,z) + d(z,y)$

(F) d(x,y) <= max(d(x,z), d(z,y))





Un grand choix mathématique

- Distance euclidienne
- Distance euclidienne généralisée
- Malahanobis
- Chi2
- Similarité en cosinus
- Combinaisons linéaires de similarités (ou distances)
- **-**



TP2 (30')

Objectifs

- Retrouver les images similaires à une image exemplaire = une approche classique depuis les années 90.
- RGB et HSV : deux espaces de couleur
 Votre avis : lequel est le plus efficace ?
- Et si vous modifiez le choix de la distance ?

08/03/2016



Quelques approches maintenant standards

SIFT, sacs de mots et la détection de visages...





Des caractéristiques de couleur

- Statistiques sur les intensités des pixels
 - Moyenne, variance, histogramme
- Changer d'espace de représentation
 - RGB -> HSV ou autre
 - Recherche d'espaces proches de notre perception

08/03/2016



Des caractéristiques de texture

- Les ondelettes de Gabor
 - Banc de filtres à différentes échelles et orientations
 - Calcul de statistiques sur le résultat du filtrage
 - Proches de notre système visuel humain
 - Présent dans la normalisation MPEG7
- D'autres ondelettes

Institut Mines-Télécom

- La Transformée de Fourier 2D
 - Capture des éléments périodiques
- Les coefficients d'Haralick

Quelques refs: http://perso.telecomparistech.fr/~campedel/MultimediaIndexing/



Descripteurs de formes

- Utilisation de dictionnaires de formes
 - Appariement en exploitant la corrélation 2D
- Segmentation des images
 - Caractérisation des régions (=segments)
 - Caractérisation des contours des régions
 - Usage fréquent des gradients
- Pour information : dans OpenCV

http://docs.opencv.org/modules/imgproc/doc/structural_analysis_and_shape_descriptors.html

SIFT



- http://www.cs.ubc.ca/~lowe/keypoints/
- Implémenté dans OpenCV
- Approche par points caractéristiques
 - SIFT (Scale Invariant Feature Transform) Lowe, 2004
 - invariance par échelle, rotation, point de vue, conditions d'éclairage
 - Extraction des points caractéristiques
 - extrema locaux du Laplacien dans le « scale-space »
 - Filtrage des points caractéristiques
 - critère de stabilité, seuillage du contraste, ...
 - Descripteurs associées aux points caractéristiques
 - histogramme local des orientations du gradient
- Exploitation :
 - Plus proches voisins (kPPV) + transformée de Hough



SIFT : Détection des points

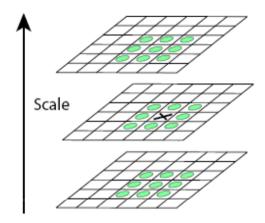


Figure 2: Maxima and minima of the difference-of-Gaussian images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3x3 regions at the current and adjacent scales (marked with circles).

Extrait de « Distinctive Image Features from Scale Invariants Keypoints » Lowe, IJCV, 2004





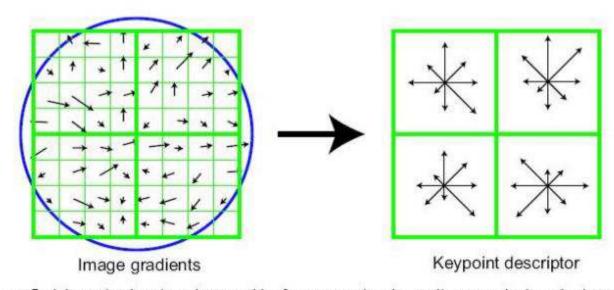
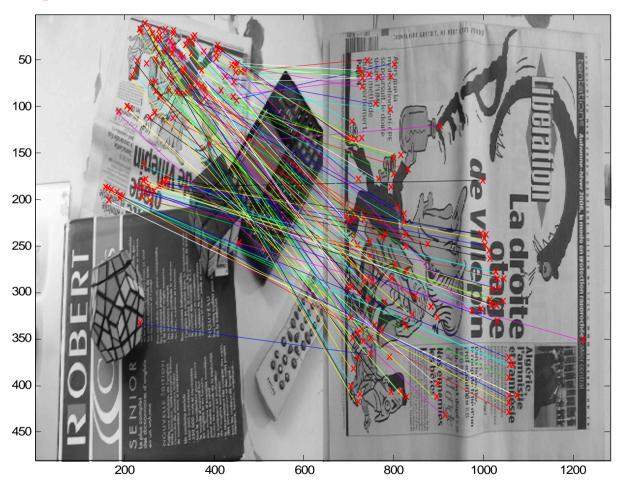


Figure 7: A keypoint descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the left. These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 2x2 descriptor array computed from an 8x8 set of samples, whereas the experiments in this paper use 4x4 descriptors computed from a 16x16 sample array.

SIFT: illustration de mise en correspondance



Robustesse des descripteurs (J. Rabin - 2006)







+ Points caractéristiques robustes

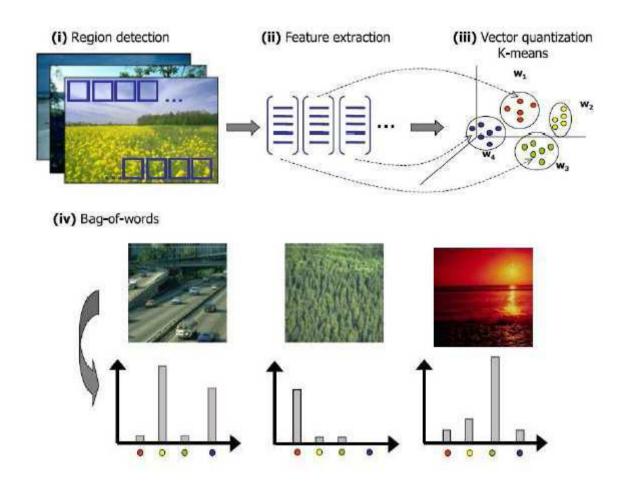
- En échelle
- En rotation
- Illumination

- ... mais génèrent une grande quantité de données

- Grande quantité de points par image
- Taille des descripteurs de chaque point
- Coût de la mise en correspondance



Une "solution": le sac de "mots visuels"





Sacs de SIFTs

- Clusterisation (kMoyenne par exemple) des signatures obtenues pour tous les points caractéristiques => clusters = mots visuels
- Signature de l'image = 1 histogramme sur les mots visuels
- La procédure pourrait s'appliquer sur d'autres caractéristiques locales.
- Une perte : localisation des points caractéristiques



Conclusion





Indexation des images



Qu'avez-vous appris cet après-midi?



Image et vidéo ?

- Vidéo = séquence d'images + audio + texte ...
- Des caractéristiques propres aux vidéos (à découvrir pendant le projet)
 - ✓ Les plans (shots)
 - Les mouvements des objets
 - Les mouvements de la caméra



Merci de votre attention!

N'hésitez pas à me contacter pour plus d'informations michel.roux@telecom-paristech.fr



Annexe Viola & Jones – Les visages

- Implémenté dans OpenCV
- http://www.vision.caltech.edu/html-files/EE148-2005-Spring/pprs/viola04ijcv.pdf
- Principe « classique »
 - Extraction de caractéristiques (Haar)
 - Apprentissage de classificateurs AdaBoost (sur une base étiquetée visage/ non visage), comme combinaison de classificateurs « faibles » (à seuil)
- Les +
 - Rapidité d'extraction grâce à *l'image intégrale*
 - La cascade de classificateurs pour éliminer rapidement les zones ne contenant pas de visage, puis réduire le bruit de détection

Image intégrale

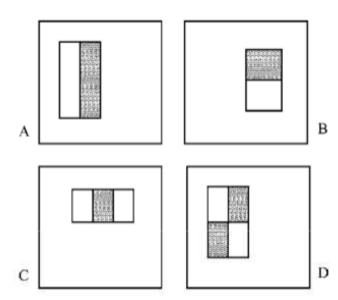


Figure 1. Example rectangle features shown relative to the enclosing detection window. The sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles. Two-rectangle features are shown in (A) and (B). Figure (C) shows a three-rectangle feature, and (D) a four-rectangle feature.

Objectif : calcul rapide des attributs De Haar – extraits de l'article de Viola&Jones

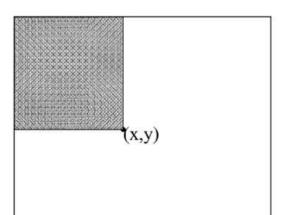


Figure 2. The value of the integral image at point (x, y) is the sum of all the pixels above and to the left.

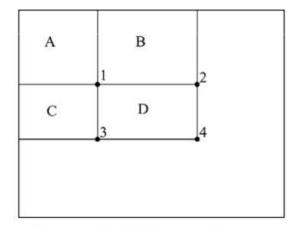


Figure 3. The sum of the pixels within rectangle D can be computed with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is A+B, at location 3 is A+C, and at location 4 is A+B+C+D. The sum within D can be computed as 4+1-(2+3).





Illustration du calcul de quelques attributs









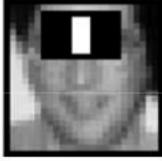


Figure 5. The first and second features selected by AdaBoost. The two features are shown in the top row and then overlayed on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.



Parcours de l'image

Différentes échelles

Il faut pouvoir détecter des visages de tailles différentes : soit en créant une pyramide d'images (avec réduction progressive de la taille de l'image), soit en modifiant la taille de la fenêtre de détection (choix de Viola & Jones)

Différentes positions

Il faut parcourir l'ensemble de l'image, ce qui pourrait être très long => idée de la cascade de classificateurs AdaBoost

- > 1er classificateur relativement léger, doit rejeter rapidement les zones ne contenant pas de visage
- > 2^{ème} plus complexe car prend en compte les éléments mal classés du 1er classificateur
- > Etc... jusqu'à atteindre un taux de fausses détections très faible.



Annexe : détection de plans

- Plan ou Shot en anglais
- La détection des plans permet de produire un résumé du contenu d'une vidéo
- Un plan: succession d'images produites dans un même flux caméra (sans interruption). Les images ont dont une même similarité visuelle du fait de l'unité de lieu et de temps.

Cette notion de plan (ou shot) n'est pas équivalente à celle des plans d'une image statique (notion de géométrie)



Annexe : détection de plans

Approche classique

- Extraction de caractéristique des images
- 2. Comparaison des caractéristiques successives afin de détecter les dissimilarités fortes.

Difficultés

- Les transitions entre plans sont plus ou moins abruptes
- Différentes caractéristiques capturent des dissimilarités différentes
- On pourrait aussi s'appuyer sur des caractéristiques sonores pour renforcer certaines détections.

Référence

http://www.videoanalysis.org/Prof._Dr._Rainer_Lienhart/Publications_files/spie99.pdf





Annexe : détection de mouvements

3 origines de mouvement :

Celui de la caméra

Celui des objets/personnages de la scène

Changements d'illumination

- Une idée de base : exploiter la redondance spatiotemporelle
- Quelques approches
 - Estimer le mouvement par mise en correspondance de points d'images successives ('block matching')
 - Compenser le mouvement : estimation du champs des déplacements
- Quelques difficultés
 - Coût de calcul
 - Occlusions