

○○
○
○
○○○○
○○○○○○○○

○○○○○○○○○○○○○○
○○○○○○○○○○○○○○
○○

Introduction au suivi d'objets visuels

Détection et mise en correspondance de points d'intérêt

20 novembre 2017



Plan

Introduction

Détection de points d'intérêt

- Définition – caractéristiques

- Fonctionnement

- Exemple de détecteur mono-échelle : Moravec/Harris

- Exemple de détecteurs multi-échelle : SIFT

Mise en correspondance

- Introduction

- Contraintes pour la mise en correspondance

- Appariement basé SIFT



Utilisation/application d'appariement de points

Suivi



Reconnaissance des formes



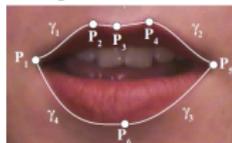
Reconstruction 3D



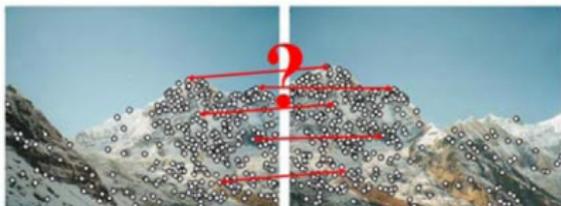
Indexation



Segmentation



Mosaïque



○○
○
○○○○
○○○○○○○○

○○○○○○○○○○○○○○
○○○○○○○○○○○○○○
○○

Éléments à considérer pour suivre/mettre en correspondance des points



Éléments à considérer pour suivre/mettre en correspondance des points

- Type d'éléments : points/pixels, régions, objets (*features*, *regions/points of interest*)
- Critères utilisés pour le suivi
- Hypothèses sur les mouvements 2D apparents
- Hypothèses sur les déformations
- Hypothèses sur les variations photométriques
- Méthode de suivi ou choix de la méthode de suivi
 - dense/éparse
 - éparse suivie de propagation (détection de points d'intérêt, d'objets d'intérêt)
 - Par opposition aux méthodes "mixtes" permettant de réaliser suivi et détection simultanément



Plan

Introduction

Détection de points d'intérêt

Définition – caractéristiques

Fonctionnement

Exemple de détecteur mono-échelle : Moravec/Harris

Exemple de détecteurs multi-échelle : SIFT

Mise en correspondance

Introduction

Contraintes pour la mise en correspondance

Appariement basé SIFT



Définition d'un point d'intérêt – principe d'un détecteur



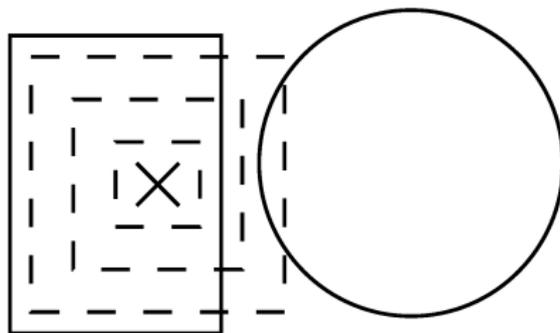
Définition d'un point d'intérêt – principe d'un détecteur

1. Points qui sont les projections de points particuliers de la scène (coins, jonctions)
2. Points ayant des caractéristiques particulières : fort contraste, texture forte, courbure forte



Type de détecteurs : mono ou multi-échelle

Influence du niveau d'analyse des objets



zones en pointillées = différentes tailles d'observation (échelles)



Fonctionnement d'un détecteur

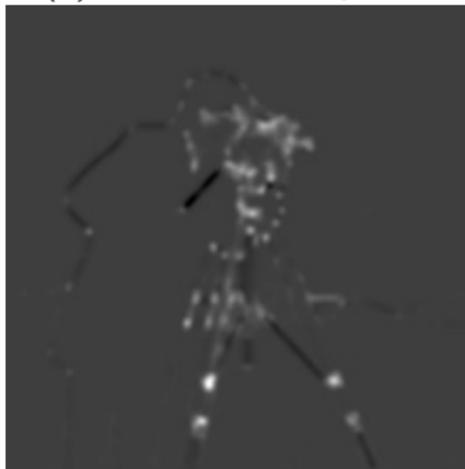
Image originale





Fonctionnement d'un détecteur

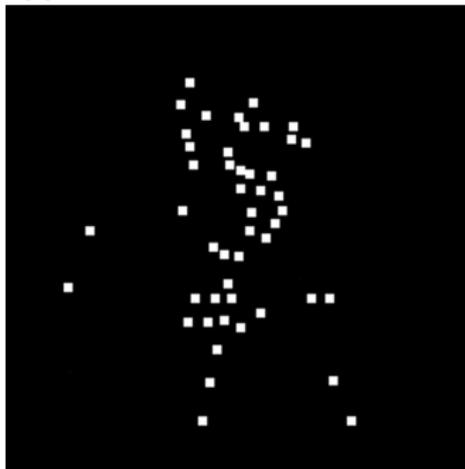
(1) Calcul d'une réponse





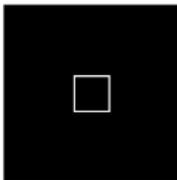
Fonctionnement d'un détecteur

(2) Suppression des non-maxima locaux





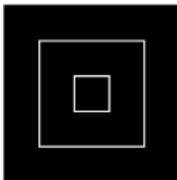
Moravec



Point quelconque



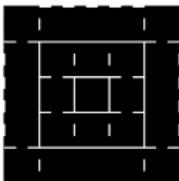
Moravec



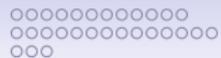
Point quelconque



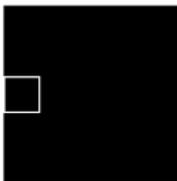
Moravec



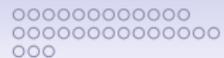
Point quelconque



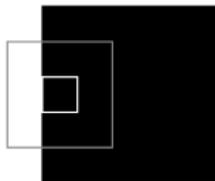
Moravec



Point contour



Moravec



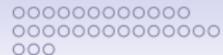
Point contour



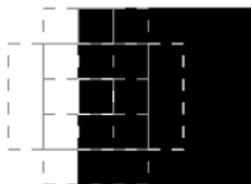
Moravec



Point contour



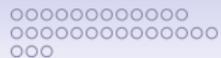
Moravec



Point contour



direction contour



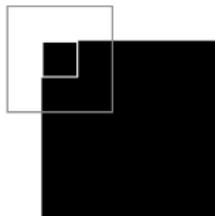
Moravec



Point de type coin



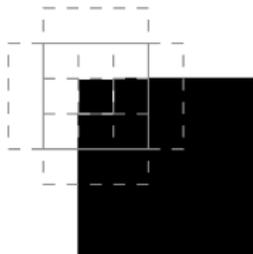
Moravec



Point de type coin



Moravec



Point de type coin



Moravec : définition

Soit (u, v) un déplacement, l'impact de ce déplacement se mesure par :

$$E_{(u,v)}(x, y) = \sum_{x,y} w(x, y) (I(x + u, y + v) - I(x, y))^2$$

où :

1. w : poids
2. $E_{(u,v)}(x, y) =$ **moyenne du changement** si on déplace la fenêtre



Moravec : définition

Soit (u, v) un déplacement, l'impact de ce déplacement se mesure par :

$$E_{(u,v)}(x, y) = \sum_{x,y} w(x, y) (I(x + u, y + v) - I(x, y))^2$$

où :

1. w : poids
2. $E_{(u,v)}(x, y) =$ **moyenne du changement** si on déplace la fenêtre

La réponse du détecteur est donc la valeur minimale parmi toutes les directions testées :

$$Mor(x, y) = \min_{(u,v)} E_{(u,v)}(x, y)$$



Harris : améliorations de Moravec

1. **La réponse est bruitée en raison des poids binaires** : utilisation des poids Gaussiens, de moyennes nulles



Harris : améliorations de Moravec

1. **La réponse est bruitée en raison des poids binaires** : utilisation des poids Gaussiens, de moyennes nulles
2. **La réponse du détecteur est anisotrope (dépendant de la direction)**



Harris : améliorations de Moravec

1. **La réponse est bruitée en raison des poids binaires** : utilisation des poids Gaussiens, de moyennes nulles
2. **La réponse du détecteur est anisotrope (dépendant de la direction)**

ET

3. **La réponse est trop forte/ambiguë au niveau des contours**



Vers le multi-échelle

- Détecteurs mono-échelle
 - Dépendant d'un paramètre : l'échelle
 - Détection des points ou régions à une échelle unique
 - Échelle *a priori* inconnue



Vers le multi-échelle

- Détecteurs mono-échelle
 - Dépendant d'un paramètre : l'échelle
 - Détection des points ou régions à une échelle unique
 - Échelle *a priori* inconnue
- Approches multi-échelle : détermination automatique de la "bonne" échelle



Vers le multi-échelle

- Détecteurs mono-échelle
 - Dépendant d'un paramètre : l'échelle
 - Détection des points ou régions à une échelle unique
 - Échelle *a priori* inconnue
- Approches multi-échelle : détermination automatique de la "bonne" échelle
- Dans la littérature, noyau gaussien = seul lissage linéaire permettant une analyse multi-échelle correcte
- Outil nécessaire : Pyramide gaussienne et localisation des extrema en x , y , σ , espace "spatial-échelle" 3D



Construction d'une pyramide

- But : produire à partir d'une image, un ensemble d'images de résolution de plus en plus faibles



Construction d'une pyramide

- But : produire à partir d'une image, un ensemble d'images de résolution de plus en plus faibles
- Plusieurs techniques développées dans le domaine de la compression d'images



Construction d'une pyramide

- But : produire à partir d'une image, un ensemble d'images de résolution de plus en plus faibles
- Plusieurs techniques développées dans le domaine de la compression d'images
- Différents types de pyramides : Pyramides **régulières** ou irrégulières



Construction d'une pyramide régulière

- *Pyramides gaussiennes* – Lissage par un filtre gaussien, puis sous-échantillonnage, en récupérant un pixel sur quatre
- *Pyramides moyennes* – Moyenne sur un voisinage 2×2



SIFT *Scale Invariant Feature Transform*



SIFT *Scale Invariant Feature Transform*

1. Détection des extremums en échelle et en espace : Détection
2. Localisation des points : Affinement de la détection
3. Caractérisation par une **orientation principale** = gradient dominant dans le voisinage du point considéré, avec prise en compte du bruit
4. Ajout d'un descripteur associé prenant en compte un **histogramme local de ces orientations**



Propriétés intéressantes

1. **Détecteur complet** : un descripteur associé permettant de mettre en correspondance de manière robuste
2. De nombreuses propriétés
 - 2.1 Localisation précise
 - 2.2 robustes aux occultations et aux bruits
 - 2.3 invariance aux changements de luminosité
 - 2.4 robustesse aux changements de points de vue

MAIS

3. **Non invariant aux transformations affines**



Détection des extremums en échelle et en espace

- Approche par **filtres en cascade**



Détection des extremums en échelle et en espace

- Approche par **filtres en cascade**
- Estimation de différences de gaussiennes = **approximation du laplacien**

$$\begin{aligned} DoG(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned}$$

où $G(x, y, \sigma)$ est la fonction gaussienne



Détection des extremums en échelle et en espace

- Approche par **filtres en cascade**
- Estimation de différences de gaussiennes = **approximation du laplacien**

$$\begin{aligned} DoG(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned}$$

où $G(x, y, \sigma)$ est la fonction gaussienne

- Pour des valeurs successives de σ : laplacien normalisé en échelle



Détection des extremums en échelle et en espace

- **Définition d'une octave** = Ensemble d'échelles

Pour passer d'une octave à l'autre = double la valeur de σ

Donc : si on souhaite n échelles par octave, alors il faut fixer

$$k = 2^{1/n}$$

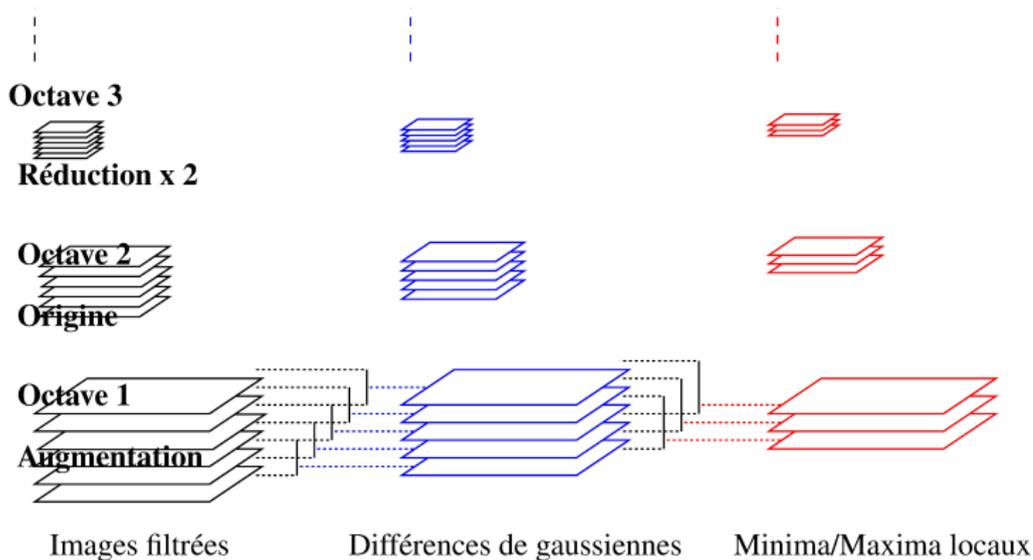


Détection des extremums en échelle et en espace

- **Définition d'une octave** = Ensemble d'échelles
Pour passer d'une octave à l'autre = double la valeur de σ
Donc : si on souhaite n échelles par octave, alors il faut fixer
 $k = 2^{1/n}$
- **maxima/minima locaux pour chaque octave**
Voisinage 3×3 + échelles inférieure et supérieure



Détection





Affinement et correction de la localisation des points

1. Calcul de la position au sous-pixel

Estimation du décalage et re-calage si nécessaire



Affinement et correction de la localisation des points

1. Calcul de la position au sous-pixel

Estimation du décalage et re-calage si nécessaire

2. Suppression des points faiblement contrastés

Différence (après filtrage) entre le point initial et le point re-localisé



Affinement et correction de la localisation des points

1. Calcul de la position au sous-pixel

Estimation du décalage et re-calage si nécessaire

2. Suppression des points faiblement contrastés

Différence (après filtrage) entre le point initial et le point re-localisé

3. Suppression des points contours

Utilisation de la matrice Hessienne (comme Harris)



Plan

Introduction

Détection de points d'intérêt

Définition – caractéristiques

Fonctionnement

Exemple de détecteur mono-échelle : Moravec/Harris

Exemple de détecteurs multi-échelle : SIFT

Mise en correspondance

Introduction

Contraintes pour la mise en correspondance

Appariement basé SIFT



Définitions

1. Trouver les entités homologues, c'est-à-dire, les éléments qui se correspondent entre deux images



Définitions

1. Trouver les entités homologues, c'est-à-dire, les éléments qui se correspondent entre deux images
2. **Propriétés attendues des points d'intérêt pour la mise en correspondance**
 - 2.1 Permettre une mise en correspondance fiable
 - 2.2 Respecter un critère de répétabilité



Éléments à définir : Coût global de mise en correspondance

$$E(I) = \sum_{\mathbf{p} \in I} \left(\underbrace{E_{sim}(\mathbf{p}, \mathbf{p}')}_{\text{Ressemblance photométrique}(I(\mathbf{p}))} + \underbrace{E_{cont}(\mathbf{p}, \mathbf{p}')}_{\text{Respect de contraintes (ordre, symétrie)}} \right)$$

avec $\mathbf{p} = \mathbf{p}' + \mathbf{d}$

\mathbf{d} = vecteur de disparité = déplacement entre les deux images



Éléments à définir : Coût global de mise en correspondance

$$E(I) = \sum_{\mathbf{p} \in I} \left(\underbrace{E_{sim}(\mathbf{p}, \mathbf{p}')}_{\text{Ressemblance photométrique}(I(\mathbf{p}))} + \underbrace{E_{cont}(\mathbf{p}, \mathbf{p}')}_{\text{Respect de contraintes (ordre, symétrie)}} \right)$$

avec $\mathbf{p} = \mathbf{p}' + \mathbf{d}$

\mathbf{d} = vecteur de disparité = déplacement entre les deux images

- Zones d'agrégation



Éléments à définir : Coût global de mise en correspondance

$$E(I) = \sum_{\mathbf{p} \in I} \left(\underbrace{E_{sim}(\mathbf{p}, \mathbf{p}')}_{\text{Ressemblance photométrique}(I(\mathbf{p}))} + \underbrace{E_{cont}(\mathbf{p}, \mathbf{p}')}_{\text{Respect de contraintes (ordre, symétrie)}} \right)$$

avec $\mathbf{p} = \mathbf{p}' + \mathbf{d}$

\mathbf{d} = vecteur de disparité = déplacement entre les deux images

- Zones d'agrégation
- Méthode d'optimisation



Approches locales versus globales

- Méthode d'optimisation
 - Globale : prise en compte de toutes les correspondances à optimiser ensemble
 - Locale : de type **winner takes all** pour chaque point individuellement



Description des approches globales

- **Algorithme**

1. Estimer un coût de correspondance E initiale lié à une première approximation des disparités
2. Mettre à jour E et l'ensemble des correspondances grâce à la méthode d'optimisation choisie



Description des approches globales

- **Algorithme**

1. Estimer un coût de correspondance E initiale lié à une première approximation des disparités
2. Mettre à jour E et l'ensemble des correspondances grâce à la méthode d'optimisation choisie

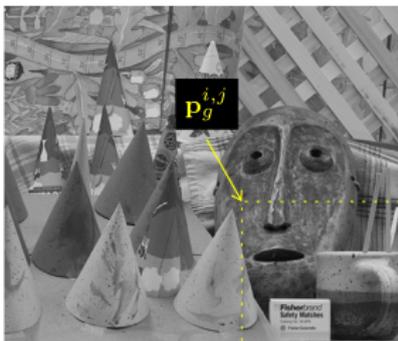
- **Optimisation**

1. Recherche exhaustive trop coûteuse
2. Méthodes utilisées : programmation dynamique, coupure de graphes, propagation de croyance, algorithmes génétiques ...

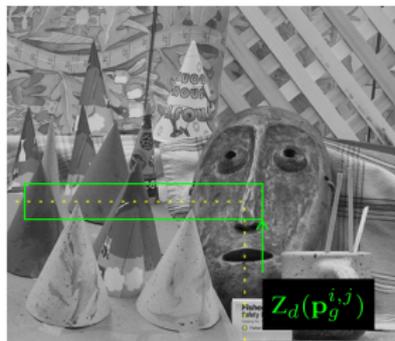


Principe des approches locales

1. Fenêtre de corrélation \mathbf{f}_l
2. Zone de recherche Z_d
3. Mesure de corrélation Mes



j

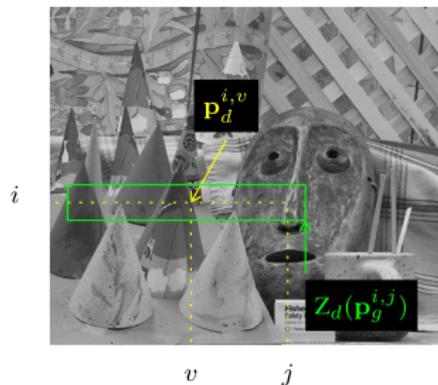
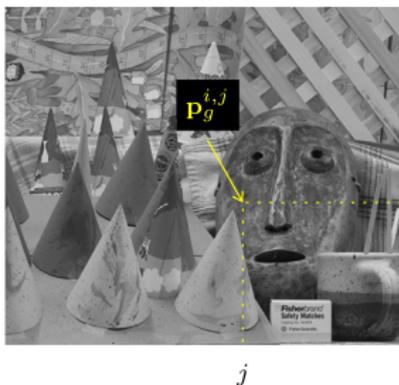


j



Principe des approches locales

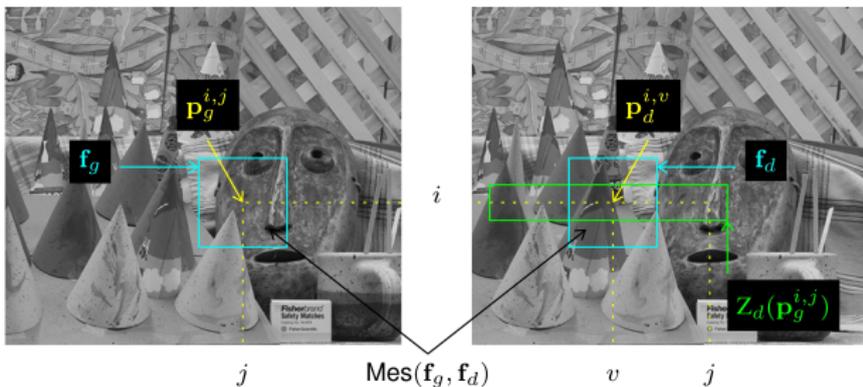
1. Fenêtre de corrélation \mathbf{f}_l
2. Zone de recherche Z_d
3. Mesure de corrélation Mes





Principe des approches locales

1. Fenêtre de corrélation \mathbf{f}_l
2. Zone de recherche Z_d
3. Mesure de corrélation Mes





Deux possibilités pour mettre en correspondance des points d'intérêt

1. Points d'intérêt vers image
2. Points d'intérêt vers points d'intérêt



Avantages des approches locales



Avantages des approches locales

- Mise en œuvre simple et rapide
- Peu coûteuse en temps de calcul (moins coûteuse que les méthodes globales)
- Dans de nombreux travaux, efficacité démontrée !



Difficultés d'appariement (liées aux deux approches)



Difficultés d'appariement (liées aux deux approches)

- Bruits
- Zones non texturées
- Textures répétitives
- Changements de luminosité
- Occultations
- Déformation des objets
- Changement d'échelle
- Points de vue trop éloignés
- Résultats erronés car certaines caractéristiques globales de l'image ne sont pas prises en compte (comme une sursegmentation)



Définition d'une mesure de similarité

Il s'agit d'une mesure qui évalue à quel point deux ensembles de données se ressemblent



Définition d'une mesure de similarité

Il s'agit d'une mesure qui évalue à quel point deux ensembles de données se ressemblent

- **But** : trouver pour chaque pixel $\mathbf{p}_g^{i,j}$ de l'image gauche, le pixel \mathbf{p}_d^{i,v^*} qui lui correspond dans l'image droite



Définition d'une mesure de similarité

Il s'agit d'une mesure qui évalue à quel point deux ensembles de données se ressemblent

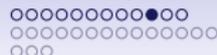
- **But** : trouver pour chaque pixel $\mathbf{p}_g^{i,j}$ de l'image gauche, le pixel \mathbf{p}_d^{i,j^*} qui lui correspond dans l'image droite
- **Hypothèse** : niveaux de gris des pixels correspondants et des pixels de leurs voisinages se ressemblent



Définition d'une mesure de similarité

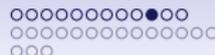
Il s'agit d'une mesure qui évalue à quel point deux ensembles de données se ressemblent

- **But** : trouver pour chaque pixel $\mathbf{p}_g^{i,j}$ de l'image gauche, le pixel \mathbf{p}_d^{i,v^*} qui lui correspond dans l'image droite
- **Hypothèse** : niveaux de gris des pixels correspondants et des pixels de leurs voisinages se ressemblent
- **Ensembles comparés** : le pixel gauche et son voisinage, \mathbf{f}_g , et le pixel de l'image droite et son voisinage, \mathbf{f}_d



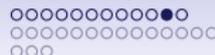
Difficultés

- **Hypothèse de similarité dans le voisinage** = tout le voisinage possède la même disparité
- **Hypothèse fausse en présence de**



Difficultés

- **Hypothèse de similarité dans le voisinage** = tout le voisinage possède la même disparité
- **Hypothèse fautive en présence de**
 - Bruits
 - Changements de luminosité
 - Déformations (dus à la projection perspective) des objets de la scène
 - Occultations



Prise en compte de ces difficultés

- **Bruit** : préfiltrage
- **Changements de luminosité** : normalisation et centrage
- **Projection perspective** : corrélation fine (estimation des paramètres d'une transformation locale)
- **Occultations** : deux solutions
 - *Fenêtres adaptatives* : Forme de la fenêtre non fixe
 - *Mesures robustes aux occultations* : Intégration direct du problème des occultations dans le calcul du score de corrélation



Corrélation croisée

Corrélation centrée normalisée, Zero-mean Normalised Cross Correlation

$$\text{ZNCC}(\mathbf{f}_g, \mathbf{f}_d) = \frac{(\mathbf{f}_g - \overline{\mathbf{f}_g}) \cdot (\mathbf{f}_d - \overline{\mathbf{f}_d})}{\|\mathbf{f}_g - \overline{\mathbf{f}_g}\| \|\mathbf{f}_d - \overline{\mathbf{f}_d}\|}$$

\mathbf{f}_w , $w = g, d$ = vecteurs des niveaux de gris dans les voisinages respectifs des points appariés



Contraintes pour la mise en correspondance

- Intervention dans le terme de coût global ou en affinement de l'appariement
- **Buts**
 - Limiter le nombre de correspondants potentiels pour chaque pixel
 - Choisir entre plusieurs correspondants potentiels
 - Vérifier et supprimer des correspondances

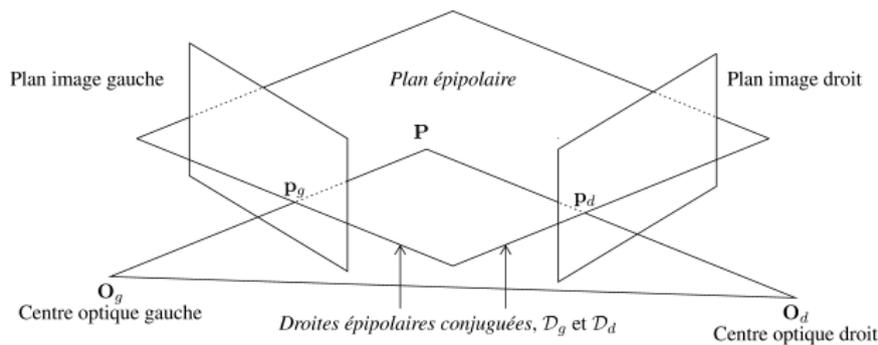


Origine

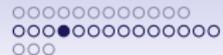
- **Propriété découle d'hypothèses faites sur :**
 - Géométrie du capteur (Contrainte épipolaire)
 - Géométrie de la scène (Contraintes de compatibilité)
 - Réflectance de la surface des objets ainsi que le type et la position de la source lumineuse (Contrainte de dissimilarité)



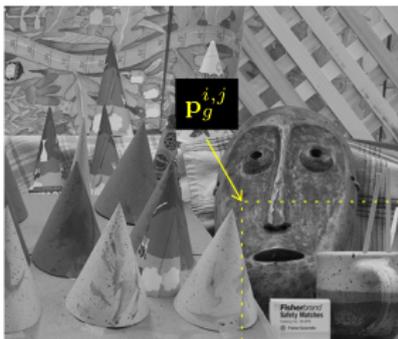
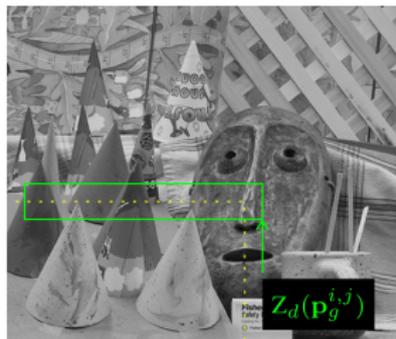
Contrainte géométrique : Géométrie épipolaire



La contrainte épipolaire permet d'obtenir une relation entre les points de l'image gauche et les points de l'image droite qui **ne dépend que des paramètres des caméras et des coordonnées image et qui est indépendante des coordonnées du point de la scène : matrice fondamentale**

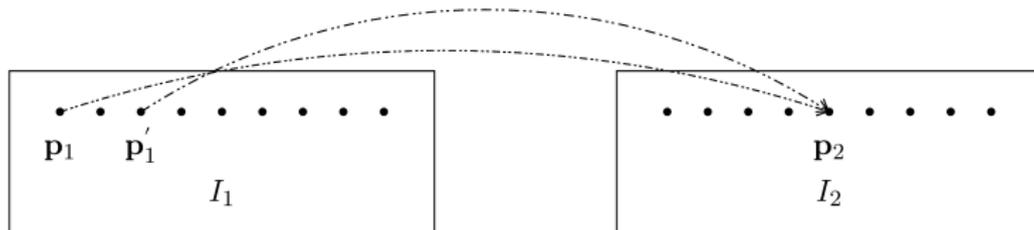


Rectification épipolaire

 j i  j

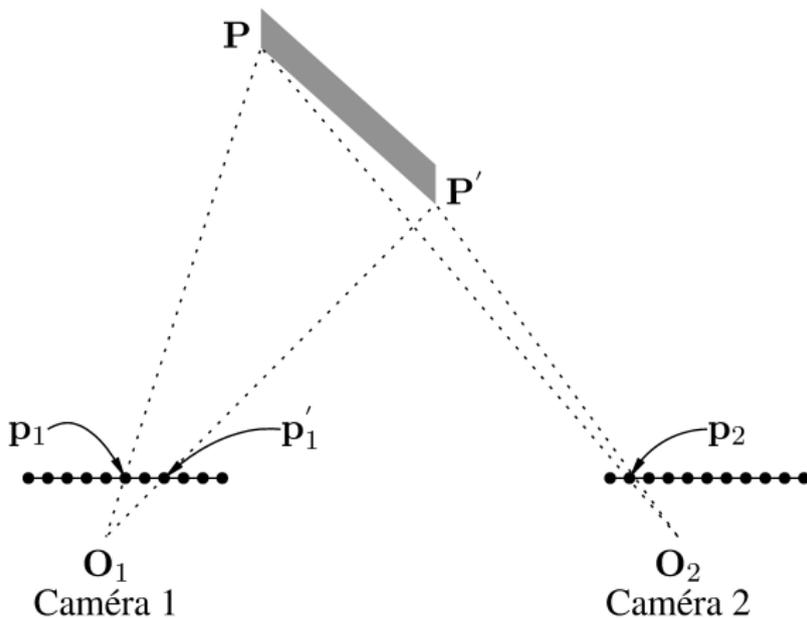


Contrainte d'unicité



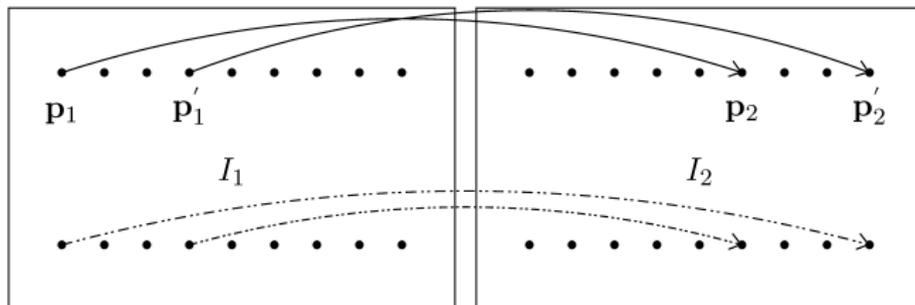


Non respect de la contrainte d'unicité



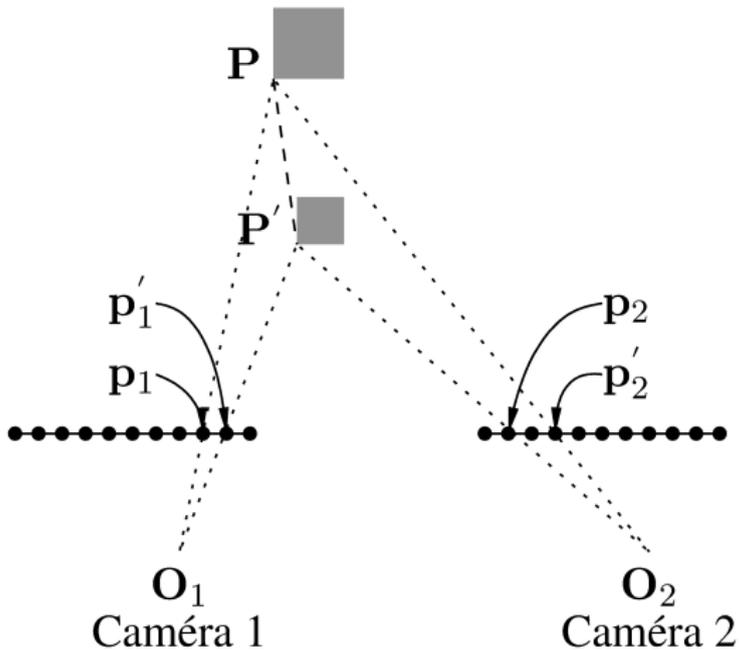


Contrainte d'ordre



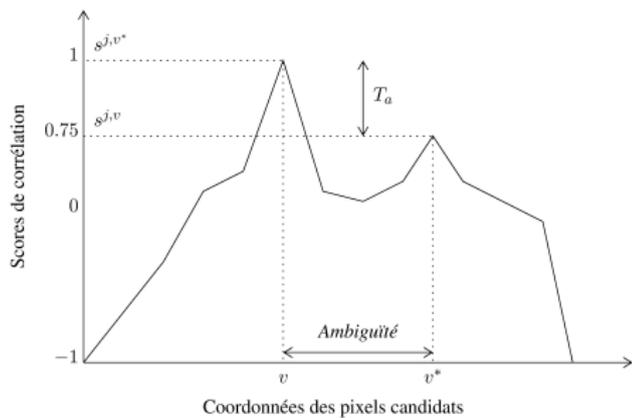


Non respect de la contrainte d'ordre



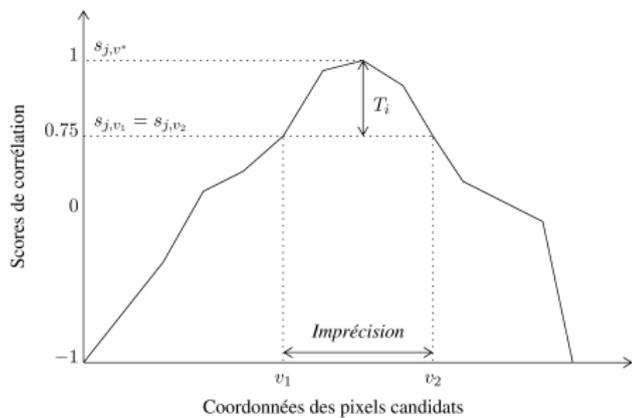


Ambiguïté





Imprécision





Contraintes liées à la réflectance

$$\text{Si } \mathcal{C}(\mathbf{p}_1) = \mathbf{p}_2 \text{ alors } \text{Mes}(\mathbf{p}_1, \mathbf{p}_2) > T \quad (1)$$

- T relatif au type de critère utilisé et à la qualité des résultats attendue
- Plus le seuil est restrictif, plus de correspondances erronées éliminées

MAIS également plus de correspondances justes éliminées



Prise en compte d'une contrainte

1. Toutes les correspondances ne respectant pas la contrainte sont éliminées : disparité éparses



Prise en compte d'une contrainte

1. Toutes les correspondances ne respectant pas la contrainte sont éliminées : disparité éparses
2. Une partie des correspondances est éliminée/l'autre est conservée, voire corrigée



Prise en compte d'une contrainte

1. Toutes les correspondances ne respectant pas la contrainte sont éliminées : disparité éparses
 2. Une partie des correspondances est éliminée/l'autre est conservée, voire corrigée
 - 2.1 Éliminer le plus petit sous-ensemble de manière à respecter la contrainte
- Exemple** Contrainte figurale : on va conserver la correspondance qui se trouve sur le contour principal



Prise en compte d'une contrainte

1. Toutes les correspondances ne respectant pas la contrainte sont éliminées : disparité éparses
 2. Une partie des correspondances est éliminée/l'autre est conservée, voire corrigée
 - 2.1 Éliminer le plus petit sous-ensemble de manière à respecter la contrainte
- Exemple** Contrainte figurale : on va conserver la correspondance qui se trouve sur le contour principal
- 2.2 Une autre solution consiste à vérifier ou compléter cette contrainte par une seconde contrainte pour faire le choix
- Exemple** Contrainte de seuil sur le critère de similarité pour déterminer quelle correspondance est conservée dans le cas où la contrainte d'ordre n'est pas respectée



Appariement basé SIFT

Affectation d'une orientation principale

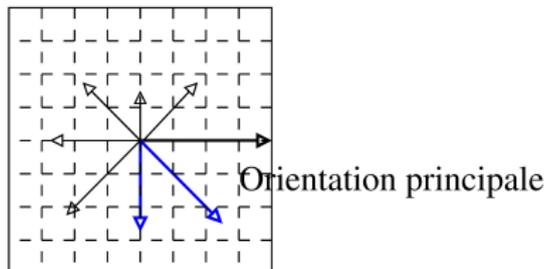
1. Estimation de la norme et de l'orientation du vecteur gradient



Appariement basé SIFT

Affectation d'une orientation principale

1. Estimation de la norme et de l'orientation du vecteur gradient
2. Construction d'un histogramme local d'orientation
 - 2.1 36 orientations
 - 2.2 Pondération par la norme et des poids gaussiens



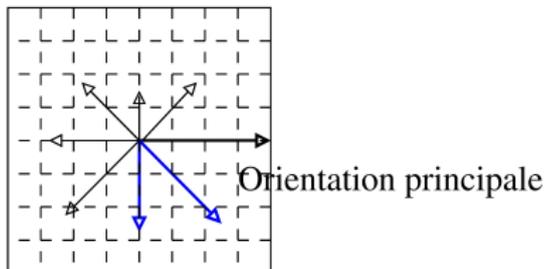
Orientations secondaires



Appariement basé SIFT

Affectation d'une orientation principale

1. Estimation de la norme et de l'orientation du vecteur gradient
2. Construction d'un histogramme local d'orientation
 - 2.1 36 orientations
 - 2.2 Pondération par la norme et des poids gaussiens



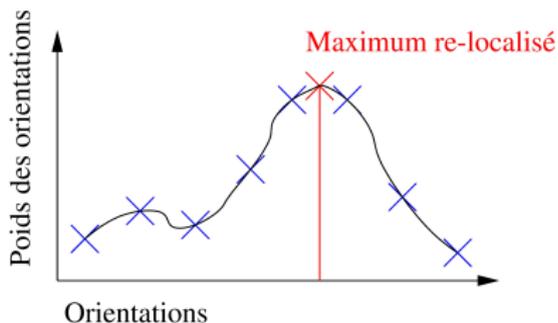
Orientations secondaires

3. Sélection des orientations principale et secondaires



Affectation d'une orientation principale

- (4) Interpolation des orientations (pour localiser correctement le maximum local)





Calcul des descripteurs

1. **Utilisation de l'orientation et de l'histogramme déjà estimés avec trois différences**



Calcul des descripteurs

1. **Utilisation de l'orientation et de l'histogramme déjà estimés**
avec trois différences
 - 1.1 8 orientations



Calcul des descripteurs

1. **Utilisation de l'orientation et de l'histogramme déjà estimés** avec trois différences
 - 1.1 8 orientations
 - 1.2 Rotation, relative à l'orientation du point d'intérêt (invariance aux rotations)



Calcul des descripteurs

1. **Utilisation de l'orientation et de l'histogramme déjà estimés** avec trois différences
 - 1.1 8 orientations
 - 1.2 Rotation, relative à l'orientation du point d'intérêt (invariance aux rotations)
 - 1.3 Pondération par la distance au point d'intérêt



Calcul des descripteurs

1. **Utilisation de l'orientation et de l'histogramme déjà estimés** avec trois différences
 - 1.1 8 orientations
 - 1.2 Rotation, relative à l'orientation du point d'intérêt (invariance aux rotations)
 - 1.3 Pondération par la distance au point d'intérêt
2. **Normalisation du vecteur-descripteur** : Invariance aux changements affines d'illumination



Calcul des descripteurs

1. **Utilisation de l'orientation et de l'histogramme déjà estimés** avec trois différences
 - 1.1 8 orientations
 - 1.2 Rotation, relative à l'orientation du point d'intérêt (invariance aux rotations)
 - 1.3 Pondération par la distance au point d'intérêt
2. **Normalisation du vecteur-descripteur** : Invariance aux changements affines d'illumination
3. **Seuillage des forts gradients** : Prise en compte des changements d'illumination non-linéaires, ceci implique une **deuxième normalisation du vecteur-descripteur**



Calcul des descripteurs

1. **Utilisation de l'orientation et de l'histogramme déjà estimés** avec trois différences
 - 1.1 8 orientations
 - 1.2 Rotation, relative à l'orientation du point d'intérêt (invariance aux rotations)
 - 1.3 Pondération par la distance au point d'intérêt
2. **Normalisation du vecteur-descripteur** : Invariance aux changements affines d'illumination
3. **Seuillage des forts gradients** : Prise en compte des changements d'illumination non-linéaires, ceci implique une **deuxième normalisation du vecteur-descripteur**

Mise en correspondance

1. Descripteurs de taille 128 : voisinage 4×4 et 8 orientations
2. Mise en correspondance par ZNCC