

La stéréovision

Elodie Lucas

Table des matières

1	Introduction	2
2	Principe de la stéréovision	2
3	La Normalized Cross Correlation	3
3.1	Notations, modèle et hypothèses	3
3.2	Normalized Cross Correlation	3
4	Formulation analytique de la corrélation	4
5	Conséquences de la méthode d'appariement par corrélation	4
6	Résultats	5
7	Conclusion	8

1 Introduction

Nous nous intéressons dans ce rapport à la stéréovision et notamment à son application sur des images satellitaires ou aériennes. Ce rapport est grandement inspiré de l'article "Small Baseline Stereovision" de Julie Delon. Nous verrons également l'application d'un algorithme permettant le calcul de la carte de disparité par block-matching sous pixelique.

Tout d'abord, nous verrons le principe de la stéréovision puis nous expliquerons les méthodes utilisées pour calculer la carte de disparité.

2 Principe de la stéréovision

La stéréoscopie est un procédé de reconstruction de profondeur à partir de deux images de la même scène. Cette méthode est apparue presque en même temps que la photographie. Le procédé de la stéréoscopie est calqué sur la perception humaine du relief grâce aux deux images planes que l'on perçoit de chaque oeil. Expliquons son fonctionnement : si deux images de la même scène sont acquises avec des angles différents, alors la profondeur de la scène crée une disparité géométrique entre elles. Notons ε cette fonction de disparité. Alors, si l'altitude des caméras est suffisamment haute pour que le modèle de projection parallèle soit précis, ε et la fonction de profondeur z sont liées par la relation

$$z = \frac{\varepsilon}{b/h}$$

où b/h est un coefficient de stéréoscopie dépendant seulement des conditions d'acquisition. b est la distance entre les centres des caméras et h est la distance entre la scène et le système de caméras.

La précision dz de la mesure de profondeur est donc liée à la précision $d\varepsilon$ de la mesure de disparité :

$$dz = \frac{d\varepsilon}{b/h}$$

Tout ceci est représenté sur la figure 1.

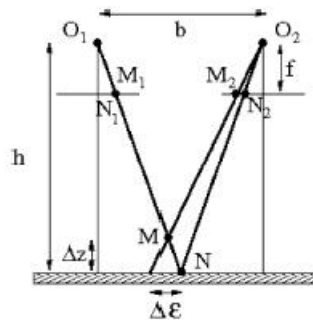


FIGURE 1 – Principe de la stéréoscopie

Pour une précision $d\varepsilon$ donnée, plus le coefficient b/h est grand et plus l'erreur de profondeur dz est petite. Mais un grand coefficient b/h signifie aussi qu'il y a beaucoup de changements entre les deux images de la paire stéréoscopique, ce qui implique plus de difficultés dans le processus d'appariement. Il faut donc faire un compromis entre une erreur de profondeur petite et une erreur de disparité petite, c'est-à-dire un processus d'appariement fiable.

Cependant, quand les deux vues sont séparées par un angle très petit, les disparités observées sur les images sont très faibles en comparaison avec la taille d'un pixel. C'est pourquoi les méthodes d'appariement pixeliques ne sont pas en mesure de trouver une information de profondeur intéressante.

On s'intéressera dans notre cas à des méthodes d'appariement subpixeliques. Mais apparier deux images avec un petit coefficient b/h a un sens seulement si la précision perdue à cause de l'angle entre les caméras est compensée par une meilleure précision sur ε . Pour atteindre une telle précision, les images doivent être parfaitement interpolées et donc échantillonnées grâce au théorème de Shannon.

Il existe deux types de méthodes utilisées pour apparier deux images : les méthodes locales et les méthodes globales. Les méthodes locales consistent à calculer la disparité d'un point donné en observant uniquement ses plus proches voisins. Parmi ces méthodes, la méthode block-matching consiste à estimer la disparité en x en comparant une vignette autour de x dans la première image avec des vignettes similaires dans la deuxième image. Mais ces méthodes sont difficilement fiables dans des zones non texturées de l'image, et elles peuvent engendrer des artéfacts d'adhérence.

Les méthodes globales résolvent des problèmes d'optimisation sur la carte de disparité entière. Elles impliquent des méthodes de minimisation d'énergies compliquées.

On se concentrera par la suite sur la méthode d'appariement la plus standard, la Normalized Cross Correlation, qui consiste à estimer en chaque point l'erreur d'appariement.

3 La Normalized Cross Correlation

3.1 Notations, modèle et hypothèses

On note u et \tilde{u} les images de la paire stéréoscopique. On suppose, sans perte de généralité, que ces images sont $2\pi \times 2\pi$ périodiques et connues sur $[-\pi, \pi] \times [-\pi, \pi]$. Seules les versions discrètes de ces images sont disponibles et on les suppose bien échantillonnées sur une grille régulière $2N \times 2N$. Cette hypothèse n'est pas satisfaite dans le cas d'un système d'acquisition réel mais devient valide dans le cas des satellites SPOT5.

Supposons que u et \tilde{u} satisfont le modèle suivant :

$$\tilde{u}(x) = \lambda(x)u(x + \varepsilon(x))$$

où λ varie peu et la fonction de disparité ε , supposée bornée, décrit la déformation géométrique entre u et \tilde{u} .

Ce modèle est faux si l'angle entre les deux vues est trop grand, mais cela reste raisonnable si b/h est petit. En effet, le modèle suppose que les différences entre u et \tilde{u} sont purement géométriques jusqu'à une fonction multiplicative λ avec des variations faibles, et qu'aucune occultation ou changement de radiométrie se produit. Le modèle devient de plus en plus précis lorsque b/h devient petit. La vision humaine satisfait aussi ces hypothèses.

3.2 Normalized Cross Correlation

Considérons φ une fonction lisse, positive, normalisée et à support compact. On introduit les notations suivantes :

- $\varphi_{x_0} : x \mapsto \varphi(x_0 - x)$
- $\int_{\varphi_{x_0}} f = \int_{\varphi_{x_0}} f(x)dx = \int \varphi(x_0 - x)f(x)dx$ pour toute fonction intégrable f
- $\|f\|_{\varphi_{x_0}}$ la norme pondérée $\sqrt{\int \varphi_{x_0}(x)f^2(x)dx}$ pour toute f^2 intégrable
- $\langle, \rangle_{\varphi_{x_0}}$ le produit scalaire tel que $\langle f, g \rangle_{\varphi_{x_0}} = \int \varphi_{x_0}(x)f(x)g(x)dx$
- $\tau_m u : x \mapsto u(x + m)$

Pour chaque point x_0 de \tilde{u} , la normalized cross correlation calcule la disparité $m(x_0)$ entre u et \tilde{u} en x_0 en maximisant un coefficient de similarité locale entre u et \tilde{u} :

$$m(x_0) = \arg \max \rho_{x_0}(m)$$

$$\rho_{x_0}(m) = \frac{\langle \tau_m u, \tilde{u} \rangle_{\varphi_{x_0}}}{\|\tau_m u\|_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}}$$

La fonction ρ_{x_0} est appelée le produit de corrélation en x_0 et φ est appelée la fenêtre de corrélation. La valeur $\rho_{x_0}(m)$ mesure la similarité entre les voisins de x_0 dans \tilde{u} et les voisins de $x_0 + m$ dans u .

Il n'est cependant pas garanti que le $m(x_0)$ où ρ_{x_0} est maximal soit exactement égal à la disparité réelle $\varepsilon(x_0)$ en x_0 .

4 Formulation analytique de la corrélation

Dans toute la suite, nous faisons l'hypothèse que les images ont été rectifiées par géométrie épipolaire. La géométrie épipolaire est la géométrie qui s'intéresse à deux images prises à deux vues différentes et elle dépend des paramètres internes des caméras.

On note

$$d_{x_0}^u : x \mapsto \frac{\|u\|_{\varphi_{x_0}}^2 u'(x)^2 - \langle u, u' \rangle_{\varphi_{x_0}} u(x) u'(x)}{\|u\|_{\varphi_{x_0}}^4}$$

la densité de corrélation de u en x_0 . Elle dépend seulement de l'image u , de la fenêtre φ et du point de corrélation x_0 .

On peut alors établir l'équation centrale de corrélation :

Supposons que la fonction de disparité ε et que $m(x_0)$ soient tels que $|\varepsilon(x) - m(x_0)| \ll 1$ sur le support de φ_{x_0} . Alors $m(x_0)$ est lié à ε par le premier ordre d'approximation :

$$\langle d_{x_0}^{\tau_{m(x_0)} u}, m(x_0) \rangle_{\varphi_{x_0}} \cong \langle d_{x_0}^{\tau_{m(x_0)} u}, \varepsilon \rangle_{\varphi_{x_0}}$$

Si ε n'est pas constante sur le support de φ_{x_0} , cette propriété montre que les valeurs $\varepsilon(x)$ qui comptent le plus dans la mesure $m(x_0)$ sont prises aux points x tels que $d_{x_0}^{\tau_{m(x_0)} u}(x)$ est grand. Cette propriété peut être interprétée comme un phénomène d'adhérence. Dans les zones où u et \tilde{u} sont constantes, la densité de corrélation est nulle, ce qui signifie qu'il n'y a pas de relations entre m et ε pouvant être déduites de la propriété précédente. Ceci tend à montrer que la corrélation a besoin d'informations de texture pour fonctionner.

5 Conséquences de la méthode d'appariement par corrélation

La méthode d'appariement par corrélation engendre des effets d'adhérence. Cet effet apparaît dans le voisinage d'une profondeur discontinue, spécialement quand cette discontinuité est amplifiée par une discontinuité de niveau de gris. Ceci résulte d'une dilatation de l'upper-ground dans la carte de disparité. Expliquons ce phénomène à l'aide d'un exemple :

Considérons un bâtiment avec de la texture qui "ment" sur la texture du fond, c'est-à-dire qu'une partie du fond est occultée par le bâtiment dans l'image de gauche. On suppose que la différence de niveau de gris entre le fond et le bâtiment est plus grande que les variations d'intensité dans les zones de textures. Soit Q un point dont la distance au bâtiment est plus petite que la moitié de la fenêtre d'appariement. Si on regarde dans l'image de droite pour le meilleur correspondant pour Q , une méthode de block-matching choisirait probablement P . Ce qui signifie que la disparité accordée à Q sera la même que celle du bâtiment. En conséquence, le bâtiment reconstruit sera dilaté par la moitié de la taille de la fenêtre.

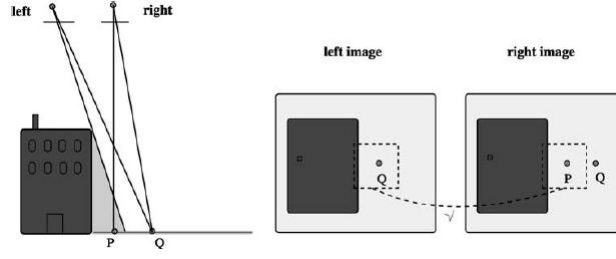


FIGURE 2 – Exemple de l'effet d'adhérence

La propriété précédente fournit une explication sur ce phénomène. En effet, supposons que la fonction de densité de corrélation en x_0 soit en réalité concentrée autour du point x_1 tel que $d_{x_0}^{T_{m(x_0)}u}$ puisse être approchée par la fonction δ_{x_1} , alors on obtient

$$\varepsilon(x_1) \simeq m(x_0)$$

Ce qui signifie que le $m(x_0)$ mesuré par corrélation en x_0 est en réalité la disparité du point x_1 . En 2D, si le voisinage de x_0 est composé de zones plates sur les deux côtés du contour, alors le $m(x_0)$ mesuré en x_0 est une moyenne de la disparité réelle du contour. Ceci produit un effet d'adhérence si le niveau de gris du contour diminue avec l'augmentation de la discontinuité. Cet artéfact est inhérent à toute méthode de block-matching.

6 Résultats

Un algorithme multi-échelles est proposé dans l'article de Julie Delon mais je n'ai pas pu le coder entièrement par faute de temps. Dans cette algorithme, elle détermine de façon automatique la taille de la fenêtre de corrélation.

J'ai programmé un algorithme, non multi-échelles, permettant le calcul de la carte de disparité par corrélation. Son fonctionnement est le suivant :

Pour chaque pixel x_0 de l'image u , on crée une vignette de taille 11. On crée cette même vignette sur l'image \tilde{u} que l'on déplace le long de la droite épipolaire. On calcule le coefficient de corrélation entre chaque vignette et on cherche le maximum. On récupère la distance séparant les deux points où ce maximum est atteint. Ce qui va nous donner une carte de disparité. Ce procédé engendre des effets de bord puisque l'on se sert de vignettes.

Voici la paire d'image sur laquelle j'ai travaillé :



FIGURE 3 – Pentagone, image de gauche



FIGURE 4 – Pentagone, image de droite

Le calcul de la carte de profondeur à l'échelle pixellique donne les résultats suivants :



FIGURE 5 – Carte de profondeur, $b/h = 0.02$



FIGURE 6 – Carte de profondeur, $b/h = 0.05$

On peut alors reconstruire une figure 3D grâce à Matlab et choisir de ne garder que les points dont l'intensité est supérieure à 230 sur la carte de profondeur. Pour les autres, on choisit une profondeur égale à 0, ce qui permet de ne faire ressortir que le pentagone. On supprime également les points blancs situés sur le bord de l'image.

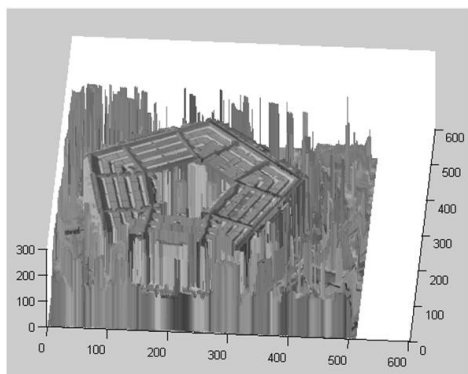


FIGURE 7 – Reconstruction 3D, $b/h=0.02$

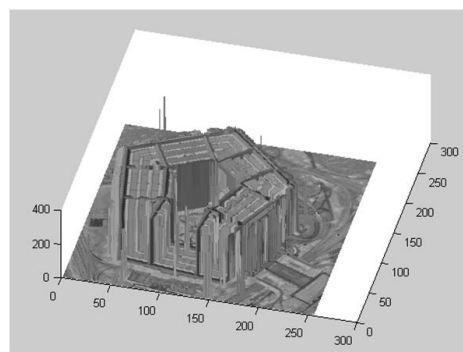


FIGURE 8 – Reconstruction 3D, seuillage de la carte de profondeur, $b/h = 0.02$

En sur-échantillonnant l'image par un facteur 2 en se servant de la fonction `interp2` de Matlab, on obtient :

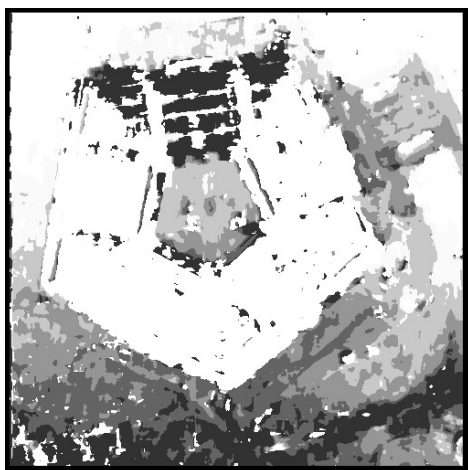


FIGURE 9 – Carte de profondeur, $b/h = 0.02$

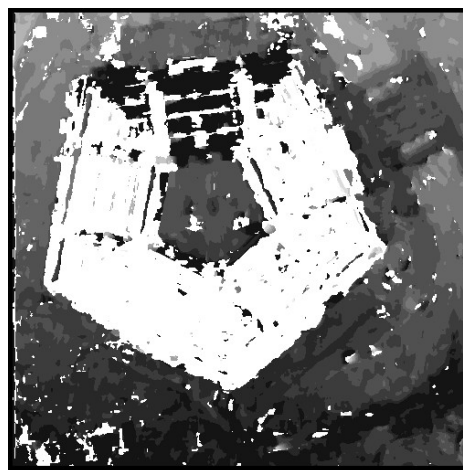


FIGURE 10 – Carte de profondeur, $b/h = 0.05$

On constate bien une amélioration de la carte de disparité à la précision sub-pixellique. La reconstruction 3D sera donc légèrement meilleure mais le seuillage ne permet pas de garder seulement le pentagone. Je choisis alors de garder la plus grande composante connexe, ce qui engendre une perte d'information sur le côté du pentagone situé en haut de l'image originale.

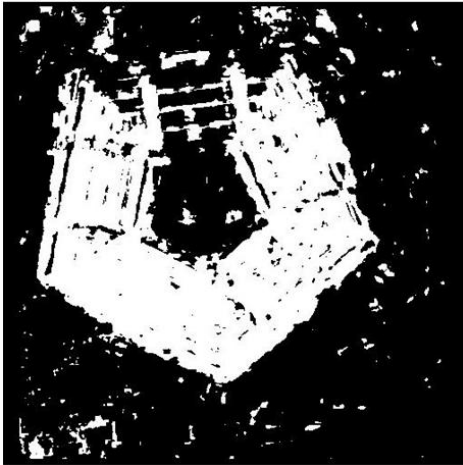


FIGURE 11 – Carte de profondeur après seuillage, $b/h = 0.05$

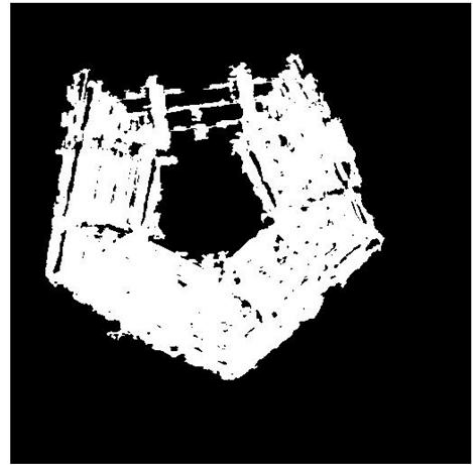


FIGURE 12 – Carte de profondeur, plus grande composante connexe, $b/h = 0.05$

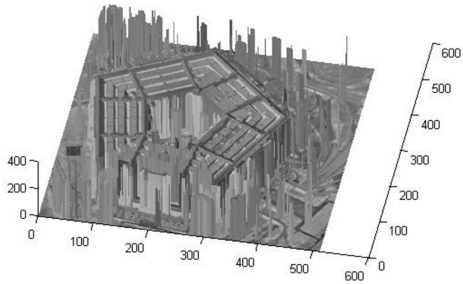


FIGURE 13 – Reconstruction 3D, seuillage de la carte, $b/h = 0.05$

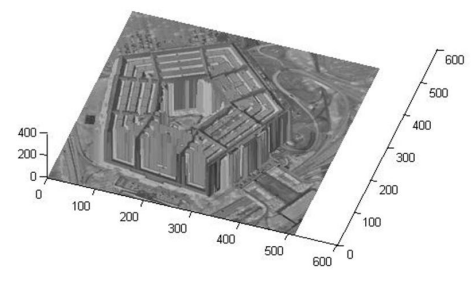


FIGURE 14 – Reconstruction 3D, plus grande connexe de la carte, $b/h = 0.05$

7 Conclusion

Ce projet m'aura permis d'approfondir les notions de stéréovision aperçues en cours. Il aurait été intéressant de voir des résultats avec un facteur d'interpolation plus grand mais le temps de calcul est trop important sous Matlab.