

La photogrammétrie appliquée au récolement des réseaux enterrés : retour d'expérience d'une méthode industrialisée

■ Jérôme LEROUX - Maxime CHAUVIN - Valentin POITEVIN

De nouvelles solutions existent aujourd'hui pour mesurer les réseaux enterrés. L'acquisition d'images avec un smartphone permet de reconstruire en 3D une fouille ouverte grâce aux algorithmes de photogrammétrie par corrélation dense. Les enjeux de fiabilité et d'industrialisation pour ce type de production de données nous ont poussés à améliorer une méthode d'acquisition préexistante. L'étude menée ici vise principalement à déterminer la meilleure technique de captation d'images. Les éléments clés de la réussite de la méthode reposent sur 1. L'utilisation de marqueurs dimensionnels à travers la pose des balises et 2. La détection de ces marqueurs dans les images via des algorithmes basés sur de l'Intelligence Artificielle. Les données 3D produites et accessibles via un portail web permettent de dessiner les infrastructures existantes en réduisant le plus possible l'intervention humaine.

Contexte de l'étude

La photogrammétrie par corrélation dense permet de produire de la donnée numérique 3D à partir de n'importe quel vecteur (avion, drone, homme) équipé d'appareil photo numérique (APN). Cette méthode de relevés permet de répondre aux problématiques actuelles du relevé numérique de l'existant (fouilles de réseaux enterrés, mobiliers urbains...). L'essor de cette technologie est dû en partie aux capacités actuelles des ordinateurs à traiter les images par "computer vision" en exploitant la puissance de calcul graphique, ainsi qu'à la miniaturisation des APN sur smartphone qui rend la captation de photos rapide, à portée de main et dont les capacités optiques permettent d'en extraire des informations 3D (Yilmazturk and Gurbak, 2019).

De nombreuses applications se développent aujourd'hui pour la production de plans de récolement géoréférencés. Généralement mesurées indirectement par détection et géolocalisation des réseaux enterrés, de nouvelles solutions permettent de relever les réseaux

enterrés par photogrammétrie ou LiDAR directement à l'aide d'un smartphone (e.g. iPhone 12 Pro) lorsque la fouille est encore ouverte. Une méthode de mesure basée sur un calcul de triangulation d'image par alignement de faisceaux a été développée il y a dix ans (RAPH® : Récolement Automatisé par PHotogrammétrie). Cette méthode se basait sur la reconnaissance automatique de points de calage (environ 50 points par image) grâce à la disposition de balises identifiées sur la scène à reconstruire (figure 1, A). Avec plus de 60 000 calculs réalisés en neuf ans, la méthode a prouvé son efficacité, mais plusieurs contraintes opérationnelles rendent la tâche chronophage et coûteuse. Le principal problème étant l'utilisation excessive des balises au sol (figure 1, A). Face à ce constat, une étude a été réalisée pour améliorer le service sur tous ces aspects au regard des nouveaux algorithmes de photogrammétrie par corrélation dense. Les objectifs sont multiples : 1. Simplifier la captation photos pour l'opérateur. 2. Réduire le nombre de balises sur

MOTS-CLÉS

Photogrammétrie, nuage de points, smartphone, Intelligence Artificielle, fouilles ouvertes

le terrain. 3. Augmenter la distance linéaire de captation. 4. Fournir des données plus exhaustives : production de nuage de points, de modèle maillé texturé ou encore d'orthophotographie.

Méthodologie générale

Le service développé avait pour résultat final la digitalisation géoréférencée de la position du câble enterré au fond de la fouille. Ce résultat était obtenu en suivant deux grandes étapes : 1. La captation des images par le technicien de terrain responsable du déploiement du nouveau câble. 2. Le géoréférencement de la scène et la digitalisation du câble par le géomètre. Entre ces deux étapes, le calcul photogramétrique devait être totalement automatisé via un serveur de calcul déporté. L'automatisation des calculs de triangulation des images passait par la reconnaissance automatique de points de calage sur les prises de vues convergentes. Un système de balises a donc été développé pour automatiser la reconnaissance de ces points (figure 1, A) et le calcul d'alignement des faisceaux.

Lors de l'étape 1, le technicien de terrain disposait vingt balises en forme de tétraèdres autour de la scène et marquait plusieurs points remarquables sur les images qui servaient de points de géoréférencement (figure 1). À l'aide d'un APN, l'opérateur prenait ensuite vingt photos de l'ensemble de la scène en veillant à prendre à la fois les balises, les marquages au sol et le fond de la fouille où se trouve le câble. Les photos étaient envoyées pour calcul sur un serveur disponible 24h/24 et le résultat était téléchargeable par le géomètre. Celui-ci était chargé du géoréférencement des



Figure 1. A. Exemple de fouille en 2012 : vingt balises disposées autour d'une scène de fouille ouverte. B. Exemple de fouille avec la nouvelle méthode : seulement trois balises entourant la fouille.

points remarquables quelques jours après le rebouchage de la fouille. L'outil mis à disposition pour cette étape 2 nécessitait le pointage de chaque point de géoréférencement et de la position du câble au minimum sur trois photos différentes, ce qui pouvait impliquer un temps de traitement au bureau important et parfois complexe.

Le succès de cette méthodologie reposait sur la pose des vingt balises entourant la fouille, ceci pour avoir assez de points homologues "artificiels" pour résoudre le calcul de triangulation. Aujourd'hui, les algorithmes SIFT [Lowe, 2004] permettent la reconnaissance de points homologues grâce aux gradients locaux sur

les images ce qui rend le calcul plus robuste tout en s'affranchissant de la connaissance *a priori* de la calibration optique de l'APN. Devant toutes ces contraintes terrain et de post-traitement, un nouveau protocole a été mis à l'étude pour réduire le temps d'opération pour tous les intervenants et produire des données plus exhaustives.

Nouvelles approches d'acquisition par photogrammétrie : Photo ou vidéo ?

Les algorithmes de photogrammétrie par corrélation dense permettent aujourd'hui de traiter en masse

des images venant de tout type de capteurs, notamment venant de smartphone (Yilmazturk and Gurbak, 2019). Ces calculs permettent d'utiliser différentes méthodes de captation par images ou même par vidéo (avec extraction d'images). La captation vidéo est intéressante pour l'opérateur qui n'a plus à se soucier de la prise de clichés. Il est demandé à l'opérateur de maintenir l'orientation des prises de vues avec un certain angle tout en longeant les éléments à reconstruire en 3D. Si la captation vidéo simplifie la tâche de l'opérateur, elle complexifie fortement le travail de post-traitement des données pour plusieurs raisons que nous avons identifiées : 1. Ratio résolution/poids de fichier défavorable pour la vidéo comparée à un set de photos échantillonné au même intervalle. Exemple avec un smartphone Note 9 : 60 images pleine résolution (4032 x 3024 pix) : 111 Mo ; une minute de vidéo Full HD 60 fps (1920 x 1080 pix) : 201 Mo ; une minute de vidéo 4K 60 fps (3840 x 2106 pix) : 519 Mo. 2. Un enregistrement de la vidéo à 25-30 fps entraîne un problème de flou de bougé par rapport au mouvement de l'opérateur et un problème dû au rolling shutter qui entraîne des déformations très importantes sur l'image. L'effet est réduit lors de la prise d'images avec une vitesse d'obturation supérieure à 1/200 de seconde. 3. L'encodage des images détruit la texture locale des images et impacte la recherche de point de corrélation entre images lors de la corrélation, comparé à une compression JPEG beaucoup plus avancée.

Pour ces raisons, la captation d'images par photographie a été privilégiée pour obtenir le meilleur compromis entre niveau de détails nécessaire à la détection de points homologues et poids de fichier léger.

Techniques de captation d'images en photogrammétrie

Les méthodologies de captation peuvent être aussi très diverses. Pour ce type de relevé, il est souvent conseillé d'utiliser des perches télescopiques pour prendre de la hauteur et augmenter le recouvrement entre les images. Cette technique

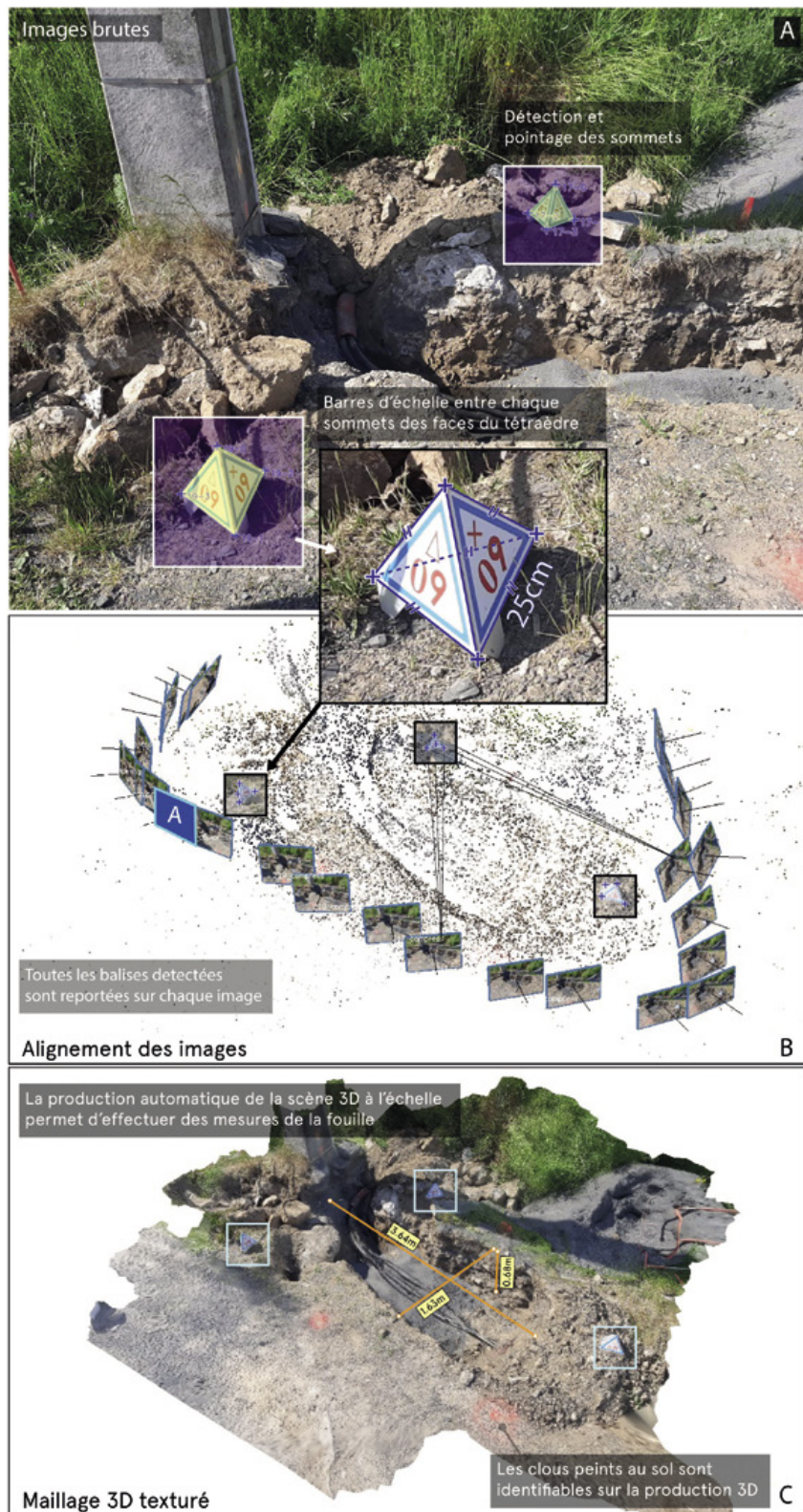


Figure 2. A. Exemple d'une image brute parmi le set d'images (25 photos) acquises autour de la fouille représentée en B. Les rectangles sur fond violet sont des aperçus de la détection des balises réalisée sur cette image. B. Orientation et position des prises de vues autour de la fouille. C. Maillage 3D texturé produit à partir du set d'images.

Pour que le service soit le plus simple possible, tout en proposant une reconstruction de l'environnement de la fouille, nous avons choisi une méthode de prise de vues simplifiée, à hauteur des épaules avec une inclinaison vers le bas d'environ 30° avec l'orientation de la prise de vue en mode paysage. Ce protocole d'acquisition permet de voir à la fois les éléments au sol et des éléments en périphérie jusqu'à une dizaine de mètres (poteau, muret, coffret, etc.). Cette méthode de prise de vues permet aussi de capturer les balises pour la mise à l'échelle des productions photogrammétriques (figure 2).

Pour compenser la réduction du nombre de balises, nous avons choisi d'augmenter le recouvrement entre les images. Ceci pour obtenir un maximum de corrélation entre les images, assurer une captation photogrammétrique exploitable et maintenir l'automatisation des calculs. L'acquisition des images se fait donc en mode "rafale" avec une prise d'image toutes les secondes automatisée par l'application sur le smartphone de l'opérateur. Cette méthode permet d'avoir un recouvrement conséquent entre prises de vues en respectant une vitesse de captation aux alentours de 1 m/s (vitesse moyenne de marche) (figure 2, B).

L'acquisition est réalisée grâce à une application mobile. Les caractéristiques des projets ainsi que les photos sont envoyées pour analyse dès que l'opérateur finalise la capture des images. Le serveur de calcul qui analyse chaque projet délivre une réponse de succès ou du non-succès de la captation dans les minutes qui suivent l'envoi.

Traitement des images

Les traitements sont faits grâce à l'API python développée par Agisoft Metashape v.1.7.1 (<http://www.agisoft.com>), logiciel

est complexe, car il est difficile de bien cadrer les prises de vues tout en maintenant un bon niveau de sécurité aux abords de la fouille ouverte. Une autre stratégie consiste à prendre seulement les prises de vues à quelques décimètres du câble avec comme cadre de photo les bords de la fouille. Ces deux méthodes nécessitent un travail manuel de pose de points de contrôle sur les images lors du traitement par photogrammé-

trie. Ces points de contrôle permettent de géoréférencer le projet et de corriger l'effet "banane" [James et al., 2020] qui se manifeste sur ce type de captation d'images en linéaire trop proche des surfaces. Cette étape de contrôle ne va pas dans le sens d'une automatisation totale des traitements photogrammétriques. De plus, l'emprise des images étant faible, il est difficile de replacer la fouille enterrée dans son environnement.

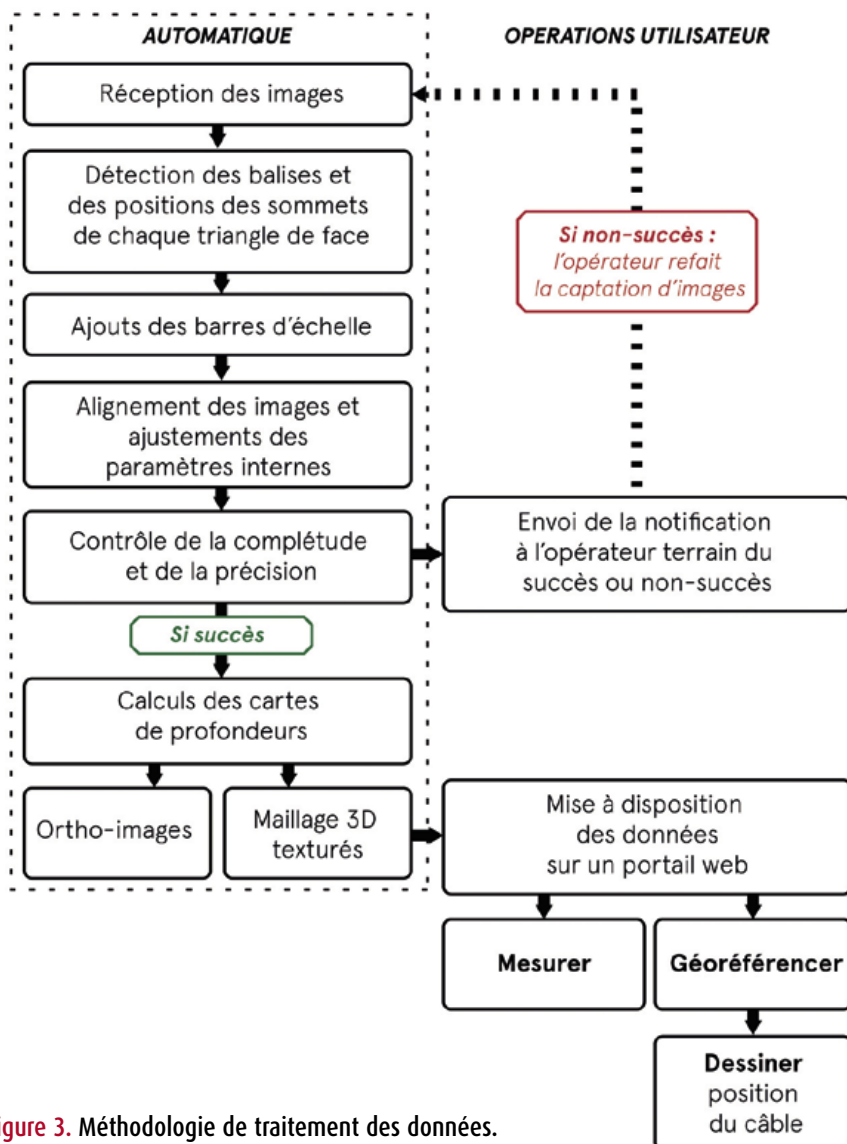


Figure 3. Méthodologie de traitement des données.

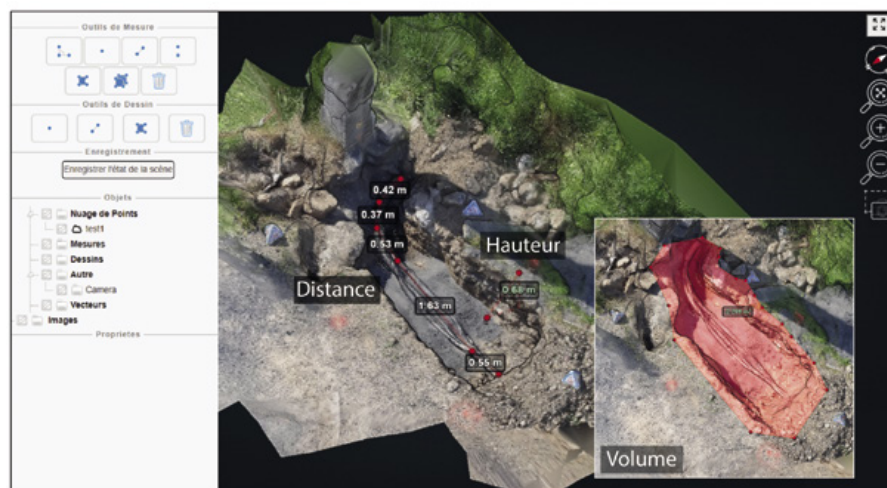


Figure 4. Interface en ligne d'exploitation des données 3D. Plusieurs outils permettent de traiter les données dont des outils de mesure sur le nuage de points qui est à l'échelle sans géoréférencement.

reconnu par la communauté scientifique pour sa qualité de reconstruction 3D par photogrammétrie [Hinge et al., 2019 ; James et al., 2020]. L'API Python permet d'établir des scripts de trai-

tements automatiques et complexes de scènes photogrammétriques. La méthodologie de traitement est résumée sur la figure 3. Après la réception des images, la première étape est la

détection des points des sommets des tétraèdres. Ces points positionnés automatiquement sur les images produisent deux effets : 1. Ajout de points de liaison entre images et consolidation de la triangulation. 2. Mise à l'échelle et optimisation des paramètres internes de la caméra grâce aux dimensions connues des arêtes des tétraèdres. L'étape suivante d'alignement des images est donc facilitée. Cette étape permet de rechercher les points homologues et les points de liaison par autocorrélation entre les images. Après la triangulation des images et un contrôle de la précision réalisé en moins de dix minutes, le script envoie un message à l'opérateur pour le notifier du succès ou du non-succès de la reconstruction des données en 3D. L'opérateur peut ainsi refaire la captation des images avant le recouvrement de la fouille.

Après calcul des cartes de profondeurs, différents produits de la photogrammétrie sont calculés dont le maillage 3D texturé et une orthophotographie. Toutes ces données sont mises à disposition sur un portail web qui permet de réaliser des mesures directement sur le nuage de points (échantillonné du maillage 3D texturé) (figure 4). Le fichier de nuage de points produit est mis à disposition sur une interface 3D interactive basée sur la technologie Potree. Cette interface customisée met à disposition trois modules spécifiques à l'exploitation des données : 1. Un module de mesures pour une analyse dimensionnelle du projet de fouille ouverte. 2. Un module de géoréférencement permettant l'import de fichier de points terrain. 3. Un module de dessin sur le nuage de points pour l'export de forme vectorielle en fichier d'export DAO.

Apport des balises sur l'industrialisation des traitements

Un autre point d'amélioration portait sur la réduction du nombre de balises à disposer sur l'emprise d'un projet de fouille ouverte. Une étude a été menée pour déterminer le nombre optimal de balises. Parmi les milliers de projets déjà traités, nous avons retenu des projets



avec un recouvrement d'images assez conséquent, semblable à une captation en mode "rafale". 207 projets ont donc été retenus pour cette analyse. Pour chaque set d'images d'un projet, plusieurs tests ont été réalisés en prenant en considération un nombre différent de balises durant le calcul d'aérotriangulation et de mise à l'échelle des projets. Le *tableau 1* indique le pourcentage d'images alignées (minimum de 1 000 points homologues recouvrant au moins 70 % de la surface de l'image) sur le total d'images des 207 projets. L'absence de balises entraîne une baisse de 15 % du nombre d'images alignées par rapport aux projets de références. La présence d'un minimum de trois balises permet d'atteindre un taux de 97,2 % d'images alignées et aucun projet en dessous de 50 % d'images alignées (ce qui assure l'exploitation des projets). Ces résultats nous indiquent qu'un minimum de trois balises est nécessaire pour assurer l'intégrité d'un projet photogrammétrique. Pour renforcer l'industrialisation des traitements, nous avons aussi travaillé à l'amélioration de la méthode de captation d'images.

Méthodologie de captation des images : aller-retour

Le protocole d'acquisition de photos sur une fouille se base sur un aller-retour de l'opérateur. L'aller-retour permet de garantir les points suivants : 1. Doubler le recouvrement d'images sur la zone à mesurer. 2. Permettre une reconstruction plus exhaustive, car les prises de vues ne sont pas toutes dirigées dans le même sens. 3. Augmenter le nombre

Nombres de balises par projet	> 15	5	3	1	0
% D'images alignées	99.6	97.9	97.2	92.0	85.0
Nb de projet avec plus de 50 % d'images non-alignées	0	0	0	9	20

Tableau 1. Statistiques de pourcentage d'images alignées pour 207 projets de reconstruction automatisés en fonction du nombre de balises prises en compte dans le calcul.

de prises de vues des balises au sol, sur au moins deux faces et sur plusieurs images. 4. Reconstruire l'intégralité de la fouille pour assurer la pertinence des mesures dimensionnelles, comme la mesure du volume. 5. Permettre aux géomètres de mesurer des points remarquables dans les environs de la fouille. Toutes ces conditions permettent d'obtenir des prises de vues exploitables pour le post-traitement des données.

Les projets de fouilles ouvertes se distinguent selon leurs géométries (*figure 5*) : Petite fouille de 5 m² où il est possible de se déplacer sur 180° autour de la fouille (Cas 1a) ou à 360° (Cas 1b). Cas 2 : Fouille linéaire le long d'un mur, possibilité de ne marcher que d'un côté de la fouille. Cas 3 : Fouille linéaire en champs ouvert avec possibilité de marcher des deux côtés de la fouille. Pour chaque cas, nous avons testé plusieurs protocoles d'acquisitions en faisant varier la vitesse de marche de l'opérateur et son parcours autour de la fouille. Pour chaque cas de géométrie de fouille, une captation d'images a été faite avec ou sans la présence de balises autour de la scène à reconstruire. Des points de contrôle ont été mesurés au sol pour géoréférencer les projets.

Les résultats pour les cas 1a et 1b des petites fouilles montrent des pourcentages d'alignement de 100 %, quel que soit le protocole d'acquisition. Les photos sont en effet toutes convergentes vers le centre de la fouille et le taux de recouvrement est supérieur à 90 % ce qui garantit que chaque image est corrélée à la reconstruction 3D de la fouille. Les résultats pour les cas 2 et 3 sont synthétisés dans le *tableau 2*. L'absence de balise entraîne une diminution de 50 % du taux d'alignement des images, quelle que soit la vitesse de marche : les images ne s'alignent que sur l'aller ou le retour, mais pas sur tout le jeu de photos. La présence de balise assure un pourcentage d'images alignées supérieur à 90 %, quelle que soit la vitesse de marche de l'opérateur. La vitesse de l'opérateur a un impact sur le nombre d'images par projet. On note une diminution de 50 % du nombre d'images acquises entre une vitesse de marche inférieure à 1 m/s ou supérieure à 1,5 m/s. Cette diminution du nombre d'images entraîne directement une diminution de recouvrement entre images.

Tests longues distances

La nouvelle méthode d'acquisition permet de reconstruire des projets sur des distances supérieures à 20 mètres.

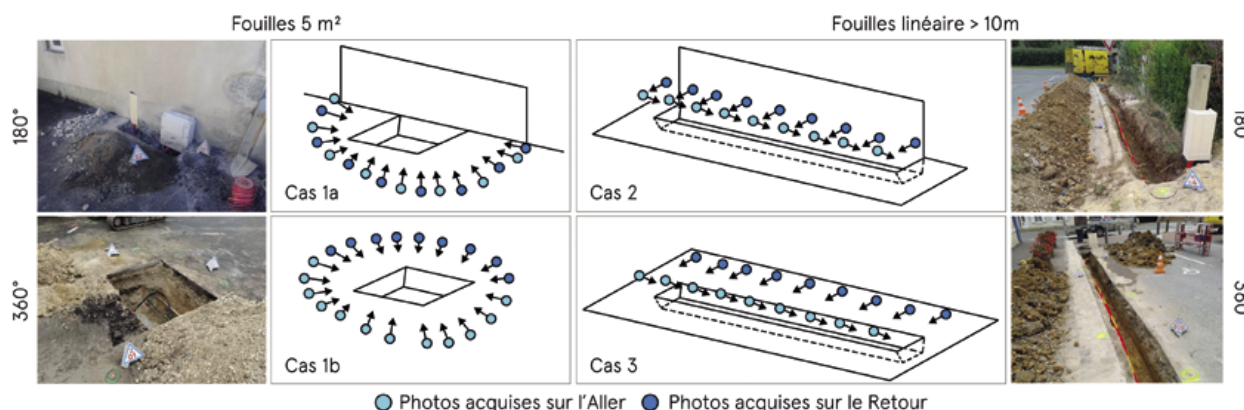


Figure 5. Classification des types de fouilles en fonction de deux paramètres : 1. Capacité à tourner autour de la fouille d'un côté (180°) ou des deux côtés (360°). 2. Géométrie de la fouille : surface de 5 m² ou linéaire supérieure à 10 m.



Vitesse	< 1 m/s				~1m/s				> 1.5m/s			
Côtés d'acquisition aller-retour	Cas 2 : Même côté		Cas 3 : 2 côtés		Cas 2 : Même côté		Cas 3 : 2 côtés		Cas 2 : Même côté		Cas 3 : 2 côtés	
Présence de balises	Avec	Sans	Avec	Sans	Avec	Sans	Avec	Sans	Avec	Sans	Avec	Sans
Nb d'images totales	51	51	55	55	33	33	36	36	21	21	24	24
Nb d'images alignées	51	26	55	55	33	18	36	16	19	9	23	11
% d'images alignées	100 % *	51 %	100 % *	100 % *	100 % *	55 %	100 % *	44 %	90 % *	43 %	96 % *	46 %

Tableau 2. Statistiques de pourcentage d'images alignées sur différents projets de tests en fonction de la vitesse de l'opérateur, de la géométrie de la scène (possibilité de se déplacer autour de la fouille ; présence/absence des balises au sol). Les * représentent les tests corrects.

L'exemple présenté en *figure 6A* montre un projet de 140 mètres de long. Sur ce projet, huit balises ont été disposées afin de garantir la présence d'une balise tous les 20 mètres. Trois tests d'aller-retour ont été effectués pour éprouver la répétabilité du protocole d'acquisition (*figure 6*) en suivant le même trajet. Trois tests ont été effectués à partir d'un Samsung

Galaxy A50 (5760x4320) (test 1 : 251 images ; test 2 : 215 images ; test 3 : 229 images). Pour chaque jeu d'images, nous avons testé le protocole de traitement en prenant en compte ou pas les balises dans le calcul de triangulation et de mise à l'échelle. Pour comparer la géométrie absolue et relative de la reconstruction 3D par photogrammé-

trie, chaque projet photogrammétrique a ensuite été "géoréférencé" en se basant sur un nuage de points acquis par LiDAR statique (Leica RTC360). Chaque nuage de points produit par photogrammétrie a été comparé à la référence LiDAR en utilisant l'algorithme M3C2 [Lague et al., 2013]. L'algorithme M3C2 est une méthode de comparaison de nuages

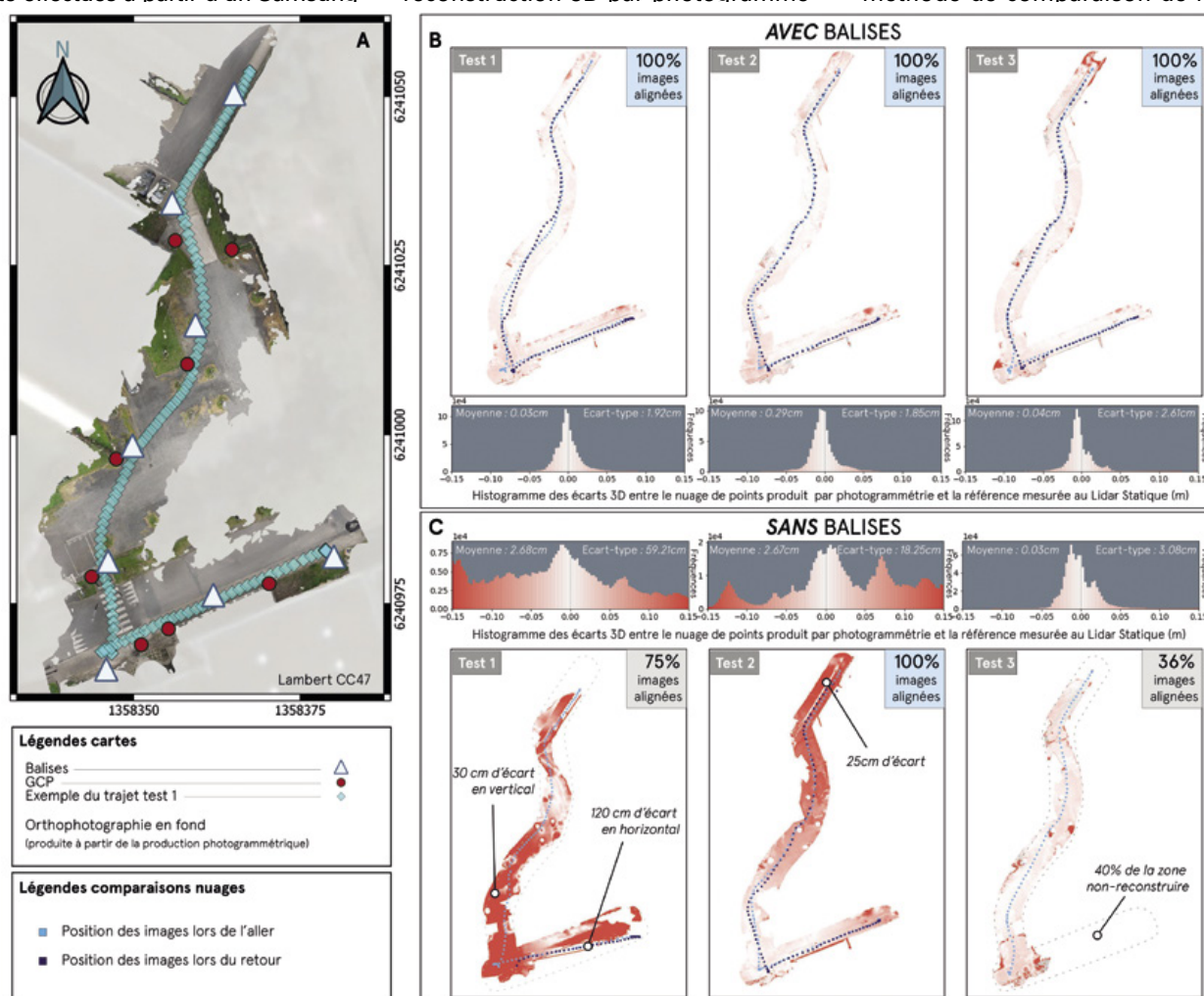


Figure 6. A. Vue aérienne du linéaire de tests avec la position des balises et des points de géoréférencement. Un exemple de trajet est représenté. L'orthoimage en fond est produite à partir de la reconstruction 3D du test 1 avec balises. B et C : Comparaison des nuages de points produits par photogrammétrie et le nuage de points LiDAR statique (B : avec balises ; C : sans balises).



de points qui permet de calculer la moyenne statistique de différences perpendiculairement aux surfaces, ce qui permet de s'affranchir du bruit du nuage de points et de comparer des surfaces obliques. Pour les tests, les différences sont calculées le long des normales aux nuages de points produites par Metashape dans un cylindre de diamètre 2 cm. La *figure 6* présente la comparaison entre des reconstructions avec balises (*figure 6B*) et des reconstructions sans balises (*figure 6C*), pour les mêmes jeux d'images de chaque test. Les trois projets tests avec balises ont tous 100 % d'images alignées, contrairement aux projets sans balises qui n'en comptent qu'un. Le test 1 (75 % d'images alignées) et le test 3 (36 % d'images alignées) présentent des zones non reconstruites pouvant atteindre 40 % du linéaire. Pour les tests avec balises, les trois histogrammes représentant les écarts entre les productions photogramétriques et la "réalité" LiDAR montrent des moyennes et écarts-types centimétriques. Les quelques écarts mesurés supérieurs à 15 cm sont présents sur moins de 1 % de la surface reconstruite. Au contraire, les mêmes images sans les balises conduisent à des reconstructions déformées géométriquement (test 2) ou totalement déplacées en planimétrie (test 1) avec des écarts horizontaux de l'ordre du mètre (*figure 6C*). Seul le test 3 montre des écarts faibles, mais la reconstruction n'est pas complète. La présence de balise consolide le protocole d'acquisition pour les captations sur des longues distances en garantissant la complétude de reconstruction grâce à un fort taux d'images alignées et une bonne précision géométrique sans géoréférencement.

Pour garantir une continuité de l'alignement des images en aller-retour sur des longues distances, nous avons estimé à 25 m la distance maximum entre deux balises sur un linéaire, ce qui représente la pose de cinq balises sur 100 m. En suivant le protocole de prise de vues en aller-retour et en disposant des balises sur la scène, il est possible de produire un nuage de points topographiques précis géométriquement à moins de 5 cm d'erreur en relatif dans la zone proche des prises de vues.

Détection des balises

La technique de captation mise en place assure une reconstruction complète des fouilles et la détection des balises assure à la fois la production d'une donnée à l'échelle permettant des mesures rapides sans géoréférencement et une consolidation de l'alignement des images. La détection des balises est un point crucial pour assurer un service automatique et industrialisé. Les balises ont été développées pour être robustes, faciles à transporter et à disposer au sol, mais aussi pour répondre aux problématiques d'automatisation des traitements. Les balises ont une forme en tétraèdre pour ressortir du sol et être plus visibles loin des positions de prise de vues. Chaque face est unique et son pictogramme est optimisé pour permettre son identification par vision par ordinateur. La reconnaissance de chaque face et la localisation de ses sommets en coordonnées pixels fournissent des points de calage automatiques précis entre les images et permettent une reconstruction précise du modèle 3D.

Les premiers algorithmes développés permettaient de localiser les sommets de tétraèdres d'une image en une minute selon trois étapes : 1. Étape de détection des tétraèdres dans l'image. 2. Traitement par vision par ordinateur pour identifier chaque face unique. 3. Ajustement du triangle bleu pour calculer les dimensions du triangle et en extraire les coordonnées des sommets.

Cette tâche nécessitait de lourds calculs et représentait une partie importante du temps total de reconstruction 3D (66 % du temps de calcul total). De plus, la méthode employée ne permettait pas l'identification de balises dans des conditions variées de prise de vues (luminosité, contraste, occultation).

Afin de réduire ce temps de prétraitement et d'augmenter son efficacité, une nouvelle méthode basée entièrement sur de l'IA (Intelligence Artificielle) a été développée. En particulier, cette nouvelle méthode utilise des réseaux DNN (*Deep Neural Network*). Un réseau DNN est un réseau de neurones comportant un nombre important de couches, chaque couche permettant une étape de transformation de la donnée d'entrée, ici les pixels de l'image. Plusieurs réseaux DNN sont utilisés pour : 1. Détecter les balises dans l'image ; 2. Identifier chaque face de balise ; 3. Identifier l'orientation de chaque face ; 4. Localiser chaque sommet de chaque face en coordonnées image (*figure 7*). Chaque réseau est entraîné sur un dataset comprenant plusieurs milliers d'images séparées en trois dossiers distincts 'train', 'valid' et 'test'. Le dossier 'train' est utilisé pour ajuster les poids du modèle, 'valid' pour ajuster les hyperparamètres et conserver le meilleur modèle et 'test' pour évaluer les performances du modèle. L'étape de détection des balises dans l'image utilise le réseau Mask-RCNN [He et al., 2020] avec comme encodeur un réseau ResNet-50-FPN [Li

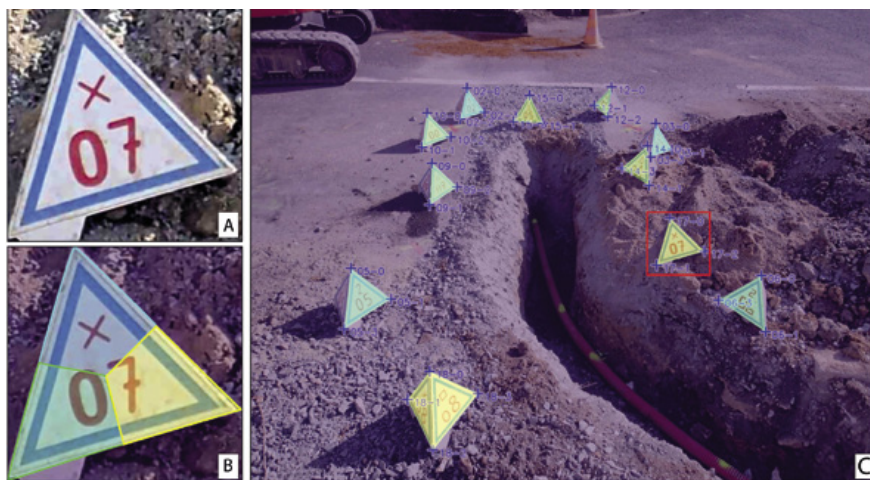


Figure 7. A. Zoom sur la balise 17 (voir dans C). B. Résultat de la segmentation des coins d'une face du tétraèdre. C. Segmentation et calcul des positions des sommets des tétraèdres sur l'image d'origine.

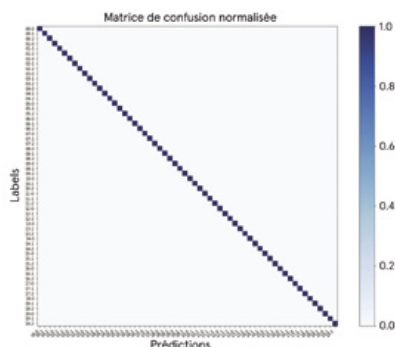


Figure 8. Matrice de confusion pour l'identification des faces de balise. Le modèle basé sur un réseau ResNet18 dépasse les 99.9 % de précision pour l'ensemble des 61 classes.

et al., 2019] pré-entraîné sur le dataset COCO (Lin et al., 2014). Ce réseau Mask-RCNN combine deux réseaux de type CNN (*Convolutional Neural Network*) pour 1. détecter des zones d'intérêt dans l'image correspondant

aux potentiels objets (les balises), 2. identifier les objets et segmenter l'image dans ces zones d'intérêt. Après un entraînement de 160 époques sur des images en résolution 960x1280, le modèle fournit une prédiction de segmentation et une boîte englobant chaque balise avec une précision de 96 % selon la métrique IoU (*Intersection over Union*). L'étape d'identification des faces utilise un réseau ResNet18 pour effectuer une classification avec 60 classes correspondant aux 60 faces uniques de balises plus une classe pour l'arrière-plan sur des images en résolution 480x480. Le réseau ResNet18 est également un réseau CNN, il est l'un des plus simples et performants pour classifier des images. Après un entraînement de 21 époques, le modèle a convergé et prédit les 61 classes avec une précision supérieure à 99.9 %. La

figure 8 montre la matrice de confusion calculée à partir du jeu 'test', seulement deux images sur 2 222 ont été mal classifiées. Ces deux images de balises correspondant à deux prises de vues dans des conditions difficiles, vue de côté ou importante occultation, pour lesquelles l'identification est très difficile et ce, même pour un humain. L'étape de détection d'orientation des faces et la localisation des sommets sont réalisées avec le même modèle, combinant le réseau ResNet18 avec en plus une étape de décodage permettant d'attribuer une classe à chaque pixel de l'image d'origine en appliquant des déconvolutions suite à l'étape d'encodage (convolutions) du ResNet18. Ce réseau est entraîné à segmenter trois classes correspondant à trois zones sur chaque face (haute, gauche et droite) de balise,



Images Brutes



Résultats de la segmentation des objets

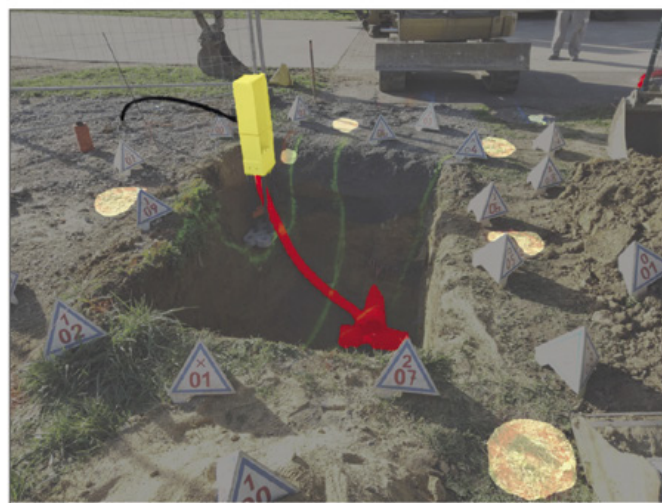
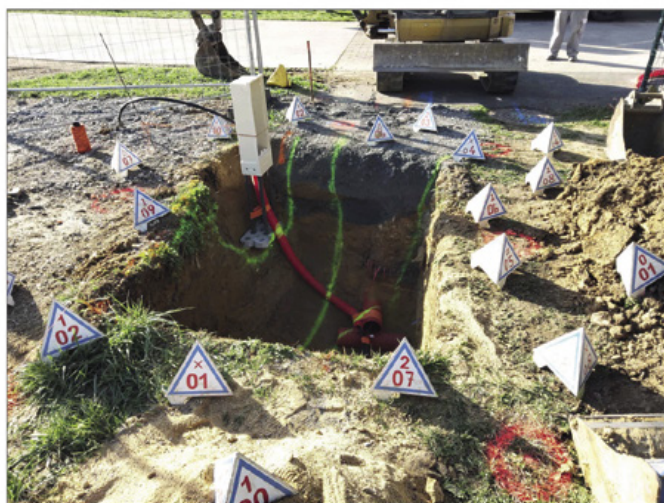
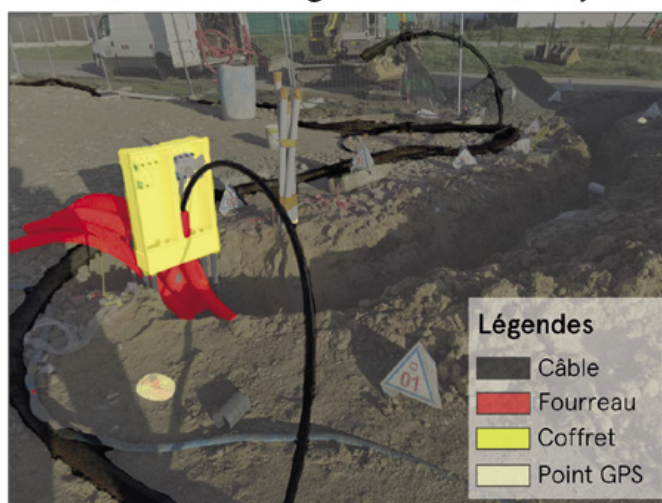


Figure 9. Résultats de la segmentation des objets par apprentissage profond pour la détection des câbles, fourreaux, coffret et points GPS présents autour d'un levé de fouille ouverte.



plus une classe pour l'arrière-plan sur des images de résolution 480x480 (figure 7, B). Après un entraînement sur 161 époques, le modèle segmente les faces avec un IoU moyen de 98 %. Grâce à cette segmentation particulière, un algorithme dédié permet de calculer la position de chaque sommet de cette face en coordonnées "image" (figure 7, C). Un quatrième modèle, basé également sur un réseau Unet-Resnet18, a été entraîné sur 88 époques pour atteindre un IoU moyen de 98 % sur la segmentation des 60 différentes faces de balise. Ce modèle supplémentaire permet de vérifier par vote les étapes 2, 3 et 4 sur les mêmes images de résolution 480x480.

Cette nouvelle méthode 100 % intelligence artificielle permet en seulement une seconde (sur une RTX 2080 Ti) d'analyser et d'extraire la position des sommets uniques de chaque balise présente dans l'image en résolution native (ex. 3096x4128). Un jeu de données sur 30 mètres linéaires est maintenant traité en moins d'une minute.

Perspectives

Dans la continuité de l'amélioration de la détection des balises, une étude a commencé en septembre 2021 pour détecter et segmenter automatiquement les objets sur un projet de fouille ouverte : les câbles enterrés, les fourreaux, les points de géoréférencement et les coffrets. La détection automatique de ces objets a nécessité l'annotation de plus de 1 000 images, soit plus de 8 000 objets. L'apprentissage se base sur le réseau Unet-ResNet18 avec des images de résolution 1920x1440. L'entraînement du réseau sur 88 époques permet d'atteindre un IoU moyen de 80 % sur la segmentation des fourreaux rouges, 60 % sur les clous, 38 % sur les coffrets et 25 % sur les câbles noirs. Les résultats sont prometteurs sur certaines configurations complexes de chantier (figure 9). Cette détection des objets sur les images permettra de proposer une digitalisation automatique du câble enterré et un pré-positionnement des points de géoréférencement pour l'utilisateur dans l'interface web pour automatiser

l'étape de dessin et de géoréférencement (figure 3).

Conclusion

La captation d'images à partir de smartphones permet un accès simple et rapide à un modèle 3D topographique de précision grâce aux avancées technologiques en photogrammétrie et intelligence artificielle. Un utilisateur sans connaissance des principes de base de la photogrammétrie par corrélation dense peut en quelques minutes produire une scène 3D d'un environnement urbain. La méthode développée ici a pour but une industrialisation du processus de production de données en livrant un nuage de points 3D et une aide à la digitalisation avec une garantie quant à l'expertise sur le traitement de la donnée. ●

Contact

Valentin POITEVIN

Directeur de l'innovation, GEOFIT Group.
v.poitevin@geofit.fr

Références

- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2020. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 386-397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- Hinge, L., Gundorph, J., Ujang, U., Azri, S., Anton, F., Rahman, A.A., 2019. Comparative analysis of 3D photogrammetry modeling software packages for drones survey. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42, 95-100. <https://doi.org/10.5194/isprs-archives-XLII-4-W12-95-2019>
- James, M.R., Antoniazza, G., Robson, S., Lane, S.N., 2020. Mitigating systematic error in topographic models for geomorphic change detection: accuracy, precision and considerations beyond off-nadir imagery. *Earth Surface Processes and Landforms* 45, 2251-2271. <https://doi.org/10.1002/esp.4878>
- Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and*

Remote Sensing 82, 10-26. <https://doi.org/10.1016/j.isprsjprs.2013.04.009>

Li, X., Lai, T., Wang, S., Chen, Q., Yang, C., Chen, R., 2019. *Weighted feature pyramid networks for object detection*. Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019 1500-1504. <https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217>

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. *Microsoft COCO: Common objects in context*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8693 LNCS, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

Lowe, D.G., 2004. *Distinctive Image Features from Scale-Invariant Keypoints*. *International Journal of Computer Vision* 60, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>

Yilmazturk, F., Gurbak, A.E., 2019. Geometric Evaluation of Mobile-Phone Camera Images for 3D Information. *International Journal of Optics* 2019, 1-10. <https://doi.org/10.1155/2019/8561380>

ABSTRACT

Today new solutions exist to measure underground networks. Image acquisition with a smartphone makes it possible to build open excavation in 3D thanks to dense correlation photogrammetry algorithms. The challenges of reliability and industrialization for this type of data production pushed us to improve a pre-existing acquisition method. The main objective of this study is to determine the best image capture technique. Key elements for the success of the method are based on 1. the use of dimensional markers through the installation of beacons and 2. the detection of these markers in images via algorithms based on artificial intelligence. The 3D data produced and available through online web application make it possible to draw existing infrastructures while reducing human intervention as much as possible.