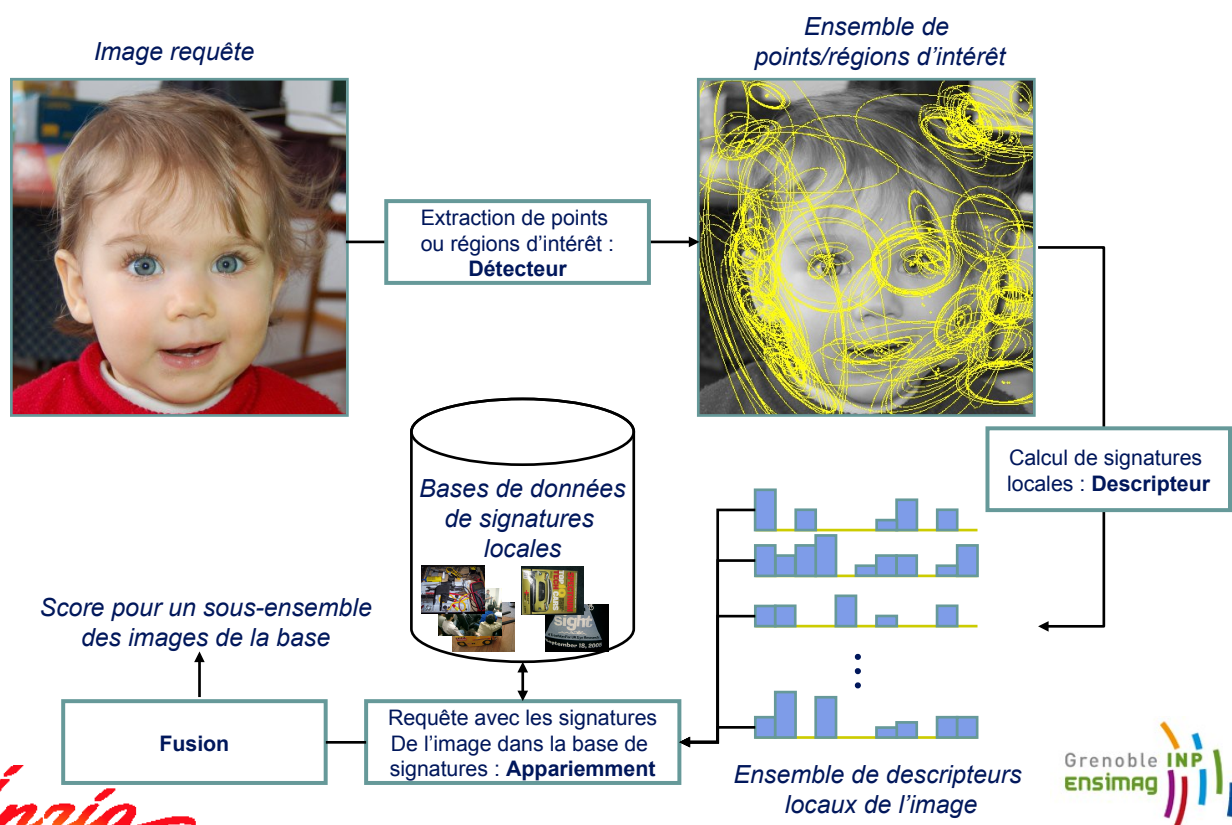


Bases de données multimédia IV – Description locale des images

ENSIMAG
2014-2015

Matthijs Douze & Karteek Alahari

Schéma type d'un système de description locale



Description locale : plan

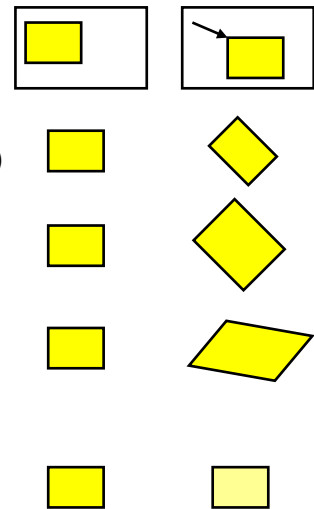
- Extraction de points/régions d'intérêt
- Descripteurs locaux
- Appariement de points

Inria



Extraction de régions d'intérêt : enjeux

- Extraire des régions qui soient invariants à de nombreuses transformations
- Transformations géométriques
 - ▶ translation
 - ▶ rotation
 - ▶ similitude (rotation + changement d'échelle isotrope)
 - ▶ affine (en 3d, textures localement planaires)
- Transformations photométriques
 - ▶ Changement d'intensité affine ($I \rightarrow a I + b$)



Inria



Base d'évaluation des points/régions d'intérêt

- Affine covariant region dataset :
 - ▶ 8 ensembles de 6 images
 - ▶ transformation 2D fournie (vérité terrain)



- Bikes et Trees : flou
- Graffiti et Wall : changement de point de vue
- Bark et Boat : zoom + rotation
- Leuven : changement de luminosité
- UBC : compression JPEG



Inria

Grenoble INP
ensimag

Principaux extracteurs de points/régions d'intérêt

- Harris
- Harris-Affine
- Hessian-Affine
- Edge-Based Region Detector (EBR)
- Intensity Extrema-Based Regions Detector (IBR)
- Maximally Stable Extremal Regions (MSER)
- Salient Region Detector

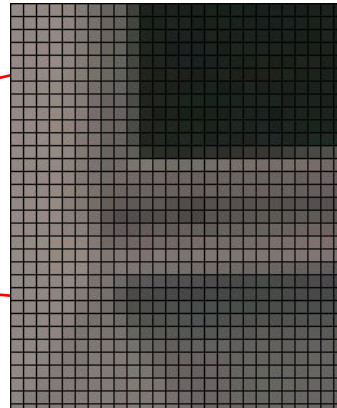
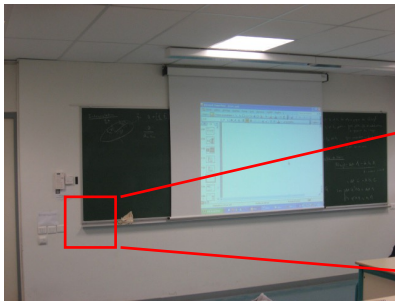
« A comparison of affine region detectors » K. Mikolajczyk et al., IJCV 2005l

Inria

Grenoble INP
ensimag

Détecteur de Harris : Introduction

- Détecter les « coins » = des structures
 - ▶ répétables de l'image
 - ▶ et localisent des endroits très discriminatifs d'une image
 - Harris
 - ▶ analyse locale
 - ▶ lorsqu'on est sur un « coin », un déplacement dans n'importe quelle direction produit un fort changement des niveaux de gris
- « A Combined Corner and Edge Detector », C. Harris et M. Stephens, 1988

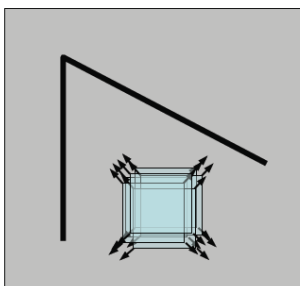


Inria

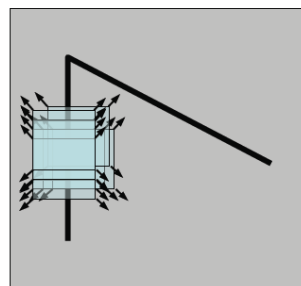


Détecteur de Harris

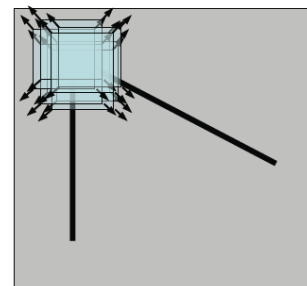
- Explication détaillée au tableau.



“flat” region:
no change in
all directions



“edge”:
no change along
the edge direction

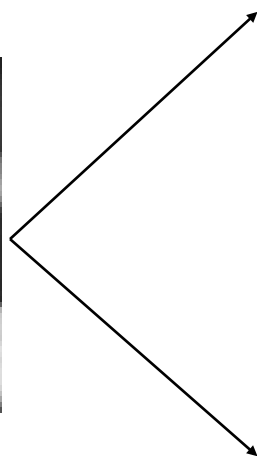


“corner”:
significant change
in all directions

Inria

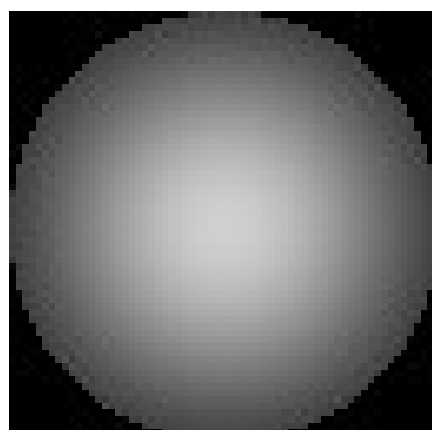


Harris : extraction du patch et calcul des gradients



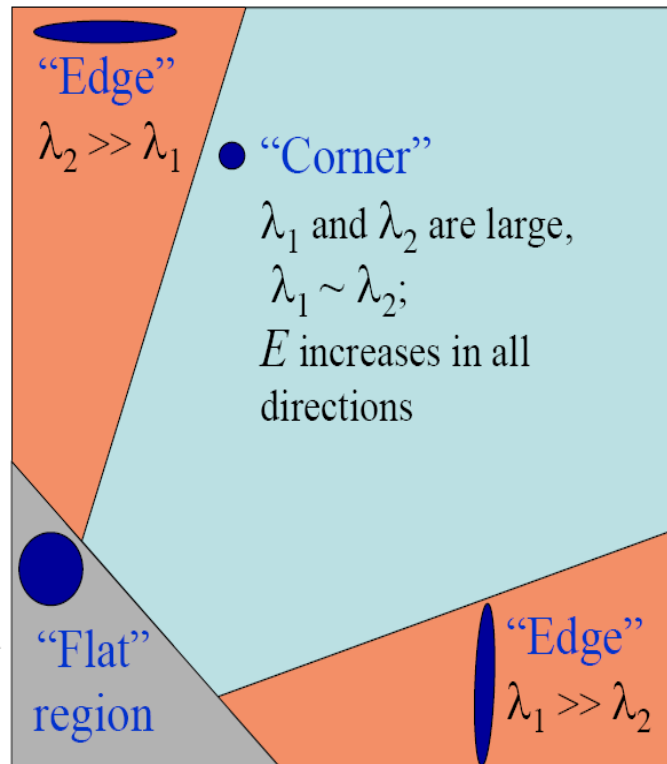
Inria

Harris : 2^e étape



Inria

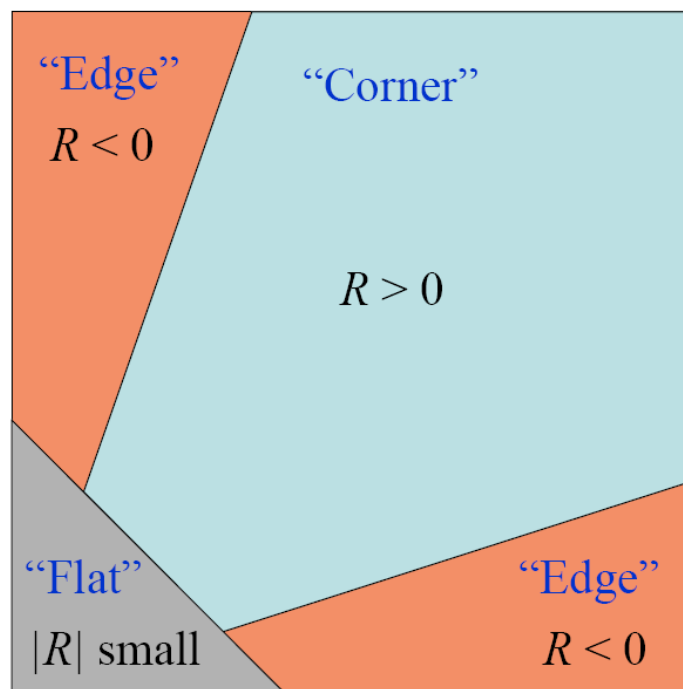
Harris : classification des types de régions



Inria



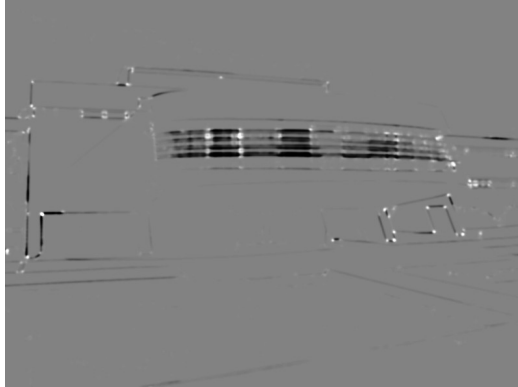
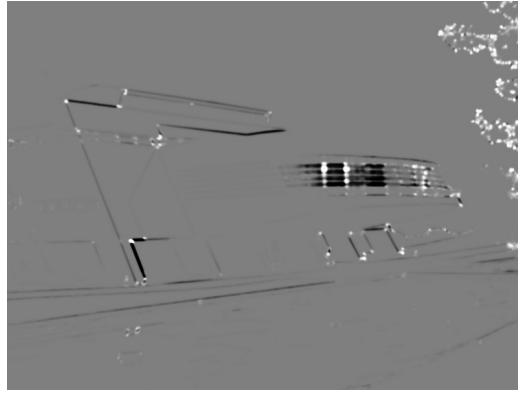
Harris : classification des régions : utilisation du seuil



Inria



Harris : exemple

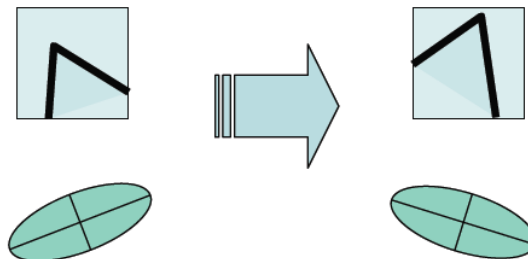


Inria

Grenoble INP
ensimag

Harris : Invariance

- Invariance à la translation : inhérente à la sélection des maxima locaux
- Invariance aux changements affines d'intensité
- Invariance à la rotation
 - ▶ l'ellipse tourne
 - ▶ mais la forme reste identique
 - ▶ les valeurs propres restent identiques



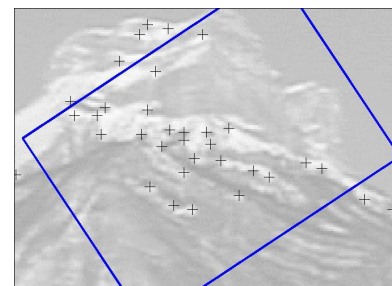
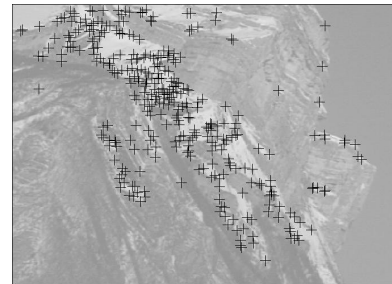
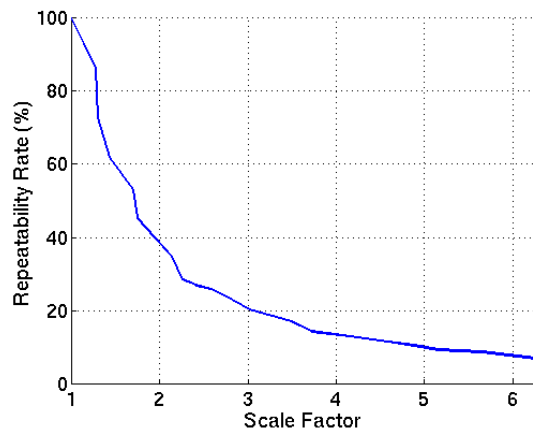
- Mais pas d'invariance à des changements d'échelle
 - ▶ exemple simple au tableau

Inria

Grenoble INP
ensimag

Harris : Non invariance à l'échelle

- Taux de répétabilité : définition à venir...



Inria

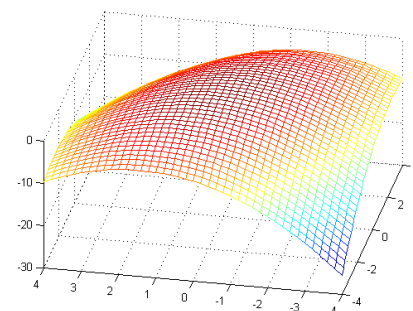
Grenoble INP
ensimag

Détecteurs hessien & DoG

- Analogue au détecteur de Harris
- Les valeurs propres de la matrice des dérivées secondes (hessienne) caractérisent la forme locale du voisinage d'un point.
- Max sur le déterminant

$$H = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}$$

- Forte réponse sur les « blobs » et les crêtes



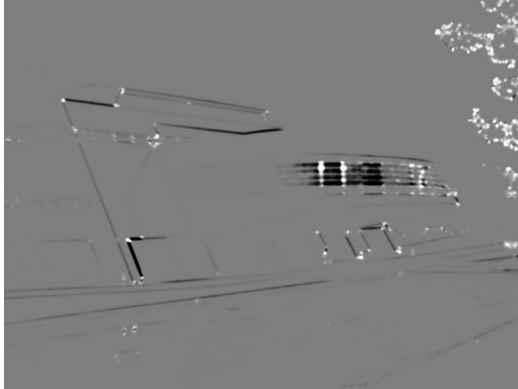
- Différence détecteur de coins/blobs
 - ▶ Blob classique : DoG

Inria

Grenoble INP
ensimag

Invariance à l'échelle (début)

- Étape suivante pour obtenir un extracteur invariant aux transformations affines
 - ▶ avec l'invariance à l'échelle, Harris devient invariant à une similitude
- Echelle de calcul de Harris/hessien/laplacien/DoG
 - ▶ Déterminée par la taille des filtres de dérivation : σ des filtres gaussiens
 - ▶ Réagit à une certaine taille de détails / une certaine plage de fréquences spatiales



Inria

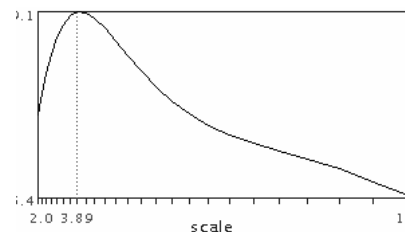
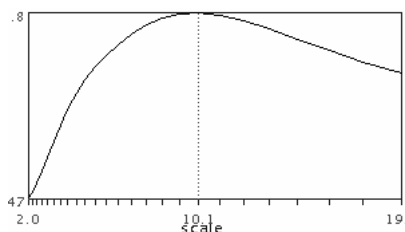


Invariance à l'échelle (suite)

- Noyau (kernel) : Laplacien ou DoG
- Sélection d'une échelle caractéristique :
 - ▶ prendre les valeurs extrêmes d'une fonction f liée à l'échelle
 - ▶ Une bonne fonction f présente des pics marqués



Laplacien



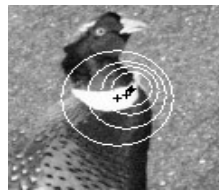
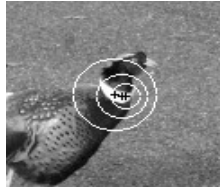
invariance de l'échelle caractéristique

Inria

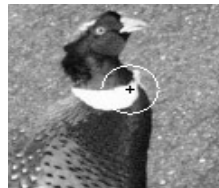


Invariance à l'échelle (suite)

- Détecteur de Harris-Laplace
 - ▶ Sélectionner le maximum de Harris en espace pour plusieurs échelles



- ▶ sélection des points à leur échelle caractéristiques avec le Laplacien



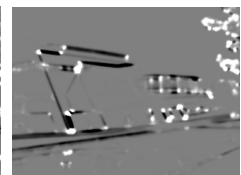
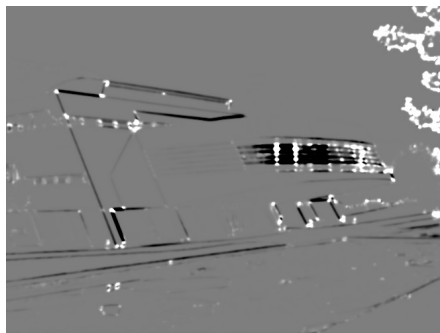
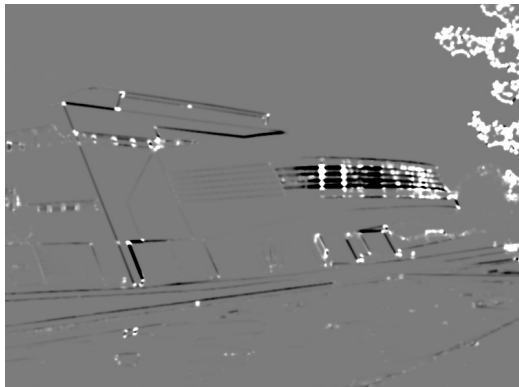
- Dans SIFT: recherche des maxima de DoG et position et en échelle

Inria



Invariance à l'échelle (fin)

- Implémentation d'un filtre gaussien
 - ▶ Valeurs coupées à w proportionnel à σ , séparable
 - ▶ Complexité en $O(w \times x_d \times y_d)$
- Multi-échelle en pratique:
 - ▶ pyramide d'images
 - ▶ Gaussiennes de taille constante



Inria



Harris-Affine et Hessian Affine (début)

- Détecteur de région de l'état de l'art

K. Mikolajczyk, C. Schmid An affine invariant interest point detector, ECCV 2002

- Algorithme : estimation itérative des paramètres
 - ▶ localisation : Harris
 - ▶ échelle : sélection automatique avec le Laplacien
 - ▶ voisinage affine : normalisation avec la matrice des seconds moments

→ jusqu'à convergence

- Initialisation à partir de points d'intérêts multi-échelle

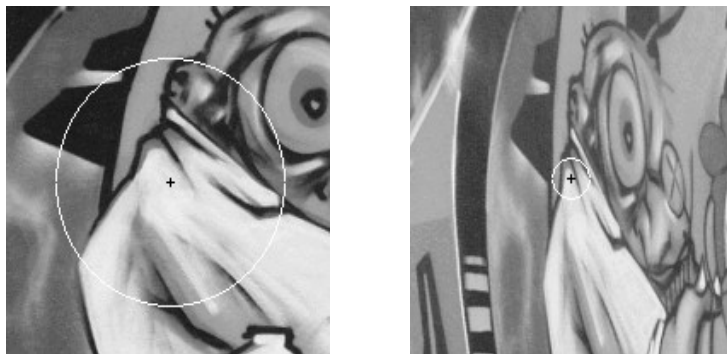
Inria



Harris-Affine et Hessian Affine (suite)

- Estimation itérative de la localisation, de l'échelle, du voisinage

Points initiaux



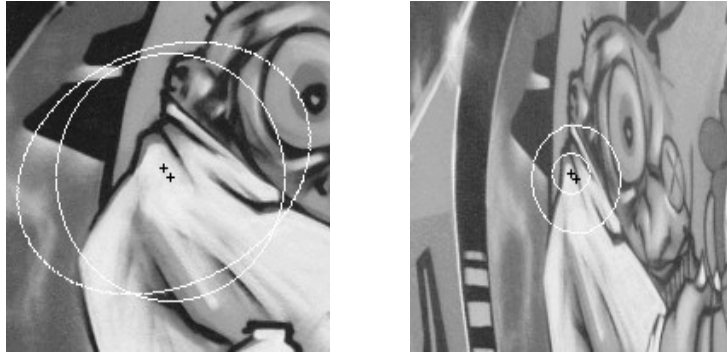
Inria



Harris-Affine et Hessian Affine (suite)

- Estimation itérative de la localisation, de l'échelle, du voisinage

Iteration #1



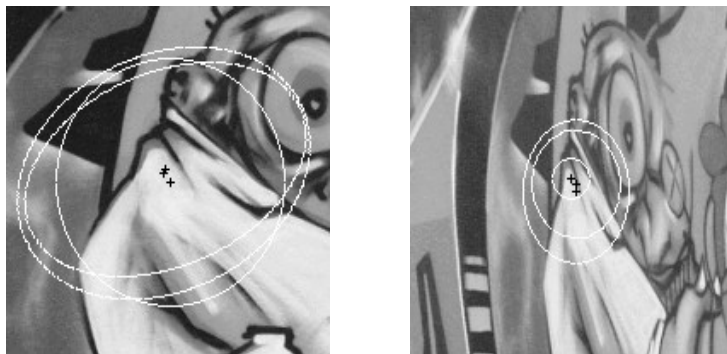
Inria



Harris-Affine et Hessian Affine (suite)

- Estimation itérative de la localisation, de l'échelle, du voisinage

Iteration #2



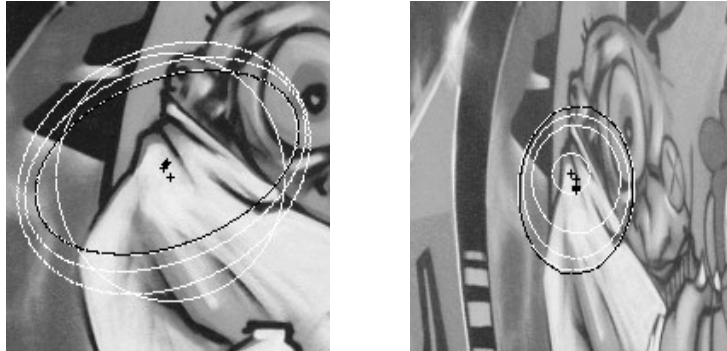
Inria



Harris-Affine et Hessian Affine (suite)

- Estimation itérative de la localisation, de l'échelle, du voisinage

Iteration #3, #4, ...

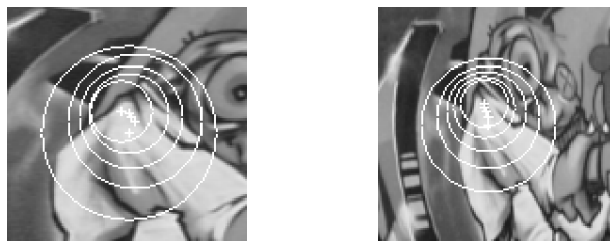


Inria



Harris-Affine et Hessian Affine : pour résumer

- Initialisation avec des points d'intérêt multi-échelle



- Estimation itérative de la localisation, de l'échelle, du voisinage

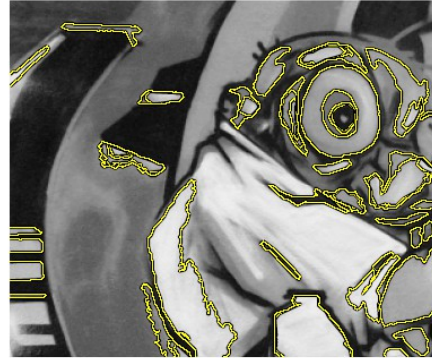


Inria



MSER

- Principe (voir séquence de seuils)
 - ▶ trier les pixels par intensité
 - ▶ tous les seuils d'intensité possibles sont testés
 - ▶ localisation des régions qui sont stables par changement du seuil
- Pourquoi ça marche (bien)
 - ▶ la binarisation locale est stable sur une large plage de seuils



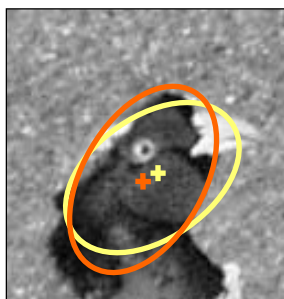
Robust wide-baseline stereo from maximally stable extremal regions, J. Matas., O. Chum, M. Urbana and T. Pajdla, Image and Vision Computing 22(10), 2004

Inria

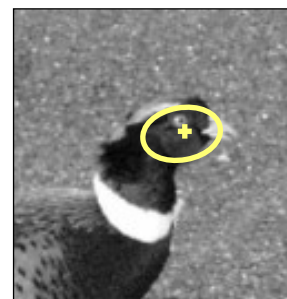


Comment évaluer un extracteur ?

- Évaluation quantitative des détecteurs de régions d'intérêt
 - ▶ les points / régions doivent être à la même position relative
- Mesure de performance : **le taux de répétabilité**
 - ▶ taux de régions qui se correspondent (vérité terrain 2D)



homographie



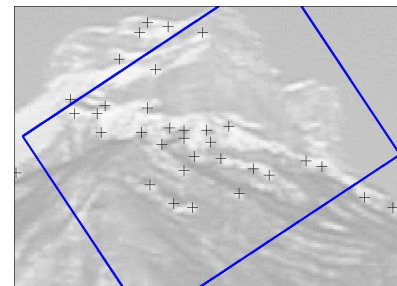
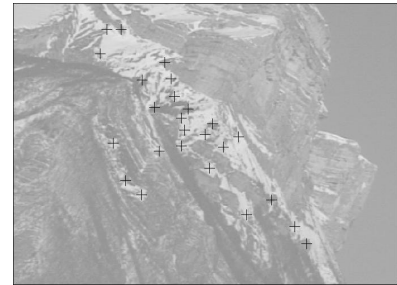
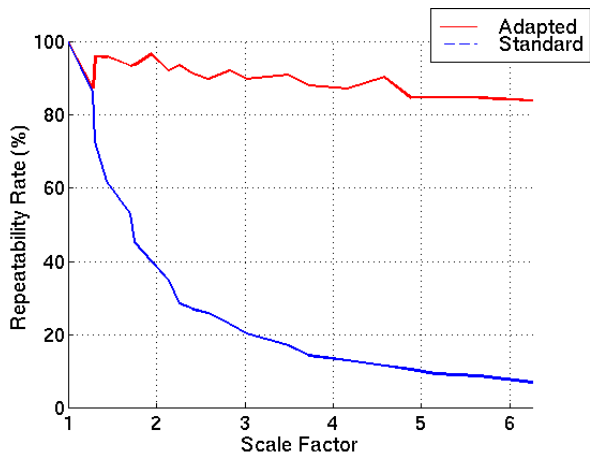
- Deux points se correspondent si
 - ▶ l'erreur de localisation est inférieure à 1.5 pixel
 - ▶ l'erreur d'intersection est inférieure à 20%

Inria

A Comparison of Affine Region Detectors, Mikolajczyk et al., IJCV 2005

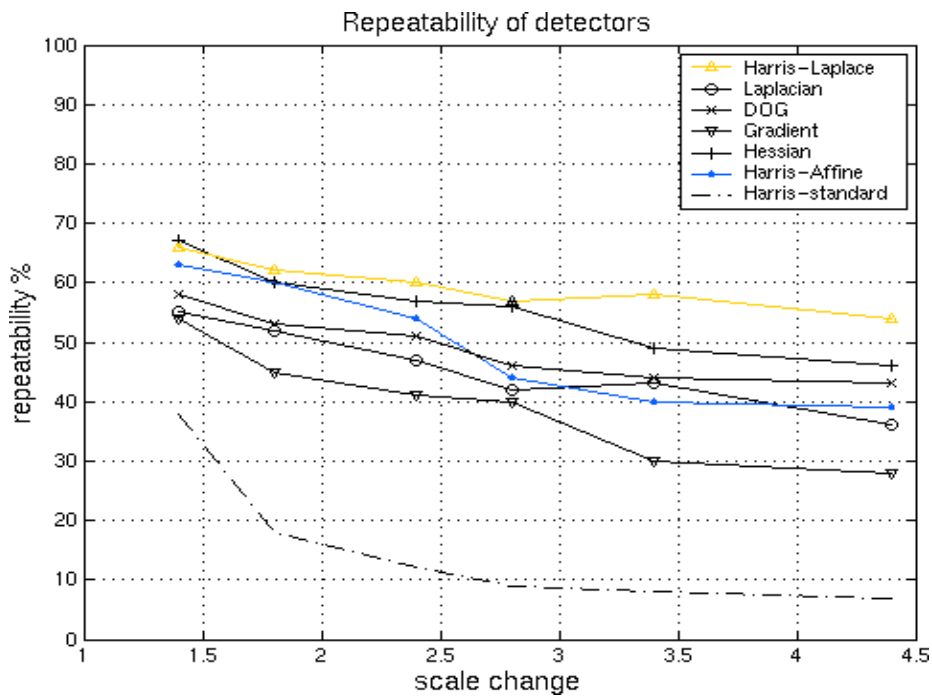


Harris vs Harris invariant à l'échelle



Comparaison des détecteurs

répétabilité- changement d'échelle

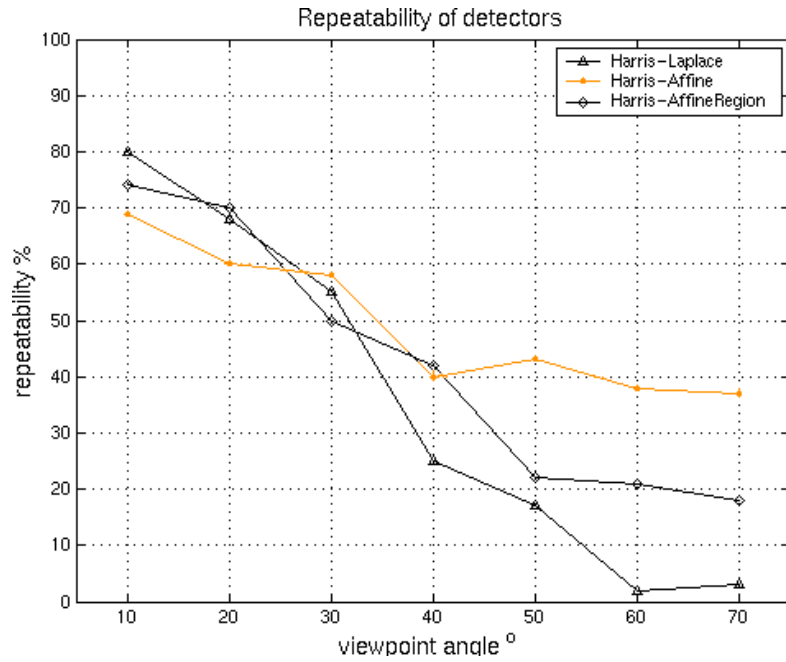


Comparaison des détecteurs



Inria

répétabilité – transformation perspective



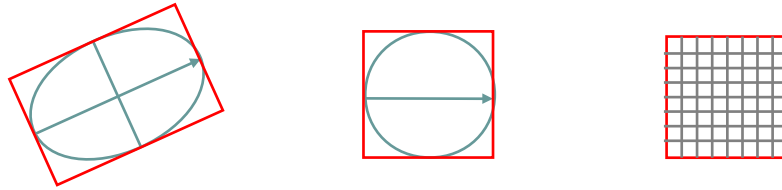
Description locale : plan

- Extraction de points/régions d'intérêt
- **Descripteurs locaux**
- Appariement de points

Inria

Et une fois qu'on a défini l'ellipse ?

- Synthèse du patch (imagerie)
 - ▶ On calcule un patch de taille donnée



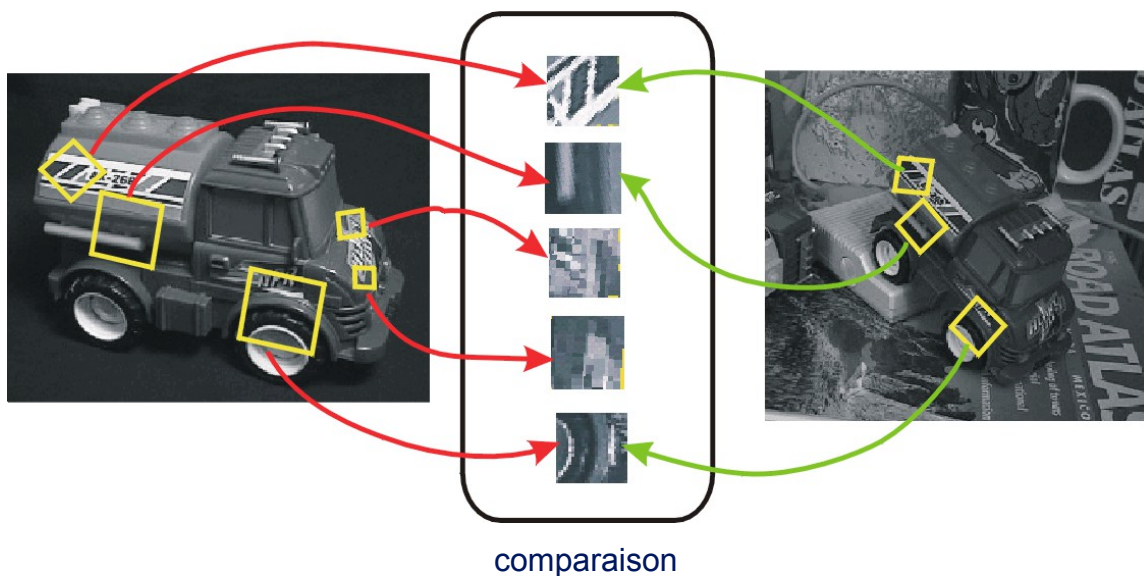
- ▶ Problème d'aliasing (apparition de haute fréquence) : application d'un filtrage (filtrage convolutif, par exemple gaussien)
- ▶ En pratique: sur la pyramide d'images

- Tailles typiques de patch : 21x21, 64x64

Inria

Grenoble INP
ensimag

Objectif (début)

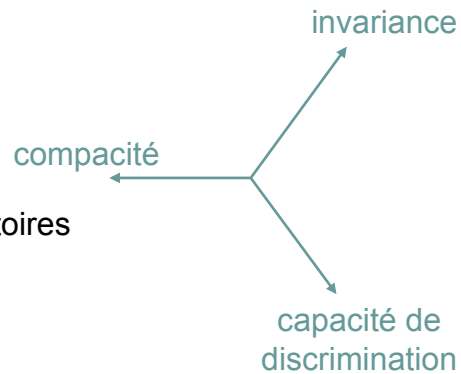


Inria

Grenoble INP
ensimag

Objectif (fin)

- En utilisant un détecteur, on est capable de détecter des zones robustes (et informatives) de l'image
- Mais à ce stade, ces zones ne peuvent pas être comparées entre elles
- Qu'est qu'un descripteur ?
 - ▶ une représentation de la zone d'intérêt
 - ▶ sous la forme d'un vecteur
 - ▶ qui appartient à un espace *muni d'une distance*
- Qu'est qu'un bon descripteur ?
 - ▶ invariant
 - ▶ discriminant
 - ▶ compact
- Remarque : ces objectifs sont contradictoires

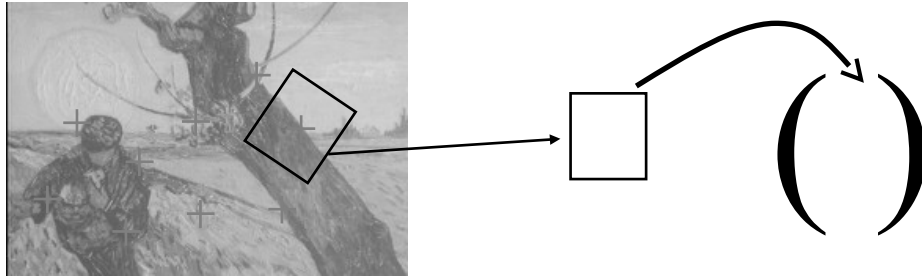


Descripteurs

- Descripteurs basiques
- Descripteurs invariants à la luminosité/rotation : invariants différentiels
- Descripteurs à base de moments couleurs affine-invariants
- État de l'art : le descripteur SIFT
- Descripteur CS-LBP
- Descripteurs de régions texturées

Descripteur pixel

- On range les pixels au voisinage du point dans un vecteur



- Calcul simple
- Volumineux (patch de 32x32 = vecteur de 1024 éléments)
- Invariance possible aux transfo affines de luminosité
- Utilisé comme « baseline » en classification

descripteur local

Inria



Invariants différentiels (début)

- Descripteurs de *points* (on ne part pas d'un patch)
- Basés sur des dérivées calculées en un point (x,y). Idée :

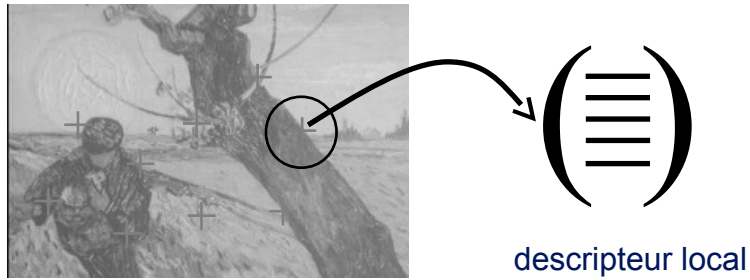
$$v(x, y) = \begin{bmatrix} I(x, y) \\ I_x(x, y) \\ I_y(x, y) \\ I_{xx}(x, y) \\ I_{xy}(x, y) \\ I_{yy}(x, y) \\ \vdots \end{bmatrix}$$

- Normalisation pour être invariant aux changements affines de luminosité

Inria



Invariants différentiels : interprétation



- Approximation locale d'une fonction par la série de Taylor
→ lien avec le détecteur Hessien
- Le descripteur : les coefficients de cette série tronquée décrivent le voisinage d'un point

Inria

Grenoble INP
ensimag

Invariants différentiels (suite)

- Calcul stable des dérivées par convolution avec dérivées gaussiennes

$$v(x, y) = \begin{bmatrix} I(x, y) * G(\sigma) \\ I(x, y) * G_x(\sigma) \\ I(x, y) * G_y(\sigma) \\ I(x, y) * G_{xx}(\sigma) \\ I(x, y) * G_{xy}(\sigma) \\ I(x, y) * G_{yy}(\sigma) \\ \vdots \end{bmatrix}$$

où σ correspond à un choix d'échelle

- ▶ σ fixé si normalisation en échelle effectué avant, (invariants différentiels développés plus tôt que la normalisation en échelle)
- En pratique: pyramide d'images (convolution en $O(\sigma^2)$)

Inria

Grenoble INP
ensimag

Invariants différentiels : invariance à la rotation

- [Koen 87]
- Invariant à la rotation (redondant avec normalisation en orientation)

$$\begin{array}{c}
 I_x I_x + I_y I_y \\
 I_{xx} I_x I_x + 2 I_{xy} I_x I_y + I_{yy} I_x I_x \\
 I_{xx} I_{yy} \\
 I_{xx} I_{xx} + 2 I_{xy} I_{xy} + I_{yy} I_{yy} \\
 \vdots \\
 \vdots
 \end{array}$$

Trace de la
matrice de Harris

Invariance à un changement
de luminosité affine :

$$\begin{array}{c}
 \frac{I_{xx} I_x I_x + 2 I_{xy} I_x I_y + I_{yy} I_x I_x}{(I_x I_x + I_y I_y)^{3/2}} \\
 \frac{I_{xx} I_{yy}}{(I_x I_x + I_y I_y)^{1/2}} \\
 \frac{I_{xx} I_{xx} + 2 I_{xy} I_{xy} + I_{yy} I_{yy}}{I_x I_x + I_y I_y} \\
 \vdots \\
 \vdots
 \end{array}$$

Inria



Invariants différentiels (fin)

- Remarque: changement d'échelle $I_2(x) = I_1(s x) \rightarrow$ dérivées liées par

$$I_1(x) * G_{i_1 \dots i_n}(\sigma) = s^n I_2(s x) * G_{i_1 \dots i_n}(s \sigma)$$

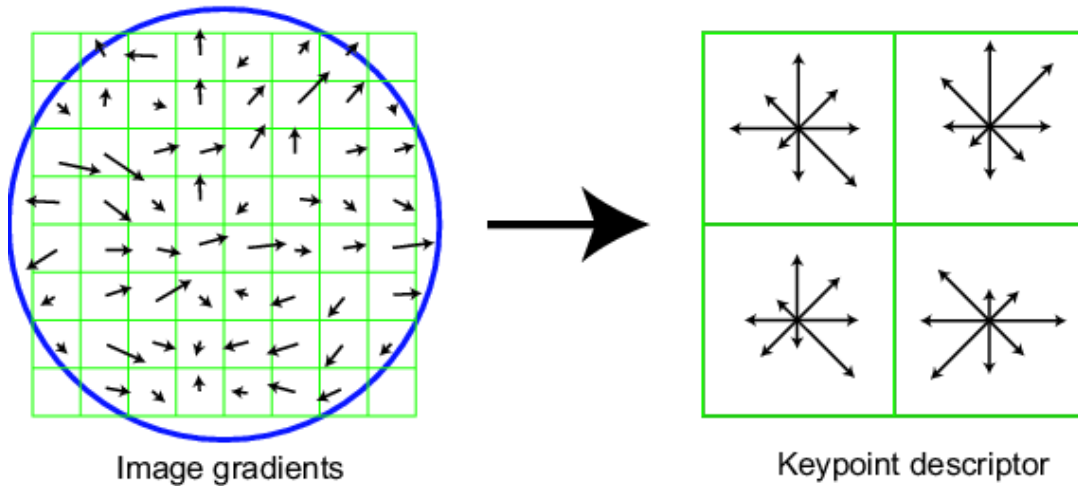
- Taille typique du descripteur : 24
- Aujourd'hui, descripteurs peu utilisés

Inria



SIFT (Scale invariant feature transform) [Lowe 04]

- ▶ Calculé sur un patch déjà normalisé : échelle (ou affine) + en orientation
- ▶ description détaillée au tableau



Inria

le INP
ensimag

CS-LBP (Center-symmetric local binary pattern)

- Description au tableau
- nombreuses autres variantes de SIFT:
 - ▶ HOG (generalization)
 - ▶ GLOH (log-polar bins)
 - ▶ SURF (approximation, Blob detector, HAAR wavelets)
 - ▶ Daisy (convolutional filters)
 - ▶

M. Heikkila, M. Pietikainen, C. Schmid, "Description of interest regions with local binary patterns"

E. Tola, V. Lepetit, P. Fua, "A fast local descriptor for dense matching". CVPR 2008

Inria

Grenoble INP
ensimag

Comparaison de descripteurs

- Il faut définir une distance
 - ▶ distances = L2, Hellinger, Hamming
- On se donne des paires d'images pour lesquelles on connaît les points qui doivent être appariés
 - ▶ vérité terrain au niveau pixel
 - ▶ Ex: base de Krystian Mikolajczyk
- On fixe une distance et un seuil μ pour lequel on considère que deux descripteurs p (dans I_1) et q (dans I_2) se correspondent :

$$\text{dist}(p,q) < \mu$$

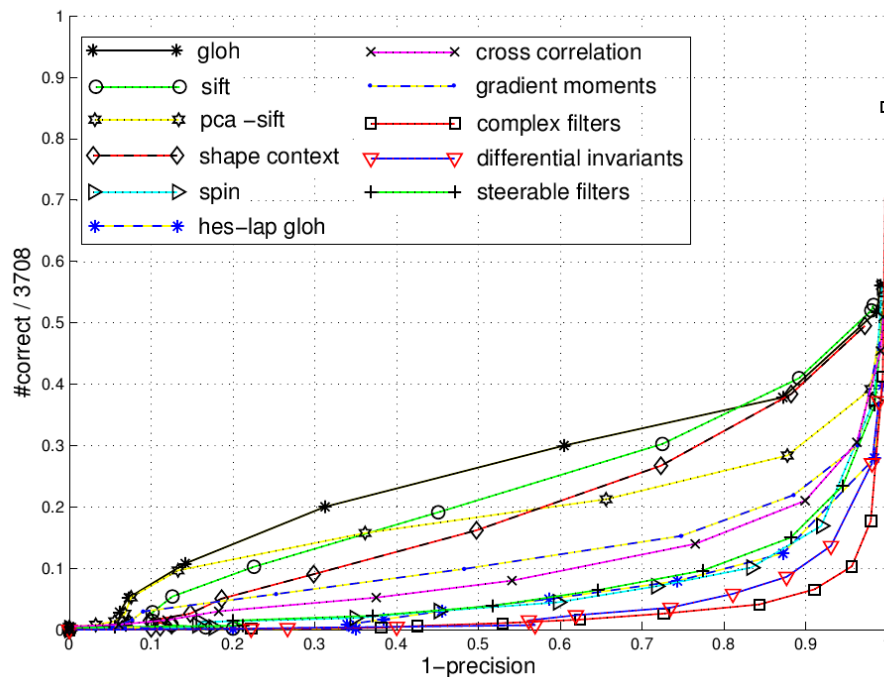
- En faisant varier le seuil μ , on obtient une courbe precision/rappel qui caractérise la qualité du descripteur

Inria

Grenoble INP
ensimag

Résultats de comparaison

A performance evaluation of local descriptors Krystian Mikolajczyk, Cordelia Schmid, PAMI, Volume 27, Number 10 - 2005



Inria

Grenoble INP
ensimag

Descripteurs : conclusion

- **Invariances** : force des méthodes génériques
 - ▶ normalisation affine de la région vers un patch de taille constante en utilisant un détecteur affine invariant
 - ▶ normalisation affine en luminosité avant le calcul du descripteur pour garantir l'invariance affine
- Utilisation d'un descripteur de relativement grande dimension (128 pour SIFT) pour obtenir une bonne capacité de **discrimination**

Inria



Description locale : plan

- Extraction de points/régions d'intérêt
- Descripteurs locaux
- **Appariement de points**

Inria



Appariement de points

- Problème : étant donné des points ou régions d'intérêt et les descripteurs associés, comment apparier de manière robuste les points de deux images ?
- Dans ce sous-chapitre : sélection des meilleurs couples de points

Inria



Comparer directement deux régions normalisées

- À partir des détecteurs uniquement : plusieurs distances
 - ▶ SSD (sum of square differences)
 - ▶ NSSD (normalized SSD)
 - ▶ ZNCC (zero-normalized cross-correlation)
- Rôle concurrent avec les descripteurs, pour lesquels la mesure de proximité entre régions donne une ressemblance entre celles-ci. On conserve le couple si cette mesure de dissimilarité est inférieure à un seuil μ :

$$d(p_{1i}, p_{2j}) < \mu, \text{ avec } p_{1i} \text{ dans } I_1 \text{ et } p_{2j} \text{ dans } I_2$$

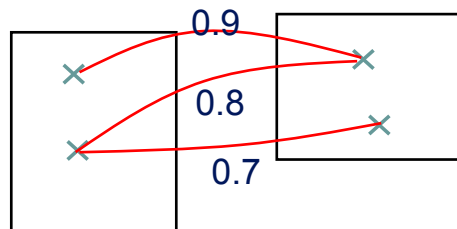
- μ est fixé à partir de la courbe précision/rappel du descripteur
- Comparaison entre descripteurs est + rapide (vecteur plus court)

Inria



Sélection des meilleurs couples

- Hypothèse : pour tout couple de points (p_{1i}, p_{2j}) , (p_{1i} dans I_1 et p_{2j} dans I_2), on a une mesure de similarité s_{ij} entre régions
- Remarque: ambiguïté



- Différentes stratégies:
 - ▶ Winner-takes-all (symétrique ou à sens unique)
 - ▶ Filtrage ultérieur par géométrie

Inria



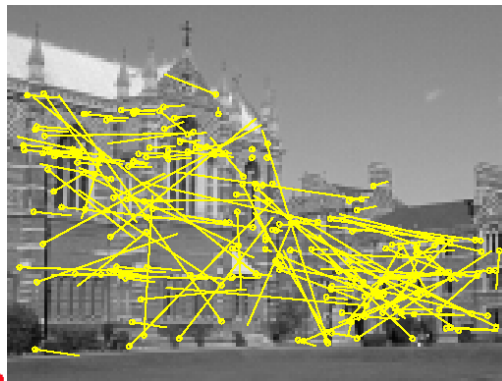
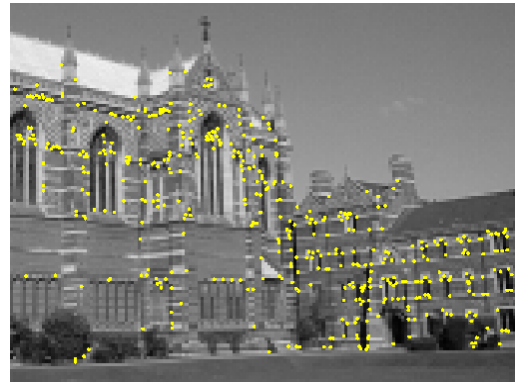
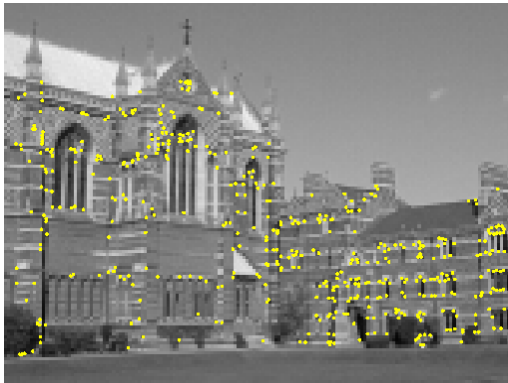
Vérification géométrique par mise en correspondance d'images

- Contraintes de voisinages
- RANSAC
- LO-RANSAC
- Transformée de Hough

Inria



Problème : appariements entre deux images



Inria



Vérification par un modèle géométrique

- Pour une paire d'images I, I'
- Il existe un vecteur de paramètres p
- tels que 2 points en correspondance (x, x') vérifient
$$F(x, x', p) = 0$$
- le plus souvent
$$x' = T(x, p)$$
- En indexation:
 - ▶ si les images sont en correspondance, alors il existe p tel que "suffisamment" de points (x, x') vérifient $F(x, x', p) = 0$
- donc estimation de p
- nb de points qui vérifient l'équation avec le meilleur p = mesure de ressemblance entre images

Inria



Exploitation de la géométrie épipolaire

- Hypothèse : I_1 et I_2 sont deux projections de la même scène 3D
- donne la localisation
- Voir le cours d'introduction à la vision par ordinateur

- Dans le cas de scènes planaires, on simplifie le problème
 - ▶ Mise en correspondance directe par homographie

Inria

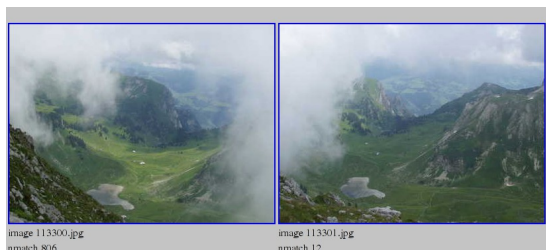


Sous-classes de transformations

- Pour diminuer l'incertitude et/ou en en fonction du problème, on réduit le nombre de degré de libertés des transformations possibles
 - ▶ Ex: copyright
- Le plus souvent, on se contente de transformées planaires
- Exemple d'utilisation dans un système de recherche de scènes :

[6] *Philbin, J. , Chum, O. , Isard, M. , Sivic, J. and Zisserman, A.*
Object retrieval with large vocabularies and fast spatial matching
<http://www.robots.ox.ac.uk/~vgg/publications/papers/philbin07.pdf>

→ différentes classes de transformations, suppression de l'invariance à la rotation



Inria



Hierarchie des transformations planaires (2D ↔ 2D)

	nombre de degré de liberté	invariants géométriques	forme
translation	2	tout, sauf les positions absolues	$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
transformée rigide	3	longueur, angle, surface, courbure	$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
similitude	4	angle, rapport de longueur	$\begin{bmatrix} X' \\ Y' \end{bmatrix} = s \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
transformation affine	6	parallélisme, rapport de surface, rapport de longueur sur une droite, coordonnées barycentriques	$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
Homographie	8	Birapport	$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \frac{1}{h_{31}X + h_{32}Y + h_{33}} \left(\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} h_{13} \\ h_{23} \end{bmatrix} \right)$

Inria

Grenoble INP
ensimag

Transformée de Hough

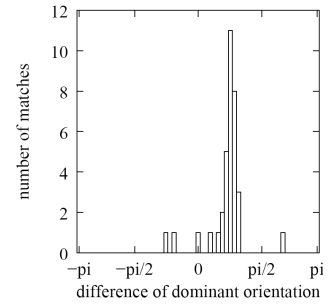
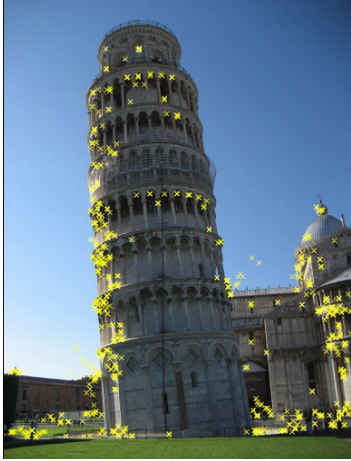
- Utilisé initialement pour détecter des segments/droites à partir de points dans les images
 - ▶ Chercher les droites possibles dans un espace de paramètre, (espace de Hough)
- Principe :
 - ▶ discrétiser l'espace de paramètres en "bins"
 - ▶ Pour chaque observation, incrémenter le(s) bin(s) consistants avec l'observation
 - ▶ Résultat = le bin de le plus peuplé
 - ▶ affinage possible avec Mean-shift
- Difficulté: dépendance à la paramétrisation et la quantification
- Dans le cas de la vérification géométrique, le nombre de degrés de liberté possibles doit être petit (Pourquoi?)

Inria

Grenoble INP
ensimag

Vérification géométrique « faible » (WGC)

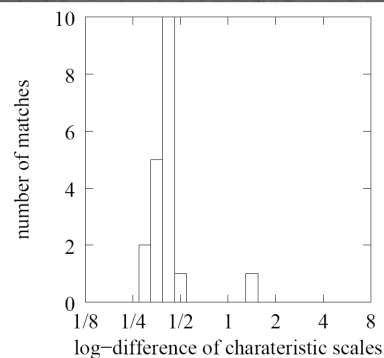
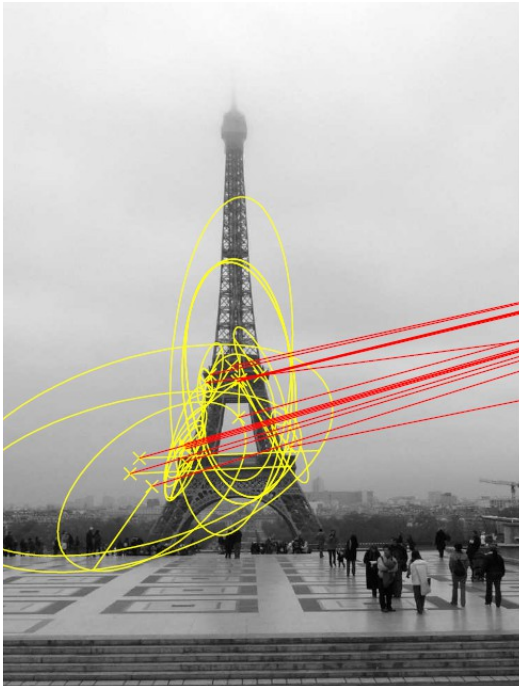
- Version très simplifiée d'un modèle géométrique appliquée au niveau point
- Utilise les infos de normalisation en rotation/échelle
- Rapide: appliqué sur les points indépendamment



Inria



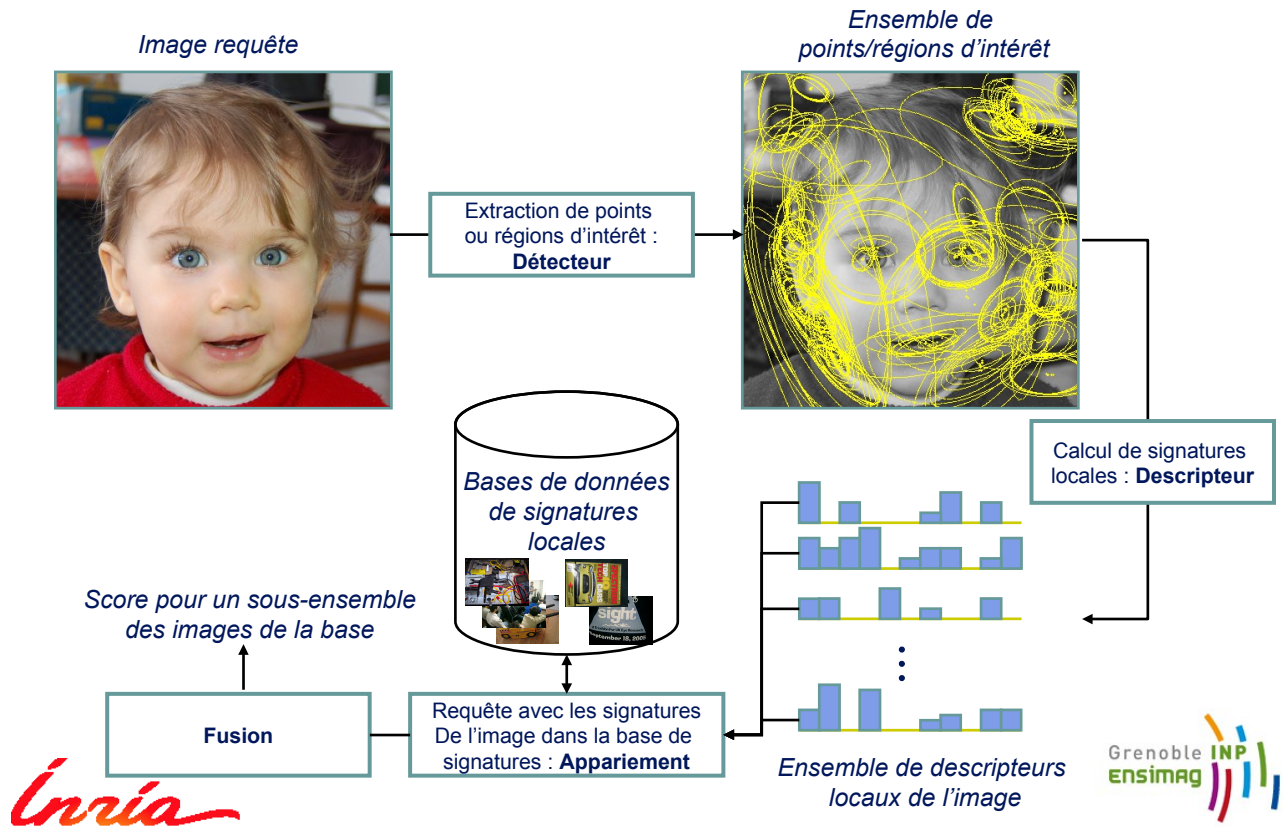
WGC: en échelle



Inria



Description locale des images : résumé et commentaires



Description locale des images : résumé et commentaires

- Étapes
 - ▶ recherche de points/régions d'intérêt
 - ▶ calcul des descripteurs pour chacune de ces régions
 - ▶ vérification de la cohérence géométrique
 - ▶ algorithme de vote
- Commentaires:
 - ▶ problème en suspens : il faut apparier les descripteurs d'une image avec tous les descripteurs d'une base d'image
→ c'est très lent !
 - ▶ la vérification de la cohérence géométrique est relativement lente
→ on ne peut l'utiliser que pour un sous-ensemble d'images *a priori* plus pertinentes