# Odométrie visuelle

**KHERDJEMIL Anis**

**Sorbonne Université**

## Résumé :

Dans ce rapport, nous allons traiter le sujet de l'odométrie visuelle qui consiste en l'estimation de la position d'un objet en mouvement. Nous allons présenter les différentes étapes qui nous permettent d'estimer ces positions à partir des photos de l'objet prises par une caméra perspective.

L'estimation des paramètres intrinsèques et extrinsèques de la caméra nous permet d'estimer cette position.

## Introduction :

Ces dernières années, l'utilisation des technologies a considérablement évoluée dans le domaine de l'industrie et les usines notamment, dans la vie active de tous les jours. On prend exemple des robots mobiles ou des voitures autonomes.

Dans notre étude, on s'intéresse à la technique utilisée par ces derniers dans leurs déplacements et leurs localisations. Pour répondre aux nombreux inconvénients des méthodes classiques de localisation, on se propose d'étudier une approche qui est basée uniquement sur la vision. L'idée est de prendre deux ou plusieurs photos d'un objet avec une seule caméra en se déplaçant, l'estimation de la pose des caméras à chaque prise nous donne la trajectoire de la caméra (ou du robot). On appelle cela l'odométrie visuelle. Une autre méthode existe pour la localisation qui consiste à comparer chaque prise à la mémoire qu'il a de son environnement, l'image la plus « proche » permettant d'estimer une position. Cette approche est, quant à elle, basée sur l'indexation d'images nécessitant une base de données (un ensemble d'images de l'environnement) construite durant une phase d'apprentissage. Le type de capteur entre aussi en jeu, les atouts n'étant pas les mêmes, que le système utilise une caméra perspective ou omnidirectionnelle.

Nous allons étudier l'odométrie visuelle avec une caméra perspective et proposer une méthode pour l'estimation des paramètres intrinsèques et extrinsèques de la caméra.

## Etat de l'art :

De nouvelles méthodes basées uniquement sur la vision sont apparues pour corriger les erreurs engendrées par les capteurs infrarouges et inertiels lors de l'estimation de la trajectoire des robots mobiles. L'odométrie visuelle est donc un sujet récent mais qui prend ses racines dans les années 1980. Actuellement, dans cette branche de la recherche, on trouve tout type de caméra, monoculaire ou stéréo, perspective, omnidirectionnelle, fish eye, etc.

En environnement connu, deux grandes familles de méthodes existent, à savoir la recherche de références géométriques dans les images pour y reconnaître un objet ou une partie d'une scène 3D connue et préalablement modélisée (modèle CAO) ou le travail direct sur l'image. C'est cette dernière approche que nous allons développer car elle ne nécessite pas de connaissances précises de l'environnement dans lequel on déplace notre caméra a priori. Ainsi, l'idée générale va être de

détecter des points d'intérêts dans deux images prises à deux positions différentes et de déterminer le déplacement de la caméra qui a permis de passer de l'une à l'autre [1].

La méthode proposée suit le schéma suivant : Détection des points d'intérêts entre deux images, mise en correspondance entre les points d'intérêt des deux images, estimation d'une relation de transformation des points d'intérêts de l'image 1 à l'image 2 (estimation de la matrice fondamentale puis essentielle ) et enfin, à partir de cette relation on en déduit une matrice de rotation et un vecteur de translation entre les deux prises de vue.

Dans notre cas, nous allons prendre des photos d'un échiquier pour estimer d'abord les paramètres intrinsèques d'une caméra. Ensuite, nous déterminons les paramètres extrinsèques de la caméra ce qui nous permettra de déterminer la position de la caméra à chaque prise. Dans le chapitre suivant, nous allons détailler ces aspects.

## Méthodologie :

Nous allons d'abord présenter le modèle géométrique d'une caméra. Ce modèle est caractérisé par un nombre de paramètres que nous allons estimer et calibrer. Ces paramètres sont les paramètres intrinsèques de la caméra et les paramètres extrinsèques qui lient le repère associé à la caméra au repère associé à la scène [2].

### 1-Modèle géométrique d'une caméra :

Le modèle d'une caméra est caractérisé par deux transformations :

1) Une projection qui transforme un point de l'espace (3D) en un point image (2D) [3].

2) Une transformation d'un repère métrique lié à la caméra à un repère lié à l'image [3].

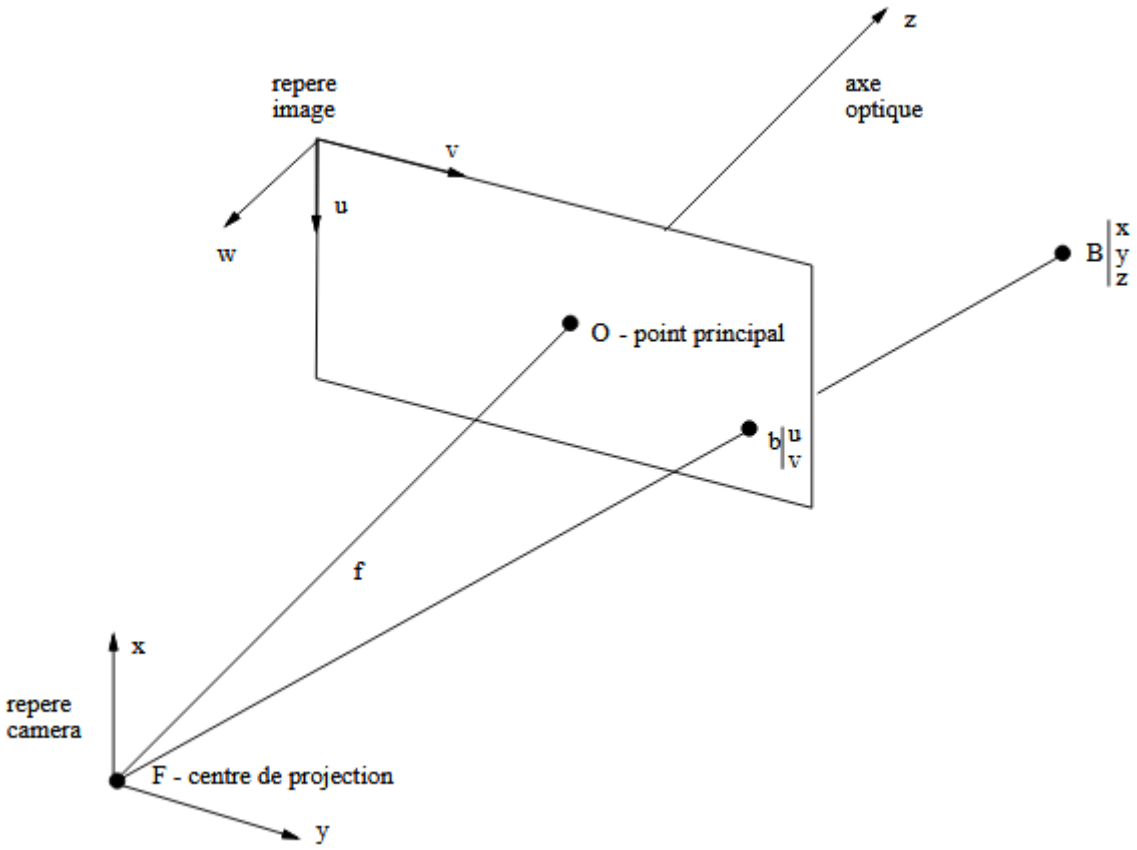


Figure 1 : Le modèle géométrique d'une caméra.

### 1.1.1-Transformation 3D/2D (Matrice de projection) :

La matrice de projection du point B dans le repère image est

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix}$$

Les coordonnées homogènes de B et de b :

$$\begin{pmatrix} sx' \\ sy' \\ sz' \\ s \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

### 1.1.2-Transformation caméra/image :

C'est une transformation qui représente un changement d'échelle, une translation et une rotation. Celle-ci est représentée par la matrice suivante :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} k_u & 0 & 0 \\ 0 & k_v & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix}$$

Tel que :

(u,v,w) est le repère image.

u0, v0 et w0 sont les coordonnées de F dans le repère image.

Ku est le facteur d'échelle vertical (pixels/mm) et Kv est le facteur d'échelle horizontal.

La composante w est toujours nul.

Cette transformation peut s'écrire sous cette forme :

**T =** $$\begin{pmatrix} -k_u & 0 & 0 & u_0 \\ 0 & k_v & 0 & v_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Et

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \mathbf{T} \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix}$$

### 1.2-Les paramètres intrinsèques :

On obtient la relation entre les coordonnées caméra et les coordonnées image en multipliant les deux matrices P et T (matrice de projection perspective et matrice de transformation)

(On multiplie aussi tous les coefficients de la matrice par f)

**K=** $$\begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Avec : αu = -Ku.f , αv= Kv.f , u0 et v0 sont les 4 paramètres à estimer pour le calibrage de la caméra.

La distance focale n'est pas calculée explicitement. On introduit les coordonnées de la caméra sans dimension telle que Zc= 1 on obtient la relation entre les coordonnées caméra et image suivante :

**Kc =**
$$\begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Et

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \text{Kc} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix}$$

### 1.3- les paramètres extrinsèques

La matrice des paramètres extrinsèques est obtenue en plaçant une mire devant la caméra. Un ensemble de points sont de coordonnées connues dans le repère de la mire, ces points sont projetés dans le repère image, on mesure ces coordonnées image.

La transformation mire/image se décompose en une transformation mire/caméra suivi d'une projection et suivi enfin d'une transformation caméra/image. La transformation mire/caméra se compose d'une rotation est d'une translation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

Cette transformation peut s'écrire sous forme d'une transformation homogène

$$\mathbf{A} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}$$

Cette matrice représente une transformation rigide (3 rotations et 3 translations) qui sont les paramètres extrinsèques.

### 2. Méthodes de détermination de paramètres extrinsèques :

A présent, nous allons présenter une méthode pour déterminer les paramètres de la pose d'une caméra. Comme cité précédemment, cette méthode consiste en :

1. Calibrage de la caméra (détermination des paramètres intrinsèques).
2. Détection des points d'intérêts dans deux images ou plus (dans notre cas nous utiliserons que deux images).
3. Appariement des points d'intérêts (matching features).
4. Estimation de la matrice fondamentale F par la méthode des 8_points et déduction de la matrice essentielle E.

5. Déduction de la matrice de rotation et de translation à partir de E

## 2.1 Calibrage de la caméra :

Pour déterminer les paramètres intrinsèques de notre caméra, nous proposons la méthode de Zhang. Cette méthode consiste à prendre plusieurs photos d'un échiquier depuis plusieurs positions (une mire planaire) ensuite on peut directement relier des points en correspondance.

Sachant :

(En supposant Z=0)

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = K \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \underbrace{K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}}_{H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

Ou H est l'homographie planaire

Chaque image nous donne une homographie H, qui calculé à partir des points en correspondance

Chaque paire de points en correspondance produit deux équations linéaires

Quatre points en correspondances suffisent à déterminer H [4].

Hm=m' -> Hm × m' = 0 $$H = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

A partir de H, nous pouvons déduire la matrice K qui est les paramètres intrinsèques de la caméra.

## 2.2 Détection des points d'intérêts et Appariement des points :

Après calibrage de la caméra, nous nous proposons de prendre deux prises d'une même scène depuis différentes positions. Ensuite nous allons utiliser la méthode SURF pour la détection de points d'intérêts dans les deux images et le matching de ces points. A noter que d'après [5] la méthode SURF est plus rapide est a de bonnes performances comparé à la méthode SIFT qui détecte plus de points d'intérêts mais qui est lente.

La méthode SURF consiste à détecter des points d'intérêts et les extraire, calcul des descripteurs et enfin appariement de ces points d'intérêts.

La méthode des SURF utilise le fast-Hessien pour la détection de points d'intérêts et une approximation des ondelettes de Haar pour calculer les descripteurs.

Le fast-Hessien se fonde sur l'étude de la matrice hessienne :

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix}$$

où Lij(x,y,σ) est la dérivée seconde suivant les directions en i et en j de L avec :

$$L(x\ ,\ y\ ,\ \sigma) = G(x\ ,\ y\ ,\ \sigma)\ *\ I\ (x\ ,\ y\ )$$

G est le gradient et I est l'image du départ. La maximisation du déterminant de cette matrice permet d'obtenir les coordonnées des points d'intérêts à une échelle donnée. Cette étape apporte une invariance des points d'intérêts par rapport à la mise à l'échelle.

Le déterminant est défini ainsi :

$$\det(H(x, y, \sigma)) = \sigma^2(L_{xx}(x, y, \sigma)L_{yy}(x, y, \sigma) - L^2_{xy}(x, y, \sigma)).$$

Cette étape permet donc détecter les points d'intérêts candidats. L'algorithme comporte ensuite des étapes intermédiaires destinées à apporter plus de précision dans leur localisation.

Le calcul des descripteurs se fait grâce aux ondelettes de Haar. Elles permettent d'estimer l'orientation locale du gradient et donc d'apporter l'invariance par rapport à la rotation. Les réponses des ondelettes de Haar sont calculées en x et y dans une fenêtre circulaire dont le rayon dépend du facteur d'échelle du point d'intérêt détecté. Ces réponses spécifiques contribuent à la formation du vecteur de caractéristique correspondant au point clé.

Pour ce qui est de la mise en correspondance des descripteurs, i.e la recherche de la meilleure similitude entre les descripteurs de deux images, le critère utilisé est le même que celui utilisé dans l'algorithme des SIFT, i.e celui de la distance euclidienne [6].

### 2.3 Estimation de la matrice fondamentale F :

La géométrie épipolaire est la géométrie projective intrinsèque entre deux points de vue. Elle est indépendante de la structure de la scène et ne dépend que des paramètres internes de la caméra et de la pose relative.
La matrice fondamentale encapsule cette géométrie intrinsèque. Il s'agit d'une matrice 3 × 3 de rang 2. Si un point de l'espace 3D en M est visualisé comme dans la première vue, et dans la seconde, les points de l'image vérifient la relation m't.F.m = 0 (m't=m' transposée).
Pour chaque paire de points en correspondances m'<->m dans les deux images, F satisfait l'équation précédente (m't.F.m = 0).

Nous aurons donc :

mt.F.m' = 0 --->

$$(u,v,1)\begin{pmatrix}\mathcal{F}_{11} & \mathcal{F}_{12} & \mathcal{F}_{13}\\ \mathcal{F}_{21} & \mathcal{F}_{22} & \mathcal{F}_{23}\\ \mathcal{F}_{31} & \mathcal{F}_{32} & \mathcal{F}_{33}\end{pmatrix}\begin{pmatrix}u'\\ v'\\ 1\end{pmatrix}=0$$

Puisque cette équation
est homogène dans les coefficients de F, nous pouvons définir F33=1 (élimination de l'ambiguïté de l'échelle) et utiliser l'algorithme des correspondances en huit points pour résoudre cette équation.

À partir de 8 correspondances de points, nous obtenons un ensemble d'équations linéaires de la forme

$$\begin{pmatrix}u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' & 1\\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' & 1\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ u_pu_p' & u_pv_p' & u_p & v_pu_p' & v_pv_p' & v_p & u_p' & v_p' & 1\end{pmatrix}\begin{pmatrix}F_{11}\\ F_{12}\\ \cdots\\ 1\end{pmatrix}=\mathbf{0}$$

Pour résoudre cet ensemble d'équations, nous allons utiliser la méthode de Hartley qui consiste en :

- Normalisation des équations à résoudre (La normalisation consiste en une translation et une mise à l'échelle de chaque image de sorte que le centroïde des points de référence soit à l'origine des coordonnées et que la distance efficace des points à partir de l'origine soit égale à √2)
- Trouver F en calculant la SVD(F)= U S V' tel que S=diag(r,s,t) ou t<=s<=r calcule F=U diag(r,s,0) V'
- La dénormalisation : F = norm2' * F * norm1 [7]

### 2.4 Estimation de la matrice essentielle E et déduction des vecteurs de translation et de rotation :

La matrice essentielle est une spécialisation de la matrice fondamentale. Elle est déduite directement de cette dernière dans le cas où la caméra est calibrée. Dans notre cas notre caméra est calibrée, nous déduisons donc E à partir de F et les paramètres intrinsèques de la caméra par l'équation suivante :

$$\mathbf{E = K' * F * K}$$

E est une matrice 3*3 qui se décompose en une rotation R et une translation T tel que R et T sont respectivement la rotation et la translation de la caméra 2 par rapport à la caméra 1.

$$\mathrm{E} = [\mathrm{t}]_{\times}\mathrm{R}$$

E peut s'écrire comme suit : E=SR tel que S est asymétrique.

Nous allons utiliser les deux matrices suivantes :

$$\mathtt{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathtt{Z} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

On peut vérifier que W est orthogonal et que Z est asymétrique d'après [8]

On note Z=diag(1,1,0)W , S=Udiag(1,1,0)WU' et E=SR=Udiag(1,1,0)(WU'R)

Il s'agit d'une décomposition en valeurs singulières de E avec deux valeurs singulières égales. Inversement, une matrice avec deux valeurs singulières égales peut être considérée comme SR de cette manière.

La matrice de rotation R peut être donnée comme suit :
Ou X est la matrice rotation

Udiag(1,1,0)V'=E=SR=(UZU')(UXV')=U(ZX)V

Duquel, on peut déduire que : ZX=diag(1,1,0)

Puisque X est la matrice de rotation, il en résulte : X=W ou X=W' [8]

Le vecteur translation T est donnée comme suit : T=U(0,0,1)' [8]

## Résultats :

L'utilisation des fonctions de Matlab ont été nécessaires pour certaines étapes de la méthodologie précédentes. L'algorithme des 8_points pour le calcul de la fonction fondamentale et le calcul de la matrice essentielle ainsi que les matrices de rotation et translation ont été implémenté.

Pour la calibration de la caméra la fonction Matlab cameraCalibrator a été utilisée pour récupérer les paramètres intrinsèques de la caméra.

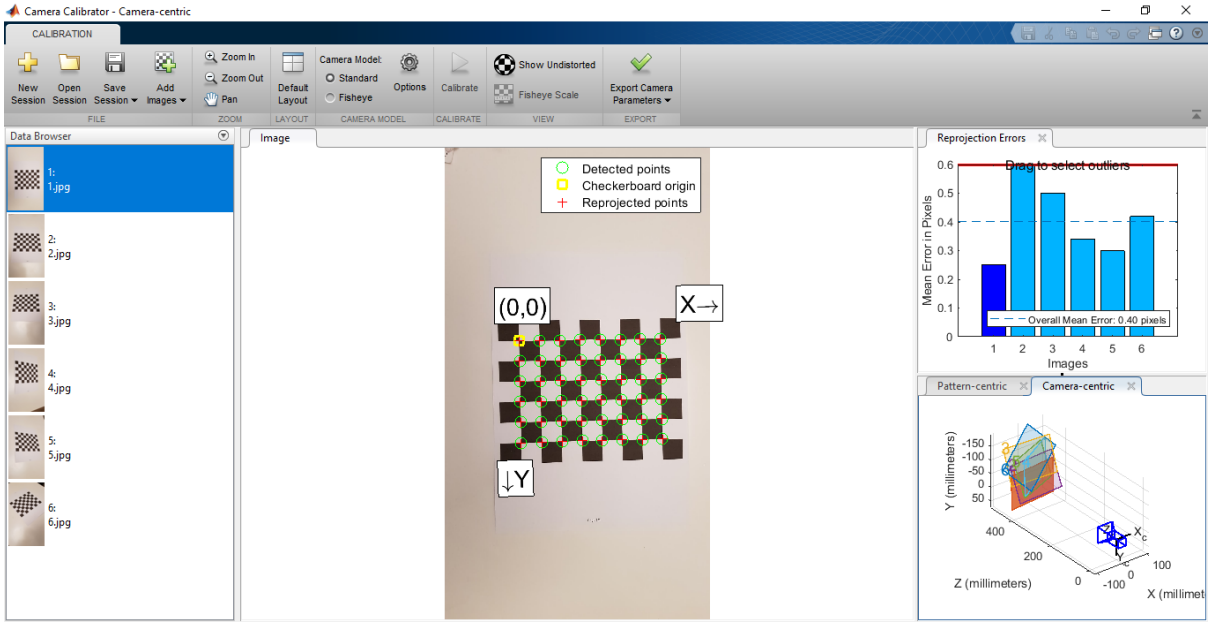


Figure 2 : Calibration d'une caméra à partir d'un ensemble de prises d'un échiquier

La matrice récupérée est :

K=

| | | |
|---|---|---|
| 2.0055e+03 | 0 | 0 |
| 0 | 1.9983e+03 | 0 |
| 714.9446 | 1.2713e+03 | 1 |

Ensuite, les fonctions detectSURFFeatures, extractFeatures et matchFeatures de Matlab ont été utilisé pour la détection et l'appariement des points d'intérêts entre deux images.

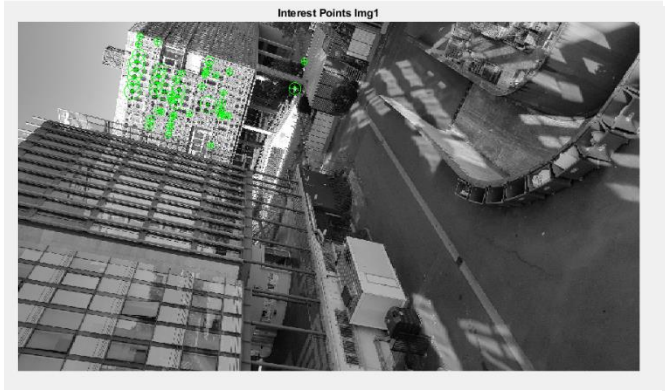


Figure 3 : points d'intérêts sur l'image 1

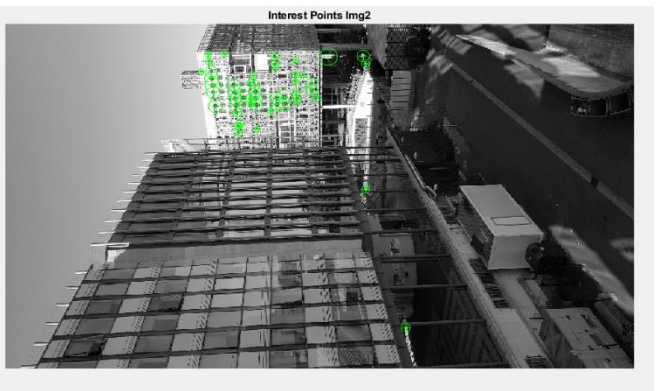


Figure 4 : points d'intérêts sur l'image 2

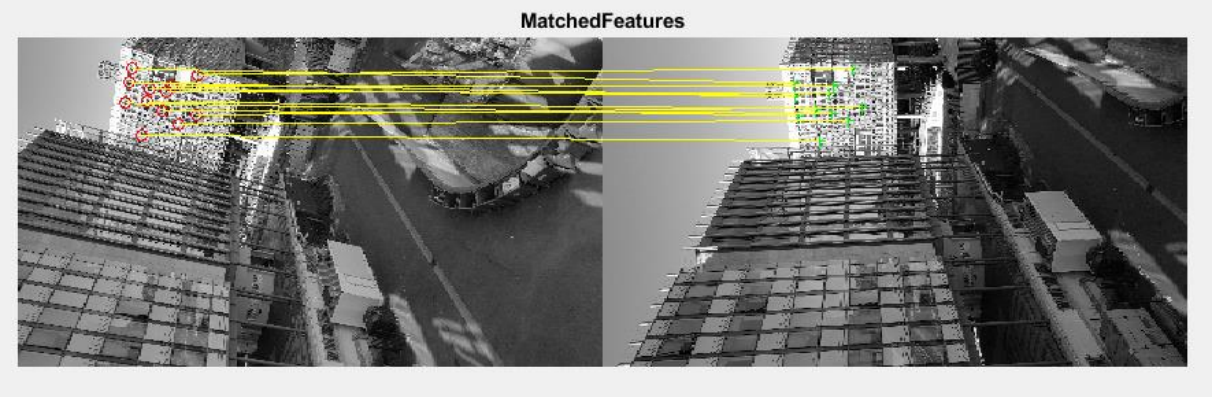


Figure 5 : Appariements des points d'intérêts sur les deux images

Pour l'estimation de la matrice fondamentale, nous avons implémenté l'algorithme des 8_points, nous obtenons la matrice fondamentale suivante :

F =

| | | |
|---|---|---|
| 5.9379e-06 | -1.9532e-05 | 7.4872e-04 |
| 1.8576e-05 | 4.6769e-06 | -0.0206 |
| -0.0126 | 0.0252 | 4.0563 |

L'estimation de la matrice de rotation R et du vecteur de translation t sont calculées néanmoins, les valeurs obtenues ne sont pas sures (possible erreur lors de l'implémentation).

```
R =

  3×3 single matrix

    0.0115    0.9999   -0.0007
   -0.9999    0.0115   -0.0014
   -0.0013    0.0007    1.0000
```

```
t =

  1×3 single row vector

   -0.0011   -0.0002    1.0000
```

## Conclusion :

L'odométrie est une technique permettant d'estimer la position d'un objet en mouvement. Pour estimer cette position, une caméra est embarquée sur l'objet en mouvement et qui prend des photos à chaque instant. Il faut estimer les paramètres de cette caméra qui sont les paramètres intrinsèques et extrinsèques pour estimer la position de l'objet. Dans ce rapport, j'ai explicité une méthodologie qui nous permet de remonter à la position d'une caméra par rapport à une position précédente. J'ai d'abord commencé par l'estimation des paramètres interne de la caméra (intrinsèques). Ensuite, j'ai détecté les points d'intérêts sur les deux images par le détecteur SURF et je les ai matchés pour pouvoir calculer la matrice fondamentale par la méthode des 8_points. Enfin, j'ai déduit la matrice essentielle à partir de laquelle j'ai calculé la matrice de rotation et le vecteur de translation entre les deux images.

## Référence :

[1] Danielle Nuzillard, Alban Goupil « Odométrie visuelle pour un robot mobile en environnement industriel » Rapport de stage de Master Recherche STIC, Juin 2007.

[2] Peter Sturm, « Quelques notes pour le cours de Vision par Ordinateur » INRIA Rhône-Alpes, Equipe-projet PERCEPTION

[3] Radu Horaud et Olivier Monga «Géométrie et calibration des caméras » in Vision par ordinateur : outils fondamentaux, Edition Hermès.

[4] Marie-Odile Berger « Calibrer avec une mire planaire [Zhang] » in Calibrage d'une caméra, novembre 2014

[5] P M Panchal, S R Panchal, S K Shah "A Comparison of SIFT and SURF "International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 2, April2013

[6] « Plus rapide que les SIFT . . . les SURF » Caractérisation des points remarquables dans les images et recalage

[7] Raul Queiroz Feitosa, "Geometry of Multiple Views" mai, 2017

[8] Hartley, Richard, and Andrew Zisserman."Two-view Geometry" in Multiple View Geometry in Computer Vision. Second Edition. Cambridge, 2000, pp 237-259