공학석사 학위논문

Camera settings and flight method for optimum photography of UAV based SfM-MVS

UAV-입체사진측량의 최적 촬영을 위한 카메라 세팅 및 비행방법

2023 년 2 월

서울대학교 대학원
에너지시스템공학부
임 준 수

Camera settings and flight method for optimum photography of UAV based SfM-MVS

지도교수 송 재 준

이 논문을 공학석사 학위논문으로 제출함

2022 년 12 월

서울대학교 대학원

에너지시스템공학부

임 준 수

임준수의 공학석사 학위논문을 인준함

2023 년 1 월

위 원 장	전 석 원	(인)
부위원장	민 기 복	(인)
위 원	송 재 준	(인)

Abstract

Over the last decade, structure−from−motion (SfM) and multi−view−stereo (MVS) techniques have proven effective in generating high−resolution and high−accuracy 3D point clouds with the possibility to integrate with unmanned aerial vehicles (UAVs). However, the SfM−MVS techniques still had the limitation that the error of point clouds could not be predetermined before point cloud generation. In this work, a theoretical error prediction model is formulated based on the propagation of 2D image errors to 3D point cloud errors, and the disruption effect of blur and noise on 2D image errors is analyzed according to camera settings, UAV flight method, camera specification, and illumination. By comparing the error predictions with those observed in experimental data, an error prediction performance of R^2=0.83 is confirmed. Based on the high performance of error prediction, this work presents a method to determine the optimum photographing settings, including camera settings and the UAV flight method, by which the point cloud errors are minimized under illumination and time constraints. The importance of the optimum photographing settings is verified by comparing the error levels of the optimum photographing setting with those of arbitrary settings. For site validation and comparison with the light detecting and ranging (LiDAR) method, the SfM−MVS method utilizing the optimum photographing settings was applied along with LiDAR to a tunnel face located in Yeoju−si, Korea, where the light and surveying time is limited. As a result, the SfM-MVS method could achieve a point cloud with 3 times better accuracy and 20 times higher resolution at a cost of 1/9 than the LiDAR method.

Keyword: SfM (structure from motion); MVS (multi-view stereo); error predetermination; camera settings; UAV flight method; Underground Digital Survey; LiDAR (light detecting and ranging)
Student Number: 2021-20978

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Structure from motion (SfM) and multi-view stereo (MVS) photogrammetry has shown great performance to generate 3D point cloud or 2.5D digital elevation model (DEM) with high resolution and precision at low prices in the past decade (Smith et al., 2016). It can be used in conjunction with moving camera platforms such as an unmanned aerial vehicle (UAV) and acquire data rapidly even in places where it is difficult to access (Aasen et al., 2015; Micheletti et al., 2015; Smith and Vericat, 2015; Smith et al., 2016; Eltner et al., 2016; Tonkin and Midgley, 2016; James et al., 2019).

SfM-MVS can generate its high dimensional output from a subsequent set of 2D photos as input by bundle adjustment (BA) and dense reconstruction processes. However, the quality of the SfM-MVS generated point clouds or DEM is an unknown function of diverse parameters including scale (Ohnishi et al., 2006; Wenzel et al., 2013; Micheletti et al., 2015; Smith and Vericat, 2015; Smith et al., 2016; Eltner et al., 2016; Bakker and Lane, 2017; James et al., 2019; Javadnejad et al., 2021), ground control points (GCP) (Wheaton et al., 2010; Smith et al., 2016; Eltner er al., 2016; Tonkin and Midgley, 2016; Javadnejad et al., 2021), focal length (Ohnishi et al., 2006; Wenzel et al., 2013; Tonkin and Midgley, 2016; James et al., 2019), angle of incidence (Javadnejad et al., 2021), texture (Furukawa and Ponce, 2009; Wenzel et al., 2013; Aasen et al., 2015; Micheletti et al., 2015; Smith et al., 2016; Bakker and Lane, 2017), number of images (Micheletti et al., 2015; Smith et al., 2016; Eltner et al., 2016; Bakker and Lane, 2017; Javadnejad et al., 2021) and camera calibration (Wenzel et al., 2013; Smith et al., 2016; Eltner et al., 2016; Bakker and Lane, 2017; James et al., 2019). In addition to the fact that numerous parameters affect the output quality, the usage of different SfM-MVS software makes it more complex to analyze (Smith et al., 2016).

Evaluating the error of the SfM-MVS generated point cloud or DEM is important because it can affect the results of such analysis like surface roughness estimation (Smith et al., 2016; Smith and Vericat, 2015; Eltner et al., 2016; James et al., 2019), topography change assessment (Lane et al., 2003; Wheaton et al., 2010; Smith and Vericat, 2015; Bakker and Lane., 2017), slope stability analysis (Ohnishi et al., 2006), vegetation monitoring (Aasen et al., 2015) and rock mass characterization (Vollgger et al., 2016; Becker et al., 2018; Zhang et al., 2019). General error change of each parameter has been investigated from various research. Smith and Vericat (2015) showed linear error increases according to scale increases. Tonkin and Midgley (2016) identified that the use of 4 or more GCPs exhibits a similar error level. Lack of texture may reduce the performance of bundle adjustment and dense matching (Furukawa and Ponce, 2009; Wenzel et al., 2013; Aasen et al., 2015; Micheletti et al., 2015; Smith et al., 2016; Bakker and Lane, 2017), and this should be solved by obtaining sufficient texture within images by photographing at an appropriate scale. Micheletti et al. (2015) found a non-linear error trend in the manner of image number used to generate a point cloud and they stated an error convergence above a certain image number. The use of fixed camera calibration was recommended for a better and more accurate BA in the SfM process by Wenzel et al. (2013).

Also, there have been attempts to theoretically predict SfM-MVS generated point clouds or DEM error. Lane et al. (2003) considered image scale and resolution to predetermine error level but it has a drawback to underestimate the actual error amount because it represents ideal minimum errors. Wheaton et al. (2010) adopted a fuzzy inference system to predict DEM elevation error using GPS point quality, slope, and point density as input, though it needs calibration between input and output (DEM elevation error) under different circumstances making it impractical. Wenzel et al. (2013) took the precision of parallax between images into account for error prediction, yet it is challenging to know the precision of parallax. However, most of the previous studies have paid little attention to image quality, though images are the main input for the SfM-MVS technique.

Images are photographed at a certain illuminance and they should be neither overexposed nor underexposed. It is important because inappropriate image brightness can disturb the feature extraction and matching process (Micheletti et al., 2015, Javadnejad et al., 2021). Image brightness is related to the camera settings including shutter speed (exposure time), F-number (aperture), and ISO, and they have a tradeoff between motion blur, out-of-focus blur, and noise respectively (Wenzel et al., 2013). In still photographing, exposure time can be increased to overcome the problem, but when the camera is carried with a moving platform like UAV, image quality is degraded due to blurring and vibrations (Dunford et al., 2009). Although the gimballed camera can diminish the vibration issue (Smith et al., 2016), still shutter speed is restricted to avoid serious motion blur in the captured image. Images can be severely distorted when they are shot in a dark environment like night photography (Burdziakowski and Bobkowska, 2021) and photographing in underground mines (Slaker and Mohamed, 2017) or tunnels (García−Luna et al., 2019) therefore selecting proper camera settings is significantly important. According to a survey conducted by the Korea institute of construction technology (KICT, 2017), there have been concerns that the quality of the SfM−MVS−generated point clouds is unreliable in a tunnel construction site due to light deficiency. For such reason, workers preferred the light detecting and ranging (LiDAR) technique still it had concerns of high capital cost, shadow zones, texture distortion, and excessive surveying time. Total surveying time is limited in general cases to be used for practical purposes. To keep the time constraint, the flight method, including distance from the object or flight height and UAV speed, needs to be planned with care as it indirectly affects the image quality. Nonetheless, little research has been conducted on the selection of the optimum photographing settings, including camera settings and UAV flight method, to acquire the best image quality.

This work suggests a procedure to derive the optimum photographing setting to obtain the best image quality which results in the minimum 3D point cloud error under given conditions of illumination and time constraint. A theoretical model is formulated to pre-determine the 3D point cloud error by the propagation of 2D image error and it is explained in Section 2.1. The model is validated with indoor experiments and its validation result is described in Section 3.1. Based on the derived error prediction model, Section 2.2 explores the decision process for optimum photographing setting and its application at an underground tunnel face in Yeoju-si, Korea. Section 3.2 covers the importance of the optimum photographing settings by comparing the error levels between the optimum and arbitrary photographing settings, and the comparison result with the LiDAR at the tunnel site. Section 4 discusses the implication of the theoretical error prediction model (Section 4.1) and the feasibility of deriving the optimum photographing settings (Section 4.2).

Chapter 2. Theory and methodology

2.1. Theoretical model for error prediction

In this work, photographs were shot where the camera is mounted on a moving platform such as a UAV, and the term UAV hereafter will refer to all of the moving platforms of this work. To predict the SfM−MVS output error, a theoretical model was formulated under three stages (Fig. 2−1). First, all of the SfM−MVS output errors were assumed to be propagated from 2D image errors, which images are the input of the SfM−MVS technique. Then, four types of image quality factors were chosen that have a major impact on image errors. Finally, the theoretical error prediction model was formulated by analyzing the parameters that determine the image quality factors.

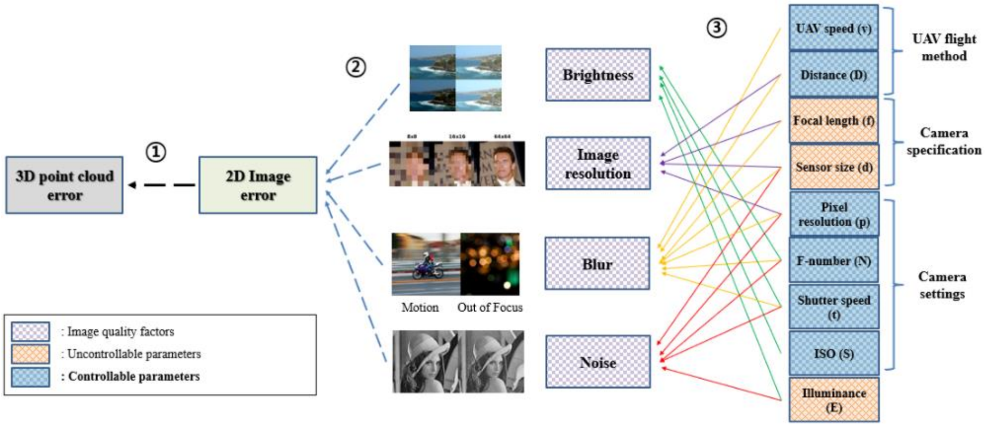


Fig. 2−1. Three stages of formulating the theoretical model for error prediction.

2.1.1. Error propagation from 2D image to 3D point cloud

Prior to the theoretical model formulation, the definition of error in this work is clarified. Three types of methods are generally used to assess the error of SfM−MVS output: point−to−point (PP), point−to−raster (PR), and raster−to−raster (RR) (Smith and Vericat, 2015). PP method compares two point clouds, the PR method compares a DEM against reference points from TS or dGPS, and the RR method compares two DEMs. In this work, the PP method was chosen because additional error can inherit from sampling strategy and interpolation methods when using DEMs (Wheaton et al., 2010), and also the accuracy of SfM−MVS may outperform the accuracy of the reference point itself (Smith and Vericat, 2015; Smith et al., 2016; Eltner et al., 2016). For the representative value for error assessment, various studies have reported root mean squared error (RMSE), mean error (ME) and mean absolute error (MAE) (Smith et al., 2016). Their trend behaves similarly, hence in this work, RMSE was selected since it has the advantage of being statistically manageable compared to the other two values. In conclusion, RMSE between two point clouds, one being an error−free reference point cloud, is used as the error of SfM−MVS output in this work.

This work assumed the error of a point cloud is originating from image error (Fig. 2−2), i.e. systematic error due to factors including invalid BA, lens distortion, inappropriate use of GCPs, insufficient image overlap, unfixed camera calibration, lack of texture is neglected. BA appears to be reliable in general cases because feature detection methods like SIFT (Lowe, 2004) and SURF (Bay et al., 2008) are robust to distortion and the keypoint correspondence process removes outliers by applying techniques like RANSAC (Fischler and Bolles, 1981), maximum likelihood estimation sample consensus (Torr and Zisserman, 2000) or the Hough transform (Ballard and Brown, 1982). Since recent SfM software well calibrates lens distortion, it was regarded to have no major impact on the error in the resultant point cloud (Bakker and Lane, 2017). In addition, the use of more than 4 GCPs, sufficient image numbers, and fixed camera calibration makes the prior assumption more reliable.

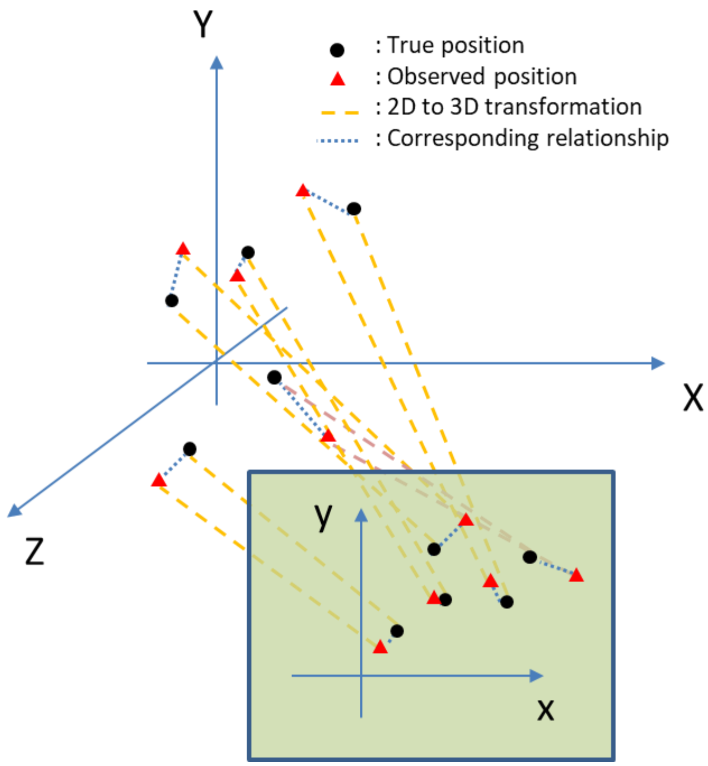


Fig. 2-2. 2D image errors propagation to 3D point cloud errors.

A point (x,y) in a 2D image is transformed into a 3D point (X,Y,Z) by Helmert transformation (Eq. (1)).

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \mathrm{sR}^{\mathrm{T}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathrm{T} \qquad (1)$$

$$R^{\mathrm{T}} = \begin{pmatrix} \overrightarrow{r_1} \\ \overrightarrow{r_2} \\ \overrightarrow{r_3} \end{pmatrix} = \begin{pmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{pmatrix} = \mathrm{f}(\omega, \phi, \kappa) \qquad (2)$$

s is the scale factor which is about the same as the ratio of the distance D to the focal length f of the camera lens, R is the rotation matrix and its elements are a function of angles between the camera coordinate and the real world coordinate (Eq. (2)), T is the translation vector, and z of the image plane is set to −f. Assumptions have been made that s, R, and T are accurately determined during BA in the SfM process (James and Robson, 2012) and the f value is properly put by using fixed and correct camera calibration. By combining Eq. (1) and the definition of the 3D point cloud error (Eq. (3)) with the error propagation theory by Taylor (1997), the relation between 2D image distortion (i.e. $\mathrm{dx_i}$ and dy_i) and 3D point cloud error becomes as Eq. (4).

$$\mathrm{RMSE_{3D}} = \sqrt{\frac{\sum_{i=1}^{N}\left(dX_i^2 + dY_i^2 + dZ_i^2\right)}{N}} \qquad (3)$$

$$\mathrm{RMSE_{3D}} = \left[\frac{1}{N} \times \sum_{i=1}^{N} \left\{ \left(\frac{\partial X_i}{\partial_{xi}} dx_i + \frac{\partial X_i}{\partial_{yi}} dy_i\right)^2 + \left(\frac{\partial Y_i}{\partial_{xi}} dx_i + \frac{\partial Y_i}{\partial_{yi}} dy_i\right)^2 + \left(\frac{\partial Z_i}{\partial_{xi}} dx_i + \frac{\partial Z_i}{\partial_{yi}} dy_i\right)^2 \right\}\right]^{\frac{1}{2}}$$

$$= \left[\frac{1}{N} \times \sum_{i=1}^{N} \{(s \times r_{11} dx_i + s \times r_{21} dy_i)^2 + (s \times r_{12} dx_i + s \times r_{22} dy_i)^2 + (s \times r_{13} dx_i + s \times r_{23} dy_i)^2\}\right]^{\frac{1}{2}} \tag{4}$$

$$= s \times \sqrt{\frac{\sum_{i=1}^{N}\left\{|\overrightarrow{r_1}|^2 dx_i^2 + |\overrightarrow{r_2}|^2 dy_i^2 + 2(\overrightarrow{r_1} \cdot \overrightarrow{r_2}) dx_i dy_i\right\}}{N}}$$

Since rotation matrix R is an orthonormal matrix, $|\overrightarrow{r_1}| = |\overrightarrow{r_2}| = 1$, $\overrightarrow{r_1} \cdot \overrightarrow{r_2} = 0$. Therefore, considering the definition of the 2D image error (Eq. (5)), Eq. (4) is simplified to Eq. (6).

$$\mathrm{RMSE_{2D}} = \sqrt{\frac{\sum_{i=1}^{N}\left(dx_i^2 + dy_i^2\right)}{N}} \tag{5}$$

$$\mathrm{RMSE_{3D}} = s \times \sqrt{\frac{\sum_{i=1}^{N}\left(dx_i^2 + dy_i^2\right)}{N}} = s \times \mathrm{RMSE_{2D}} \cong \frac{D}{f} \times \mathrm{RMSE_{2D}} \tag{6}$$

Eq. (6) indicates that 2D image error is magnified by s, the scale factor when it is transformed into a 3D point cloud. Hence, SfM−MVS generated 3D point cloud error $\mathrm{RMSE_{3D}}$ could be predicted by calculating 2D image error $\mathrm{RMSE_{2D}}$.

2.1.2. Image quality factors

In this work, four types of image quality factors that take a significant role in the amount of 2D image error were selected: Image brightness, image resolution, blur, and noise. Other image quality attributes were excluded because they were regarded to have little impact on the image error. For example, white balance (WB) was excluded because it can be adjusted perfectly by image post-processing if the image is photographed in RAW format, and its effect would have little impact on image error if an extreme WB setting was not used. Another factor, such as lens distortion, may result in a severe image error but it was also excluded since recent SfM-MVS software are well capable of lens distortion correction (Micheletti et al., 2015).

Image brightness (C) is determined by the illuminance (E, [lux]) and camera settings such as ISO linear speed (S, [s]), shutter speed (t), and F-number (N), and their relationship is expressed as Eq. (7) (International Organization for Standardization, 1974).

$$\mathrm{C} = \frac{\mathrm{ESt}}{N^2} \tag{7}$$

Image brightness was considered an important image quality factor for three reasons. First, if the image is not appropriately exposed, i.e. overexposed or underexposed, it would have a negative effect on feature extraction and dense matching (Micheletti et al., 2015, Javadnejad et al., 2021). Second, image brightness has a tradeoff relation between blur and noise considering that it is related to camera settings (Wenzel et al., 2013). If one has to increase image brightness under given illuminance, either blur or noise amount of the image is also inevitably increased. Lastly, the resultant point cloud would have the same color or texture as the image used, thus appropriate image brightness is necessary where the resultant point cloud texture plays an important role in the subsequent analysis. According to the international organization of standardization (1974), the use of C=250 is recommended, therefore, this work analyzed the image error which is caused by the blur or noise when using the camera settings to maintain C=250 in Eq. (7).

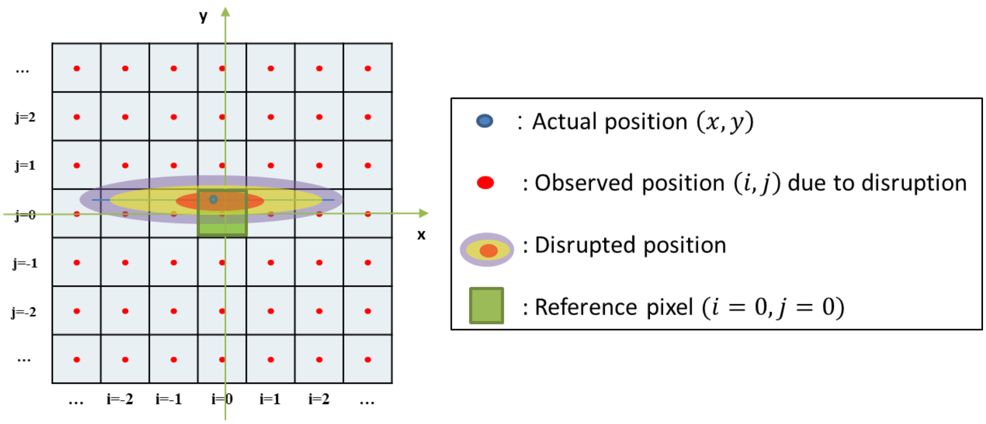


Fig. 2-3. Theoretical error prediction model: a point (x, y) in the reference pixel is disrupted due to blur and noise, and it is observed in (i, j) pixels rather than (0, 0).

Image resolution or sharpness was regarded as important because the precise position of the points within a pixel is lost

(Wenzel et al., 2013). It was considered that image error arises when positions of multiple points within a pixel are merged into the center of the pixel. The model in Fig. 2−3 represents an image plane and (i, j) are the pixels where i=0 and j=0 is the reference pixel. Points in a reference pixel can be disrupted from their true positions by blur and noise. Due to the disruption, they have a possibility to be observed in other pixels rather than the reference pixel. When a point, where its true position is (x, y), is observed in (i, j) pixel, then the error of the point $\epsilon_{((i,j)|(x,y))}$ is expressed as Eq. (8).

$$\epsilon^2_{((i,j)|(x,y))} = (x-i)^2 + (y-j)^2 \; [pixel^2] \tag{8}$$

$P((i,j)|(x,y))$ is the probability of a point (x, y) to be observed in (i, j) pixel, and there are multiple possible pixels to be observed, the error of the point $\epsilon_{(x,y)}$ is calculated as Eq. (9).

$$\begin{aligned} \epsilon^2_{(x,y)} &= \sum_i \sum_j \{P((i,j)|(x,y)) \times \epsilon^2_{((i,j)|(x,y))}\} \\ &= \sum_i \sum_j \left\{P((i,j)|(x,y)) \times \{(x-i)^2 + (y-j)^2\}\right\} \; [pixel^2] \end{aligned} \tag{9}$$

The point (x, y) is an arbitrary point within the reference pixel following the probability distribution p(x,y). Considering that there are innumerable points inside the reference pixel, image error $RMSE_{2D}$ calculated as Eq.(10).

$$\mathrm{RMSE}_{2\mathrm{D}}^2 = \int_{-0.5}^{0.5}\int_{-0.5}^{0.5} p(x,y) \times \epsilon_{(x,y)}^2 \; dx\, dy \; [pixel^2] \tag{10}$$

Assuming that all of the points inside the reference pixel have the same weight, i.e. $\mathrm{p(x,y)}=\frac{1}{1\ \mathrm{pixel}^2}$, Eq. (10) is simplified as Eq. (11).

$$\begin{aligned}\mathrm{RMSE}_{2\mathrm{D}}^2 &= \int_{-0.5}^{0.5}\int_{-0.5}^{0.5} \epsilon_{(x,y)}^2 \; dx\, dy \\ &= \int_{-0.5}^{0.5}\int_{-0.5}^{0.5} \sum_i \sum_j \Big\{ \mathrm{P}\big((\mathrm{i,j})\big|(\mathrm{x,y})\big) \\ &\qquad \times \{(x-i)^2 + (y-j)^2\}\Big\} \; dx\, dy \; [pixel^2]\end{aligned} \tag{11}$$

Therefore, $\mathrm{RMSE}_{2\mathrm{D}}$ can be calculated by knowing the $\mathrm{P}\big((\mathrm{i,j})\big|(\mathrm{x,y})\big)$ and it lets the calculation of $\mathrm{RMSE}_{3\mathrm{D}}$ from Eq. (6), which is the SfM−MVS generated point cloud error. $\mathrm{P}\big((\mathrm{i,j})\big|(\mathrm{x,y})\big)$ can be computed statistically by estimating the disruption amount caused by blur and noise.

2.1.3. Effects of parameters on image disruption

The exterior orientation of the camera is measured during BA, and dense matching is conducted along the relative orientation between images. Without losing generality, UAV was set to move along +x direction, dense matching would be also conducted along +x direction, facilitating subsequent analyses. In this work, zoom lenses were not considered since it changes the focal length which should be avoided (Micheletti et al., 2015; Smith et al., 2016; Eltner et al., 2016).

2.1.3.1. Motion blur

Image is exposed for a time of shutter speed t during UAV motion, and light from a point object is spread out in a form of point spread function (PSF) on the image plane In this study, it is assumed that the UAV is moving at a uniform speed v along the x−axis, then its PSF would be a uniform distribution along the x−axis having the inverse of the blur amount as its magnitude while no spread along the y−axis. The blur amount B on an image plane can be calculated as Eq. (12) using proportional expressions described in Fig. 2−4, and PSF due to blur is expressed in Eq. (13).

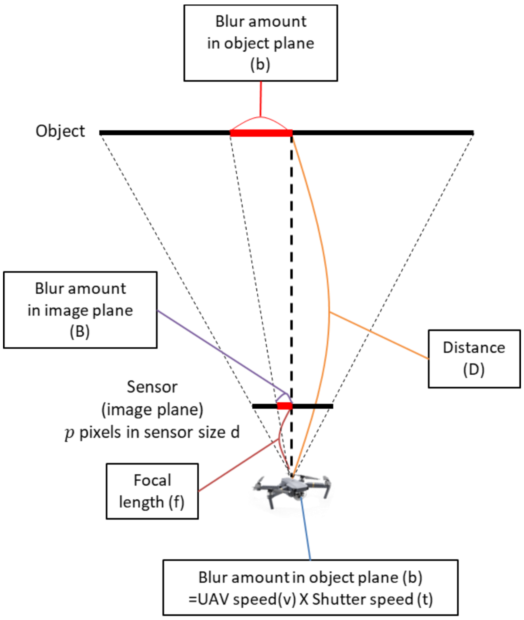


Fig. 2−4. Motion blur is captured in an image due to UAV motion (v) and shutter speed (t).

$$B = \frac{\mathrm{bf}}{\mathrm{D}} = \frac{vtf}{D}\ [m] = \frac{vtf}{D} \times \frac{p}{d}\ [\mathrm{pixel}] \tag{12}$$

$$\Delta \mathrm{x_{mb}} \sim \mathrm{U}\left(-\frac{\mathrm{B}}{2}, \frac{B}{2}\right) \\ \Delta y_{mb} = 0 \tag{13}$$

D is the distance from the object, f is the focal length, p is pixel resolution or number of pixels on the sensor, and d is sensor size. Note that p/d was multiplied to convert the meter unit into a pixel unit.

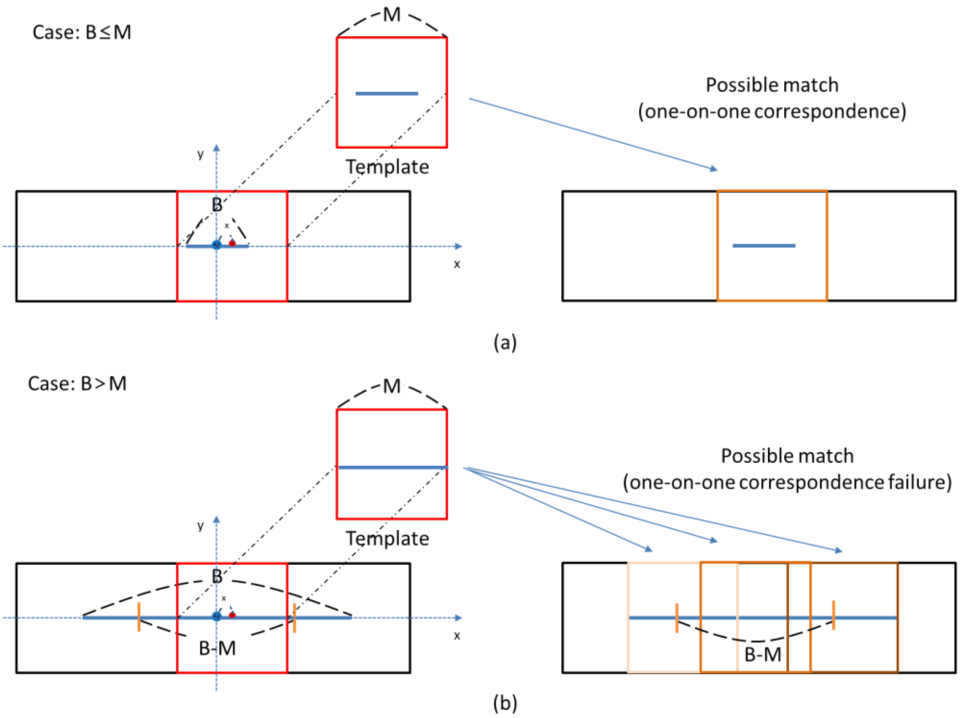


Fig. 2−5. Maximum window size M effect on the dense matching of a point disturbed by motion blur amount of B: (a) B≤M case: One−to−one correspondence between two images; (b) B>M case: A template from one image has multiple correspondences in the other image.

Most dense matching algorithms of the recent SfM-MVS software are patch-based or window-based methods (Smith et al. 2016). The disruption effect in Eq. (13) is valid only until B exceeds the maximum window size M used in the dense matching process because the search window could not capture all of the blurs at once. As seen in Fig. 2-5(a), a reference point is disrupted due to motion blur B while any disrupted position x can correspond to the same position relative to the reference point 0 of the other image. While as seen in Fig. 2-5(b), when B is larger than M, the disrupted position in one image has multiple possible matches in the other image.

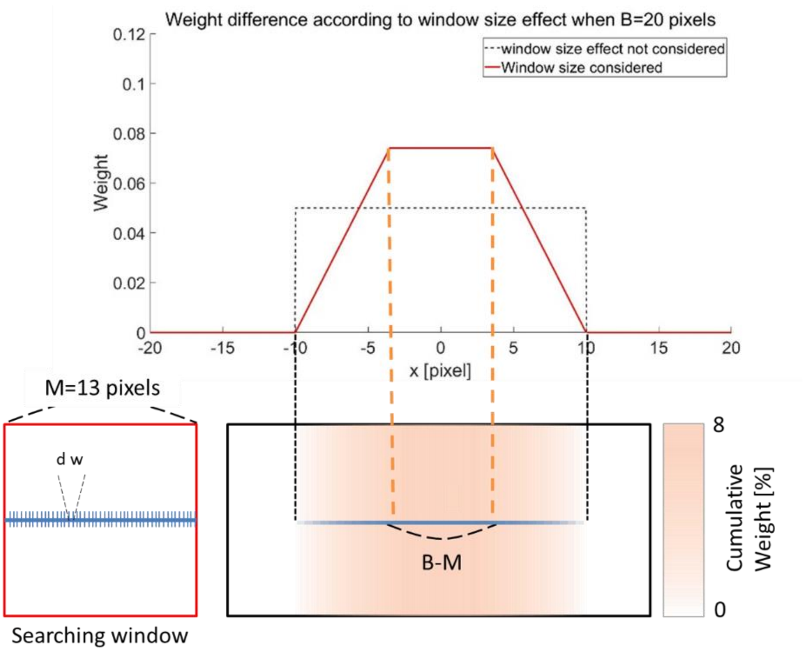


Fig. 2-6. Weight difference due to window size effect when B=20 pixels and M=13 pixels. As higher weight is more concentrated near the reference point when the window size effect is considered (red line of the upper figure), the error expectation is lower than the case when the window size effect is not considered (black dashed line of the upper figure).

All blurred points within the searching window would have the same weight of dw to be chosen instead of the reference point O. Considering the window size effect, i.e. multiple correspondences of the searching window, each possible match would cumulate the weight dw, resulting in isosceles trapezoidal distribution. Center point of the searching window $x \in [-\frac{B-M}{2}, \frac{B-M}{2}]$, and assuming all possible matches have the same possibility, the cumulative weight would have the shape of an isosceles trapezoid (Eq. (14)). Weight increases linearly at the region of $x \in [-\frac{B}{2}, -\frac{B-M}{2}]$, ceases during $x \in [-\frac{B-M}{2}, \frac{B-M}{2}]$, and decreases at the region of $x \in [\frac{B-M}{2}, \frac{B}{2}]$, and Eq. (13) is modified as Eq. (15).

$$\Delta x_{mb} \sim \mathrm{Trap}\left(-\frac{B}{2}, -\frac{B-M}{2}, \frac{B-M}{2}, \frac{B}{2}\right) \tag{14}$$

$$\begin{cases} \Delta x_{mb} \sim U\left(-\frac{B_{eff}}{2}, \frac{B_{eff}}{2}\right) \cdots B \le M \\ \Delta x_{mb} \sim \mathrm{Trap}\left(-\frac{B}{2}, -\frac{B-M}{2}, \frac{B-M}{2}, \frac{B}{2}\right) \cdots B > M \end{cases} \tag{15}$$
$$\Delta y_{mb} = 0$$

Since the higher weight region of the trapezoidal distribution is more concentrated near the reference point, the error expectation is lower than the uniform distribution case. Namely, the disruption effect of motion blur reduces when B is larger than M and the maximum window size effect should be considered to predict the error. However, different SfM−MVS software adapts different dense matching algorithms using different maximum window sizes (Scharstein, 2002), thus, the M value should be calibrated according to the software used.

2.1.3.2. Out−of−focus blur

The camera lens converges light from a point object to the image plane but the lens cannot clearly focus an object at all distances and it causes out−of−focus blur. The diameter of the out−of−focus blur circle $\mathrm{D_{off}}$ is calculated as Eq. (16) (Born and Wolf, 1965).

$$\mathrm{D_{off}} = \frac{|D-H|}{D}\frac{f^2}{N(H-f)}\frac{\mathrm{p}}{\mathrm{d}}\ [pixel] \tag{16}$$

H is the focus distance of the lens and it may be adjusted manually or automatically, however, it was set to hyperfocal distance in this work. The main reason was to use a consistent focal length since the listed focal length represents when the lens is focused at a hyperfocal distance and it changes according to the focus distance. In addition, using hyperfocal distance was beneficial because it keeps objects at a far distance in focus which is important in aerial photogrammetry, and it gives a maximum depth of field which is required in such situations like low light photography where lower F−number is used. If one inevitably has to use other focus distances, camera calibration should be done with care.

Out−of−focus blur has PSF in form of 2D Gaussian distribution with a standard deviation σ_{off} equal to half the blur circle diameter rather than pillbox form (Pentland, 1987). Therefore, σ_{off} is as Eq. (17) and the disruption effect due to out of focus is as Eq. (18).

$$\sigma_{off} = \frac{|D-H|}{2D}\frac{f^2}{N(H-f)}\frac{\mathrm{p}}{\mathrm{d}}\ [pixel] \tag{17}$$

$$\begin{aligned} \Delta x_{off} &\sim N\left(0,\ \sigma_{off}\right) \\ \Delta y_{off} &\sim N\left(0,\ \sigma_{off}\right) \end{aligned} \tag{18}$$

2.1.3.3. Noise

Image noise is an undesired pixel intensity variation that can be generated from multiple sources. Unlike motion blur and out-of-focus blur, the noise itself is not a spatial disruption on the image plane but it is obvious that a lower signal-to-noise ratio (SNR) has a negative effect on SfM-MVS procedures while its impact has not been fully identified. Photon shot noise and read noise are the two major noise sources where read noise includes dark current noise and readout noise. Dark current noise in visible ray region (Aasen et al., 2015) and readout noise of modern camera sensors (Boukhayma et al., 2015) contribute a small portion to total noise if the image is not photographed in an extremely low light environment. In this work, photon shot noise limited region was assumed, and therefore noise in an image is proportional to the square root of the signal (Farooque and Rohankar, 2013). Signal per pixel is proportional to illuminance

E, exposure time (shutter speed t), the square of aperture (F−number N), and pixel size. Pixel size refers to the square of the sensor size d divided by the pixel resolution p and lowering pixel resolution by pixel binning has the same effect as increasing pixel size (Zhou et al., 1997). The signal is expressed as Eq. (19) and the noise−to−signal ratio σ_{noise} which is the inverse of SNR is calculated as Eq. (20).

$$\text{Signal} \propto \frac{Etd^2}{N^2p^2} \tag{19}$$

$$\sigma_{\text{noise}} = \frac{\text{Noise}}{Signal} = Q\frac{Np}{d\sqrt{Et}} \tag{20}$$

The proportional constant Q is related to parameters such as sensor quantum efficiency and light wavelength, and because it is impractical to estimate such features, Q is determined through experimental calibration.

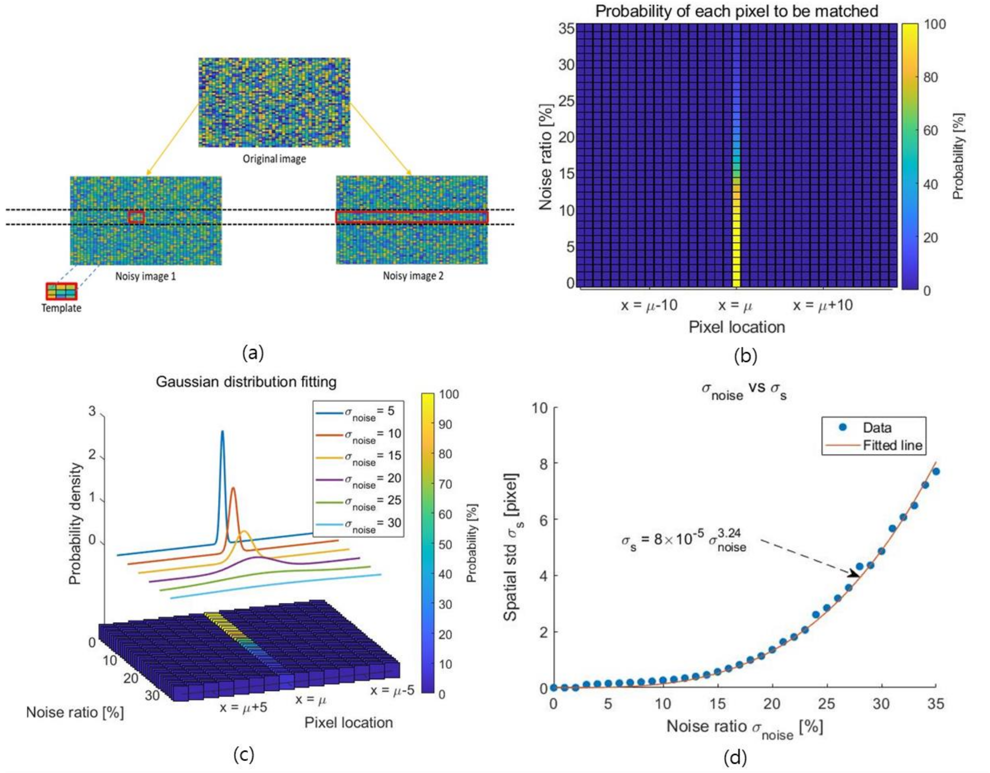


Fig. 2-7. Statistical simulation procedures: (a) Two noisy images from an identical image are generated and a template from one image is matched to another image; (b) Probability of correct matches (x= μ) is estimated which reduces according to $\boldsymbol{\sigma_{noise}}$; (c) 1D Gaussian probability density function is assumed where its mean is x= μ and its standard deviation is $\boldsymbol{\sigma_s}$; (d) Relation is fitted between $\boldsymbol{\sigma_{noise}}$ and $\boldsymbol{\sigma_s}$. The reader should refer to the colored version of this figure.

In order to identify the noise effect on dense matching, a statistical simulation was conducted. Two noisy images were generated each added by the noise amount of $\sigma_{\mathbf{noise}}$ to an identical image of a completely random shape (Fig. 2-8(a)). A template from one noisy image and its center being $\mathbf{x} = \mu$ [pixel] was matched along the x direction at another noisy image. It is assumed to be a square shape and its window size varies from a minimum of 3 pixels to a maximum of 21 pixels. Most of the SfM−MVS software adapt one of the matching costs of normalized cross correlation (NCC), the sum of squared difference (SSD), or binary matching cost, and their degree behaves similarly to each other (Scharstein, 2002). Therefore, NCC between the template and matching part was calculated (Scharstein, 2002; Furukawa and Ponce, 2009), and the pixel location with the maximum NCC score was regarded as a match point. The process was iterated 1000 times to obtain statistical convergence, and the probability of each pixel being matched was calculated (Fig. 2−8(b)). Understandably, the probability of a correct match P(x=μ), i.e. probability of maximum NCC score attained at x=μ [pixel], decreased according to σ_{noise}. Since the original image had a completely random shape, the probability of incorrect matches was distributed evenly along the x−axis. In reality, however, the image generally has features within it, probability of incorrect matches would likely be higher near the correct location. Therefore, 1D Gaussian probability density function was assumed where its mean is x=μ and the standard deviation is σ_s (Fig. 2−8(c)). σ_s is calculated as Eq. (21) and Eq. (22) by integrating the probability density function at the region of the correct location (x∈[μ−0.5, μ+0.5]).

$$\begin{aligned} \mathrm{P}(\mathrm{x} = \mu) &= \int_{\mu-0.5}^{\mu+0.5} \frac{1}{\sigma_s\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_s}\right) dx \\ &\cong \mathrm{erf}\left(\frac{0.353553}{\sigma_s}\right) \end{aligned} \tag{21}$$

$$\sigma_s = erfinv\big(P(x = \mu)\big)\ [pixel] \tag{22}$$

As seen in Fig. 2−8(d), σ_s increases according to σ_{noise} and fitted relation is shown in Eq. (23).

$$\sigma_s = 8 \times 10^{-5} \times \sigma_{noise}^{3.24}\ [pixel] \tag{23}$$

Like the PSF of blurs, the standard deviation of the 1D Gaussian probability density function would exhibit the disrupted amount of a point location due to noise during dense matching. Combining Eq. (20) and Eq. (23), the disruption effect due to noise is expressed as Eq. (24). Conducted statistical simulation cannot represent entire SfM−MVS dense matching algorithms, however, the discrepancies may reduce while calibrating Q.

$$\begin{gathered}\Delta x_{noise} \sim N\left(0{,}8 \times 10^{-5} \times \left(Q\frac{Np}{d\sqrt{Et}}\right)^{3.24}\right) \\ \Delta y_{noise} = 0\end{gathered} \tag{24}$$

2.1.3.4. $\mathbf{P}\big((\mathbf{i},\mathbf{j})\big|(\mathbf{x},\mathbf{y})\big)$ calculation

A point (x,y) in the reference pixel is disrupted due to blur and noise (note Section 2.1.2). The disruption effect of motion blur (Eq. (21)), out−of−focus blur (Eq. (18)), and noise (Eq. (24)) is integrated as Eq. (25).

$$\Delta x = \Delta x_{mb} + \Delta x_{off} + \Delta x_{noise}$$
$$\Delta y = \Delta y_{mb} + \Delta y_{off} + \Delta y_{noise} \tag{25}$$

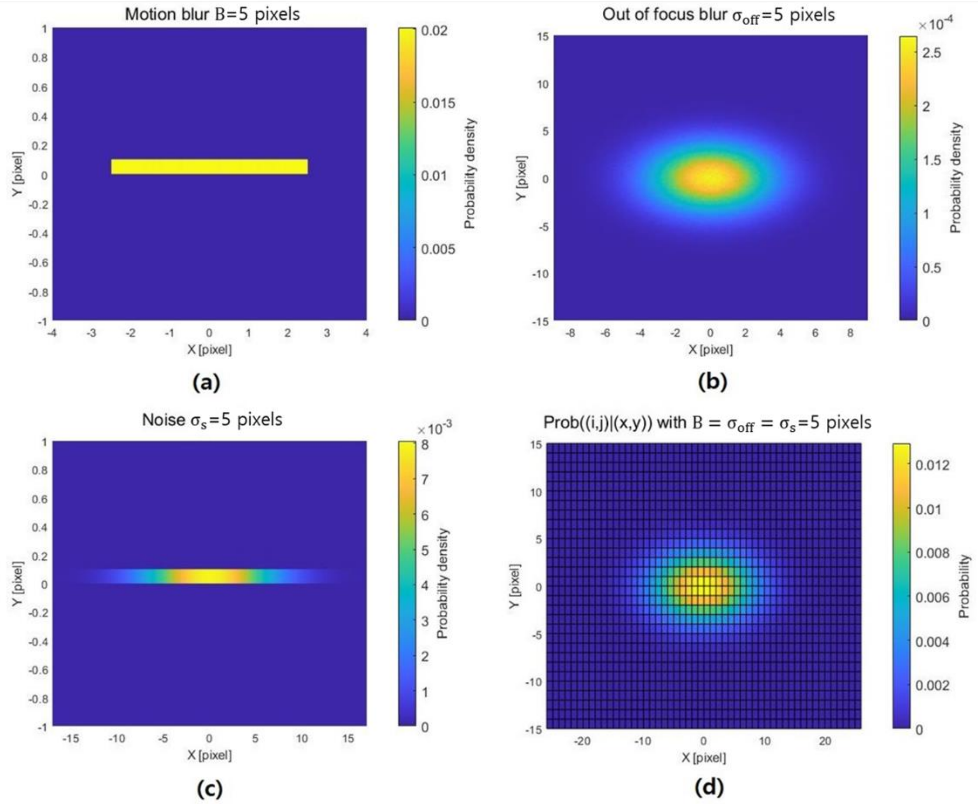


Fig. 2-8. $\boldsymbol{P}\big((\mathbf{i},\mathbf{j})\big|(\mathbf{x},\mathbf{y})\big)$ calculation using Monte Carlo simulation when $(\mathbf{x},\mathbf{y}) = (\mathbf{0},\mathbf{0})$: (a) Motion blur disruption when $\mathbf{B} = \mathbf{5\ pixels}$; (b) Out-of-focus blur disruption when $\boldsymbol{\sigma}_{\mathbf{off}} = \mathbf{5\ pixels}$; (c) Noise disruption when $\boldsymbol{\sigma}_{\mathbf{s}} = \mathbf{5\ pixels}$; (d) Integration of (a)~(c) and $\boldsymbol{P}\big((\mathbf{i},\mathbf{j})\big|(\mathbf{x},\mathbf{y})\big)$ at every pixel is calculated. The reader should refer to the colored version of this figure.

In order to calculate $P\big((i,j)\big|(x,y)\big)$, a Monte Carlo simulation was performed. A great number of points were generated and they were disrupted according to the blur and noise and corresponding disruption effect. $P\big((i,j)\big|(x,y)\big)$ was calculated by counting the number of disrupted positions $(x+\Delta x, y+\Delta y)$ within the pixel (i,j) at every pixel and by dividing it by the number of total generated points, (Fig. 2−9). Then, 2D image error could be calculated using Eq. (11) and 3D point cloud error could be predicted using Eq. (6).

2.1.4. Methodology for theoretical model validation

2.1.4.1. Experimental setup

An experimental environment was set up for the purpose of comparing theoretically predicted errors with actually observed errors. A reference target was configured which consisted of three planes and is rich in texture (Fig. 2−10). To avoid systematic errors, six GCPs were precisely placed, a sufficient number of images was utilized and cameras with their calibrations known were used. DJI’s Mavic 2 Pro and Air 2S type drones with gimbals were operated to photograph the target, and their camera specs are listed in Table 1 and Table 2 respectively.

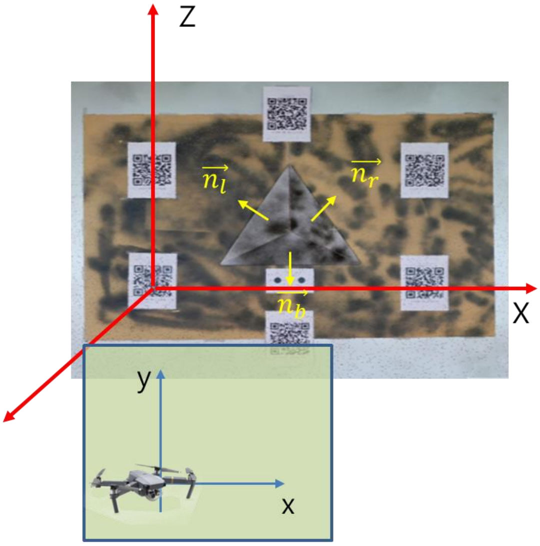


Fig. 2-9. Experimental setup and axis settings. The target consists of three planes where the normal vectors are $\overrightarrow{\boldsymbol{n_l}} = (\mathbf{0.423}, \mathbf{0.876}, \mathbf{-0.217})$, $\overrightarrow{\boldsymbol{n_r}} = (\mathbf{0.438}, \mathbf{0.869}, \mathbf{-0.229})$, $\overrightarrow{\boldsymbol{n_b}} = (\mathbf{-0.001}, \mathbf{0.857}, \mathbf{0.516})$ respectively. Angles between camera coordinate (x-y-z) and experimental environment coordinate (X-Y-Z) are $\boldsymbol{\omega} = \frac{\mathbf{3\pi}}{\mathbf{2}}$, $\boldsymbol{\phi} = \mathbf{0}$, $\boldsymbol{\kappa} = \mathbf{0}$.

Sensor	1″ CMOS Effective pixels: 20M
Lens	Field of view (FOV): about 77° Focal length: 28mm (35mm equivalent) Aperture: f/2.8-f/11 Shooting range: 1 m to ∞
ISO range	100-12800 (for still image) 100-6400 (for video)
Shutter speed	8~1/8000 s
Still image resolution	5.4K: 5472 × 3648
Video resolution	4K: 3840 × 2160 @ 24/25/30 fps 2.7K: 2688 × 1512 @ 24/25/30/48/50/60 fps FHD: 1920 × 1080 @ 24/25/30/48/50/60/120 fps

Table 1 DJI Mavic 2 Pro specification (https://www.dji.com/mavic-2)

Sensor	1″ CMOS Effective pixels: 20M
Lens	Field of view (FOV): about 88° Focal length: 22mm (35mm equivalent) Aperture: f/2.8 Shooting range: 0.6 m to ∞
ISO range	100−12800 (for still image) 100−6400 (for video)
Shutter speed	8~1/8000 s
Still image resolution	5.4K: 5472 × 3648
Video resolution	5.4K: 5472 × 3648 @ 24/25/30 fps 4K: 3840 × 2160 @ 24/25/30 fps 2.7K: 2688 × 1512 @ 24/25/30/48/50/60 fps FHD: 1920 × 1080 @ 24/25/30/48/50/60/120 fps

Table 2 DJI Air 2S specification (https://www.dji.com/air−2s/specs)

Since the camera is mounted on the UAV, Parallel axis acquisition schemes were applied (Eltner et al., 2016) and drones flew in the +X−direction with uniform speed. The target was photographed by changing the illuminance from 30 to 500 lux, distance from 2.5m to 10m, drone speed from 0 (still image) to 1.1 m/s, and camera settings which satisfy Eq. (7). Total of 261 data with different parameter combinations were obtained and Bentley's ContextCapture (https://www.bentley.com/en/products/brands/contextcapture) was used to generate point clouds from the data. Considering

2.1.4.2. Indirect measurement of point-to-point error

As pointed out by Smith et al. (2016), directly estimating point-to-point error is limited because it is difficult to identify which point from one point cloud corresponds to which point from another point cloud. In this work, considering the target consists of flat planes, the difference (note Fig. 2-11) between the predicted point-to-point error $\mathrm{RMSE}_{\mathrm{point,pre}}$ (Eq. (26)) and the predicted point-to-plane error $\mathrm{RMSE}_{\mathrm{plane,pre}}$ (Eq. (27)) could be used to correct the observed point-to plane error $\mathrm{RMSE}_{\mathrm{plane,obs}}$ into point-to-point error $\mathrm{RMSE}_{\mathrm{point,obs}}$.

$$\begin{aligned} RMSE_{point,pre} &= \sqrt{\frac{1}{N}\sum_k d_{t,k}^2} = \sqrt{\frac{1}{N}\sum_k (s\,dx_k\widehat{r_1} + s dy_k\widehat{r_2})^2} \\ &= s \times \sqrt{\frac{1}{N}\sum_k (\,dx_k\widehat{r_1} + dy_k\widehat{r_2})^2} \end{aligned} \tag{26}$$

$$\begin{aligned} RMSE_{plane,pre} &= \sqrt{\frac{1}{N}\sum_k d_{o,k}^2} = \sqrt{\frac{1}{N}\sum_k (\hat{n}\cdot\overrightarrow{d_{t,k}})^2} \\ &= s \times \sqrt{\frac{1}{N}\sum_k \{dx_k(\hat{n}\cdot\widehat{m_1}) + dy_k(\hat{n}\cdot\widehat{m_2})\}^2} \end{aligned} \tag{27}$$

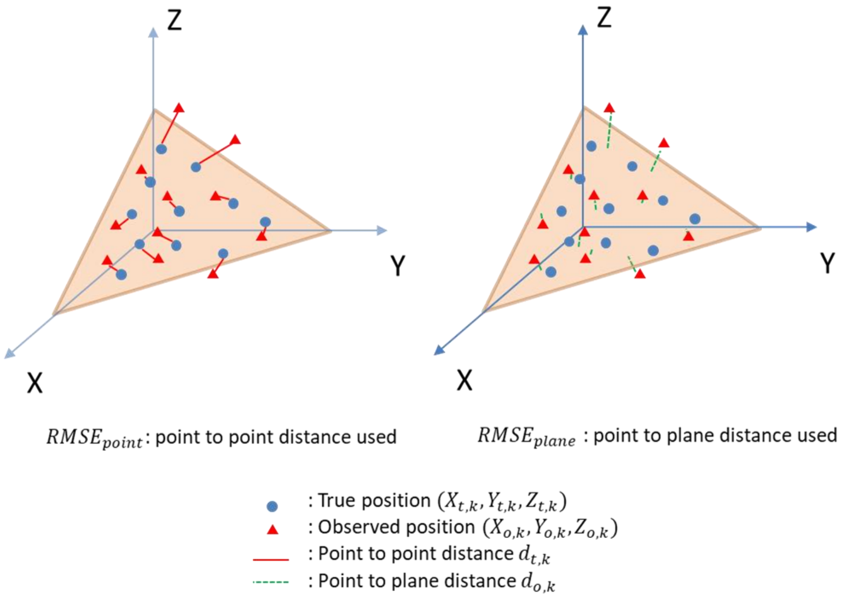


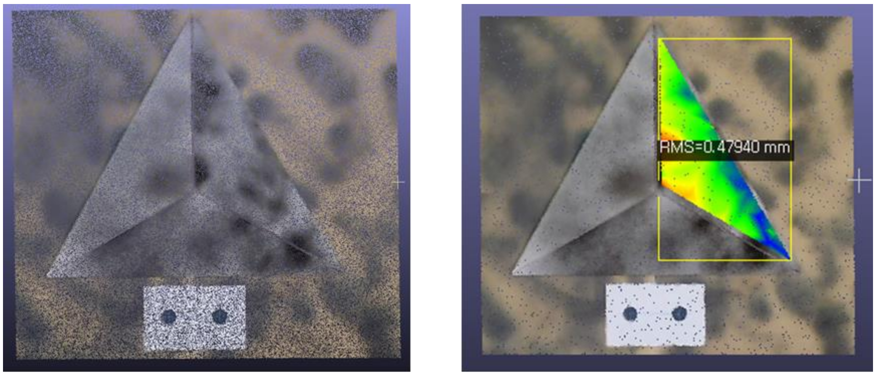
Fig. 2−10. Difference between point−to−point error $\mathbf{RMSE_{point}}$ and point to plane error $\mathbf{RMSE_{plane}}$.

Since angles between coordinates of the experimental environment were set to $\omega = \frac{3\pi}{2}$, $\phi = 0$, $\kappa = 0$, hence, $r_{11} = 1$, $r_{12} = r_{13} = r_{21} = r_{22} = 0$, $r_{23} = -1$ which simplifies Eq. (27) to Eq. (28).

$$RMSE_{plane} = s \times \sqrt{\frac{1}{N}\sum_{k}\{dx_k(n_1) + dy_k(-n_3)\}^2} \tag{28}$$

$\mathrm{RMSE_{point,obs}}$ could be indirectly measured by correcting the $\mathrm{RMSE_{plane,obs}}$ as in Eq. (29).

$$RMSE_{point,obs} = \mathrm{RMSE_{plane,obs}} \times \frac{RMSE_{point,pre}}{RMSE_{plane,pre}} \tag{29}$$



(a) (b)

Fig. 2-11. **$RMSE_{plane}$** measurement using CloudCompare: (a) Point cloud generated from ContextCapture; (b) **$RMSE_{plane}$** calculation on the best fitting plane.

Considering that all of the disruption effects had zero mean value (note Eq. (21), Eq. (18) and Eq. (24)) and neglecting the systematic errors, it was assumed that the best fitting plane of generated points is equal to the real plane (Fig. 2-12). Using the best-fitting plane as the reference has an additional benefit that it may avoid registration errors pointed out by Bakker and Lane (2017). $RMSE_{plane,obs}$ of each plane was calculated on the best-fitting plane and CloudCompare software (http://www.danielgm.net/cc/) was used to carry out the process. Observed errors $RMSE_{point,obs}$ were obtained by correcting $RMSE_{plane,obs}$ (Eq. (29)) and they were compared with the predicted errors $RMSE_{point,pre}$ in Section 3.1.

2.2. Derivation of optimum photographing settings

On the basis of the theoretical error prediction model described in Section 2.1, the error of specific parameter combinations could be predicted and compared. Two constraints, which are illumination and time (Section 2.2.1), limit the selection of parameter combinations, and this work demonstrates a procedure to select the optimum photographing settings among limited combinations (Section 2.2.2). In section 2.2.3, an underground tunnel site is introduced where the optimum photographing settings were applied and compared with LiDAR data.

2.2.1. Constraints: illumination, time constraints

2.2.1.1. Illumination

Among parameters that predetermine the error (Note Fig. 2-1), one can select controllable parameters, which are camera settings and UAV flight methods, when given uncontrollable parameter conditions, which are illumination and camera specs. Since camera specs adjustment is not available or should be avoided (Wenzel et al., 2013; Bakker and Lane, 2017; James et al., 2019), illumination is the only varying uncontrollable parameter making it a constraint of selecting controllable parameters.

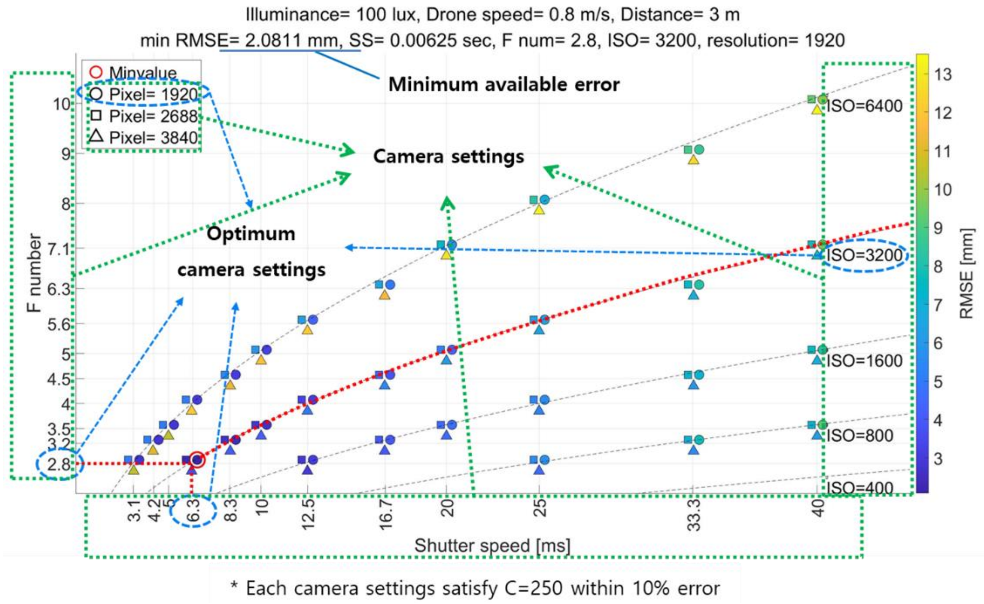


Fig. 2-12. Camera settings optimization under E-D-v condition. Optimum camera settings are the combination that makes the minimum error among all combinations which maintains proper image brightness. For example, when utilizing DJI Mavic 2 Pro drone and its camera, if the given illumination constraint is E=100 lux and the drone flew at v=0.8 m/s and D=3m, then the minimum achievable error among combinations is 2.0811 mm. And the optimum camera settings combination is to use N=2.8, t=1/160 s, S=3200, and p=1920 pixels. The reader should refer to the colored version of this figure.

When flying a UAV in a certain way under the conditions of a given illumination and camera specification used, the error created by the camera setting combination that maintains proper image brightness (satisfying Eq. (7) where C=250) varies depending on the chosen combination. As shown in Fig. 2-13, errors vary depending on the camera settings used, and the minimum error and the optimum camera settings accordingly could be analyzed. It is noteworthy that error levels can differ significantly by just changing the camera settings. The minimum achievable error and corresponding optimum camera settings differ according to the illumination and the UAV flight method.

2.2.1.2. Time

In most cases, total photographing time is limited and target regions should be photographed within a limited surveying time. UAV flight methods can be controlled by a user, however, in order to meet the time constraint, UAV speed and distance from the object are restricted.

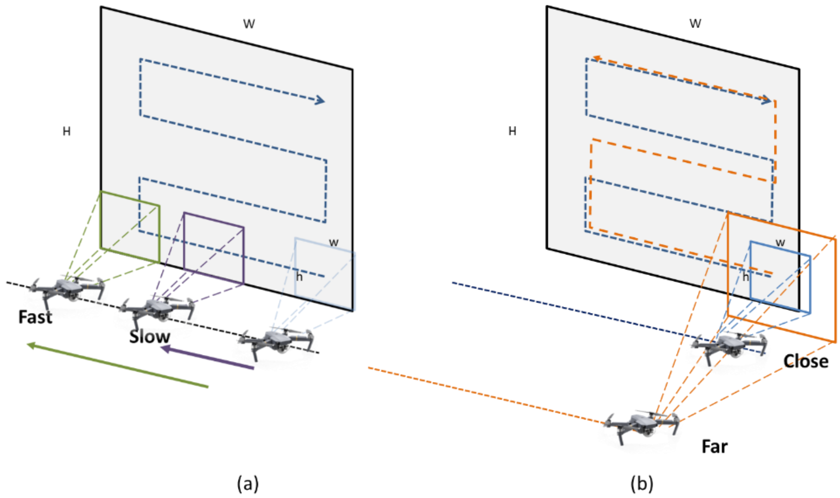


Fig. 2-13. Effect of UAV flight method on photographed area per unit time: (a) UAV speed; (b) Distance from the object.

The faster the UAV speed (Fig. 2-14(a)) and the farther the distance from the target (Fig. 2-14(b)) the larger the area to be photographed per unit time. When a UAV moves in a zigzag path parallel to the target, the faster the speed, the less time it takes to move horizontally. As the distance from the object increases, the area contained in a scene increases, so the number of horizontal movements decreases. The angle of view affects the area of a scene at a certain distance, but if a zoom lens is not used then the angle of view of the same camera is constant. Neglecting acceleration and deceleration near the boundaries and assuming the target region is large enough, then the area photographed per unit time A is proportional to UAV speed v and distance D (Eq. (36)).

$$A \propto vD \tag{36}$$

Since the minimum error in Fig. 2-13 is achievable at conditions of distance and UAV speed under illuminance constraint, it has to be compared with other D-v conditions. In Fig. 2-15, each point represents the minimum errors at different D-v conditions under a fixed illuminance constraint. By comparing points having the same photographing area per unit of time (dashed lines in Fig. 2-15), the D-v combination having the minimum error is acquired (red dots in Fig. 2-15). The trend of the optimum UAV flight methods behaves in a complicated manner because the camera settings which minimize the error at the specific UAV flight method are different from each other.

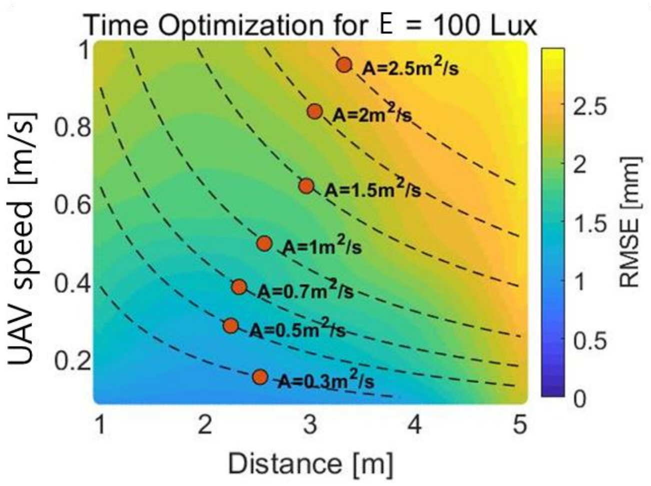


Fig. 2−14. Time optimization for E=100 lux when utilizing DJI Mavic 2 Pro. Each point of the D−v chart is the minimum achievable error under the E−D−v condition (note Fig. 2−13). Each dashed line represents points having the same A and the red dot on the dotted line represents the minimum point along the line. The reader should refer to the colored version of this figure.

2.2.2. Optimum photographing settings derivation

This work derives optimum photographing settings of camera settings (N, t, S, p) and UAV flight methods (D, v) under illumination and time constraints. Optimum photographing settings refer to the photographing settings that make the minimum error under the constraints and the derivation procedure is composed of 4 main steps.

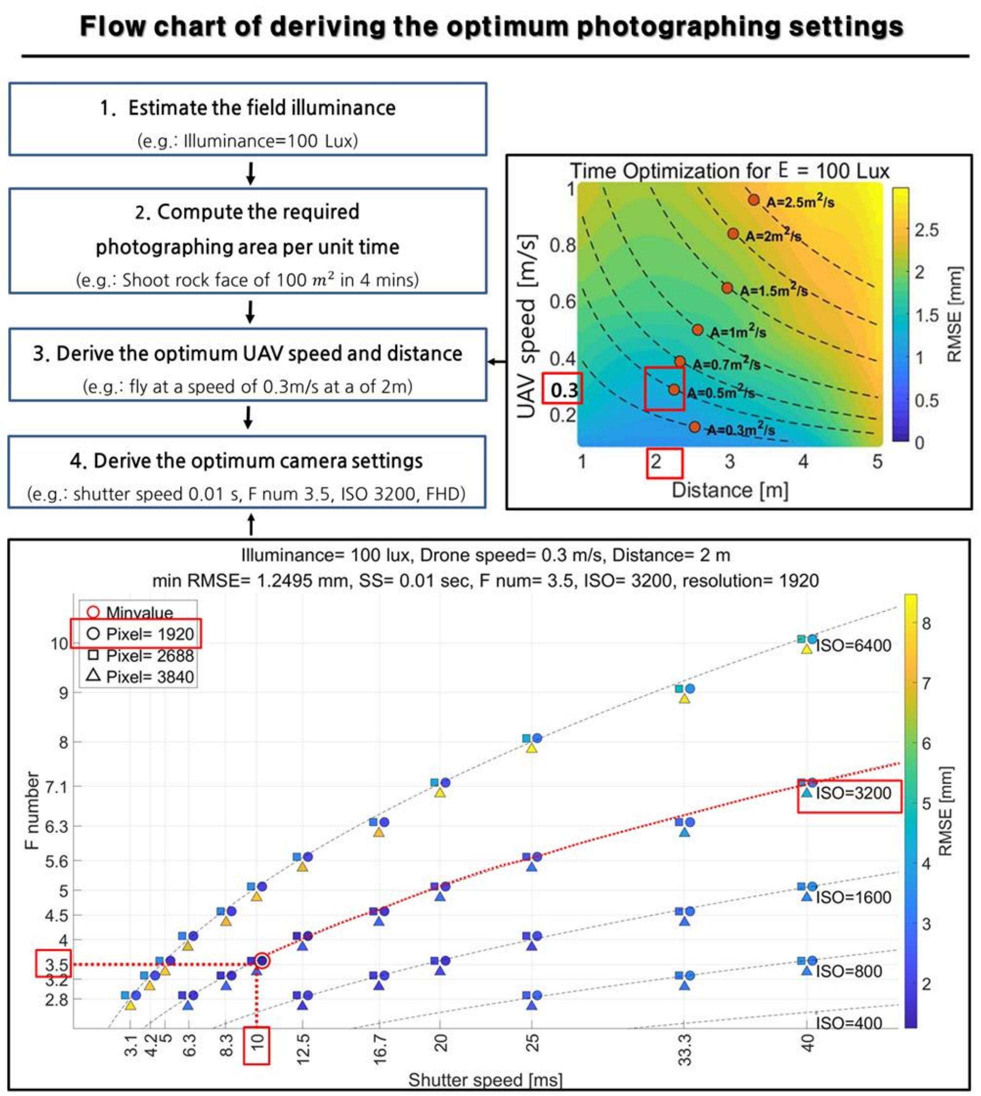


Fig. 2-15. Optimum photographing settings derivation of UAV based SfM-MVS and an example when using DJI Mavic 2 Pro drone and its camera. When measured illuminance is 100 lux and it is required to shoot the entire target area of 100 $\mathbf{m^2}$ in 4 minutes, flying a drone at v=0.3 m/s at D=2 m from the target and using camera settings set N=3.5, t=1/100 s, S=3200 and p=1920 pixels are the optimum settings. The reader should refer to the colored version of this figure.

Fig. 2-16 describes the 4 main steps: 1) measure the illuminance, 2) compute the required photographing area per unit time, 3) find the optimum UAV flight methods, and 4) find the optimum camera setting. For steps 1 and 2, it is irrelevant if the order is changed, but for steps 3 and 4, the previous two and three steps respectively must be preceded. It may be confusing that the D−v chart (note Section 2.2.1.2) used in step 3 is originating from minimum values of the camera settings chart under E−D−v conditions (note Section 2.2.1.1) used in step 4, however, since optimum camera settings are dependent on UAV flight methods, D and v should be determined (step 3) prior to the camera settings (step 4). In order to emphasize the importance of shooting at the optimum photographing settings, point cloud error using the optimum photographing settings from the flow chart is compared with arbitrary photographing settings cases in Section 3.2.1.

2.2.3. Field application of the derivation procedure

A tunnel construction site has a low−light environment and surveying time is limited due to subsequent construction processes. In order to find out the applicability of the SfM−MVS method in a such challenging environment, the SfM−MVS method using the optimum photographing settings was applied to an underground tunnel construction site in Yeoju−si, Korea and it was compared with the LiDAR method. As seen in Fig. 2−17, the target rock face area was roughly 70 m^2 and with help of field lighting, the field illuminance was about 25 lux. Since the field survey team operates the LiDAR instrument only for 5 minutes, the required time constraint was calculated as $A=70\ m^2$ per 4 mins considering some tolerance. Given $E=25$ lux and $A=0.3\ m^2/s$ constraints, optimum photographing settings were obtained for the case when using DJI Mavic 2 Pro camera. The LiDAR instrument used by the surveying team was Leica BLK360 imaging laser scanner, and its specification is listed in Table 3. Note that the ranging accuracy is the error in laser beam direction and it is different from the 3D point accuracy. The error term in this work would correspond to the 3D point accuracy of the LiDAR specification. The quality of the point clouds from both methods is compared in Section 3.2.2.

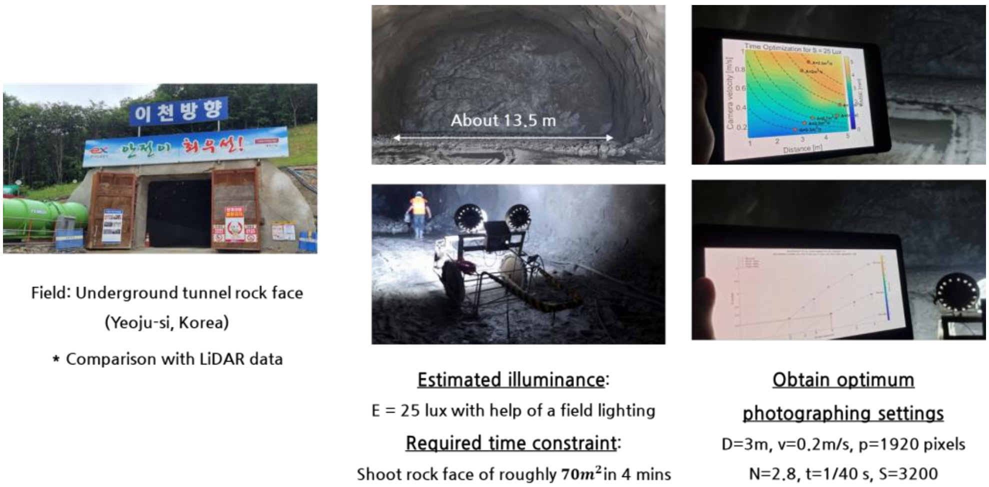


Fig. 2-16. Field application of the derivation procedure. Obtained optimum photographing settings were D=3 m, v=0.2 m/s, p=1920 pixels, N=2.8, t=1/40 s, S=3200.

Field of view	**360° (horizontal)/300° (vertical)**
Range	min 0.6 – up to 60 m
Point measurement rate	up to 360,000 pts/s
Ranging accuracy	4 mm @ 10 m / 7 mm @ 20 m
Measurement speed	Less than 3 mins for complete fulldome scan
3D point accuracy	6 mm @ 10m / 8 mm @ 20 m

Table 3 Leica BLK360 imaging laser scanner specification (https://leica-geosystems.com/products/laser-scanners/scanners/blk360)

Chapter 3. Validation and comparison

3.1. Validation of the theoretical model

3.1.1. M value calibration

In this work, the Bentley's ContextCapture software was used as the SfM−MVS software and since it is commercial software, its dense matching algorithm is kept confidential. This leaves M to be calibrated from experimental data, hence, 106 image sets photographed by DJI Mavic 2 Pro camera with different motion blur but little out−of−focus blur and noise were used to calibrate the M value

Fig. 3−1 represents the importance of considering the maximum window size effect for error prediction. If inappropriate M values are used, then the errors of images with severe motion blur are overestimated as can be seen in Fig. 3−1(a) and (c). On one hand, when lower M is used (Fig. 3−1(a)), it has the negative effects of slightly underestimating the errors of images with weak motion blur and fails to correct the overestimated errors of images with severe motion blur. On the other hand, when higher M is used (Fig. 3−1(c)), it has the drawbacks of overestimating the errors of images with severe motion blurs and has the possibility to underestimate the errors of images with much severe motion blurs.

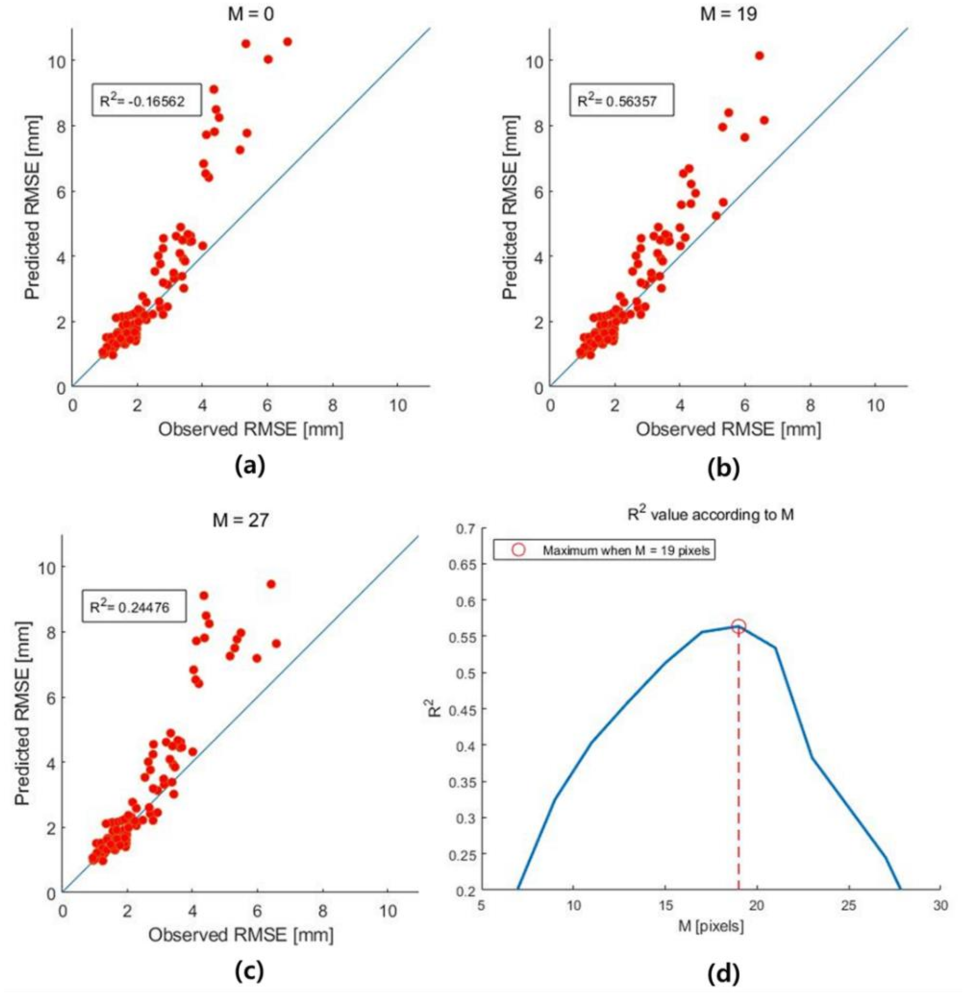


Fig. 3-1. Prediction performance according to M values: (a) M=0 pixel, equivalent to not considering the maximum window size effect; (b) M=19 pixels, best prediction performance; (c) M=27 pixels; (d) $\mathbf{R^2}$ according to M, maximum $\mathbf{R^2}$ at M=19 pixels.

The predicted errors matched the observed errors well when an appropriate M value is used and, in this work, M=19 pixels showed the best matching result between the predicted and observed errors (Fig. 3-1(b). High R^2 value was achieved in the relationship between observed and predicted errors, indicating the validation of the motion blur effect on point cloud errors.

3.1.2. Q value calibration

Since Q is relevant to the camera sensors, two different types of cameras on DJI Mavic 2 Pro and DJI Air 2S were utilized. In order to calibrate the Q value, 50 image sets for DJI Mavic 2 Pro and 36 image sets for DJI Air 2S with different noises but little blurs were used.

Fig. 3-2 and Fig. 3-3 show the prediction performance according to the Q value used for the noise-to-signal ratio σ_{noise} in Eq. (20). Higher Q exaggerates the σ_{noise} and corresponding σ_s in Eq. (23) which results in the overestimation of predicted errors, and it is completely opposite for the lower Q cases. As can be seen in Fig. 3-2(d) and Fig. 3-3(d), each camera has one appropriate Q that maximizes the prediction performance and they are respectively $\mathbf{Q}$=2.62×10^{-5} $lux^{0.5}s^{0.5}$m/pixels for DJI Mavic 2 Pro camera and $\mathbf{Q}$=2.18×10^{-5} $lux^{0.5}s^{0.5}$m/pixels for DJI Air 2S camera. In both cases, high R^2 values were achieved and it can be said that the noise effect on point cloud errors is validated.

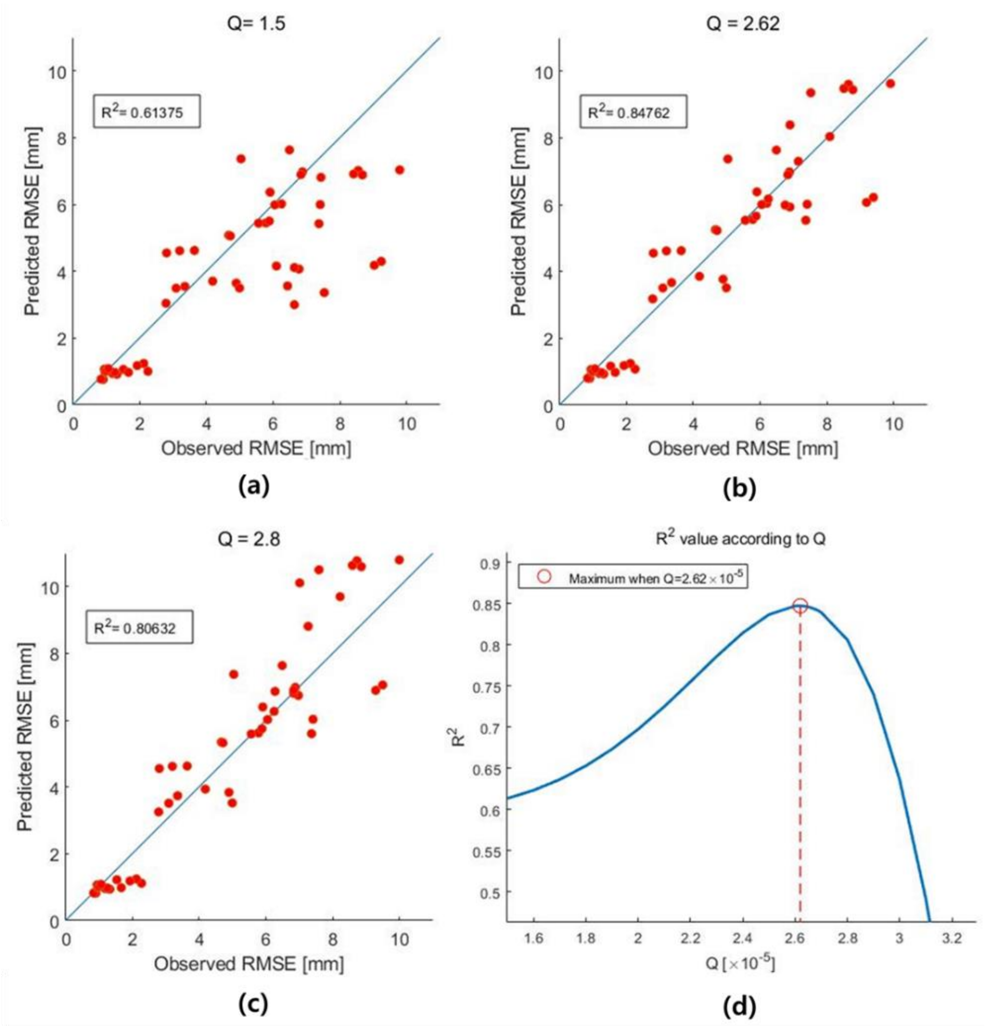


Fig. 3−2. Prediction performance according to Q values for DJI Mavic 2 Pro camera: (a) $\mathbf{Q = 1.5 \times 10^{-5}\ lux^{0.5}s^{0.5}m/pixels}$, underestimates the predicted errors; (b) $\mathbf{Q = 2.62 \times 10^{-5}\ lux^{0.5}s^{0.5}m/pixels}$, best prediction performance; (c) $\mathbf{Q = 2.8 \times 10^{-5}\ lux^{0.5}s^{0.5}m/pixels}$, overestimates the predicted errors.

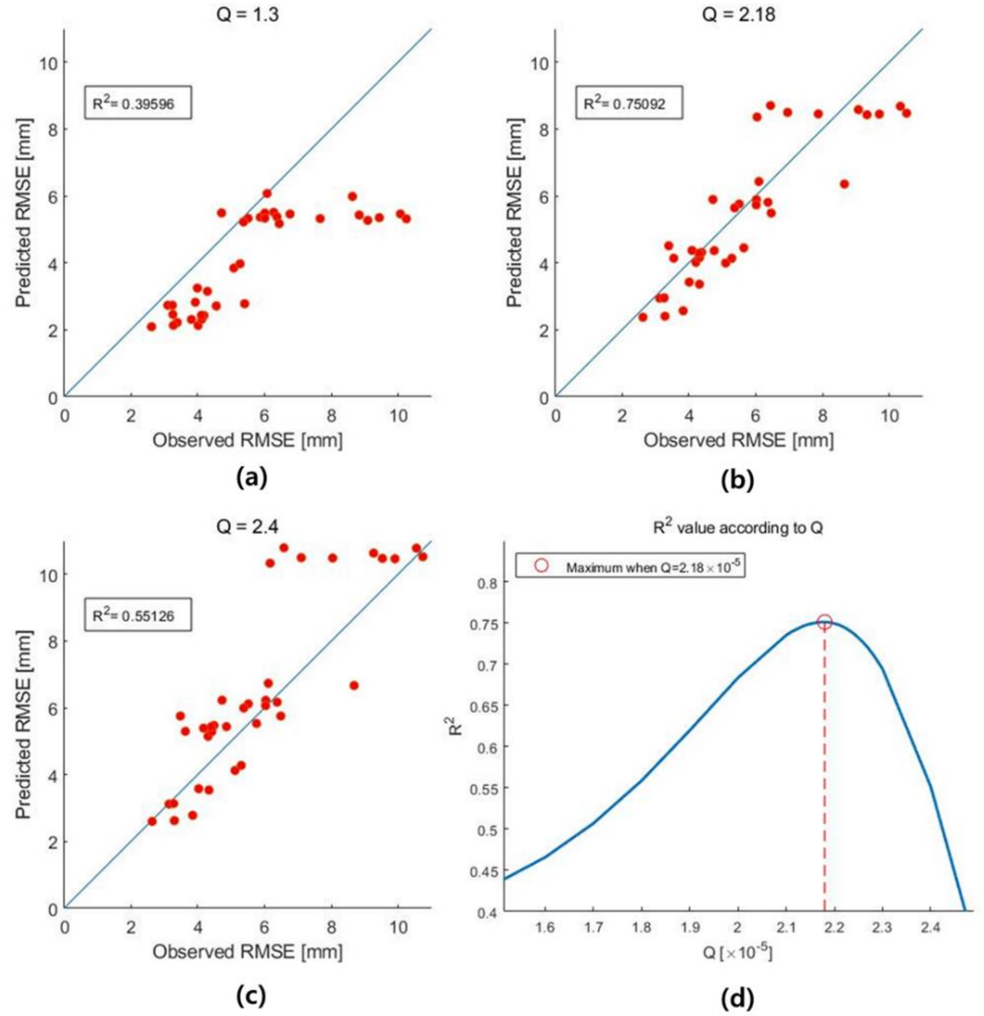


Fig. 3−3. Prediction performance according to Q values for DJI Air 2S camera: (a) $\mathbf{Q = 1.3 \times 10^{-5}\ lux^{0.5}s^{0.5}m/pixels}$, underestimates the predicted errors; (b) $\mathbf{Q = 2.18 \times 10^{-5}\ lux^{0.5}s^{0.5}m/pixels}$, best prediction performance; (c) $\mathbf{Q = 2.4 \times 10^{-5}\ lux^{0.5}s^{0.5}m/pixels}$, overestimates the predicted errors.

3.1.3. Validation result

Three experimental data were classified as outliers since they exhibited periodic noise in the image set. As seen in Fig. 3−4(a), dark image sections did not only have inappropriate image brightness but also the location of the sections was not constant within the image sets, degrading the dense matching performance. In fact, the observed errors were about 3 to 4 times higher than the expected errors (Fig. 3−4(b)). Excluding the three outliers, 258 experimental data were validated using M=19 pixels and Q values of the DJI Mavic 2 Pro camera and DJI Air 2S camera 2.62×10^{-5} $lux^{0.5}s^{0.5}m/pixels$ and 2.18×10^{-5} $lux^{0.5}s^{0.5}m/pixels$ respectively.

(a) (b)

Fig. 3−4. (a) Periodic noise in an image; (b) Periodic noise has a negative effect on error prediction.

In Fig. 3−5, observed errors were compared with predicted errors from the theoretical model. Fig. 3−5(a)~(c) show the validation result when motion blur, out of focus blur, or noise is the dominant image disruption effect respectively. Fig. 3−5(a) and Fig. 3−1(b) are the same figures since the M value is calibrated by utilizing the data which have less out of focus blur and noise. Likewise, Fig. 3−5(c) is the concatenation of Fig. 3−2(b) and Fig. 3−3(b) because the Q value is calibrated by dealing with the data having a small amount of blurs. Fig. 3−5(d) shows the comprehensive validation result and the R^2 value was about 0.83 indicating the high prediction performance of the model. It includes not only the data of Fig. 3−5(a)~(c) but also the data not dominated by one of the three disruption effects.

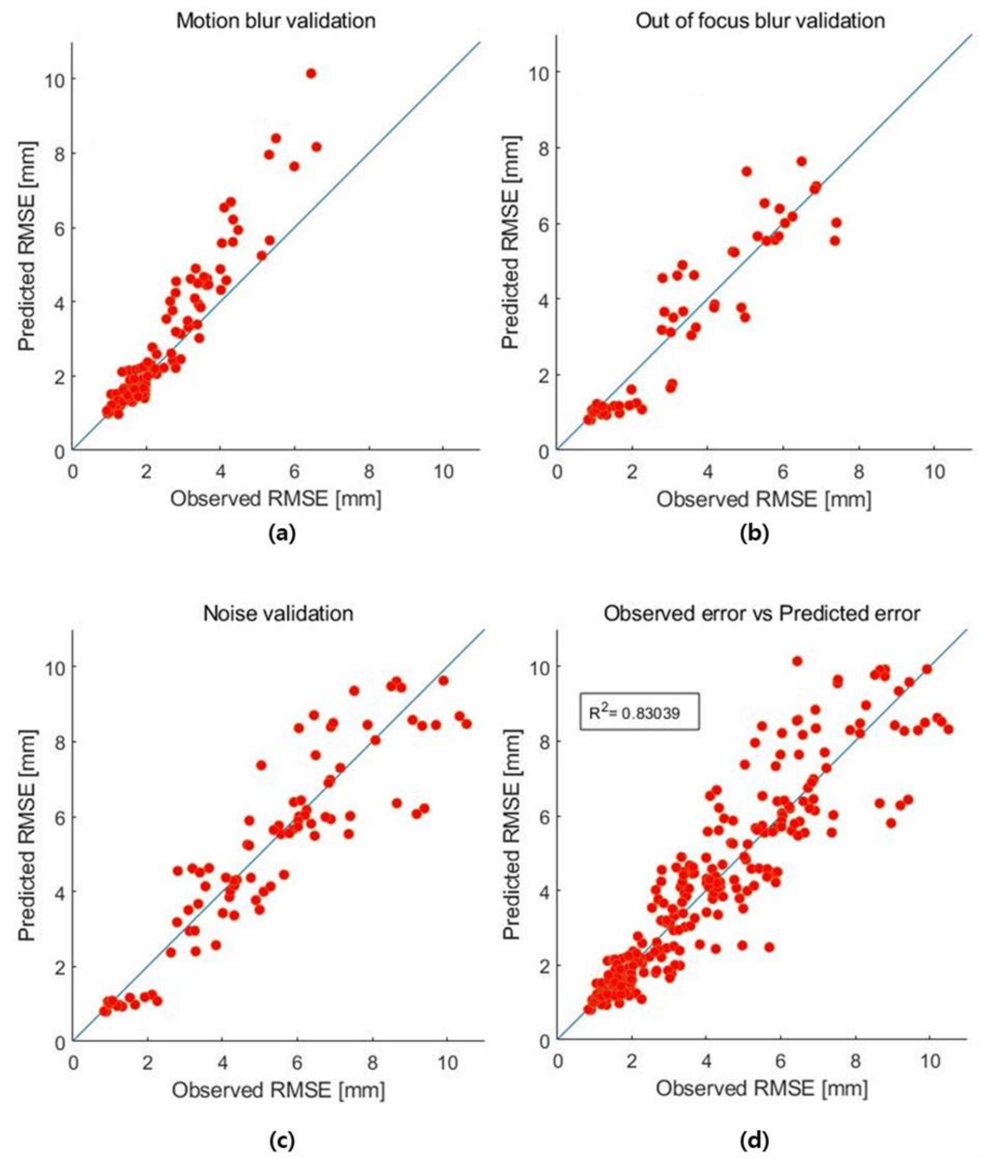


Fig. 3−5. Validation result: (a) Motion blur with 106 data; (b) Out of focus blur with 52 data; (c) Noise with 86 data; (d) Total 258 experimental data

3.2. Optimum photographing settings application

3.2.1. Importance of optimum photographing settings

The importance of the optimum photographing settings can be seen in Fig. 3−6 by comparing the case of optimum photographing settings obtained from the flowchart (Fig. 2−16) and cases of two arbitrary photographing settings. Two constraints were given as $E=100$ lux and $D \times v=1.8\,m^2/s$, and as expected, there was an extensive error level difference between the optimum case and arbitrary cases. Without additional cost or the need of changing the constraints or camera model, considerably better point cloud quality could be acquired by just changing the camera settings and UAV flight methods. Therefore, it could be said that utilizing the derived optimum photographing settings can provide significantly better point cloud quality.

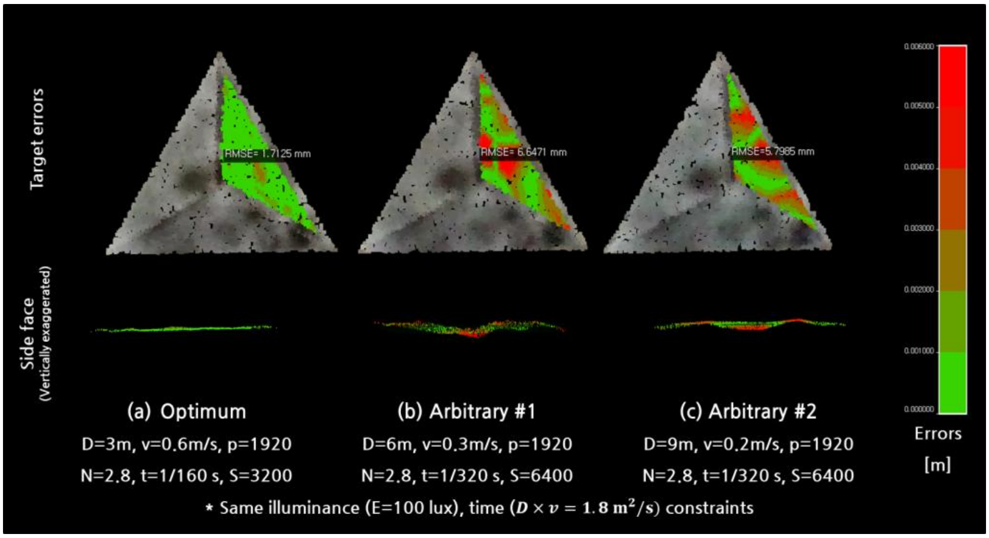


Fig. 3-6. Comparison between optimum photographing settings and arbitrary cases when using DJI Mavic 2 Pro camera under E=100 lux and $\mathbf{D} \times \mathbf{v} = \mathbf{1.8\ m^2/s}$ constraints: (a) Optimum photographing settings (D=3m, v=0.6m/s, p=1920 pixels, N=2.8, t=1/160 s, S=3200) resulted observed error level of 1.7125 mm; (b) Arbitrary photographing settings #1 (D=6m, v=0.3m/s, p=1920 pixels, N=2.8, t=1/320 s, S=6400) resulted observed error level of 6.6471 mm; (c) Arbitrary photographing settings #2 (D=9m, v=0.2m/s, p=1920 pixels, N=2.8, t=1/320 s, S=6400) resulted in observed error level of 6.6471 mm.

3.2.2. Comparison with the LiDAR

The actual surveying time for both SfM-MVS and LiDAR methods was equally 5 minutes when the parallel axis acquisition scheme (Eltner et al., 2016) for the SfM-MVS method was applied as illustrated in Fig. 3-7(c). The LiDAR-generated point cloud is shown in Fig. 3-7(a), and it had a low point resolution of roughly 1,700,000 points (equal to 150 pts/m). Due to the active nature of laser scanning (Baltasvias, 1999), intensity data were available, however, it could not express the texture of rock mass well while the texture is one of the important rock mass characteristics. Although the instrument was capable of higher point resolution, however, due to the limited time of 5 mins degraded the resolution of the LiDAR-generated point cloud. In contrast, SfM-MVS generated point cloud had a high point resolution of roughly 35,100,000 points (equal to 670 pts/m) which is about 20 times higher than the LiDAR case and it is shown in Fig. 3-7(b). It could well express the texture of the rock mass by reason of SfM-MVS is based on images while it could not obtain intensity data like LiDAR. The passive nature of images (Baltasvias, 1999) and the fact that SfM-MVS is based on images allow the SfM-MVS method to obtain high-resolution data from a wide range of areas within a short period.

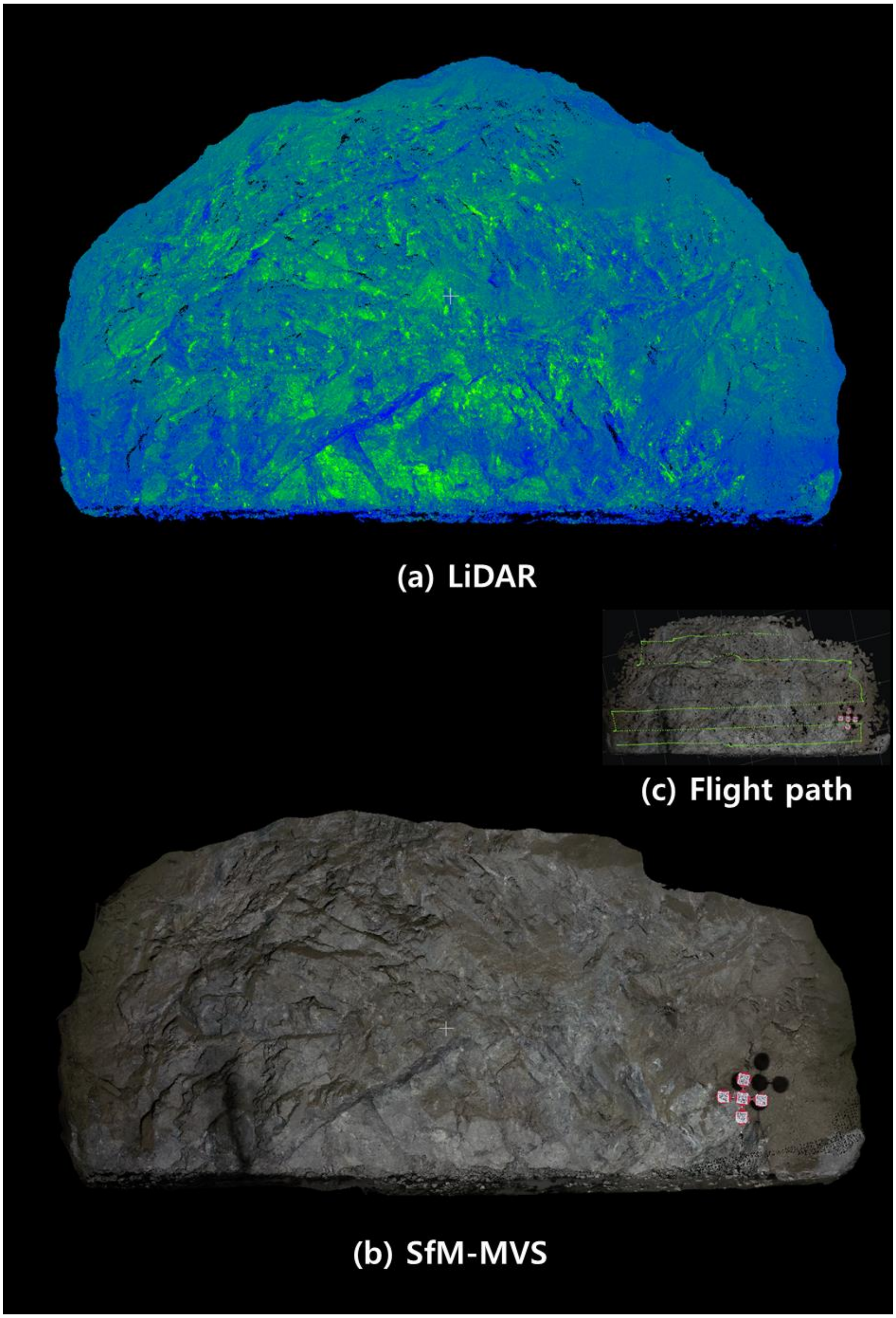


Fig. 3-7. Generated point clouds from LiDAR and SfM-MVS method: (a) LiDAR generated point cloud has low point resolution (roughly 1,700,000 points); (b) SfM-MVS generated point cloud has high point resolution (roughly 35,100,000 points); (c) UAV flight path.

Considering the target area is a tunnel construction site and the rock mass characteristics are required for stability analysis, expression capabilities of both methods for joint planes and joint traces, which are particularly important features, were qualitatively compared. Fig. 3−8(b) compares the joint plane expression capabilities. Both methods had lower point resolutions than the generally used laboratory value of 1000 (Ge et al., 2017)−10000 pts/m (Park and Song, 2013) for measuring surface roughness purposes, still, LiDAR had more concerns to distort surface roughness than SfM−MVS. As the LiDAR instrument used was a terrestrial laser scanner (TLS) type, it emits light from a fixed position, hence having shadow zones where little or no light can reach. Shadow zone did not appear in the SfM−MVS point cloud because images from various locations could be acquired due to the combination with the drone. It should be noted that dark−colored points due to occlusion are different from the shadow zone. There could be a doubt that LiDAR can overcome the shadow zone problem in the same manner, however, other problems arise when combined with UAV including worse accuracy and resolution (Smith et al., 2016). Fig. 3−8(c) compares the joint trace expression capabilities. LiDAR resulted in an overly sparse point cloud to depict the joint traces while SfM−MVS had the ability to express them.

According to the theoretical error prediction model derived in this work, SfM−MVS utilizing the optimum photographing settings would have resulted in a point cloud error of 2 mm under E=100 lux and A=0.3 m^2/s constraints while LiDAR had an accuracy of 6 mm when measured at a 10 m distance. Such outperforming is similar to what numerous studies (Smith and Vericat, 2015; Smith et al., 2016; Eltner et al., 2016) have stated. The overall performance comparison between the two methods is summarized in Table 4, and it could be stated that SfM−MVS can generate higher−quality point clouds at a lower cost than the LiDAR technique in an underground tunnel site.

	SfM−MVS	LiDAR
Instrument	DJI Mavic 2 Pro	Leica BLK 360
Capital cost	$ 1770	$ 15545
Survey time	5 mins	5 mins
Resolution	35,100,000 pts	1,700,000 pts
Accuracy	2 mm (@ $\mathrm{E} = 25\ \mathrm{lux},\ \mathrm{A} = 0.3\ m^2/s$)	6 mm (@ D=10 m)
Features	Joints expressed No shadow zones Need of drone control Search process for optimum settings required	Intensity data Shadow zones Simple scanning (∵TLS)

Table 4 Summary of performance comparison

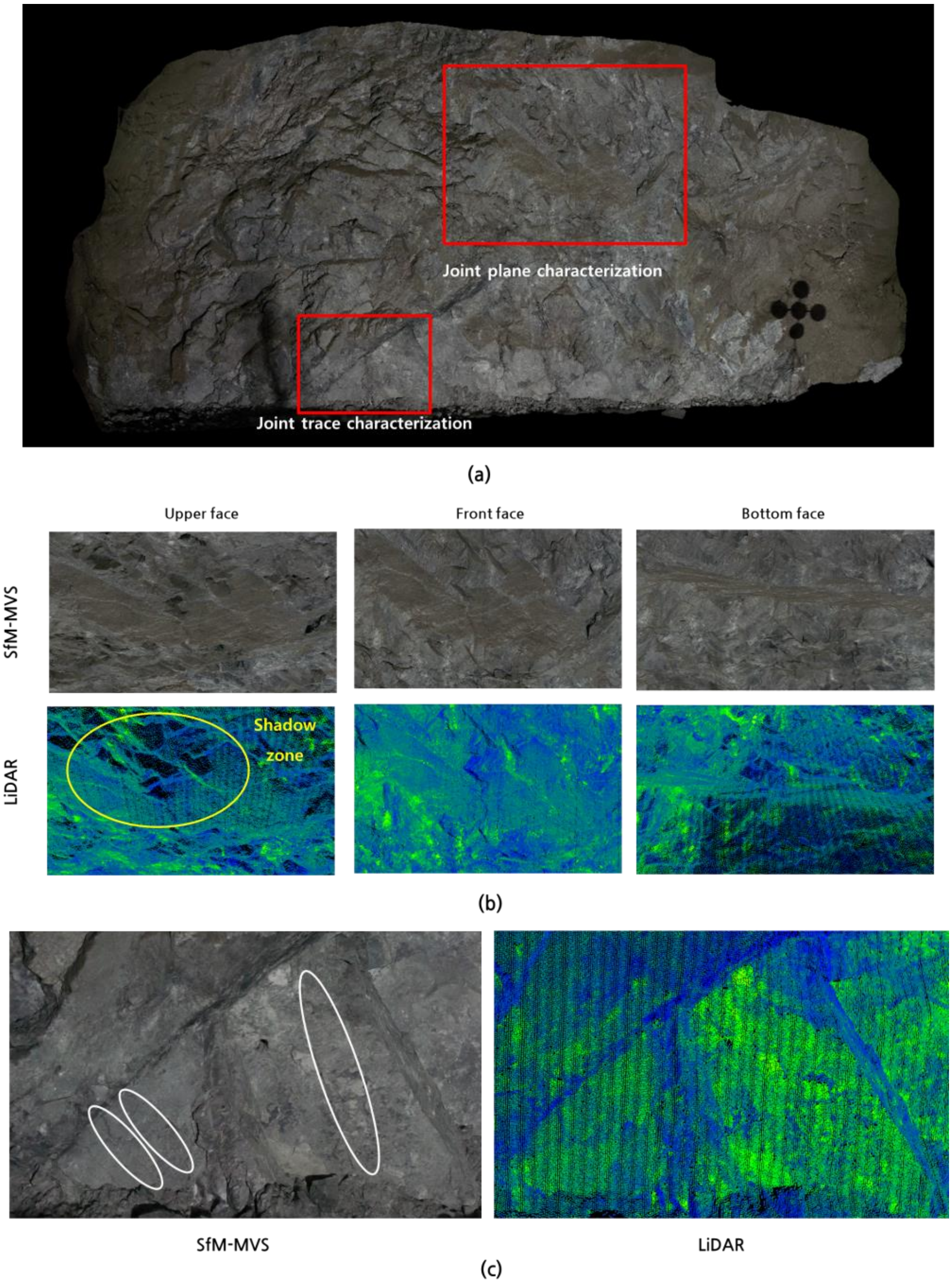


Fig. 3-8. Comparison of rock mass characteristics expression capabilities between SfM-MVS and LiDAR: (a) Location of a joint plane and joint traces used for comparison; (b) Joint plane expression capabilities comparison; (c) Joint trace expression capabilities comparison.

Chapter 4. Discussion

4.1. Implication of the error prediction model

In this work, a theoretical error prediction model was formulated and M and Q values were calibrated to precisely assess the effect of motion blur and noise respectively. When Bentley's ContextCapture was used for the SfM-MVS software, M is calibrated as 19 pixels. It is interesting because a window size larger than 19 pixels displayed similar dense matching performance as reported in Fig. 11 of Scharstein (2002) while using a larger maximum window size is more computational. Although considering the maximum window size effect enhanced the prediction performance, there still existed slight deviations of predicted errors being higher than the observed errors where error levels are significant in Fig. 3-1(b). The deviations are thought to be originating from under-observed errors due to the best-fitting plane of point clouds not being the actual reference plane (note Section 2.1.4.2). Hence, it is thought that a higher R^2 value would be achieved when $RMSE_{plane}$ is calculated based on the actual reference plane rather than the best-fitting plane while systematic errors in the coordinates should be carefully considered.

Two types of cameras, DJI Mavic 2 Pro and DJI Air 2S, were utilized for Q value calibration, and the best prediction performances were achieved when $Q=2.62\times10^{-5}\,lux^{0.5}s^{0.5}m/pixels$ and

$Q=2.18\times10^{-5}$ $lux^{0.5}s^{0.5}m/pixels$ respectively. The difference in appropriate Q value between the two cameras is noteworthy since lower Q represents higher sensor efficiency which indicates a higher signal given the same amount of light. Considering the fact that DJI Mavic 2 Pro was released in 2018 while DJI Air 2S was released in 2021, the difference indirectly implies an improvement in sensor efficiency between the time difference.

While some scholars (Aasen et al., 2015; Eltner et al., 2016) insisted sensor quality is important, some (Micheletti et al., 2015; Smith et al., 2016) argued that it has little impact on the error that consumer-grade cameras do not have any significant difference to professional cameras. Such discordance might be explained by the exponential relationship between σ_s and σ_{noise} (note Eq. (23)). On one hand, as seen in Fig. 2-8(d), at the low σ_{noise} region by the sufficient light environment, the difference in σ_{noise} has little effect on the change of σ_s. Thus, a small difference in sensor quality should not have a noticeable difference in error, i.e. consumer-grade cameras should not have a big difference from professional cameras in such circumstances. On the other hand, in the high σ_{noise} region such as photographing in a low light environment, a slight difference in σ_{noise} might result in a significant σ_s difference which directly changes the resultant error. In such cases, using higher-quality cameras or high-efficiency sensors should output better-quality point clouds.

A total of 258 among 261 experimental data were validated with a significantly high R^2 value of about 0.83 when applying M=19 pixels for Bentley's ContextCapture and $Q=2.62\times10^{-5}$ $lux^{0.5}s^{0.5}m/pixels$ for DJI Mavic 2 Pro camera data or $Q=2.18\times10^{-5}$ $lux^{0.5}s^{0.5}m/pixels$ for DJI Air 2S data. M and Q values are calibrated by utilizing the data having only motion blur and noise being the dominant disruption effect respectively. However, using the identical M and Q values for the same SfM-MVS software and camera model could also fit the data with the combined disruption effect (Fig. 3-5(d)).

Not only does the validation result of $R^2=0.83$ in Fig. 3-5(d) verify the theoretical model, but also previous studies support the model. Since the image error is amplified by scale factor s (note Eq. (6)) which is proportional to distance D or scale and inversely proportional to focal length f, the model can explain the empirical linear relation between error and scale in Smith and Vericat (2015) and empirical linear relation between error and D/f in Ohnishi et al. (2006). Wenzel et al. (2013) theoretically support the effect of pixel resolution which is considered by the model (see Section 2.1.2) and its effect is also mentioned in Lane et al., (2003), Aasen et al. (2015) and Eltner et al. (2016). The scale and pixel resolution of the image would be the dominant error sources in a bright environment where little blurs and noise exist in the image. Although there existed some dispersion from the perfect matches in the validation result, however, they are thought to be originating from factors including slightly inaccurately measured illumination, UAV speed or distance, improperly corrected $RMSE_{plane}$ and incompletely removed systematic errors. Nevertheless, the result is significantly meaningful as one can now predetermine output quality or error by just estimating only nine parameters (See Fig. 2-1) which can be practically measured.

Using lens magnification as one of the parameters was not attempted (Micheletti et al., 2015; Smith et al., 2016; Eltner et al., 2016) in this work, however, the magnification effect on the error may be identified by having full knowledge about the change in camera calibration when using zoom lenses. If magnification is considered as one of the parameters, then it could affect the optimum photographing settings, especially the optimum distance. Future work may include the impact of considering magnification as one of the parameters in error prediction and accordingly derivation of optimums photographing settings.

Most notably, this is the first study to our knowledge to propagate image error in order to predetermine point cloud error where the prediction performance has shown its superiority. The predicted error may be used for such analysis including surface topography change assessment, roughness estimation, slope stability analysis, and rock mass characterization. Especially for the topography change assessment, spatially varying minimum level of detection (minLoD) values is used as thresholds to distinguish actual topography changes from the errors (Wheaton et al., 2010). However, conventional single-point error assessment at multiple locations had to choose either the accuracy of minLoD by averaging the estimated error values or the spatial variability by using each error value as the minLoD, while this work can provide spatially varying minLoD with high accuracy. As stated by Smith et al. (2016) the difference between SfM−MVS processing software makes difference in generated point cloud errors even though the same image set has been used. Nonetheless, this work could handle the error differences by calibrating appropriate M and Q values for each software since the derived theoretical model does not have any customized features just for the ContextCapture software used in this work.

4.2. Feasibility of the derivation

The optimum photographing settings are derived under illuminance and time constraints and their importance has been verified in Section 3.2. According to the flow chart (Fig. 2-16), the optimum UAV flight method is determined first, however, in reality, UAV speed or distance may be restricted for some reason, e.g. speed regulation or altitude limit, and they would act as an upper or lower limit for UAV speed or distance. Yet the optimum D-v combination can be found within the restricted D-v zone unless there is no possible D-v combination satisfying the required A. For example, if the minimum UAV speed is restricted to 0.6 m/s, one cannot utilize the optimum UAV flight method of v=0.5 m/s and D=2.5 m when required A=1 m^2/s in Fig. 2-15. In such a case, one can alternatively utilize the second-best UAV flight method of roughly v=0.6 m/s & D=2m at a cost of slightly higher error. There may be no available UAV flight method within the restricted region, for example, not only minimum UAV speed is restricted to 0.6 m/s but also the minimum distance is restricted to 3 m in Fig. 2-15. In such case, an error increase due increase of A is inevitable but still one can use v=0.6 m/s and D=3 m instead of the original optimum UAV flight method.

In Section 3.1.3, three of the data were excluded since their image sets exhibited periodic noise and the phenomenon occurs when the sensor fails to capture the oscillating light at once due to the rolling shutter. It rarely occurs under natural lighting while it may happen under artificial lighting including LED strobe lights which are commonly used in underground mines or tunnels. The negative effect of the periodic noise has not been explored in this work, however, it can be avoided by selecting alternative photographing settings. If the images using the optimum photographing settings displayed periodic noise, a user can select different photographing settings with different shutter speeds which can remove the noise at a cost of slightly higher error. For instance, if using the optimum camera settings in Fig. 2-13 showed periodic noise in the images, then a user can check whether the camera settings like N=3.2, t=1/120 s, S=3200 or N=2.8, t=1/80s, S=1600 shows the same problem. If not, the settings can be used at a cost of slightly higher error, or if so, other settings are checked until the problem is solved.

Although the camera is considered to be mounted on a moving platform like UAV, however, this work can also predict errors when still images are used especially when utilizing UAVs is unnecessary due to the small target area. Despite the ability to predict the error, a derivation procedure of the optimum photographing settings for still imaging is beyond the scope of this study since convergent acquisition schemes might be better for still images (Eltner et al., 2016) and acceleration or deceleration between photographing locations should be considered to satisfy the time constraint. Even so, since motion blur does not appear in still images, a slower shutter speed can be utilized, and hence high SNR should be achieved even at high ISO making out-of-focus blur the most priority to be avoided, i.e. highest F−number, slowest shutter speed, and ISO maintaining appropriate image brightness are recommended.

Since this work assumed consistent parameter values within the entire image sets, the prediction performance may be limited when mixed image sets are used to generate a point cloud. For example, if the target has brightly illuminated regions and darkly illuminated regions like a shadow, one has to use two optimum photographing settings for each bright and dark region. There will be a transition zone where the settings change and prediction performance might degrade due to inconsistent disruption on image sets, e.g. one image has 5 pixels of motion blur while another image has 10 pixels. A study about orthophoto (Aasen et al., 2015) reported that details in overlapping images are mixed rather than using the best detail, similarly, error for the transition zone might be between errors of the bright region and dark region. Nonetheless, conservative choices can be made by deriving optimal photographing settings for the dark area if the settings do not severely overexpose the bright area. Optimum camera settings for a dark area might overexpose the bright area, however, since image brightness constant C (see Eq. (7)) was set to 250, a lower C value can be instead used which prevents underexposing the dark area and overexposing the bright area. Therefore, a minimum C value that does not damage the SfM−MVS matching performance and accordingly change in optimum photographing settings should be further investigated.

There have been some concerns that the quality of the SfM−MVS−generated point clouds would not be sufficient in the underground tunnel since limited lighting and surveying time are likely to disrupt photographed images, however, it turned out to be vague concerns (note Section 3.2.2). When the optimum photographing settings are utilized for SfM−MVS, it could generate a 3 times more accurate and 20 times higher resolution point cloud at a cost of 1/9 than LiDAR confirming the quality of the SfM−MVS−generated point cloud. The derivation procedure for the optimum camera settings and UAV flight method might also be used as a manual for how the workers should shoot the images or utilize the UAVs and it could be automated through app development.

Still some limitations may come up when applying UAV−derived SfM−MVS in underground spaces. For example, excessively poor illuminance conditions may be overcome by attaching lights to drones, but the illuminance in the scene should be kept homogeneous by using lighting that can illuminate a larger area than the FOV of the camera. Due to the subsequent construction process, one might not satisfy the time constraint but it may be possible to overcome by operating multiple UAVs at the same time considering the low price of the equipment.

Chapter 5. Conclusion

In this work, a theoretical error prediction model based on image error propagation was formulated for the first time and validated with a high prediction performance of R^2=0.83. The model includes the analysis of 2D image error propagation to 3D point cloud error, four image quality factors of brightness, resolution, blur, and noise, and the effect of nine practically measurable parameters on image quality factors. Based on the error prediction model, this work derives the optimum photographing settings of the UAV flight method and camera settings which minimizes the error under illuminance and time constraints. The importance of the optimum photographing settings was demonstrated by comparing the error when utilizing optimum settings and arbitrary settings. SfM−MVS utilizing the optimum photographing settings was applied and compared with LiDAR in an underground tunnel construction site with low illuminance and short surveying time constraints. It could generate a 3 times more accurate and 20 times higher resolution point cloud at a cost of 1/9 than LiDAR confirming the quality of the SfM−MVS−generated point cloud and solving the existing vague concerns about the SfM−MVS quality in underground space.

References

Aasen, H., Burkart, A., Bolten, A., Bareth, G., 2015. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. ISPRS Journal of Photogrammetry and Remote Sensing 108, 245–259.

Bakker, M., Lane, S. N., 2017. Archival photogrammetric analysis of river–floodplain systems using Structure from Motion (SfM) methods. Earth Surface Processes and Landforms 42 (8), 1274–1286.

Ballard, D. H., Brown, C. M., 1982. Computer Vision Prentice–Hall Inc. Englewood Cliffs, New Jersey, pp. 313–437.

Baltsavias, E. P., 1999. A comparison between photogrammetry and laser scanning. ISPRS Journal of photogrammetry and Remote Sensing 54 (2–3), 83–94.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded–up robust features (SURF). Computer vision and image understanding 110 (3), 346–359.

Becker, R. E., Galayda, L. J., MacLaughlin, M. M., 2018. Digital photogrammetry software comparison for rock mass characterization. In 52nd US Rock Mechanics/Geomechanics Symposium. OnePetro.

Born, M., Wolf, E., 1965. Principles of Optics, Third (revised) edition. Pergamon Press, pp. 544−555.

Boukhayma, A., Peizerat, A., Enz, C., 2015. Temporal readout noise analysis and reduction techniques for low−light CMOS image sensors. IEEE Transactions on Electron Devices 63 (1), 72−78.

Burdziakowski, P., Bobkowska, K., 2021. UAV Photogrammetry under Poor Lighting Conditions—Accuracy Considerations. Sensors 21 (10), 3531.

Dunford, R., Michel, K., Gagnage, M., Piégay, H., Trémelo, M. L., 2009. Potential and constraints of Unmanned Aerial Vehicle technology for the characterization of Mediterranean riparian forest. International Journal of Remote Sensing 30 (19), 4915−4935.

Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F., Abellán, A., 2016. Image−based surface reconstruction in geomorphometry−merits, limits and developments. Earth Surface Dynamics 4 (2), 359−389.

Farooque, M. A., Rohankar, J. S., 2013. Survey on various noises and techniques for denoising the color image. International Journal of Application or Innovation in Engineering & Management (IJAIEM) 2 (11), 217−221.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24 (6), 381–395.

Furukawa, Y., Ponce, J., 2009. Accurate, dense, and robust multiview stereopsis. IEEE transactions on pattern analysis and machine intelligence 32 (8), 1362–1376.

García–Luna, R., Senent, S., Jurado–Piña, R., Jimenez, R., 2019. Structure from Motion photogrammetry to characterize underground rock masses: Experiences from two real tunnels. Tunnelling and underground space technology 83, 262–273.

Ge, Y., Tang, H., Ez Eldin, M. A. M., Wang, L., Wu, Q., Xiong, C., 2017. Evolution process of natural rock joint roughness during direct shear tests. International Journal of Geomechanics 17 (5), E4016013.

International Organization for Standardization, 1974. Photography — General purpose photographic exposure meters (photoelectric type) — Guide to product specification (ISO Standard No. 2720:1974). https://www.iso.org/standard/7690.html. (Accessed 30 November 30, 2022)

James, M. R., Robson, S., 2012. Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application. Journal of Geophysical Research: Earth Surface 117 (F3).

James, M. R., Chandler, J. H., Eltner, A., Fraser, C., Miller, P. E., Mills, J. P., Lane, S. N., 2019. Guidelines on the use of structure-from-motion photogrammetry in geomorphic research. Earth Surface Processes and Landforms 44 (10), 2081−2084.

Javadnejad, F., Slocum, R. K., Gillins, D. T., Olsen, M. J., Parrish, C. E., 2021. Dense point cloud quality factor as proxy for accuracy assessment of image−based 3D reconstruction. Journal of Surveying Engineering 147 (1), 04020021.

KICT, 2017. Tunnelsigong jung digital mappingeul tonghan online ampanjeong gisul mit unyeongmodel gaebal choejongbogoseo [Final Report on the Development of Online Cancer Determination Technology and Operation Model through Digital Mapping in Tunnel Construction]. National Library of Korea (3), 1−578.

Lowe, D. G., 1999. Object recognition from local scale−invariant features. In Proceedings of the seventh IEEE international conference on computer vision, September, vol. 2, pp. 1150−1157.

Lane, S. N., Westaway, R. M., Murray Hicks, D., 2003. Estimation of erosion and deposition volumes in a large, gravel-bed, braided river using synoptic remote sensing. Earth surface processes and landforms: the journal of the British Geomorphological Research Group 28 (3), 249−271.

Micheletti, N., Chandler, J. H., Lane, S. N., 2015. Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smartphone. Earth Surface Processes and Landforms 40 (4), 473-486.

Ohnishi, Y., Nishiyama, S., Yano, T., Matsuyama, H., Amano, K., 2006. A study of the application of digital photogrammetry to slope monitoring systems. International Journal of Rock Mechanics and Mining Sciences 43 (5), 756-766.

Park, J. W., Song, J. J., 2013. Numerical method for the determination of contact areas of a rock joint under normal and shear loads. International Journal of Rock Mechanics and Mining Sciences 58, 8-22.

Pentland, A. P., 1987. A new sense for depth of field. IEEE transactions on pattern analysis and machine intelligence (4), 523-531.

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision 47 (1), 7-42.

Slaker, B. A., Mohamed, K. M., 2017. A practical application of photogrammetry to performing rib characterization measurements in an underground coal mine using a DSLR camera. International journal of mining science and technology 27 (1), 83-90.

Smith, M. W., Vericat, D., 2015. From experimental plots to experimental landscapes: topography, erosion and deposition in sub-humid badlands from structure-from-motion photogrammetry. *Earth Surface Processes and Landforms 40* (12), 1656–1671.

Smith, M. W., Carrivick, J. L., Quincey, D. J., 2016. Structure from motion photogrammetry in physical geography. Progress in physical geography 40 (2), 247–275.

Taylor, J., 1997. Introduction to error analysis, the study of uncertainties in physical measurements, second ed. University Science Books, New York.

Tonkin, T. N., Midgley, N. G., 2016. Ground–control networks for image based surface reconstruction: An investigation of optimum survey designs using UAV derived imagery and structure–from–motion photogrammetry. Remote Sensing 8 (9), 786.

Torr, P. H., Zisserman, A., 2000. MLESAC: A new robust estimator with application to estimating image geometry. Computer vision and image understanding 78 (1), 138–156.

Vollgger, S. A., Cruden, A. R., 2016. Mapping folds and fractures in basement and cover rocks using UAV photogrammetry, Cape Liptrap and Cape Paterson, Victoria, Australia. Journal of Structural Geology 85, 168–187.

Wenzel, K., Rothermel, M., Fritsch, D., Haala, N., 2013. Image acquisition and model selection for multi–view stereo. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci 40 (5W), 251–258.

Wheaton, J. M., Brasington, J., Darby, S. E., Sear, D. A., 2010. Accounting for uncertainty in DEMs from repeat topographic surveys: improved sediment budgets. Earth surface processes and landforms: the journal of the British Geomorphological Research Group 35 (2), 136-156.

Zhang, Y., Yue, P., Zhang, G., Guan, T., Lv, M., Zhong, D., 2019. Augmented reality mapping of rock mass discontinuities and rockfall susceptibility based on unmanned aerial vehicle photogrammetry. Remote Sensing 11 (11), 1311.

Zhou, Z., Pain, B., Fossum, E. R., 1997. Frame-transfer CMOS active pixel sensor with pixel binning. IEEE Transactions on electron devices 44 (10), 1764-1768.

Abstract in Korean

지난 10년간 structure from motion (SfM)과 multi view stereo (MVS) 기술이 측량 분야에서 고해상도와 고정밀도의 3차원 점군자료를 경제적으로 생성할 수 있고 UAV와의 결합 능력을 보여주었음에도 불구하고 아직까지 오차수준을 생성 전에 결정할 수 없다는 문제가 있다. 본 연구에선 블러와 잡음에 의한 이미지 오차의 3차원 점군자료로의 전파를 기반으로 이론적 오차예측 모델을 구성하였고, 카메라 세팅, UAV 비행방법, 카메라 사양 그리고 조도에 따른 블러와 잡음의 크기를 분석하였다. 실험을 통해 예측한 오차와 관측한 오차를 비교하였고 $R^2 = 0.83$ 라는 우수한 예측 성능을 확인하였다. 본 연구는 높은 오차 예측 성능을 기반으로 주어진 조도 및 시간 제약 조건에서 점군자료 오차를 최소화하는 카메라 세팅 및 UAV 비행 방법을 포함한 최적 촬영조건을 도출하는 방법을 제시하였다. 최적 촬영조건과 임의 촬영조건 간 오차 수준을 비교하여 카메라 세팅과 UAV 비행방법을 조정하는 것만으로도 월등히 높은 품질의 점군자료를 획득할 수 있는 것을 확인함으로써 최적 촬영조건의 중요성을 검증하였다. 본 연구에서 광량과 시간이 제한적인 여주시에 위치한 지하 터널 막장면을 대상으로 최적 촬영조건을 이용하는 SfM-MVS 기술과 light detecting and ranging (LiDAR) 기술을 비교하였다. 그 결과 SfM-MVS 기술을 이용하면 LiDAR 기술보다 3배 정확하고 20배 고해상도의 점군자료를 9배 더 경제적으로 획득할 수 있음을 확인했다.

주요어: SfM (structure from motion); MVS (multi view stereo); 오차 예측; 카메라 세팅; UAV 비행방법; 지하 디지털 조사; LiDAR (light detecting and ranging)
학번: 2021-20978