



**UNIVERSITÉ Gustave Eiffel**  
**Master II** informatique  
Parcours Sciences de l'images  
**Module** : Restauration d'image

---

# Rapport du Projet

## Flot Optique

---

Enseignante :

- **Mr François-Xavier VIALARD**

Réalisé par :

- **Mr Rachid LADJOUZI**

## 1. Introduction :

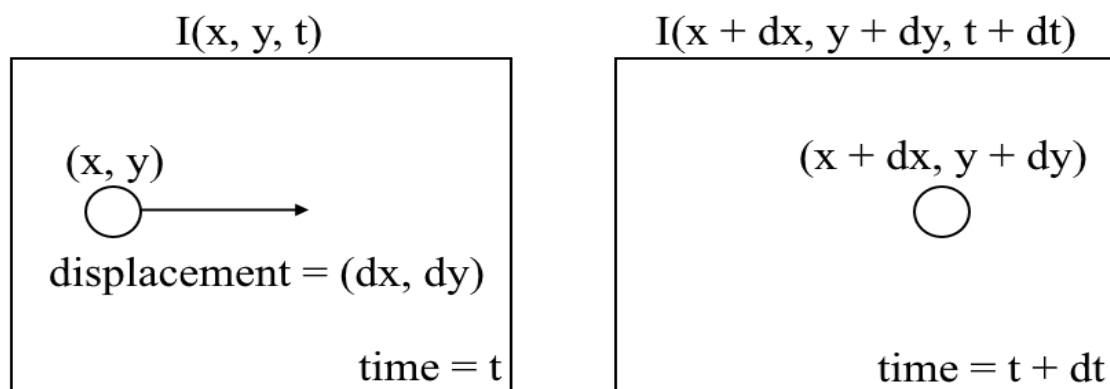
La perception est le regroupement et recoupement d'un ensemble d'informations en vue de comprendre son environnement. A cette fin, l'homme utilise plusieurs modalités sensorielles telles que la vision, le toucher, l'ouïe ou même la proprioception. Ces sens eux-mêmes peuvent convoier différents types d'information. La vision par exemple, propose des indices de texture, de couleurs, de formes ou de mouvement. Le flot optique est ainsi le déplacement de l'image sur la rétine. Avec lui seul, un homme est capable d'appréhender une partie des formes observées. Etant donné qu'une vidéo n'est qu'une collection d'images fonctionnant dans une résolution temporelle spécifique, c'est-à-dire des images par seconde. Cela signifie que les informations dans une vidéo sont encodées non seulement spatialement (c'est-à-dire dans les objets ou les personnes dans une vidéo), mais aussi séquentiellement et selon un ordre spécifique, par exemple : attraper une balle vs lancer une balle. Cette information supplémentaire est ce qui rend la classification des vidéos très intéressante et difficile en même temps ce qui ramène à définir des algorithmes permettant de donner la priorité au mouvement en tant que caractéristique clé de cette classification, et qui est réalisable par calcul dans un contexte du monde réel. Il existe plusieurs façons d'y parvenir et l'estimation du flot optique est considéré comme étant l'une des méthodes potentielles pour atteindre le résultat souhaité. En partant du fait que les différents objets se déplacent dans une scène selon leurs différentes classes, les méthodes d'estimation de flot optique existantes émettent des hypothèses génériques et spatialement homogènes sur la structure spatiale du flot. Dans ce travail, nous exploitons les progrès récents des différentes méthodes d'estimation de flot optique des scènes et faisons une comparaison avec la méthode proposée FlowNet2.0 pour la détection des différents objets en mouvement. Différents types de mouvement d'image dans différentes régions sont définis et cela en fonction du type d'objet. Ce qui nous ramène à la problématique suivante : Comment pouvons-nous améliorer l'estimation d'écoulement en exploitant les limites des modèles en couches des méthodes traditionnelles pour faire face au mouvement dans une scène complexe.

## 2. Flot optique :

Le flot optique représente une très importante opération de base dans l'analyse de vidéos. Cette technique a été utilisée pour améliorer la précision lors de la classification des vidéos et à des coûts de calcul inférieurs. Elle existe depuis les années 1980 sous la forme des approches de hand crafted. Cette méthode vise à analyser une vidéo en calculant le mouvement avec le nombre minimum de frames.

Le flot optique est une prédiction par pixel et l'idée principale est qu'il suppose une constance de luminosité, ce qui signifie qu'il essaie d'estimer comment la luminosité des pixels se déplace à travers l'écran au fil du temps.

Le flot optique est le mouvement d'objets entre des images consécutives de séquence, provoqué par le mouvement relatif entre l'objet et la caméra. Le problème du flot optique peut s'exprimer comme suit :



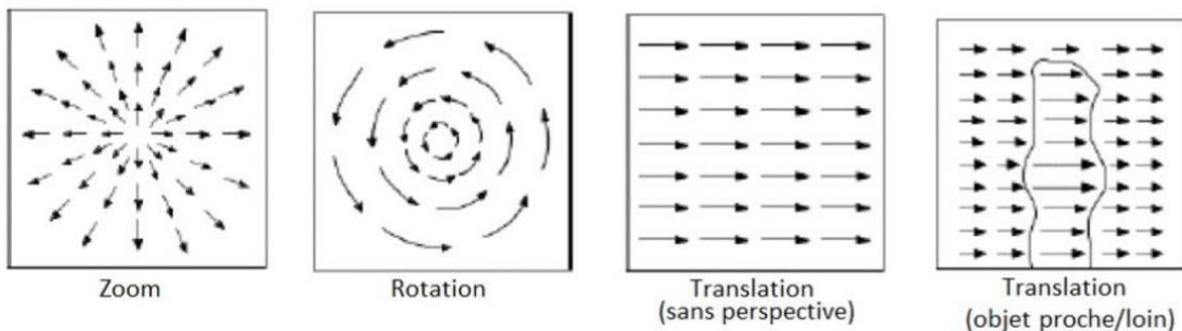
où entre les images consécutives, nous pouvons exprimer l'intensité de l'image ( $I$ ) en fonction de l'espace ( $x, y$ ) et du temps ( $t$ ). En d'autres termes, si nous prenons la première image  $I(x, y,$

t) et déplaçons ses pixels de  $(dx, dy)$  dans le temps  $t$ , nous obtenons la nouvelle image  $I(x + dx, y + dy, t + dt)$

Il suppose que  $I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$  en clair, c'est la caractéristique du pixel au temps  $t$  (c'est-à-dire les valeurs rgb) est la même que les caractéristiques du pixel à  $\Delta t$  mais à un emplacement différent (désigné par  $\Delta x$  et  $\Delta y$ ), où le changement d'emplacement est ce qui est prédit par le champ d'écoulement

L'estimation du mouvement Le flot optique peut être le résultat d'un mouvement d'objets et/ou de la caméra, d'un changement d'illumination ou d'apparence des objets, d'un changement de paramètres intrinsèques de la caméra (e.g. distance focale, ouverture, etc.)

La figure suivante nous montre différents flot optiques dû aux mouvements de caméra :



Le but du Flot optique est d'estimer, pour chaque pixel, un vecteur de déplacement  $V = (v_x, v_y)$  exprimant :

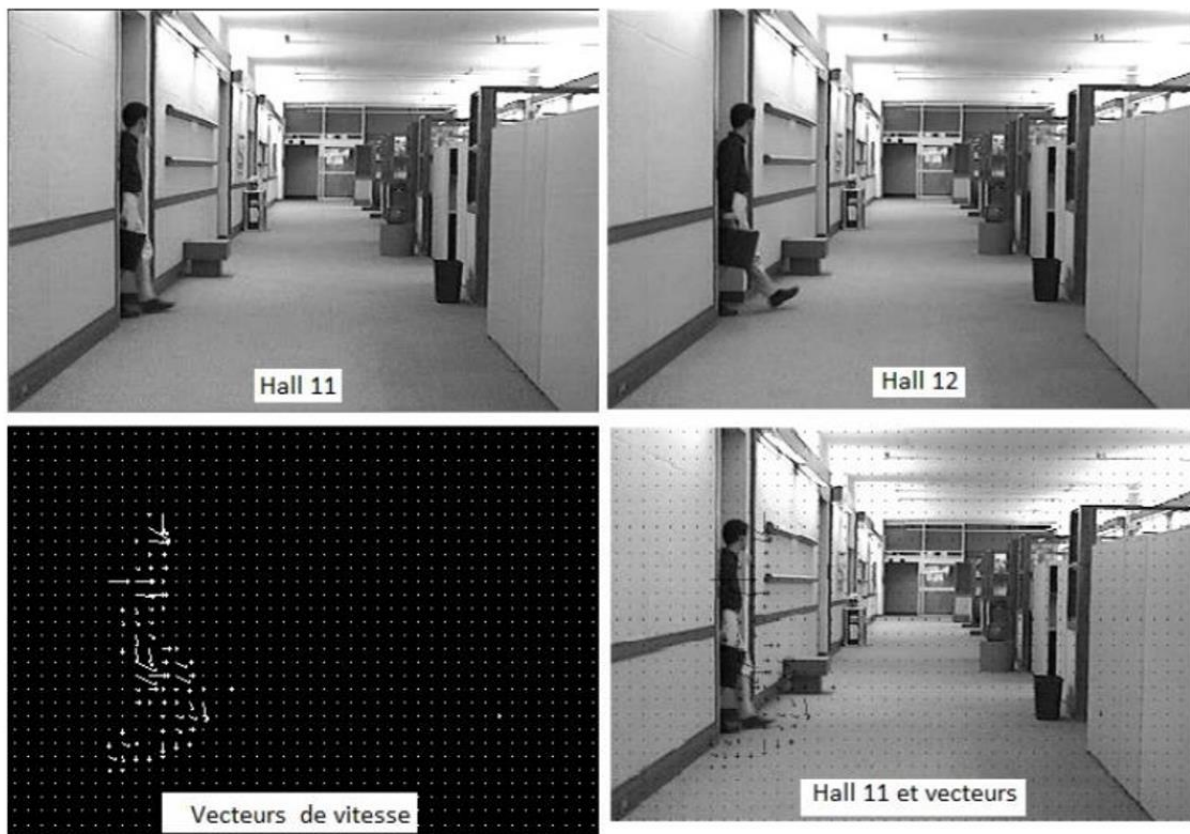
- La vitesse de déplacement d'un pixel dans le repère de l'image,
- La direction vers laquelle le pixel se déplace.

Exemple :

Supposons une illumination constante pour un point dans le temps :



En supposant que les propriétés d'illumination de l'objet sont conservées entre  $t$  et  $t + dt$ , on aura alors :  $I(x + dx, y + dy, t + dt) = I(x, y, t)$ .



### 3. Application

Il existe de nombreuses applications du flot optique :

- segmentation basée mouvement.
- La génération de formes 3D à partir du mouvement.
- La compensation globale du mouvement qui est utilisée dans la stabilisation des caméras vidéo ainsi que la compression vidéo.

### 4. Approches d'estimation du flot optique

Le flot optique représente le mouvement du contexte de la scène par rapport à un observateur. Pour un geste humain, une seule image avec seulement une scène statique peut introduire une ambiguïté qui rend difficile l'inférence de la classe de geste. Le mouvement des corps humains doit être exploité pour une bonne compréhension de la classe de geste. Il existe de nombreuses méthodes d'estimation du flot optique entre deux images, notamment les Approches variationnelles, fréquentielles, phasiques. [1]

#### 1.1. Les Approches variationnelles

Les méthodes basées sur une approche variationnelle consistent à résoudre un problème d'optimisation (local ou global) en minimisant une fonctionnelle, généralement basée sur l'équation (1) à laquelle on ajoute une contrainte pour particulariser les solutions. Nous présentons quelques unes :

$$(\nabla I)T \cdot \omega + It = 0 \quad \dots\dots\dots \text{Equation (1)}$$

##### a. Méthodes Variationnelles Globales

Ce type d'approche consiste à minimiser sur le domaine entier de l'image une fonctionnelle prenant en compte l'équation (1) du flot optique ainsi qu'un terme de lissage, c'est à dire en ajoutant une contrainte de régularisation portant sur le gradient, le laplacien (ou ordre supérieur) du champ de vitesse. La première méthode de calcul du flot optique a été développée en 1980 par **Horn & Schunck [2]** et consiste à minimiser sur l'ensemble de l'image la fonctionnelle :

$$J_{HS} = \iint [((\nabla I)^T W + I_t)^2 + a((\nabla v_x)^2 + (\nabla v_y)^2)] dx dy$$

Ce type d'approche a l'avantage de fournir un flot dense (un résultat pour chaque pixel), ce qui est requis par le cahier des charges. Toutefois, le résultat est assez lisse (la méthode étant globale, les mouvements principaux sont identifiés mais pas les variations locales) et il est difficile de régler les pondérations. A cela s'ajoute les problèmes numériques dû au choix de la méthode de minimisation (forcément itérative). Ces méthodes nécessitant un nombre assez élevé d'itérations successives (qu'il faudra également régler). Cela ne plaide pas en faveur de la parallélisation de la méthode. Par ailleurs, les résultats sont assez médiocres : en plus de la non identification des petits mouvements, le calcul est très sensible au bruit.

## b. Méthodes Variationnelles Locales

Les méthodes locales consistent à prendre en compte des hypothèses supplémentaires sur un domaine de taille réduite pour particulariser le flot optique. On minimise alors un critère sur un petit domaine, et on obtient ainsi le flot optique de ce petit domaine. La méthode locale la plus célèbre est celle de Lucas & Kanade [3] : la vitesse locale est supposée constante sur un voisinage spatial  $\Omega$ , on minimise alors la fonctionnelle

$$J_{LK} = \sum W 2[\nabla I \cdot \omega + I_t]^2$$

Sur ce domaine.

$W$  est une fenêtre locale, pouvant également être interprétée comme la pondération du critère des moindres carrés. On donne généralement une importance plus grande au pixel central (filtrage de type gaussien, facultatif). Des variantes de cette méthode existent, avec la prise en compte des dérivées secondes.

Les méthodes variationnelles locales sont intéressantes car hautement parallélisables (chaque calcul sur une petite fenêtre est indépendant des autres).

Les résultats sont par ailleurs moins sensibles au bruit et permettent le calcul de mouvements locaux. Il faut toutefois régler convenablement la taille de la fenêtre de recherche, et être conscient que les mouvements de grande taille ne peuvent être repérés si la taille de la fenêtre est trop faible (ce problème sera en partie résolu à l'aide d'une implémentation pyramidale).

## 1.2. Les Approches Fréquentielles

Les méthodes fréquentielles sont fondées sur une caractérisation du mouvement dans le domaine des fréquences. Les méthodes fréquentielles présentent de nombreux atouts tels que la simplicité de l'interprétation physique ou la cohérence avec d'autres traitements (compression, restauration d'images) utilisant les mêmes modèles.

Parmi les méthodes qu'on trouve dans cette approche, nous citons :

- a. Méthodes basées filtrage
- b. Méthode basées corrélation de phase
- c. Méthodes basées Ondelette

## 1.3. Les Approches Basées corrélation Block Matching

A partir d'un bloc de taille donnée de l'image, on cherche le déplacement de ce bloc (approximation de la vitesse) entre les deux images. On autorise pour cela un déplacement maximal et on cherche (différentes techniques d'exploration) le bloc qui correspond au mieux au bloc initial, de manière à minimiser un critère d'erreur (corrélation).

Les critères de corrélation les plus souvent employés sont les suivants :

- SSD Sum of Squared Differences.
- SAD Sum of Absolute Differences

## 1.4. Les Approches basées DeepNetwork

Les méthodes traditionnelles d'estimation du flot optique utilisent un modèle variationnel ou fréquentiel qui inclut le terme de données et le terme de lissage, qui peuvent établir une relation de contrainte entre deux images adjacentes et le flot optique. Cependant, la plupart d'entre eux sont trop lents pour être utilisés dans des applications en temps réel. Récemment, des réseaux de neurones convolutifs ont été utilisés avec succès dans la zone du flot optique.

De nombreuses méthodes d'apprentissage actuelles utilisent de grands ensembles de données qui contiennent la vérité du terrain pour la formation en réseau, qui peuvent utiliser les connaissances antérieures pour estimer directement le flot optique.

Le cœur du papier FlowNet 2.0 est basé sur les algorithmes FlowNetS, FlowNetC, FlowNetCSS, FlowNetSD, ces méthodes qui sont basées sur les DeepNetworks que nous allons introduire dans cette partie comme suit :

#### **a. FlowNetS**

FlowNetS ou précisément FlowNet Simple a été initialement présenté dans le document [4] et dont l'algorithme est résumé visuellement par le schéma qui suit :

Les données dans le FlowNetS données sont compressées spatialement dans une partie contractuelle du réseau, puis raffinées dans une partie en expansion. Il prend deux images, les concatène et les transmet à un réseau. La paire d'images est ensuite traitée et les informations de mouvement sont extraites.

La couche de raffinement est la suivante :

Dans cette partie du réseau, on effectue une série de conversions à la hausse afin d'augmenter la résolution de l'image, ce qui signifie que c'est la partie étendue du réseau.

Cela se fait en combinant des cartes d'entités de cours (d'une partie antérieure du réseau) avec les informations locales fines fournies dans des cartes d'entités de niveau inférieur.

#### **b. FlowNetC**

FlowNetC qui signifie FlowNet Correlated a été également introduit dans [4], il s'agit d'une amélioration de FlowNetS qui avait un EPE (fonction de perte pour le calcul du flot optique utilisant l'erreur de point final) élevé par rapport aux implémentations de flot optiques traditionnelles. L'algorithme FlowNetC peut être résumé selon le schéma suivant :

Il s'agit d'une structure de réseau moins générique où chaque image individuelle est introduite dans un flot de traitement séparé mais identique du réseau. Cela signifie que les représentations significatives sont apprises séparément, par exemple bords, etc. Ensuite, les représentations individuelles subissent des comparaisons multiplicatives de patch dans une couche de corrélation (idée similaire à une multiplication matricielle). Le résultat de la comparaison multiplicative des correctifs se déroule de manière similaire à FlowNetS, ce qui signifie qu'il passe par un processus de raffinement similaire.

#### **c. FlowNet 2.0**

En combinant toutes les idées ci-dessus, l'architecture FlowNet 2.0 est la suivante :

D'après le diagramme d'architecture FlowNet 2.0, nous pouvons voir que la toute première partie de la pile est le réseau FlowNetC, qui prend deux images et qui est conçu pour détecter les grands déplacements. Le champ d'écoulement résultant est appliqué à la deuxième image via la warping. Cette dernière est transmise avec la première image au réseau FlowNetS qui calculera les grands déplacements de flot et l'erreur de luminosité, qui représente la différence entre l'image déformée et l'image originale. Cette combinaison de FlowNetC et FlowNetS est appelée FlowNetCS.

Cette opération est ensuite répétée avec un second réseau FlowNetS qui produit un champ de flot et une magnitude de flot. Cette combinaison est appelée FlowNetCSS. L'idée principale de FlowNetCSS est que l'estimation des flot optiques peut être grandement améliorée en empilant les réseaux.



Un deuxième flot du réseau contient FlowNetSD (FlowNet Small Displacements) qui est alimenté par la première et la deuxième image originales. Dans l'article, il a été constaté que malgré l'empilement, le réseau était toujours incapable de produire avec précision des champs de flot pour les petits déplacements, ce qui entraînait beaucoup de bruit. Une variante de FlowNetS a donc été ajoutée au réseau avec quelques légères modifications afin de capturer des mouvements plus petits, ces modifications comprenaient le changement de la taille du noyau ainsi que l'ajout de convolutions pendant les convolutions ascendantes.

Les flots résultant de FlowNetCSS et FlowNetD sont fusionnés avec les flots des branches principales dans un réseau de fusion pour produire le champ de flot final.

#### Warping

En regardant le premier réseau FlowNetC, on calcule un flot optique. Le flot optique produit par FlowNetC est "appliqué" à la deuxième image pour décaler l'image en fonction du champ de flot optique afin d'essayer de faire correspondre l'image 1. Cette nouvelle image 2 est ensuite envoyée à la couche FlowNetS suivante. De cette façon, le réseau dans la pile peut se concentrer sur l'incrément restant entre l'image 1 et l'image 2. La raison pour laquelle cette approche fonctionne est que l'hypothèse selon laquelle si seule la première image est connue ainsi que les vecteurs de flot, alors la deuxième image peut être générée. C'est ce principe qui est utile dans la compression vidéo car une représentation de flot optique utilise moins de paramètres qu'une représentation vidéo réelle et évite ainsi la redondance.

## 5- Flot optique clairsemé vs dense

Un flot optique clairsemé donne les vecteurs de flot de certaines "caractéristiques intéressantes" (disons quelques pixels représentant les bords ou les coins d'un objet) dans le cadre tandis que le flot optique dense, qui donne les vecteurs de flot de l'ensemble du cadre (tous les pixels) - jusqu'à un vecteur de flot par pixel. Comme vous l'auriez deviné, le flot optique dense a une plus grande précision au prix d'être lent / coûteux en calcul.



### 6.1- Flot optique clairsemé

Un flot optique clairsemé sélectionne un ensemble de pixels (par exemple des éléments intéressants tels que les bords et les coins) pour suivre ses vecteurs de vitesse (mouvement). Les caractéristiques extraites sont transmises dans la fonction de flot optique d'une trame à l'autre pour garantir que les mêmes points sont suivis. Il existe différentes implémentations de flot optique clairsemé, notamment la méthode Lucas-Kanade, la méthode Horn – Schunck, la méthode Buxton – Buxton, et plus encore. Dans le code que je vais présenter, je vais utiliser la méthode de **Lucas-Kanade**.

Pour la mise en œuvre d'un flot optique clairsemé, nous suivons uniquement le mouvement d'un ensemble de pixels. Les fonctionnalités des images sont des points d'intérêt qui présentent des informations riches sur le contenu des images. Par exemple, de telles fonctions peuvent être des points de l'image invariants à la translation, à l'échelle, à la rotation et aux changements

d'intensité tels que les coins, pour cela nous allons nous baser sur le détecteur de Harris et le détecteur de Shi-Tomassi.

Le détecteur de coin Shi-Tomasi est très similaire au détecteur de coin Harris populaire qui peut être mis en œuvre par les trois procédures suivantes :

- 1- Déterminer les fenêtres (petits correctifs d'image) avec de grands gradients (variations d'intensité d'image) lorsqu'ils sont traduits dans les deux directions x et y,
  - 2- Pour chaque fenêtre, calculez un score R,
  - 3- En fonction de la valeur de R, chaque fenêtre est classée comme un plat, un bord ou un coin.
- /\*\* <https://courses.cs.washington.edu/courses/cse576/06sp/notes/HarrisDetector.pdf> \*\*/

Shi et Tomasi ont ensuite apporté une petite mais efficace modification au Harris Corner Detector dans leur document « Good Features to Track ».

/\*\* <https://ieeexplore.ieee.org/document/323794> \*\*/

La modification est à l'équation dans laquelle le score R est calculé. Dans le Harris Corner Detector, la fonction de notation est donnée par :

$$R = \det M - k (\text{trace} M)^2$$

$$\det M = \lambda_1 \lambda_2$$

$$\text{trace} M = \lambda_1 + \lambda_2$$

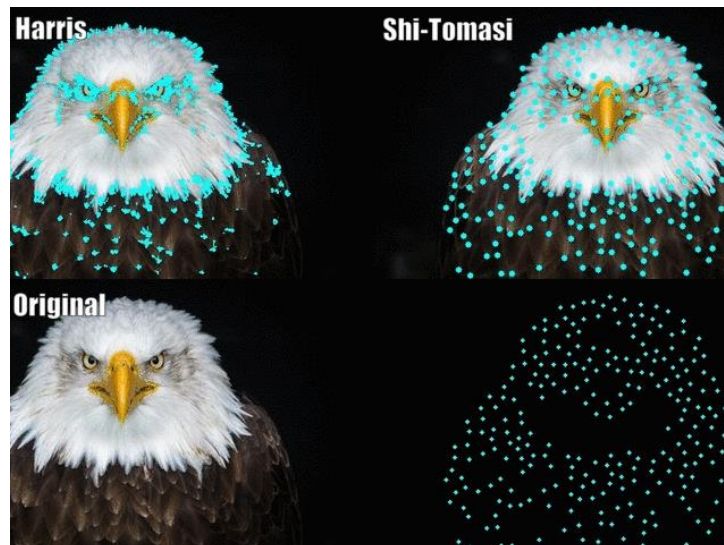
$$k = 0,02$$

Au lieu de cela, Shi-Tomasi a proposé la fonction de notation comme :

$$R = \min (\lambda_1, \lambda_2)$$

Ce qui signifie essentiellement que si R est supérieur à un seuil, il est classé comme un coin.

Ce qui suit compare les fonctions de notation de Harris (à gauche) et de Shi-Tomasi (à droite) dans l'espace  $\lambda_1 - \lambda_2$ .



Par la suite, nous allons calculer un flot optique pour un ensemble d'entités clairsemées à l'aide de la méthode itérative de Lucas-Kanade.

L'algorithme de Lucas-Kanade suit les correctifs d'image à travers les cadres.

En mathématiques, il optimise une translation  $(dx, dy)$  pour minimiser la différence entre 2 patches d'image.

Étant donné deux images A et B, nous voulons trouver une translation  $(dx, dy)$  telle que la différence entre A et B soit minimisée., nous voulons donc minimiser le coût de  $(dx, dy)$  telle que

$$\text{minimizeCost}(dx, dy) = \sum_{x,y} \|A(x, y) - B(x + dx, y + dy)\|^2$$



## 6.2- Flot optique dense

Gunnar Farneback a proposé une technique efficace pour estimer le mouvement de caractéristiques intéressantes en comparant deux images consécutives dans son article **Two-Frame Motion Estimation Based on Polynomial Expansion**.

## 6. Références Bibliographiques

- [1] Pavan Turaga, Rama Chellappa, Ashok Veeraraghavan : Advances in Video-Based Human Activity Analysis: Challenges and Approaches]
- [2] B.K.P. Horn, B. G. Schunck, Determining Optical Flow, Artificial Intelligence, 16(1--3):185-- 203, August 1981.
- [3] B.D. Lucas, T. Kanade, An Iterative Image Registration Technique with an Application to Stereo Vision In IJCAI81, pages 674--679, 1981.
- [4] FlowNet: Learning Optical Flow with Convolutional Networks