Introduction à la photogrammétrie dans PhotoSurvey

Présentation

PhotoSurvey est une application permettant de mettre en œuvre le procédé de photogrammétrie, c'est-à-dire la reconstruction du modèle numérique tridimensionnel d'une scène réelle statique, sur la base d'une acquisition photographique de cette scène.

De manière très générale, une scène correspond à un volume d'espace limité quelconque contenant une ou plusieurs entités matérielles. Néanmoins, dans le cadre de PhotoSurvey, pour des raisons de précision, nous nous limitons pour l'instant à la reconstruction de tranchées ou de fouilles de taille réduite (une vingtaine de mètres maximum), quitte à découper la zone en plusieurs scènes respectant ces limites.

La photogrammétrie est une alternative crédible aux autres systèmes de capture 3D, en premier lieu le relevé par scanner laser (lasergrammétrie), qui nécessite un équipement bien plus onéreux. En revanche, la photogrammétrie nécessite plus d'expérience d'utilisation, car pour produire de bons résultats, il est nécessaire de comprendre les principes fondamentaux de fonctionnement de cette technique et de réussir l'acquisition des clichés, qui sont déterminants pour la qualité du résultat obtenu.

Le logiciel PhotoSurvey s'efforce de masquer au maximum la complexité technique du procédé à l'opérateur, mais il ne peut s'y substituer complètement, et notamment pour ce qui concerne la phase cruciale d'acquisition des vidéos ou photos.

Ce document aspire à apporter à l'utilisateur de PhotoSurvey les grands principes qui sous-tendent la génération d'un nuage de points. Il présentera dans un premier temps les étapes fondamentales d'une production photogrammétrique, puis détaillera un ensemble de recommandations qui augmenteront les chances de réussite d'une reconstruction précise.

La photogrammétrie dans PhotoSurvey

Pour disposer du modèle tridimensionnel d'une tranchée et procéder à son récolement « virtuel », il vous faudra enchaîner l'ensemble des étapes suivantes :

1. Acquisition de vidéos ou de photos pour la tranchée d'intérêt,
2. Dans le cas d'une acquisition vidéo, extraction de photos montrant la scène sous un grand nombre de points de vue,
3. Reconstruction d'une « structure de scène », c'est-à-dire d'un nuage de points clés reconnus sur plusieurs images, qui permettent grâce à des calculs de géométrie épipolaire de retrouver la position et l'orientation des prises de vue dans l'espace, et potentiellement même les paramètres de distorsion de la lentille s'ils ne sont pas connus initialement ; cette étape est appelée *SfM*, issu directement de la terminologie anglo-saxonne *Structure from Motion*, qui fait référence dans le domaine.
4. Reconstruction d'une scène, par la corrélation stéréoscopique de l'ensemble des photos dont on connaît désormais toutes les caractéristiques. Cette étape est dénommée *MVS*, qui signifie *Multi-View Stereo*. Il en ressort un nuage de points, d'une densité plus ou moins forte selon le temps que l'on octroie à ce traitement.
5. Géoréférencement du nuage de points, afin de le transformer d'un système de coordonnées arbitraire à un système de coordonnées euclidien (ou quasi-euclidien) connu.
6. Exploitation du nuage géoréférencé dans un applicatif 3D interfacé avec TopoCalc (Geo2Cloud ou CloudCompare) pour réaliser le relevé des éléments topographiques ciblés (conduites, structures, …)

Il est à noter que la photogrammétrie à proprement parler ne concerne que les étapes 3 et 4 de cette liste. Les autres étapes servent à fournir les données en entrée du traitement et à en exploiter les résultats. Afin de comprendre au mieux les directives qui seront énoncées pour ces étapes annexes, les étapes relatives à la photogrammétrie vont être à présent détaillées.

Considérons que l'on dispose de plusieurs photos montrant une scène selon un ensemble de points de vue « suffisamment » nombreux ; nous verrons ultérieurement quelle signification nous pouvons donner à ce « suffisamment ».

La phase permettant de reconstruire la structure de la scène peut être scindée en trois étapes déterminantes :

1. L'extraction de points clés dans les photos,
2. L'association de ces points clés entre les photos,
3. La reconstruction itérative de la scène.

Extraction des points clés

La première opération consiste donc à faire l'analyse de chaque photo en entrée afin d'identifier un ensemble de points clés caractéristiques qui seront susceptibles d'être reconnus dans les autres images. Il est évident que cette opération est d'autant plus fructueuse que l'élément photographié est texturé, c'est-à-dire qu'il ne montre pas une surface totalement uniforme comme cela peut être notamment le cas avec des objets peints, homogènes ou brillants (murs, vitres, tuyaux lisses, …). Cela peut dépendre également dans une certaine mesure de la distance à la surface : un mur crépi peut apparaître comme texturé à faible distance et non texturé de loin.

Figure 1 - Exemple d'une scène très texturée

Figure 2 - Exemple d'une scène très peu texturée

Par ailleurs, on peut aisément comprendre que la netteté des images est déterminante dans la capacité à extraire des points caractéristiques.

PhotoSurvey s'appuie sur un des algorithmes les plus robustes et reconnus pour réaliser cette extraction, appelée *SIFT* (*Scale-Invariant Feature Transform*). Il présente en effet l'avantage de produire un descripteur du contenu visuel local d'une image de la façon la plus indépendante possible du zoom (et donc de la résolution) de l'image, de l'angle d'observation et de la luminosité, facilitant ainsi la reconnaissance d'éléments identiques entre deux photographies prises avec des points de vue différents, en angle, en distance et en exposition !

Ci-dessous est représenté le résultat d'une extraction, réalisée sur une photographie de tranchée ; les petits cercles rouges identifient les zones de l'image où il a été possible d'extraire un descripteur pouvant servir à une association ultérieure.

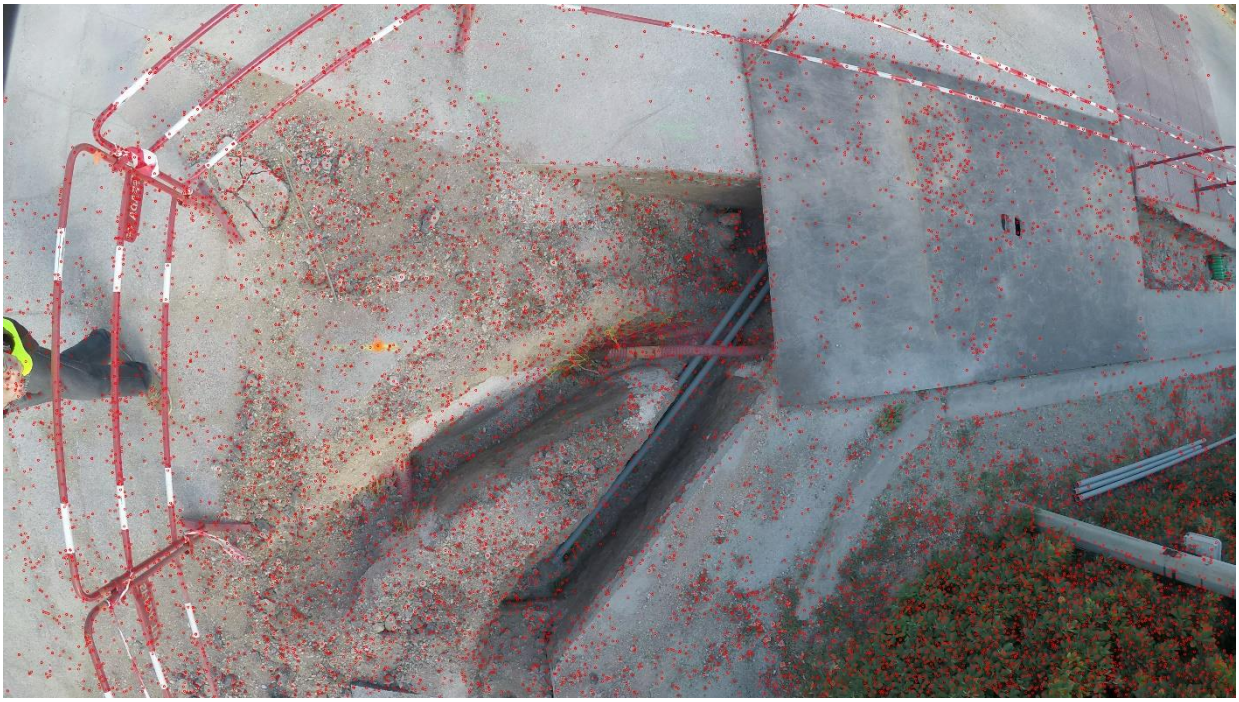


Figure 3 - Points clés SIFT extraits d'une photo

En zoomant sur le point de marquage :

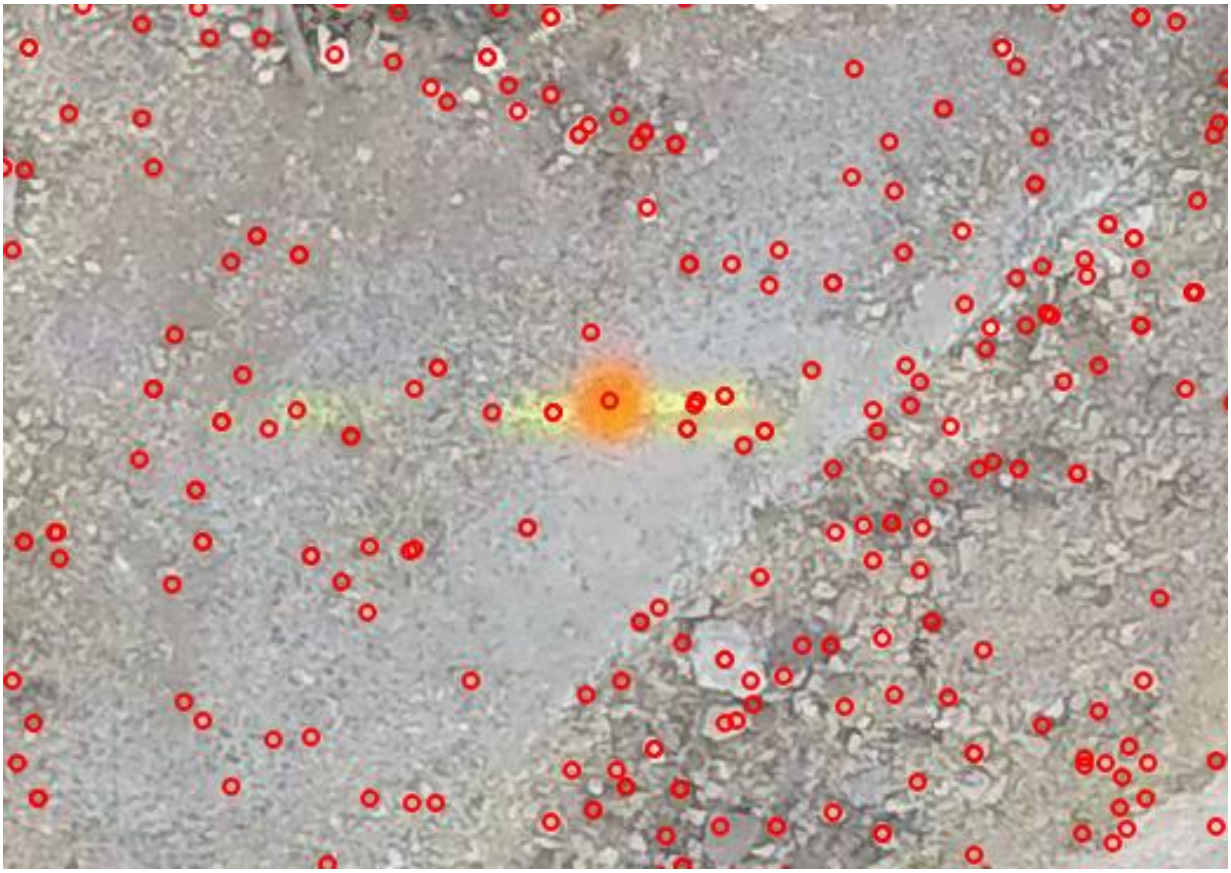


Figure 4 - Agrandissement de la zone du point de marquage

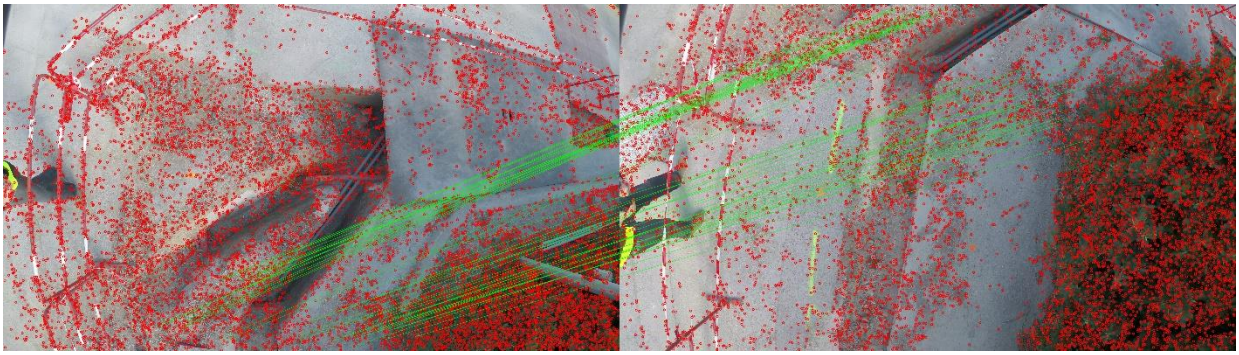
Associations des points clés

La seconde opération consiste à confronter l'ensemble des points clés trouvés d'image à image, afin de trouver des associations matérialisées par la similarité de leurs descripteurs.

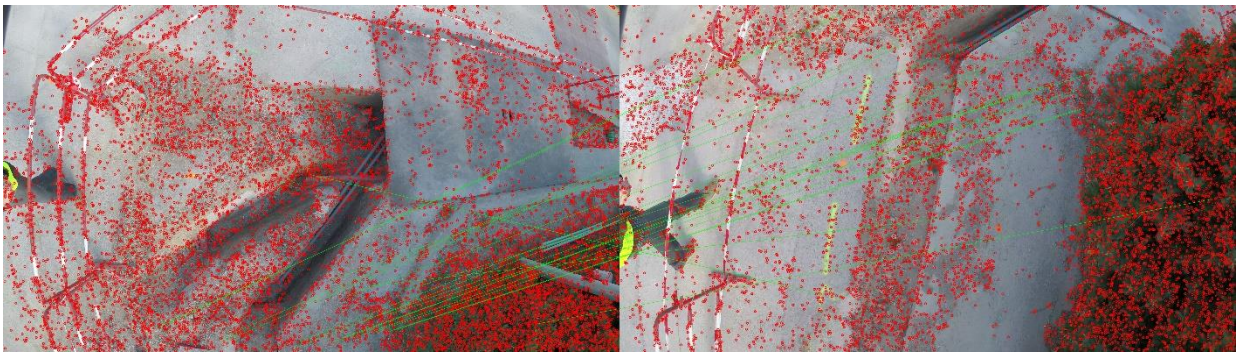
Le traitement peut donc par exemple reconnaître le point de marquage précédent sur plusieurs images, et entouré dans les trois choisies ci-dessous :

Figure 5 - Point clé avec descripteur commun à plusieurs images

On tente classiquement une représentation de ces associations, en reliant les points des images mis côte à côte, mais cette vue n'est intéressante que lorsque le nombre d'associations trouvées est faible, pour conserver une certaine lisibilité :



Cette détection est suivie d'un calcul géométrique de vraisemblance, qui s'assure que les correspondances établies sont bien crédibles au regard des lois de la géométrie épipolaire. Les associations en défaut, qui sont normalement en nombre faible, sont automatiquement rejetées. On n'est en effet pas à l'abri que deux éléments tout à fait distincts aient une apparence très proche dans deux photographies et puissent être confondus. Par exemple, en agrandissant la paire suivante,



vous pouvez voir en zoomant un trait d'association croiser l'ensemble des autres, ce qui n'a pas de sens d'un point de vue géométrique. La qualité d'un moteur de SfM est directement liée à la robustesse de l'ensemble de ses algorithmes pour résister à ce genre de confusions.

A la sortie de cette étape, pour chaque paire d'images, le système est capable de lister un ensemble de points clés qu'elles ont en commun, si bien que de proche en proche un même élément spatial peut-être associé à deux, trois images, voire bien plus.

Reconstruction itérative de la structure de scène

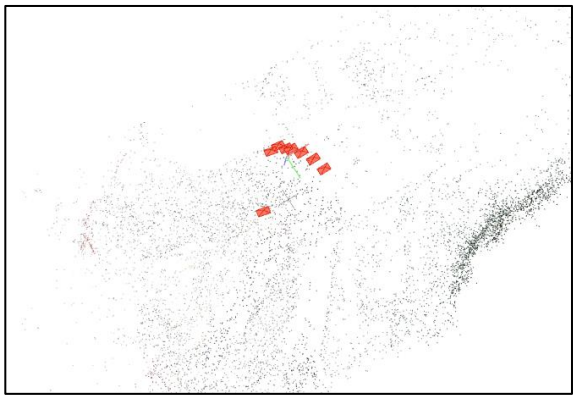
La dernière étape pour l'élaboration de la structure de scène va exploiter l'ensemble des correspondances trouvées pour reconstruire progressivement la scène.

Un couple d'images est initialement choisi de sorte qu'il possède un grand nombre de correspondances et puisse constituer un ensemble de points tridimensionnels de nombre et de qualité suffisants pour venir y raccrocher ensuite d'autres images et de nouveaux points issus de l'analyse des nouvelles correspondances incorporées.

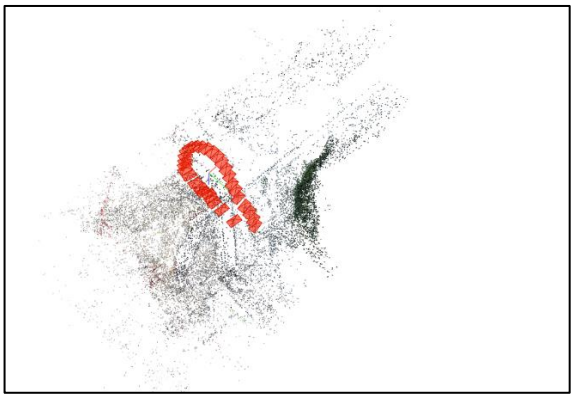
Cette étape, très technique, met notamment en jeu plusieurs algorithmes :

- d'estimation de pose, capable d'approximer la position et l'orientation des prises de vue sur la base des correspondances de points clés,
- de triangulation, capable de produire un point dans l'espace à partir de rayons issus d'une paire d'images positionnées et orientées dans l'espace,
- de consolidation de structure, réajustant régulièrement les paramètres à calculer afin de minimiser l'erreur globale calculée par la reprojection de l'ensemble des points 3D générés dans les poses estimées avec les paramètres de caméra estimés.

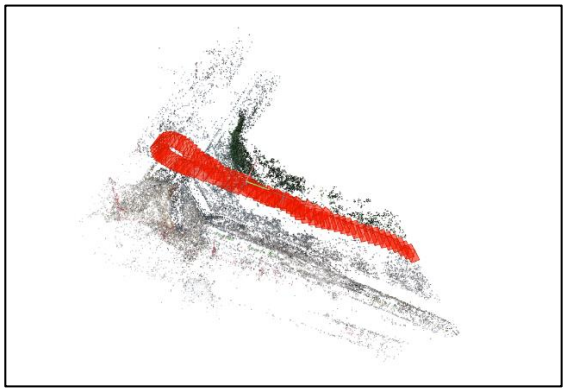
Ci-dessous quelques étapes de la reconstruction :



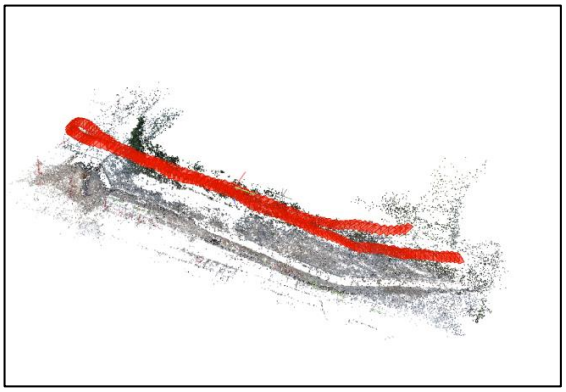
8 images



32 images



77 images



163 images

Il en résulte finalement un nuage de points 3D épars directement issus des correspondances des points clés initiales, et un ensemble de poses matérialisant le lieu et l'orientation des prises de vues fournies en entrée de traitement :

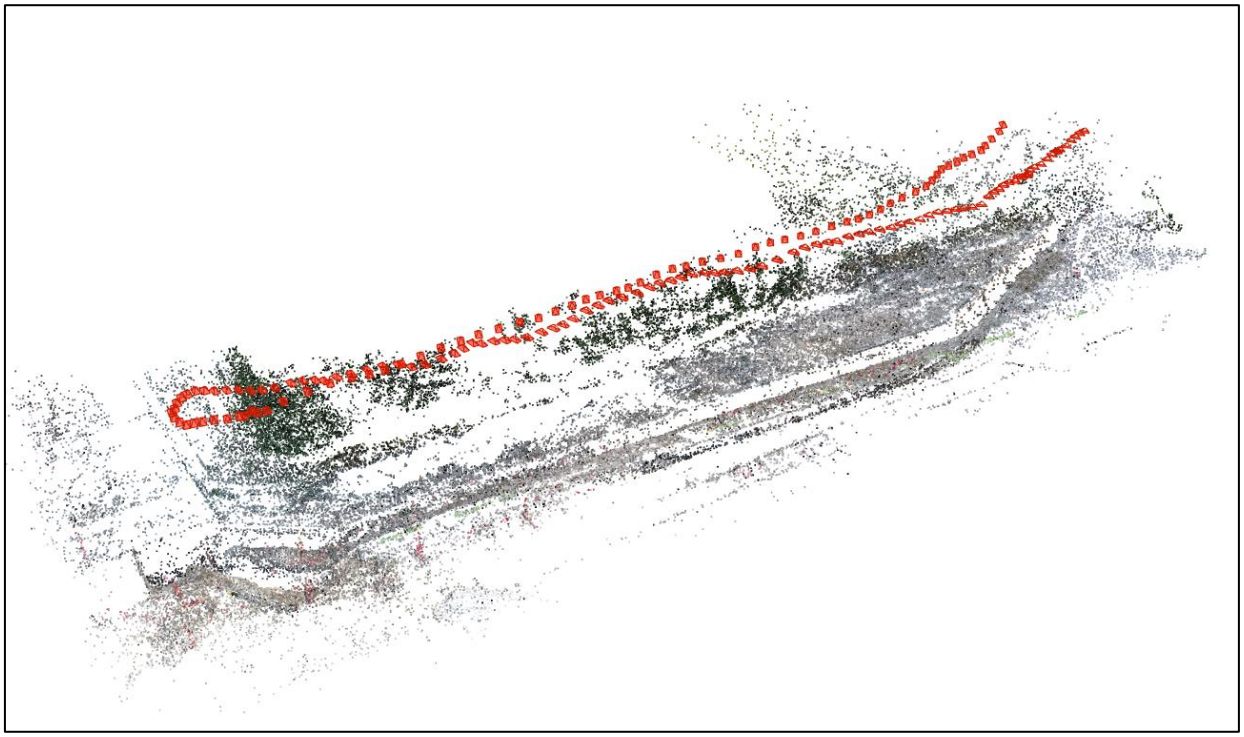


Figure 6 - Reconstruction de la structure d'une scène de tranchée capturée par 163 poses

Dès lors, il est encore plus aisé de visualiser les correspondances ayant fait émerger un point 3D (cet ensemble est appelé la piste du point 3D), par exemple pour le point de marquage visualisé précédemment :

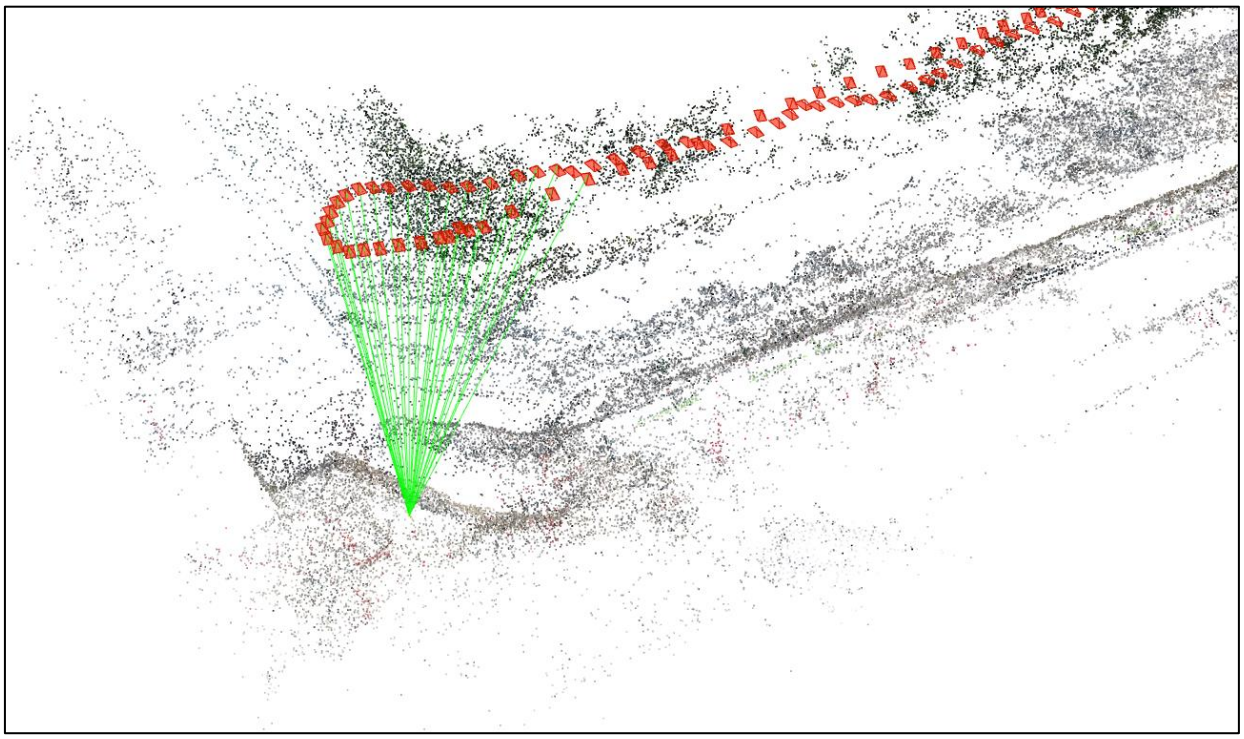


Figure 7 - Piste des points 2D ayant produit un point 3D en SfM

Et on peut de manière duale visualiser l'ensemble des points 3D auxquels a contribué une photo, ainsi que l'ensemble des autres images avec lesquelles elle a « collaboré » pour les générer :

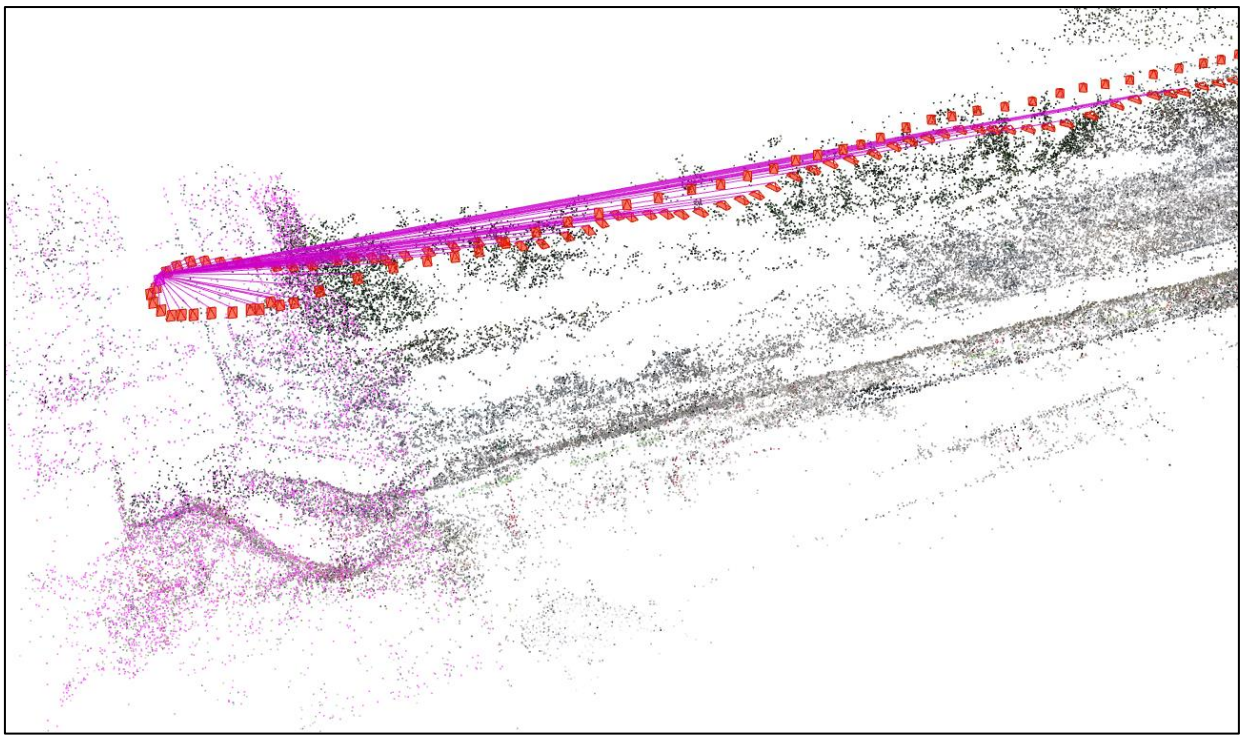


Figure 8 – Ensemble des points 3D associés à une vue en entrée

Reconstruction de la scène (MVS)

Dès lors, il est possible de réinitier des calculs d'une feuille blanche, mais avec désormais la connaissance de l'ensemble des poses (positions et orientations des clichés) et des paramètres de caméra précédemment calculés. Une procédure de calcul stéréoscopique va permettre de générer un nuage contenant beaucoup plus de points. Cette phase peut être à nouveau scindée en trois étapes fondamentales :

1. Le redressement des photos d'origine,
2. Le calcul stéréoscopique pour chaque image des matrices de profondeur et de vecteurs normaux,
3. La fusion des matrices obtenues permettant la génération des points 3D.

Redressement des photos d'origine

Outre la position et l'orientation des poses de vue, l'étape de SfM a permis d'élaborer un modèle de distorsion pour la lentille de l'appareil utilisé pour obtenir les vues de la scène. Traditionnellement, on ne prête guère attention aux déformations qu'une lentille a pu introduire dans une photographie, car elles sont le plus souvent minimes. Néanmoins, pour faire de la mesure, il devient indispensable d'en tenir compte et d'autant plus avec les lentilles grand angle qui provoquent le fameux

effet « fisheye » de distorsion radiale, où les déformations deviennent en revanche flagrantes. Pour réaliser la confrontation stéréoscopique des pixels des images et basculer totalement dans le domaine de géométrie épipolaire sous-jacent, il est nécessaire de redresser les images afin de retrouver un modèle où les rayons passant par les pixels ne sont plus déformés.

On peut visualiser ci-dessous un exemple du résultat de cette étape, sur une photo extraite du film d'une GOPro :

Figure 9 - Photo initiale contenant des distorsions de type fisheye

Figure 10 - Photo redressée après traitement

On peut constater que la résolution de l'image est modifiée lors de cette transformation. En effet, le redressement produit théoriquement une image qui n'est plus rectangulaire et on ne retient que la partie qui peut être maintenue dans un rectangle, ce qui a tronqué la photo d'origine dans le sens de la hauteur.

Calcul des matrices de profondeur et de vecteur normaux

La seconde étape, qui est la plus coûteuse en temps de calcul de toutes les étapes de la reconstruction, consiste à confronter les photos ayant un certain recouvrement de vue afin d'estimer pour chacun de leur pixel la profondeur à laquelle se situe la surface visualisée par ce pixel, ainsi que son vecteur normal, qui pourra être utilisé pour l'illumination ultérieure du nuage de points ou son éventuelle transformation en surface maillée.

Deux techniques sont employées pour retrouver aussi précisément que possible ces informations, la première photométrique, la second géométrique. Les résultats de cette étape intermédiaire de calcul sont théoriquement visualisables à l'aide de dégradés de couleur, comme l'illustrent les images suivantes :

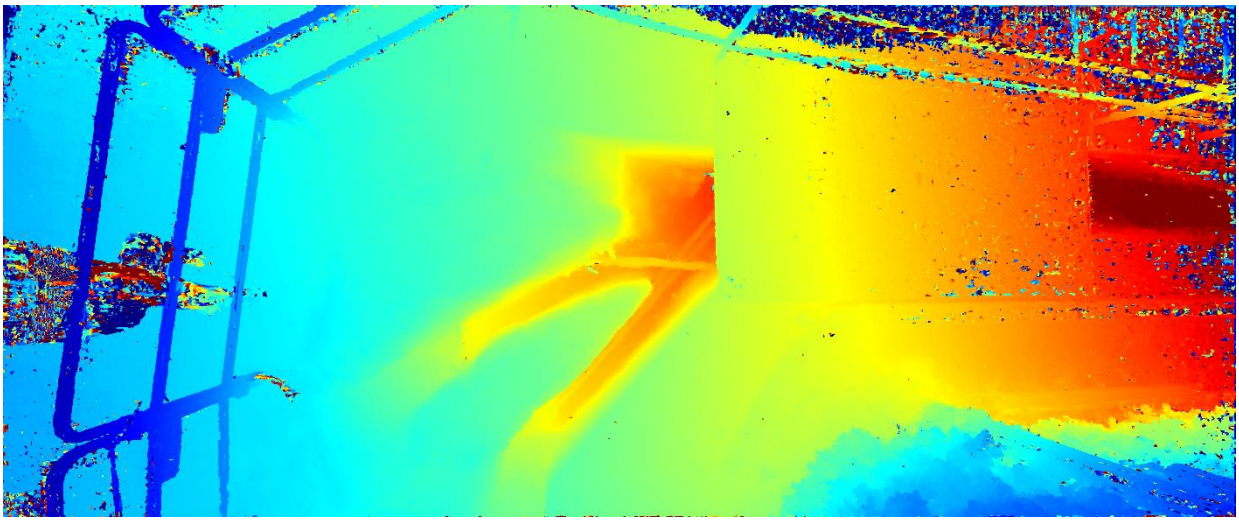


Figure 11 - Représentation colorisée d'une matrice de profondeur photométrique

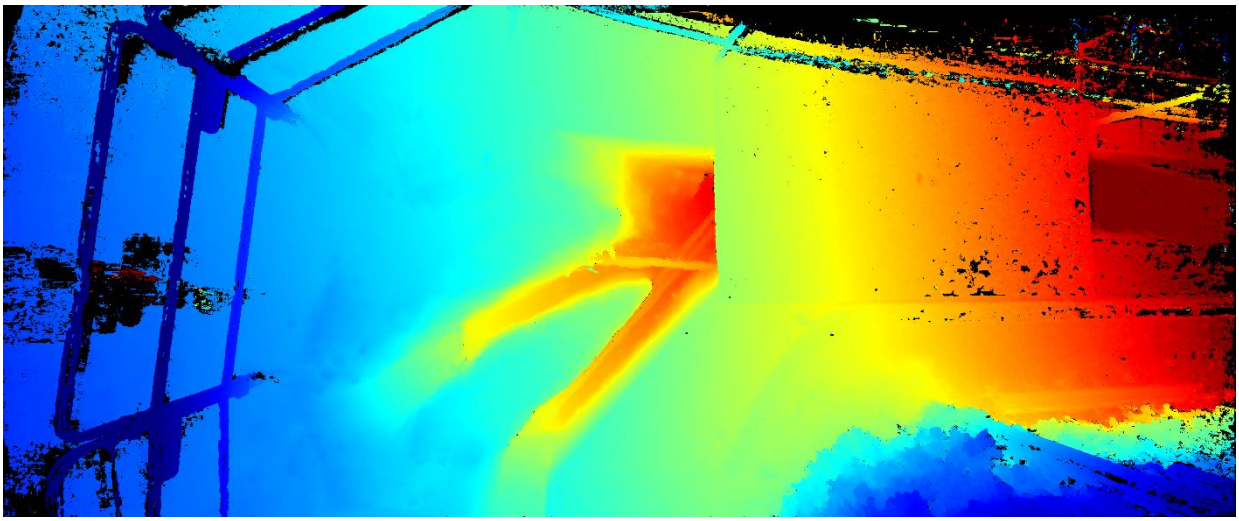


Figure 12 - Représentation colorisée d'une matrice de profondeur géométrique

Fusion des matrices de profondeur

L'ultime étape de la phase photogrammétrique s'attache à exploiter les informations de profondeur précédemment calculées pour chaque pixel de chaque photo ayant contribué à la reconstruction, afin de produire un point 3D lorsque suffisamment de pixels se retrouvent colocalisés dans l'espace. En fonction de la résolution d'image choisie pour le calcul, et de la quantité de photos en entrée, on obtient un nuage de points d'une densité plus ou moins forte, l'idée étant d'ajuster cette densité en fonction du rapport temps de calcul / granularité de nuage compte tenu de l'exploitation que l'on veut en faire ensuite, et de la précision requise.

Ci-dessous le résultat de la reconstruction d'une scène de blocs de pierre prise par 10 prises de vue, avec les trois niveaux de densité de points proposées par *PhotoSurvey*, et les temps de calcul correspondant sur une GeForce RTX2080 :

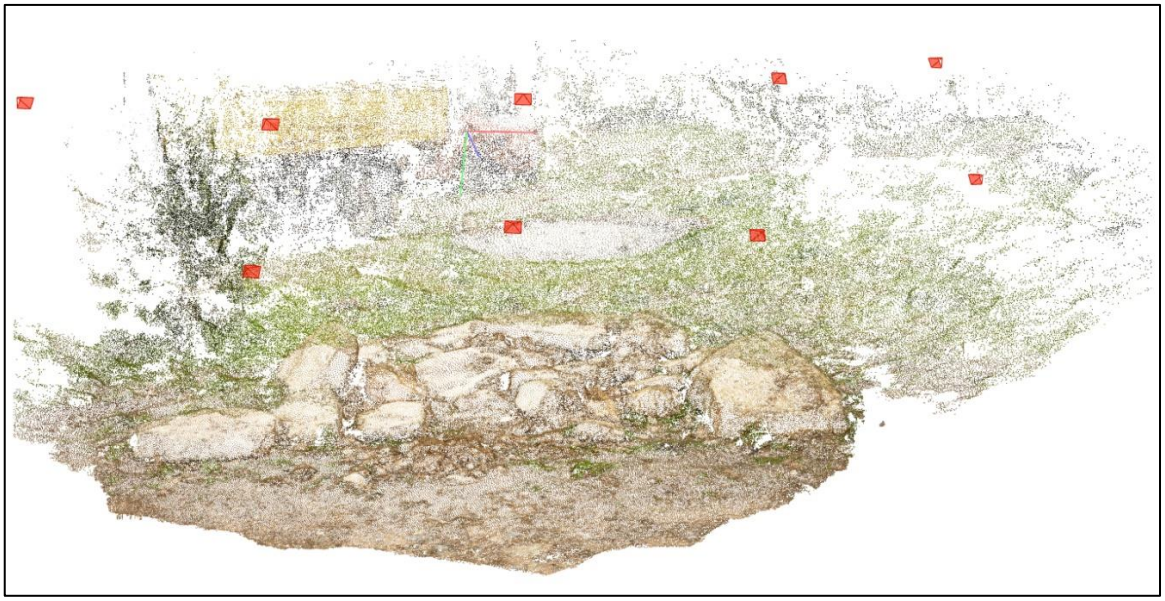


Figure 13 - Scène peu dense (281664 points calculés en 1min13)

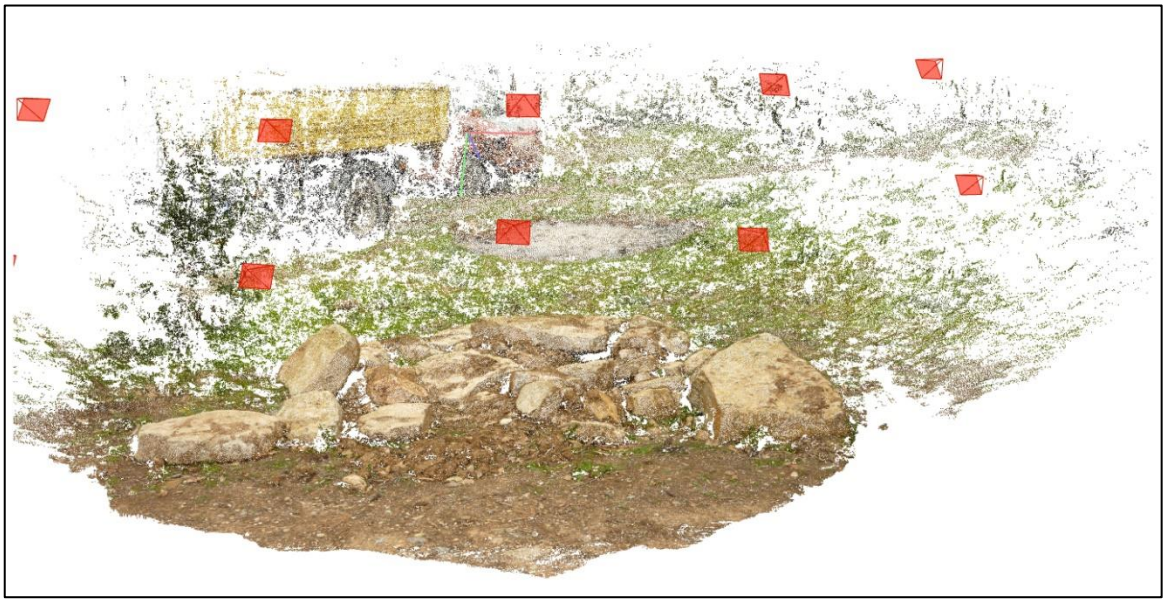


Figure 14 - Scène moyennement dense (575623 points calculés en 4m40)

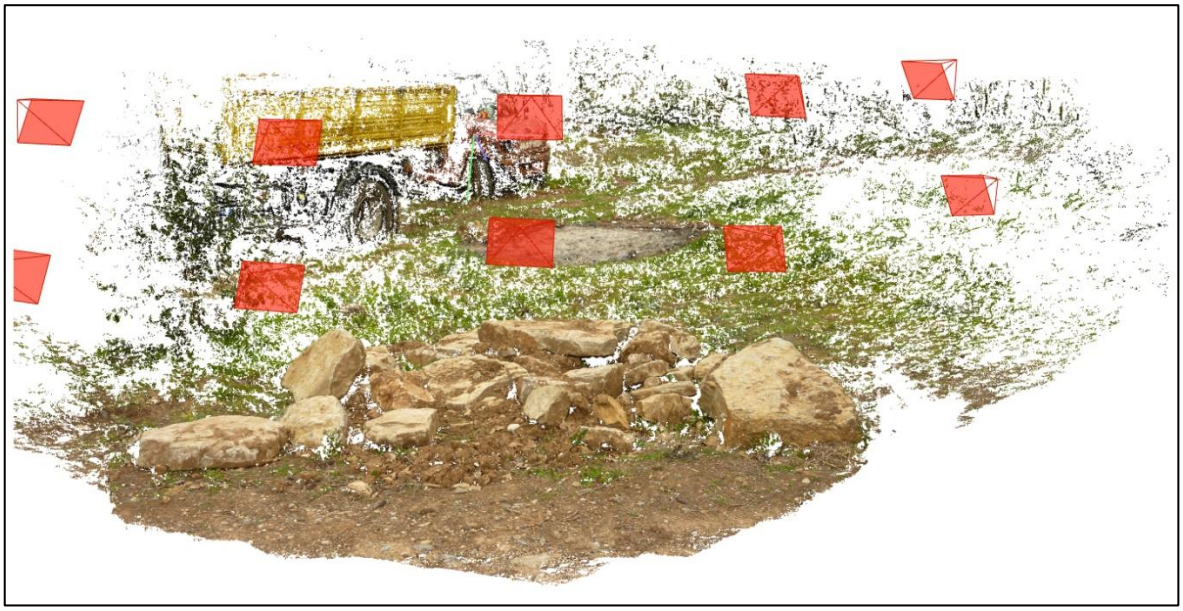


Figure 15 - Scène dense (1692652 points calculés en 10m34)

Exploiter le procédé de manière optimale

Maintenant que le principe de fonctionnement a été présenté, il est utile d'en déduire un certain nombre de recommandations générales qui permettront de tirer le meilleur parti du procédé, et de maximiser ses chances de succès, sa précision et son niveau de détail.

Phase de repérage

Avant l'acquisition, il est nécessaire d'identifier ou mettre en place des points de repère ponctuels dans la scène, qui devront être relevés de manière traditionnelle avec TopoCalc afin de géo-référencer ultérieurement le nuage de points généré pour la reconstruction. Trois points non alignés, disposés autant que possible en extrémité de scène sont théoriquement suffisants pour recaler le nuage, mais on a tout intérêt à en relever plus, afin d'une part de compenser d'éventuelles erreurs de mesure, mais également d'avoir une idée de la précision de la reconstruction obtenue.

Prochainement, ces points seront même utilisés lors des calculs de structure de scène et permettront de faire des reconstructions précises de grande étendue (plusieurs dizaines voire centaines de mètres). Il sera dès lors utile de disposer d'un ou deux points tous les 15 à 20m, qui serviront de contraintes lors de ces calculs de structure.

Phase d'acquisition

Pour qu'une mesure soit utile, il est crucial qu'elle soit interprétable avec précision, d'une part en faisant en sorte qu'elle comporte le moins de bruit possible, d'autre part en maîtrisant au mieux tous les paramètres de sa chaîne d'acquisition.

Un certain nombre de paramètres sont particulièrement importants pour réussir l'acquisition de la zone à modéliser, et franchir avec succès les étapes présentées précédemment. Trois critères sont primordiaux : le recouvrement, la netteté et la stabilité.

Le recouvrement

Il est évident que pour reconstruire précisément des points 3D, il est nécessaire d'avoir le plus possible de correspondances entre les points 2D les représentant. En théorie, il faudrait s'assurer que chaque zone de la scène à reconstruire ait été capturée par au moins trois clichés, mais il est fortement conseillé d'en avoir au

moins le double pour prendre en compte d'éventuelles difficultés d'extraction, le manque de parallaxe, les imprécisions sur le modèle de distorsion, etc. La capture vidéo présente l'avantage de pouvoir ajuster a posteriori ce nombre d'images, mais au détriment de leur netteté et de leurs déformations (telles que celles induites par les obturateurs roulants, évoquées ci-dessous) ; il est néanmoins inutile de prendre les vidéos avec une fréquence supérieure à 24 images par seconde.

La netteté

La netteté de l'ensemble du cliché est directement issue de ce que l'on appelle le piqué en photographie, c'est-à-dire la capacité d'un équipement à faire ressortir les détails, et cela en dehors même de la notion de résolution.

Ce critère ramène donc à la qualité d'une production photographique classique. Il faut à tout prix éviter les grandes erreurs amenant du flou ou une luminosité incorrecte : mouvement brusques, à-coups, chocs lors de l'acquisition, réglages inadaptés de la caméra lorsque l'on est en réglage manuel…

Les zones floues perturberont sensiblement le traitement, il vaut mieux disposer de moins de clichés mais parfaitement nets que d'en ajouter un maximum sans considération de leur netteté. Ainsi, la précision et l'association des points clés trouvés dans chaque image seront optimales.

Sur le plan de l'exposition, il faut évidemment éviter la lumière directe dans l'objectif (soleil de face). Il faut également éviter de mélanger des zones fortement sombres avec des zones fortement éclairées sur une même prise de vue (ce qui peut notamment arriver lorsque figure l'ombre projeté de gros bâtiments à proximité), le risque étant de sous-exposer ou surexposer considérablement les parties importantes de l'image. Le contexte idéal, (qu'on ne maîtrise malheureusement pas complètement), serait une prise de vue avec soleil au zénith masqué par un léger voile nuageux (ce qui diffuse la lumière, atténue les ombres projetées et limite les trop forts contrastes selon les directions de prise de vue).

Il peut du coup y avoir des cas extrêmes où les conditions de prise deviennent assez problématiques, par exemple une tranchée étroite et profonde avec un soleil puissant incliné et transverse à la tranchée. Le cœur de tranchée devient alors très sombre, ce qui encouragerait pour ajuster l'exposition à ne cibler que le strict contenu de la tranchée, mais qui n'est pas totalement possible puis qu'on a besoin de points de repère extérieurs pour le géoréférencement. Ce sont des conditions qui génèreront plus de risque lors de la reconstruction.

La stabilité

Elle est garante de l'intégrité du modèle à reconstituer. L'acquisition se faisant naturellement sur plusieurs secondes voire minutes, il y a un risque de mouvement et donc de modification de la scène, ne serait-ce même qu'induit par les phénomènes naturels (le vent, la pluie, les ombres portées, …). L'opérateur devra s'efforcer de les réduire autant que possible, y compris en arrière-plan (circulation de véhicules, piétons, animaux, …) pour ne pas fragiliser la précision et la robustesse

des algorithmes, d'autant qu'avec la caméra et le mode d'acquisition recommandé, il est déjà très probable de voir la perche et l'opérateur dans les images.

Autres points importants

1. Permettre le calcul stéréoscopique

Ces trois premiers critères étant autant que possible respectés, il est ensuite nécessaire de « créer de la parallaxe » entre les prises de vue pour que les principes de calcul stéréoscopique puissent être mis en œuvre (d'où le nom de « structure from motion »). A priori, le relevé photographique d'une tranchée, par définition linéaire, induit assez naturellement cette parallaxe. Il s'agit en fait d'éviter la simple rotation du capteur sur lui-même lors de l'acquisition, et de s'efforcer autant que possible de ne pas colocaliser les prises de vue, ce qui empêche la triangulation. Le déplacement transversal du capteur le long de la surface d'intérêt, ou sa rotation autour d'une zone cible (par exemple un trou) sont à privilégier, en faisant éventuellement plusieurs passes à différentes incidences (voire distances), pour améliorer la précision. Un nombre plus élevé d'images fiabilisera et densifiera la reconstruction mais ralentira son calcul, l'opérateur devra donc trouver le meilleur compromis entre d'une part ses contraintes de délai et de disponibilité de sa station de travail, et d'autre part la précision et la qualité requises pour réaliser son récolement.

2. Réserver la reconstruction à des surfaces adaptées

Par nature, le procédé ne permet pas de relever précisément les surfaces d'aspect totalement uniforme, ou brillant, ou réfléchissant (vitrages). En effet dans ces cas-là, la nature des points rendus soit diffère d'une image à l'autre, soit ne permet pas le calcul stéréoscopique faute d'association précise de points entre les images.

3. Limiter le mouvement

Enfin, l'opérateur doit être conscient d'un effet néfaste provoqué par le procédé d'acquisition des appareil numériques les plus courants (dont la GoPro) appelé « rolling shutter », c'est-à-dire l'obturation déroulante. A la différence d'un appareil photographique traditionnel où le mécanisme d'obturation est quasi instantané, ce type d'appareil réalise l'obturation sous la forme d'un balayage vertical très rapide, ne masquant à un moment donné que la ligne de pixels en cours d'acquisition. Cela convient pour la plupart des usages, mais peut induire une déformation dans l'image, dont les lignes ne sont pas toutes prises strictement au même moment. Cette déformation est d'autant plus importante que l'appareil est en mouvement relatif rapide par rapport à la scène photographiée. C'est ainsi que pris d'un véhicule se déplaçant rapidement, des structures proches peuvent apparaître obliques :

Figure 16 - Effet du rolling-shutter sur une prise de vue depuis une voiture (source Wikipedia)

Et inversement la prise de vue d'un élément en rotation très rapide peut provoquer des représentations surprenantes :

Figure 17 - Effet du rolling-shutter sur une prise de vue d'un élément rotatif rapide (source Wikipedia)

Même si les vitesses de mouvement en jeu ne sont pas du même ordre que ces exemples, l'exigence de précision de la mesure imposerait théoriquement de ne prendre des clichés que statiquement, ce qui est relativement incompatible avec un relevé de la scène vidéo. Il est donc hautement recommandé de réaliser le relevé lentement, et de placer sa caméra en hauteur, à l'aide d'une perche afin de limiter la vitesse de défilement des pixels. L'ajout d'une passe en sens de déplacement inverse (et sans retourner l'appareil !), en modifiant éventuellement son incidence est théoriquement de nature à compenser les déformations initiales.