

## Photogrammétrie et vision par ordinateur

Mahzad KALANTARI

Laboratoires Irceyn (Nantes) et Matis (IGN)

Michel KASSER

ENSG (IGN)

### Résumé

Une comparaison est effectuée entre les façons employées en photogrammétrie et en vision par ordinateur pour traiter les problèmes liés à l'acquisition de la 3D à partir d'images stéréoscopiques. Le formalisme adopté, assez différent, est présenté, et quelques perspectives d'évolution en sont déduites.

### Abstract

*A comparison is done between the ways used in photogrammetry and in computer vision to process the problems bound to the acquisition of 3D from stereoscopic pictures. The formalisms adopted, somewhat different, are presented, and some perspectives of evolution are deduced.*

### Mots-clés

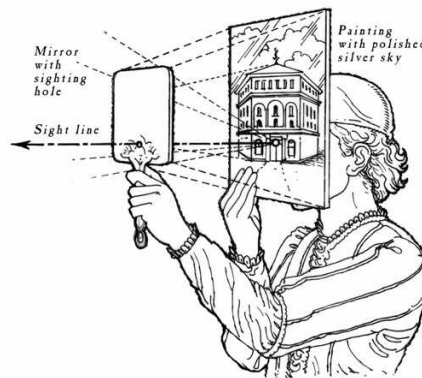
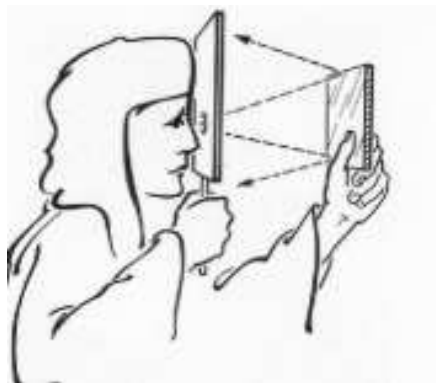
Photogrammétrie, vision par ordinateur, matrice essentielle, matrice fondamentale, orientation relative.

### Introduction

Au cours des dernières années, il est devenu de plus en plus apparent que la communauté de vision par ordinateur (*Computer Vision* en anglais), après avoir plus ou moins ré-inventé tout le corpus de connaissances de la photogrammétrie en partant de zéro, en plus d'un très grand nombre d'autres choses, était en train de marginaliser nettement celle-ci. Le présent article vise donc à évaluer les situations respectives de ces deux domaines techniques, vus par des géomaticiens, et à présenter leurs liens les plus apparents, afin d'essayer d'anticiper ce que sera la photogrammétrie de demain.

### 1/Des domaines techniques très proches, et d'histoires très différentes

La photogrammétrie est une technique qui a suivi très directement l'invention de la photographie au XIX<sup>ème</sup> siècle. Déjà au cours des siècles précédents, différents peintres avaient étudié la perspective en utilisant des dispositifs optiques simples : par exemple, Brunelleschi (Figure ci-dessous), Albert Dürer, etc... L'idée était déjà de fixer sur le papier une image aussi neutre, objective et conforme à la réalité que possible.



Ensuite, l'exploitation d'images pour mesurer les distances de différents objets n'était qu'un simple remploi de techniques de topographie, de type triangulation et intersection. Et comme les publications de ces techniques, dès le XVI<sup>ème</sup> siècle, le montraient bien, les applications envisagées étaient d'abord de type militaire : comment ajuster le tir d'un canon, comment cartographier une place forte ennemie sans s'en approcher, etc...

Dès que Nadar a produit au le milieu du XIX<sup>ème</sup> siècle les premières images aériennes, depuis un aérostat, ce sont des applications militaires qui ont encore été le moteur principal de la photogrammétrie naissante. Et A. Laussedat, l'inventeur de cette technique, était polytechnicien et officier du Génie. A cette époque, la cartographie nationale était elle aussi sous tutelle militaire, comme dans pratiquement dans tous les pays. C'est donc très logiquement que cette technique a été rapidement orientée vers des applications de cartographie, sur des surfaces très étendues, et donc de façon de plus en plus industrielle. En parallèle, les applications architecturales ont aussi été développées dès le début, mais sans rencontrer des débouchés commerciaux de même niveau, et ce ne sont donc pas elles qui ont été prépondérantes.

Ceci pour expliquer que la principale orientation de la photogrammétrie, quasiment dès sa naissance, a été la cartographie à partir de vues aériennes. Beaucoup de conséquences en découlent, en particulier :

- une recherche de précision au meilleur niveau, afin de faire la meilleure cartographie possible avec un nombre minimal de photos,
- un travail avec un axe optique quasi-vertical, et donc des photos à axes presque parallèles,
- un travail de restitution qui peut prendre beaucoup de temps, le délai entre la prise de vues et la cartographie résultante pouvant se chiffrer en mois, voire en années.
- des développements qui progressivement ont quitté le domaine académique (avec de nombreuses publications), pour devenir quasi-exclusivement du domaine industriel, sans aucune publication. Ceci s'est traduit dans les dernières décennies, lors du passage au numérique, par de véritables boîtes noires sans aucun moyen pour l'utilisateur de savoir ce qu'il s'y faisait exactement. On a d'ailleurs assisté régulièrement à des travaux de recherche, dans des domaines connexes potentiellement usagers de cette technique, qui en ré-inventaient tout ou partie, par faute de publications accessibles. Quand une technique est ainsi portée par les seuls industriels, c'est un écueil fréquent, les étudiants cherchent des publications de recherche et n'en trouvent pas, alors qu'il s'agit de sciences de l'ingénieur, la partie récente étant presque totalement couverte par le secret industriel.

La vision par ordinateur est par contre un domaine qui n'a guère plus de trois décennies d'existence. Il s'est développé dès qu'on a su numériser des images vidéo, et il couvre de nombreuses applications orientées vers le temps réel, ceci incluant l'extraction automatique d'éléments dans l'image. D'abord, simplement les contours, puis des éléments de plus en plus évolués, tels que des objets connus (des pièces mécaniques empilées en vrac), ceci allant jusqu'à des objets très complexes (reconnaitances de visages). Et puis ensuite, la volumétrie des objets visibles, à partir d'images prises de deux points de vues différents, et permettant un effet stéréoscopique. Une utilisation évidente a été le domaine de la robotique, l'objectif étant de permettre à une plateforme autonome de cartographier en temps réel son environnement immédiat, avec une exigence de précision assez modeste, mais variable : comme pour tout être vivant, le besoin de précision est d'autant plus grand que les objets sont proches, et la vision humaine est parfaitement adaptée à ce besoin.

Fig. 1 : les applications inattendues de la vision par ordinateur ne manquent pas, du temps réel, des résultats robustes, mais pas vraiment de besoin de grande précision (RoboCup2007-Day3-06, crédit photo Rob Felt, 2007 Georgia Tech).



Dans cette communauté règne une intense activité de recherche, poussée par des demandes industrielles très fortes, qui atteignent actuellement le grand public : citons par exemple, dans ce domaine, pour les appareils photos actuels les mises au point automatiques qui localisent d'elles-mêmes la zone d'image sur laquelle elle doivent s'exercer, ou même mieux, la photo qui ne se déclenche que quand le sujet photographié sourit et ne ferme pas les yeux : véritable prouesse, quand on y réfléchit. Mais dans les domaines techniques professionnels, la multiplication des surveillances vidéo a contribué, à son tour, à susciter une demande considérable pour trouver de façon automatique, parmi des millions d'heures d'enregistrements, tel type d'objet, de véhicule, de visage, etc...

Donc la photogrammétrie et la vision par ordinateur partagent indiscutablement une même recherche de mesure 3D à partir d'images permettant la stéréoscopie, mais leurs passés respectifs et leurs clientèles très différentes les ont amenées à se développer de façon complètement parallèle, avec peu de zones communes.

## 2/ Une même géométrie, deux formulations

Une étape importante, pour ces deux communautés, est la mise en place des images dans l'espace, c'est-à-dire le calcul de la position et l'orientation des images au moment des prises de vues. En photogrammétrie cette étape est appelée l'orientation externe, en symétrique de l'orientation interne qui cherche à déterminer les paramètres de calibration de la caméra comme la focale, le centre principal d'autocollimation, le polynôme de distorsion ainsi que le centre principal de symétrie. L'orientation externe est basée sur les équations de colinéarité. Comme le nom l'indique, celles-ci consistent à dire que le point terrain  $M$ , le sommet de prise de vue  $S$  (qui correspond au centre optique), et  $m$  la projection de  $A$  sur l'image, sont sur la même droite [1].

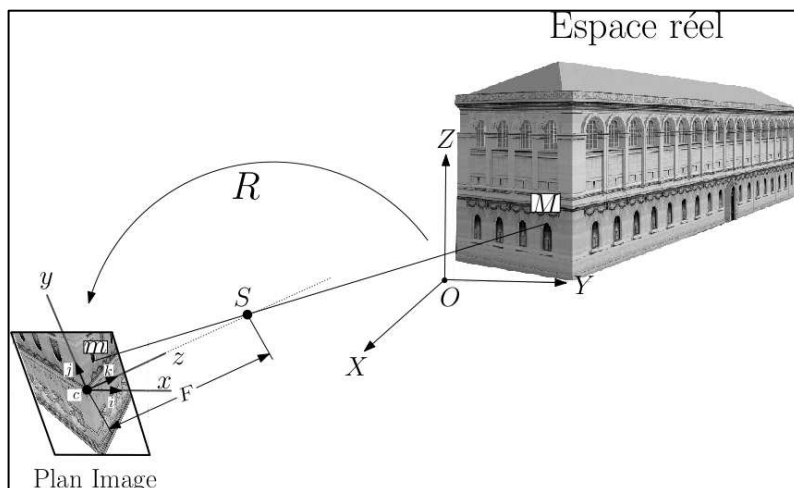


Figure 2 : La projection conique utilisée en photographie, formalisée par l'équation de colinéarité (illustrée par un modèle 3D de la bibliothèque Ste Geneviève réalisé par Leonhard Pröttel)

En entrée de l'équation de colinéarité on a comme paramètres les coordonnées images des points, et la calibration de la caméra, un point terrain étant nécessairement vu au moins sur deux images. Certains des points terrain sont exprimés dans un référentiel connu (les points d'appui), ce qui permet de réaliser l'orientation absolue. Mais on peut très bien dans un premier temps ne pas disposer de points d'appui, auquel cas on travaille dans un espace euclidien, mais sans échelle (orientation relative). En sortie des équations de colinéarité, nous pouvons obtenir l'orientation des images, les sommets de prises de vues, ainsi que les coordonnées terrain des points de liaison. Ces coordonnées sont obtenues soit dans un référentiel connu et avec une échelle si nous avons fait l'orientation absolue, soit dans le référentiel de la première image et sans échelle si seule l'orientation relative a été effectuée. Il est tout à fait possible, à tout moment, de faire basculer les points 3D, obtenus à partir de l'orientation relative, dans un référentiel connu, et pour cela donc il faut disposer de points connus dans le référentiel voulu.

Revenons maintenant à la résolution de ces équations de colinéarité. Il faut au minimum 5 points homologues pour pouvoir effectuer l'orientation relative, or en général bien plus de 5 points homologues peuvent être disponibles (par exemple en utilisant des outils d'extraction automatique de points d'intérêt, cf § 3 ci-après). Dans ce cas il faudra minimiser par moindres carrés le carré de la distance entre le point théorique terrain et celui que l'on calcule. Le problème de ces équations est leur non linéarité : Une étape de linéarisation est donc nécessaire. Et une fois la linéarisation effectuée il faut posséder toutes les valeurs initiales des inconnues recherchées. En photogrammétrie aérienne, ces valeurs sont faciles à déterminer. Outre le fait que les prises de vues sont quasi verticales et donc à axes parallèles, la hauteur de vol ainsi que la position des sommets de prises de vues peuvent être aisément connus de manière approchée. Ce qu'on peut retenir ici est que la photogrammétrie dans son utilisation classique sur vues aériennes simplifie de manière significative la résolution des équations de colinéarité. Et par la suite la compensation par faisceaux, traitement qui s'applique lorsqu'on traite en bloc un grand nombre d'images.

Le problème est plus compliqué quand la photogrammétrie s'attaque à des prises de vues terrestres. La première grande différence est que les prises de vues ne sont plus à axes quasi parallèles, mais deviennent convergentes. D'autre part, à moins d'avoir des points d'appui connus, le calcul des paramètres approchés devient lui aussi plus compliqué qu'en photogrammétrie aérienne.

Dans un contexte de vision par ordinateur et de robotique des années 80, il y avait ce besoin de pouvoir déterminer de manière directe et surtout linéaire les paramètres d'orientation et de position. Pensons à un robot qui doit se déplacer et voir le monde en trois dimension pour obtenir des informations de type topologique, du genre "la table est derrière la chaise" : premièrement il n'est pas

nécessaire pour ce genre d'application d'avoir des points d'appui, et ensuite une grande précision telle que celle que l'on cherche en photogrammétrie n'est pas nécessaire.

C'est pour cela qu'une nouvelle modélisation a été proposée, basée non pas sur les équations de colinéarité, mais sur une autre contrainte très simple elle aussi, appelée contrainte de coplanarité.

Comme on le voit mieux sur la figure 3, la condition de coplanarité entre deux images exprime le fait que le vecteur de visée depuis le premier sommet de prise de vues  $\vec{V}_1$ , le vecteur de visée depuis le deuxième sommet de prise de vues (et exprimé dans le référentiel du premier)  $\vec{V}_2$ , ainsi que le vecteur de la translation (entre les deux sommets de prises de vues)  $\vec{T}$  se trouvent dans le même plan, appelé le plan épipolaire.

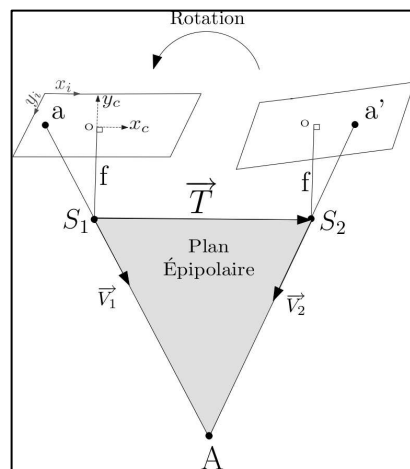


Figure 3 La matrice rotation de la seconde caméra par rapport à la première est appelée  $R$ , et le vecteur de translation  $T$  est la base qui relie les centres optiques des caméras ( $S_1$  et  $S_2$ ).  $O$  est le point principal d'autocollimation (ppa). Les images du point terrain  $A$  sur les 2 images sont  $a$  et  $a'$ .

On peut traduire cette condition par un produit mixte nul entre ces 3 vecteurs. En d'autres termes :

$$\vec{V}_2^t \cdot (R\vec{V}_1 \wedge \vec{T}) = 0 \quad (1)$$

En exprimant le produit vectoriel de manière algébrique (la translation étant exprimée sous forme d'axiateur), l'équation (1) peut être simplifiée dans sa forme matricielle :

$$\begin{bmatrix} x_{a2} & y_{a2} & f \end{bmatrix} \begin{bmatrix} 0 & Tz & -Ty \\ -Tz & 0 & Tx \\ Ty & -Tx & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_{a1} \\ y_{a1} \\ f \end{bmatrix} = 0 \quad (2)$$

Le produit des deux matrices  $3 \times 3$ , l'axiateur formé sur le vecteur translation  $\vec{T}$ , et la matrice rotation  $R$ , est une matrice  $3 \times 3$ , qui a été dénommée, un peu pompeusement certes, mais c'est maintenant un terme d'usage, matrice Essentielle (E) [2]. Pour définir la matrice E, les paramètres de calibration doivent être connus. Dans ce cas, et c'est comme avec l'orientation relative en photogrammétrie, nous travaillons dans un espace euclidien à une échelle près.

L'autre façon d'aborder le problème par cette communauté a été de gérer au mieux des cas où la calibration est inconnue : on ne sait rien de la caméra, de sa focale, du centrage de l'optique, etc..., un peu comme quand on regarde une photo d'album [3]. Dans ce cas là, avec les mêmes équations que (2), au lieu d'avoir la matrice  $E$ , nous obtenons ce qu'il a été convenu d'appeler la matrice Fondamentale ( $F$ ), qui exprime simplement la stéréoscopie d'images acquises en géométrie conique, ce qui conduit à travailler dans l'espace projectif, parfaitement adapté à ce cas : Un espace où les angles ne sont pas préservés et où les droites parallèles se coupent au point de fuite.

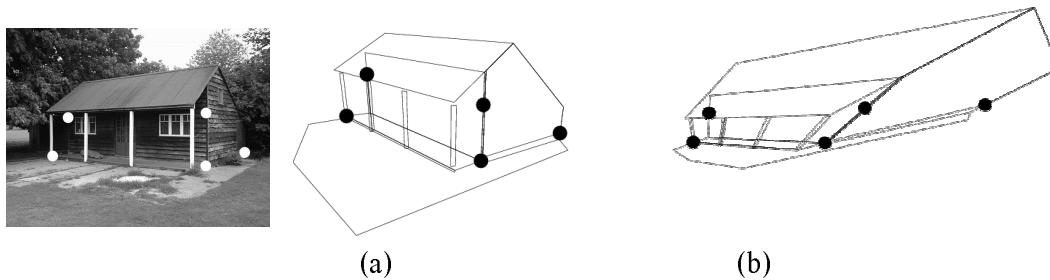
Ici nous insistons sur le fait que beaucoup de photogrammètres pensent que la communauté de vision par ordinateur ne travaille que dans l'espace projectif. Ce qui est inexact car, dans bien des cas, c'est la matrice  $E$  qui est employée, et nous travaillons alors dans un espace euclidien classique.

Un autre point très important auquel il faut faire attention est que, pour la résolution de la matrice  $E$  ou  $F$ , la notion de moindres carrés est sans objet. Car dans ces équations, comme on peut le voir, les points terrains ne figurent pas, à l'inverse des équations de colinéarité, et il n'y a donc rien à minimiser. La résolution de ce genre d'équations homogènes se fait à l'aide d'une décomposition SVD (*Singular Value Decomposition*), que nous ne détaillerons pas ici.

La première résolution de la matrice  $E$  dans les années 80 nécessitait 8 points [2]. Rappelons que, comme dans un contexte d'orientation relative il n'y a que 5 inconnues (3 paramètres de la rotation et 2 paramètres pour la base, car on travaille à une échelle près), donc 5 points suffisent. Or avec la nouvelle résolution à partir de 8 points, 3 degrés de liberté qui ne correspondent pas à la physique réelle ont été rajoutés.

C'est pour cela qu'avec cette résolution, quand l'objet est plan, par exemple un mur, cela devient un cas dégénéré. Pendant les vingt dernières années, une recherche de la communauté de vision par ordinateur a été de donner une résolution directe avec le minimum de points possible, c'est-à-dire 5, avec une contrainte de temps réel [4][7].

Figure 4 Reconstruction basée avec l'aide la matrice Essentielle (a) et Fondamentale (b). Tiré de "Multiple View Geometry in Computer Vision" de Hartley & Zisserman [3].



### 3/ Quelle place pour le traitement d'image ?

Comme nous l'avons déjà évoqué, la notion de traitement d'image couvre un champ bien plus vaste que celui de la photogrammétrie, avec en particulier tous les domaines d'extraction automatique de segments, de formes plus ou moins complexes, où les aspects 3D n'interviennent que peu, voire pas du tout. Mais par ailleurs, dans l'activité du photogrammètre, interviennent toute une série d'aspects qui sont du même registre, sauf qu'elle n'est pas encore bien automatisée compte tenu de sa grande complexité : c'est la phase de photo-identification, où on ne parvient pas encore à se passer de l'intelligence humaine. Extraire un segment, oui, mais est-ce un bord de chemin, un mur, une limite de parcelle, le bord d'une meule de foin ? Jusqu'ici l'homme est nécessaire, et il reste encore une grande marge de progrès possibles pour des démarches d'automatisation. On n'a pas encore tiré tout le profit de l'extraordinaire puissance de calcul disponible, ainsi que de la très grande dynamique des images

que l'on obtient couramment, avec un bruit extrêmement faible, créant des conditions pourtant très favorables aux traitements numériques.

Néanmoins, certains algorithmes de traitement d'image sont déjà très largement employés, ce sont ceux qui permettent l'extraction automatique de points d'intérêt. Une phase critique, qui intervient au cours de l'orientation relative de deux images, est en effet celle qui consiste à identifier, sur deux photographies différentes, les deux points qui sont les images d'un même point du terrain. Pour automatiser cette phase, on commence par extraire de chaque image des points faciles à pointer, et logiquement, si ces points ont été bien choisis, on parvient ensuite à mettre en correspondance la plupart des points d'une image avec ceux de l'autre dans la zone vue en stéréoscopie : on appelle ces points les points d'intérêt. Bien évidemment il est important d'éviter les points mal définis (p. ex. pris le long d'une bordure), ceux qui n'ont pas de définition géométrique stable (p. ex. bordure d'une ombre, intersection de lignes qui ne sont pas dans un même plan), etc. L'automatisation de l'extraction des points d'intérêt est directement liée à l'emploi d'images numériques, et l'un des premiers outils utilisés a été publié par Harris en 1986 [5] : son détecteur est basé sur l'extraction automatique de coins, et si assez rapidement ses insuffisances ont été connues, sa simplicité d'implémentation en a fait un outil très employé. Néanmoins il a fallu attendre les résultats de deux décennies de recherche pour disposer d'une méthode réellement plus fiable, la méthode SIFT.



Figure 5 : exemple de détection de points avec l'aide du détecteur de Harris

La méthode SIFT de Lowe, 2004 (*Scale Invariant Feature Transform*) [6] permet d'obtenir des points d'intérêt dont la détermination est très peu sensible à des changements, même importants, de facteurs d'échelle et d'orientation, et aussi assez peu sensibles aux variations locales de radiométrie (différences d'éclairage, différences de point de vue, etc.), toutes sortes de situations rencontrées très fréquemment en vision par ordinateur et où le détecteur de Harris est souvent tenu en échec. Sur des images de taille réduite (500 x 500 pixels), on peut généralement identifier plusieurs milliers de points d'intérêt, ce nombre dépend du réglage plus ou moins critique de nombreux paramètres, à toutes les étapes du calcul. Il est clair qu'avec un nombre aussi important de points trouvés, la mise en correspondance peut être aisément assortie de critères de qualité très stricts, et même ainsi, il reste couramment plusieurs centaines de points d'intérêt correctement extraits [7].



Figure 6 : exemple de détection et d'appariement avec l'aide de l'extracteur SIFT

En résumé, il ne pourrait plus y avoir de photogrammétrie sans traitement d'image, maintenant que tout y est numérique, et il reste encore beaucoup à faire pour épuiser tous les progrès possibles, et automatiser tout ce que fait un opérateur de saisie actuellement. Mais dès qu'il s'agit de recherche des plus grandes précisions possibles, les photogramètres forment probablement la communauté qui dispose de la plus importante compétence opérationnelle, même si elle n'est pas assez identifiée comme telle.

#### 4/ Conclusions

On peut donc se faire maintenant une idée un peu plus précise du paysage autour des deux domaines techniques évoqués ici.

La photogrammétrie s'est construite entièrement sur la recherche de la meilleure précision possible et, science ancienne, elle est surtout entre les mains d'acteurs industriels. Si elle est enseignée dans diverses universités étrangères, c'est comme une branche de la géomatique. Et en France, elle ne figure dans aucun cursus universitaire, depuis un demi-siècle que la chaire ouverte au CNAM, alors tenue par G. Poivilliers, est restée vacante : elle n'a été enseignée que dans les quatre grandes écoles de géomatique (ENSG, ESGT, ESTP et INSA-S), l'essentiel des développements étant menés à l'IGN, usager majeur et presque exclusif du domaine pendant plusieurs décennies. En outre, l'immense majorité des thèses soutenues dans le domaine l'ont été à l'université, donc à l'étranger, ce qui a progressivement effacé les traces pionnières de la photogrammétrie française de la scène mondiale. Par ailleurs, les développements menés chez les industriels, et les regroupements de ceux-ci en un nombre extrêmement réduit au plan mondial, ont conduit à des outils très performants, mais complètement fermés, et dont le détail est lourdement couvert par le secret industriel. C'est une science qui a atteint sa maturité, et qui évolue essentiellement par ses matériels et logiciels industrialisés.

A l'opposé, la vision par ordinateur est encore en pleine phase de croissance juvénile, s'attaquant sans états d'âme à des quantités de problèmes nouveaux. Les congrès internationaux couvrant ce domaine, même les plus sélectifs, sont plusieurs dizaines de fois plus nombreux que ceux consacrés à la photogrammétrie. Cette effervescence se traduit aussi au niveau des enseignements, et de très grandes quantités de cursus en universités et grandes écoles font une large place à ce domaine. En parallèle, un nombre considérable de thèses y sont menées, de sorte que la plupart des éléments d'actualité sont disponibles sur Internet et largement publiés. En outre, comme nous l'avons vu, certains résultats

commencent à entrer dans des applications réellement grand public, ce qui est un accélérateur de progrès considérable.

Alors, cela veut-il dire que la photogrammétrie est condamnée à ne plus être qu'une sous-branche secondaire de la vision par ordinateur ? Sous sa forme actuelle, peut-être, mais seulement si elle cessait de progresser, ce qui apparaît plutôt improbable. Tout de même il est clair que face aux recherches et progrès dont elle a besoin, elle n'a pas une puissance d'attraction comparable à celle dont jouit la communauté de vision par ordinateur. Elle va donc probablement progresser désormais comme la plupart des domaines de la topométrie, et par exemple comme le GPS : de façon opportuniste, en ré-employant de façon appropriée des résultats obtenus pour des clientèles bien plus puissantes. Ce sera peut-être moins gratifiant, mais c'est très efficace, et c'est une chance à ne pas manquer !

Nous terminerons en montrant, à cheval entre les deux domaines, l'application PhotoSynth de Microsoft et l'Université de Washington, qui permet d'assembler automatiquement et intelligemment des photos quelconques trouvées sur Internet, en fabriquant une 3D partielle mais rigoureuse... nous sommes dans le domaine grand public, mais les outils développés sont directement ré-employables en photogrammétrie. [8]



Figure 7 : Illustration tirée de PhotoSynth [9]

### Références

- [1] Kasser M, Egels Y., 2001. Photogrammétrie Numérique. Hermès-Sciences.
- [2] Longuet-Higgins, H., 1981. A Computer Algorithm for Reconstructing a Scene from Two Projections, *Nature*, 293(10):133-135.
- [3] Hartley R., Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN 0-521-62304-9.
- [4] Nister, D., 2004. An Efficient Solution to the Five-Point Relative Pose Problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756-770.
- [5] Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, Manchester, UK, pp. 147-151
- [6] Lowe, D. G. 2004, 'Distinctive image features from scale-invariant keypoints.', *Int. Journal of Computer Vision* 60(2), 91-110.
- [7] M. Kalantari, F. Jung. Estimation automatique de l'orientation relative en imagerie terrestre. *Revue XYZ*, n°114.
- [8] Noah Snavely, Steven M. Seitz, Richard Szeliski. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision*, 2007.
- [9] <http://photosynth.net/>