EFFECT OF KEYFRAMES EXTRACTION FROM THERMAL INFRARED VIDEO STREAM TO GENERATE DENSE POINT CLOUD OF THE BUILDING'S FACADE

S. Motayyeb [1*], F. Samadzadegan [1], F. Dadrass Javan [1,2], H.R. Hosseinpour [1]

[1] School of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran - (Soroush.motayyeb, samadz, fdadrasjavan, hosseinpoor)@ut.ac.ir
[2] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7522 NB Enschede, the Netherlands

Commission IV, WG IV/3

KEY WORDS: 3D Reconstruction, Keyframes Extraction, Standard Baseline, Degeneracy Condition, GRIC.

ABSTRACT:

Keyframes extraction is required and effective for the 3D reconstruction of objects from a thermal video sequence to increase geometric accuracy, reduce the volume of aerial triangulation calculations, and generate the dense point cloud. The primary goal and focus of this paper are to assess the effect of keyframes extraction from the thermal infrared video sequence on the geometric accuracy of the dense point cloud generated. The method of keyframes extraction of thermal infrared video presented in this paper consists of three basic steps. (A) The ability to identify and remove blur frames from non-blur frames in a sequence of recorded frames. (B) The ability to apply the standard baseline condition between sequence frames to establish the overlap condition and prevent the creation of degeneracy conditions. (C) Evaluating degeneracy conditions and keyframes extraction using Geometric Robust Information Criteria (GRIC). The performance evaluation criteria for keyframes extraction in the generation of the thermal infrared dense point cloud in this paper are to assess the increase in density of the generated three-dimensional point cloud and reduce reprojection error. Based on the results and assessments presented in this paper, using keyframes increases the density of the thermal infrared dense point cloud by about 0.03% to 0.10% of points per square meter. It reduces the reprojection error by about 0.005% of pixels (2 times).

1. INTRODUCTION

Today, 3D information obtained from close-range images that overlap has a wide range of applications, including 3D modeling of urban environments (Bakogiannis, 2020), urban mapping and planning (Peleshko, 2020), virtual reality (Caciora, 2021), change detection (Han, 2021) and damage assessment (Chowdhury, 2020), architecture (Liu, 2021), and digital tourism (Poux, 2020). Based on advancements in camera technology and image processing algorithms, the use of Unmanned Aerial Vehicles (UAVs) has piqued the interest of researchers and activists in the fields of computer vision and photogrammetry in recent years.

In this regard, UAVs have become a standard and reliable platform for 3D model data collection due to advantages such as their high maneuverability in urban environments, the ability to obtain images with high overlap and different viewing angles from a close distance to the object, and the ability to use image processing algorithms such as Structure from Motion (SfM) (Jarzabek-Rychard, 2016). However, the accuracy of the models generated by the UAV photogrammetry method is dependent on factors such as flight path design, capturing overlapped images with a standard baseline, avoiding dead areas, and having high contrast in the images (Koch, 2019; Motayyeb, 2022).

Today, in addition to traditional imaging with metric and non-metric cameras, video as a source for capturing overlapped images in photogrammetry applications has been proposed. Because video recording captures information in a stream faster than imaging; it contains a large volume of images or frames; as a result, this feature creates a very high overlap in sequence frames and can reduce possible dead areas in imaging. Furthermore, when imaging with the camera, it is possible to capture blur images due to the impulses of the UAV platform; this is although in video recording, due to the high volume of frames obtained, this problem will be overcome by extracting non-blur frames.

However, the use of video frames in the process of 3D modeling of objects is associated with imaging geometry challenges and high calculations mathematical. For example, a short distance between two sequence frames' baselines leads to degeneracy conditions, and the fundamental matrix is not generated during the modeling process; if the volume of images increases, it is difficult for the processing system to process them concurrently (Zhang, 2017). Therefore, keyframe extraction as a representative of all captured frames capable of overcoming the issues mentioned above is critical in the 3D modeling process.

As previously stated, to reconstruct the 3D model using video frames, keyframes must be extracted for accurate geometric estimation of the 3D model (Choi, 2016). The following section reviews related works in keyframes extraction from a video frame sequence. Keyframes extraction has been investigated from two perspectives: radiometric and geometric.

From a radiometric standpoint, the radiometric quality of each frame is evaluated, and low-quality or blurry frames are eliminated. Various types of studies and methodologies have been offered to determine the rate of blurred frames (Ming-Chao, 1997; Frederic, 2002; Ong, 2003). BLUR METRIC (BLuM) is one of these metrics (Crete, 2007). This method's primary objective is to blur the original image and study the behavior of nearby pixels. Using the threshold (obtained by averaging the BLuM measurement across a set of high-quality frames), the blurred frames are eliminated (Crete, 2007). An approach based on the quantification of blurriness in the image, which is an automatic processing to determine blur frames and is introduced with the abbreviation SIEDS, is one of the other criteria for checking blur frames (Sieberth, 2016). This method

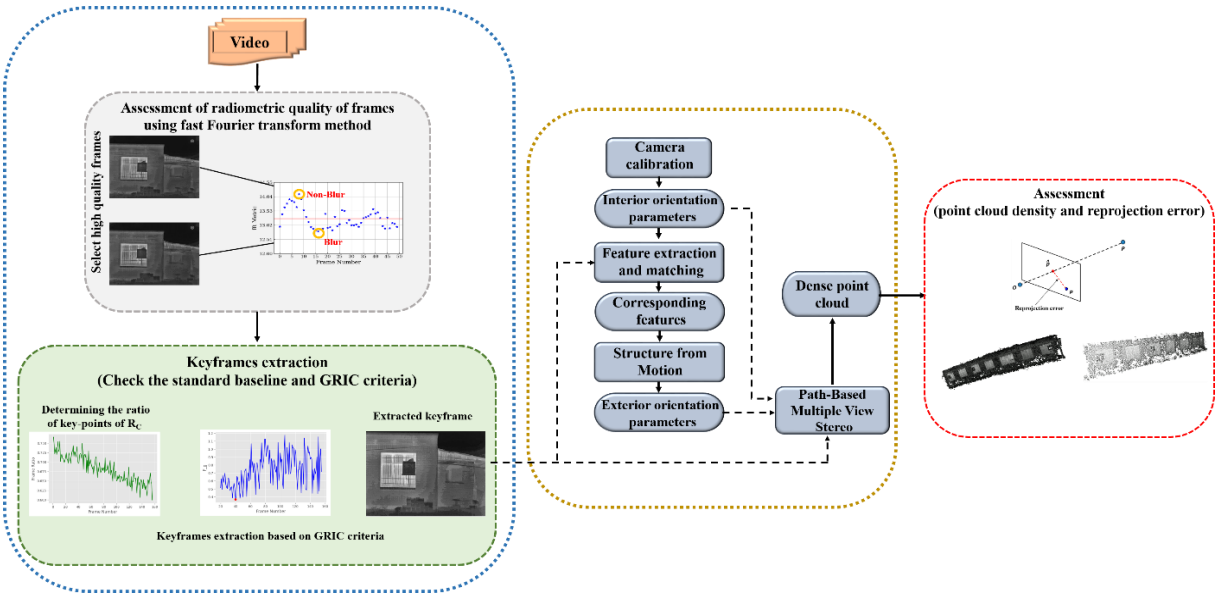


Figure 1. Flowchart of the proposed method.

digitally processes images with a certain degree of blurriness to determine the quantitative degree of blurriness of the image.

The frames with low modeling geometry are identified and removed from the sequence of frames based on the geometric aspect of keyframes extraction. As previously stated, keyframes extraction from the geometric standpoint is related to the examination of criteria such as the number of frames, the position of the extracted frames, their stability, and the baseline of the overlap between the frames to avoid degeneracy conditions and obtain Epipolar geometry suitable for 3D reconstruction. Following is a review of related work from a geometric standpoint.

Xie et al. (2015) proposed a hierarchical approach for keyframes extraction. To extract keyframes accurately, this method considers only one reference keyframe and several candidate frames adjacent to it as a sequence to calculate local correspondences. The keyframe extraction criterion is then calculated using the ratio of corresponding points and the Geometric Robust Information Criteria (GRIC) (Torr, 1998) criterion for selecting and extracting keyframes (Xie, 2015). Hossein pour et al. (2016) present a method for keyframes extraction from video sequences while minimizing reprojection error in their study. To avoid degeneracy conditions, the proposed method includes removing blur frames, applying an overlapping filter between frames, selecting an appropriate baseline between two frames, and utilizing the GRIC criterion (Hossein Pour, 2016). Following that, Choi et al. (2016) present a method for extracting frames containing helpful information from a video captured by a handheld camera in their study. In their research, they propose an approach that combines extraction criteria based on determining the appropriate baseline between frames, frame jumping for fast search in the movie, GRIC geometric information criteria to calculate frame by frame homography and fundamental matrix, and removing blur frames (Choi, 2016).

Zhang et al. (2017) present a fast approach to keyframe extraction and an optimal matching method based on geometric constraints of the path and flight direction. This method is primarily offered to improve keyframe extraction efficiency and obtain more accurate corresponding points. Therefore, the frame is extracted as the keyframe if it meets the degeneracy conditions and the corresponding ratio requirements by calculating the GRIC value and the number of corresponding points (Zhang, 2017). Dadras Javan et al. (2019) used the BluM metric as a measure to assess the radiometric quality and method (Seo, 2003; Seo, 2008) for geometric keyframe extraction from a sequence of thermal video frames in their study (Dadras Javan, 2019). Azimi et al. (2022) proposed a method for keyframes extraction. This method divides the angle between the normal to the surface and the observation vector of each point in each image into four distinct patches. The keyframe is chosen as the camera frame that covers most areas of all points (Azimi, 2022).

In this paper, the thermal infrared video sequence was used instead of the visible range to investigate the application of GRIC in keyframes extraction. The use of thermal infrared cameras is due to the limitations of visible cameras in adverse weather and at night. The main disadvantages of thermal infrared images are their low spatial resolution and geometric accuracy. In other words, because of the relatively large pixel dimensions of thermal infrared cameras and the short focal length, 3D models reconstructed from thermal infrared images have a low spatial resolution (Dadras Javan, 2019).

Therefore, one of the limitations of previous studies is determining the appropriate threshold to identify and remove blur frames, which is optimized in this paper's proposed method by transferring images to the frequency space. Also, given that the video data used in this paper were recorded from the facades of a building with very low textures, among other challenges, we can mention feature extraction and matching algorithms in thermal infrared frames. In this regard, instead of using Kanade-Lucas-Tomasi (KLT) feature tracker algorithms in keyframes extraction methods, the proposed method utilizes the Scale-Invariant Feature Transform (SIFT) algorithm and matching key points (Suhr, 2009; Kumar, 2018; Wang, 2022). The steps in the keyframe extraction method presented in this paper are as follows: (1) the ability to recognize and remove blur frames from a sequence of thermal infrared video recorded frames. (2) The ability to apply the standard baseline condition between sequence frames to establish the overlap condition and avoid degeneracy.

As the second goal of this paper is to evaluate the method of keyframe extraction, the role of this method in the generation

of the dense point cloud from building facades has been investigated further below. This goal is associated with three-dimensional modeling of buildings based on thermal infrared images to evaluate building thermal properties, heat loss, air leakage, and humidity (Kylili, 2014; Dahaghin, 2021).
This paper is organized as follows: Section 2 presents the proposed paper method, and Section 3 discusses the implementation and the results of the proposed algorithm's evaluation. Section 4 concludes with conclusions and future suggestions.

2. PROPOSED METHOD

To improve the geometric accuracy and calculation speed of the thermal infrared dense point cloud, keyframes extraction from the video is required as a pre-processing step. The main goal and focus of the paper are to investigate the effect of keyframes extraction from the thermal infrared video sequence on the geometrical accuracy of the thermal infrared dense point cloud. Figure 1 depicts the proposed method for keyframes extraction from thermal infrared video.
According to Figure 1, in the proposed paper method, a non-metric thermal infrared camera geometric calibration step is performed to reduce relative orientation error and bundle adjustment to generate the dense point cloud with optimal geometric accuracy. The quantity of blurriness, the overlap of baseline, the condition of degeneracy, and the extraction of keyframes have been examined. Finally, the reprojection error and density of the thermal infrared point cloud have been utilized to evaluate the effect of keyframes extraction from the thermal infrared video sequence. The proposed method's steps are outlined below.

2.1 Geometric Calibration of Thermal Infrared Camera

In this paper, the geometrical calibration of the thermal infrared camera using a combination of photogrammetry and computer vision methods has been examined. The calibration pattern is a rectangular plate with hollow circles. Using the calibration pattern with circular targets and extracting the two-dimensional coordinates of the center of the circles to relate with the ground space and estimate the calibration parameters yields acceptable results (Usamentiaga, 2017). Furthermore, because of the flexible geometry of the circle, the optimal ellipse can be identified in the images captured from the calibration pattern (Datta, 2009; Usamentiaga, 2017). The calibration pattern used in this paper is shown in Figure 2.

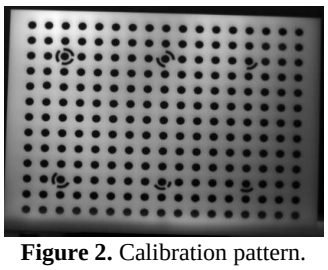


Figure 2. Calibration pattern.

Because of the spatial resolution and low contrast of thermal infrared cameras, circular targets are captured as ellipses in the image; therefore, the Hough Transformation (Chia, 2007) algorithm was used to fit and extract the exact two-dimensional coordinates of the focal center of ellipse targets in the image space. Finally, using the geometric calibration mathematical model, the association between two-dimensional and three-dimensional space is established, and the geometric calibration parameters are estimated. The equations of the collinearity condition are defined by equations (1) and (2).

$$x_a = x_p - c\frac{r_{11}(X_A - X_O) + r_{21}(Y_A - Y_O) + r_{31}(Z_A - Z_O)}{r_{13}(X_A - X_O) + r_{23}(Y_A - Y_O) + r_{33}(Z_A - Z_O)} \quad (1)$$

$$y_a = y_p - c\frac{r_{12}(X_A - X_O) + r_{22}(Y_A - Y_O) + r_{32}(Z_A - Z_O)}{r_{13}(X_A - X_O) + r_{23}(Y_A - Y_O) + r_{33}(Z_A - Z_O)} \quad (2)$$

Where c = principal distance
r = rotation matrix elements
x_a, y_a = image coordinates
x_p, y_p = principal point coordinates
X_0, Y_0, Z_0 = coordinates of projection centre
X, Y, Z = object coordinates

Additionally, lens distortion parameters are iteratively calculated using Brown's equations (Brown, 1971). Equations (3) and (4) define Brown's equations.

$$x' = x\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + p_2\left(r^2 + 2x^2\right) + 2p_1 xy\right) \quad (3)$$

$$y' = y\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + p_1\left(r^2 + 2y^2\right) + 2p_2 xy\right) \quad (4)$$

Where x', y' = corrected image coordinates
k = radial distortion coefficients
p = tangential distortion coefficients

2.2 Keyframes extraction

2.2.1. Evaluating the Radiometric Quality and Removing the Blur Frame

The presence of low radiometric quality frames, which appear as image motion in the frames, is one of the significant limitations in the processing of video frames (Cai, 2009). Because keyframe extraction is considered an essential step in the process of 3D reconstruction from video, the idea of identifying and removing motion and blur frames is proposed (Rashidi, 2013). Changes in the intensity of pixels along the edges have been extensively studied to quantify the blurriness effect (Marziliano, 2002; Yun-Chung, 2004; Varadarajan, 2008).
The Fast Fourier Transform (FFT) metric was used in this paper to identify and remove blur frames. This metric expresses the radiometric quality of an image based on its blurriness in frequency space (De, 2013; Pagaduan, 2021). The magnitude spectrum image of the FFT is frequently displayed to assess the geometric and radiometric quality of the frames because it contains more information about the geometric structure and radiometric quality of the image in spatial space (Gonzalez, Woods, 2002). FFT mathematical equations are expressed in equations (5) and (6) (Abdel-Qader, 2003).

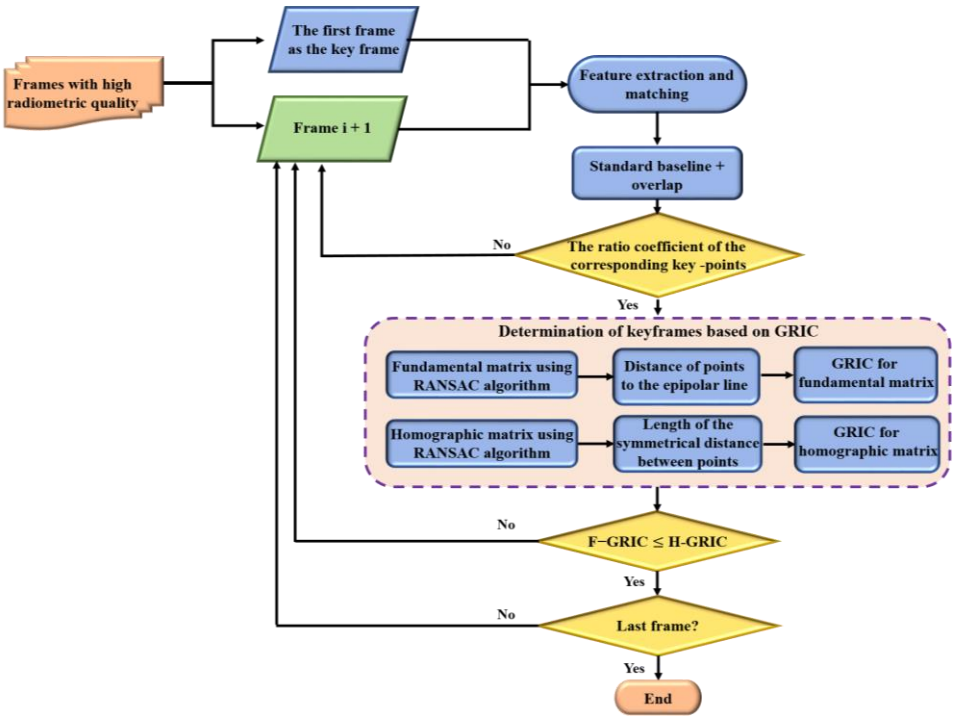


Figure 3. Flowchart of keyframes extraction with optimal geometry.

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-2\pi j (xu/M + yv/N)} \quad (5)$$

$$F(x,y) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(u,v) e^{-2\pi j (xu/M + yv/N)} \quad (6)$$

Where f (x, y) = image in the spatial space
f (u, v) = image in the frequency space
(x, y) = pixels in spatial space
(u, v) = pixels in frequency space
M×N = image dimensions

The proposed method for determining the radiometric quality of frames consists of three steps: (a) removing image noise to prevent frame blur detection, (b) determining the optimal threshold value for image blur detection (c) based on the optimal threshold value, classifying the frames into the blur and non-blur frames. Because thermal infrared images contain noise by definition, removing the noisy frames is a necessary step.

In this paper, to remove noise, the bilateral filter has been used (Tomasi, 1998; Paris, 2009). This image filter preserves the main edges of the image, and its output is by preserving the edges and reducing noise. The FFT algorithm calculates image frequencies at various points and decides on image blurriness based on the frequency level. As a result, image blur or non-blur quality is measured using high and low-frequency values. If the values with a low frequency have a high number, the image is considered blurred, and vice versa (Abdel-Qader, 2003).

Following the conversion of the frames to the frequency domain and the generation of the magnitude spectrum image, the optimal threshold value is determined using high-frequency values in the form of 50-frame intervals. The average magnitude spectrum is calculated to identify and remove blur frames in the desired intervals. As a result, frames with average magnitude spectrum values less than the threshold are identified as blurred and removed from the keyframes extraction process to improve geometric and radiometric accuracy.

2.2.2. Keyframes Extraction with Optimal Geometry

Keyframes extraction with optimal geometry is a method for extracting frames containing acceptable geometric information for 3D reconstruction from high radiometric quality (non-blur) video frames to improve geometric accuracy and reduce calculation volume in the 3D reconstruction process. Figure 3 illustrates the process of extracting keyframes with optimal geometry. According to Figure 3, the method of keyframes extraction with optimal geometry includes checking the standard baseline between sequence frames by evaluating the appropriate overlap between extracted features in sequence frames and the GRIC criterion to obtain frames with optimal geometry during the 3D reconstruction process.

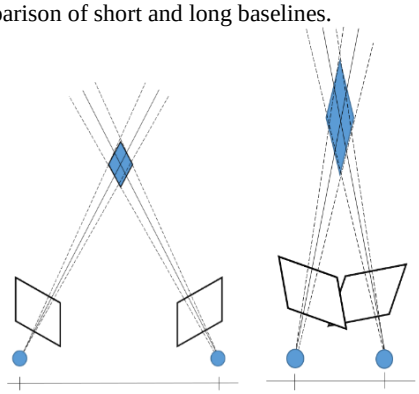
2.2.3. Key Points Extraction and Matching

The SIFT algorithm is used to extract features in this paper (Lowe, 2004). The descriptors of key points are then matched using the kd-tree method of the Approximate Nearest Neighbors (ANN) algorithm for each pair of frames to perform image matching (Arya, 1998). To match the key points of two frames *I* and *J*, a kd-tree is constructed from the feature descriptors of frame *J*. For each feature in the frame *I*, the kd-tree is used to locating the nearest neighbor in frame *J*. Using ANN's priority search method, each search is restricted to visiting no more than 200 trees to increase efficiency (Snavely, 2008). Instead of classifying false matches based on the distance to the nearest neighbor, the ratio test

described by (Lowe, 2004) was used in this paper. The two nearest neighbors in frame *J* with distances of d_1 and d_2 are found for the feature descriptor in frame I. If the ratio of d1 to d2 is less than 0.6, the correspondences are matches.

2.2.4. Standard Baseline between Sequence Frames

One of the most critical steps in the 3D reconstruction process is determining the amount of overlap and the standard baseline of sequence frames. Furthermore, the standard baseline between sequence frames should be sufficient to reduce the uncertainty of the depth calculation resulting from the triangulation method of the corresponding features in the 3D reconstruction process. Figure 4 depicts the comparison of short and long baselines.



A. long B. short

Figure 4. Standard base length.

Figure 4 illustrates that a short standard baseline increases measurement error compared to a long baseline (Ahmed, 2010; Choi, 2016). In this paper, the ratio coefficient of the corresponding key points between the two frames was used to evaluate the standard baseline between the frames to keyframes extraction, according to equation (7).

$$R_c = \frac{T_c}{T_f} \quad (7)$$

Where R_C = ratio of corresponding key points
T_C = number of corresponding key points
T_f = total number of key points extracted

The camera's movement is inversely proportional to the numerical value of R_C. As a result, in the first few frames where the camera is fixed or moves slightly, this numerical value is close to one. As a result of moving the camera, the numerical value of the ratio coefficient decreases; therefore, this criterion is used as a suitable solution to estimate the camera movement to establish an appropriate standard baseline between two sequence frames. Finally, to determine the search range for keyframe extraction, two thresholds, maximum and minimum, must be chosen. The minimum and maximum permissible thresholds in this paper are 0.6 and 0.8. Therefore, frames with a standard baseline are permitted if the numerical value of the ratio factor R_C falls between these two thresholds.

2.2.5. Keyframes Extraction Based on the Prevention of Degeneracy Conditions

The fundamental matrix is used to investigate the overall structure of the camera in various locations, as well as the connection of the corresponding features between two frames. However, in the case of degeneracy, estimating the position of the camera is impossible. In degeneracy conditions, two essential modes are motion degeneracy and structure degeneracy (Torr et al., 1999). In the case of motion degeneracy, the epipolar geometry is not established if the camera rotates around its axis without transmission. However, the camera's homography matrix can be calculated using known control points in three-dimensional space. Structure degeneracy occurs when all three-dimensional points of an object are placed on a flat plane. In this case, it is impossible to estimate the fundamental matrix using the corresponding features, and the epipolar geometry is not established, as in the case of motion degeneracy.
In degeneracy modes, the homography matrix is used to match frame pairs. As a result, using the modes above, a comparison between homography and fundamental matrices is made. Finally, the GRIC optimal geometric information criterion is used to compare two homography and fundamental matrices. The desired value is calculated by adding the two optimal fit components and the saving model. The optimal GRIC criterion is defined Using equation (8).

$$GRIC = \sum \rho\left(e_i^2\right) + \left(\lambda_1 nd + \lambda_2 k\right) \quad (8)$$

Where n = number of corresponding extracted features
e_i = vector of residuals
r = dimension of the measurement data
k = motion model parameters

In equation (8), the parameter *n* expresses the number of corresponding extracted features in solving the fundamental matrix and homography, e_i is the vector of residuals, the standard deviation of the measurement of points, and *r* is the dimension of the measurement data (for two frames, *r*=4, which is equivalent to the coordinates of the corresponding points in the two frames), and *k* is equal to the motion model parameters for the homography and fundamental matrices (for example, the number 8 for homography and 7 for fundamental) and the dimensions of the model structure (2 for homography matrix and 3 for fundamental). Based on equation (9), the residual vector is considered while evaluating the optimal fit component.

$$\rho\left(e_i^2\right) = \min\left(\frac{e_i^2}{\sigma^2}, \lambda_3(r-d)\right) \quad (9)$$

Where σ^2 = variance of the measurement error
d = model structure's dimensions

In equation (9), σ^2 represents the variance of the measurement error in calculating the extracted features. Residual values are computed using the symmetric transfer error obtained from the estimation of the homography matrix by the RANSAC algorithm (Fischler, 1981). In addition, the number of residuals is estimated using the Simpson error coming from the RANSAC algorithm's estimation of the fundamental matrix. Regarding this, the parameter controlling the error value of the residuals (λ_3) is used to manage the high values of the estimated error of the homography and fundamental matrices, which is equivalent

to 2 in this paper. In Table 1, the needed optimal geometric parameters for computing the GRIC criterion in two homography and fundamental models are compared (Torr, 1999).

Description	General model	Degeneracy conditions
Parameter	$F=\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$	$H=\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$
Constraint	$x^T Fx = 0$	$x^T = Hx$
k	7	4
d	3	2
Model	Fundamental matrix	Homography matrix

Table 1. Comparison of criteria parameters of GRIC.

The saving model component $(\lambda_1 d + \lambda_2 k)$ has two parts: the structural component $\lambda_1 nd$ with value $\lambda_1 = \ln(r)$ and the model component $\lambda_2 k$ with the value $\lambda_2 = \ln(rn)$. The numerical value of the GRIC criterion for the fundamental matrix is continuously computed to be less than that of the homography matrix, regardless of the optimal fit component in equation (7). If the residuals in the fundamental matrix have high values, degeneracy requirements exist, and the homography matrix must be employed. Finally, using the condition provided by equation (10), a keyframe will be extracted that has the lowest numerical value of the GRIC criterion of the fundamental matrix model between two frames compared to the homography matrix model.

$$f_G(i,j) = \frac{GRIC_F(i,j) - GRIC_H(i,j)}{GRIC_H(i,j)} \quad (10)$$

2.3 Generation of Thermal Infrared Point Cloud

In this paper, keyframes extracted from thermal infrared video and photogrammetry and computer vision techniques such as SfM and Multi-view stereo (MVS) were used to generate a dense point cloud (Ullman, 1979; Furukawa, 2010). Following the extraction of keyframes from the thermal infrared video sequence, the SIFT technique extracts the corresponding features between the keyframes. In this regard, the fundamental matrix can be computed by using the corresponding key features between pairs of frames (Longuet-Higgins, 1981). The candidate fundamental matrix is then evaluated using the RANSAC algorithm, and any incorrect features are removed. The SfM algorithm then converts the two-dimensional coordinates of the corresponding features of keyframe pairs into three-dimensional coordinates by using the correct corresponding key features and the bundle adjustment. Then sparse thermal infrared point cloud is generated. The MVS algorithm is then applied to increase the density of the sparse point cloud. The implementation and evaluation of the paper's results are presented in the following sections.

3. EXPERIMENTAL RESULTS

In this paper, two thermal infrared video data sets were recorded from the facade of the patriarchal palace in Aliabad village, Aradan-Garmsar city, Semnan province, to implement and assess the proposed method. The study area is located at longitude 52.3034 and latitude 35.1600. The flight path in this paper is designed to collect data from the building's facade vertically at a distance of 11 meters and a flight altitude of 1.70 meters. The data were collected from two different facades of the building with the same flight parameters settings during the early winter season, early night hours, and the same weather conditions. The second data is used to evaluate the performance of the proposed method, and it was recorded by a UAV under the same conditions as the first data (test). A vertical flight UAV with a low flight altitude and an MC1-640s thermal infrared camera produced by KeiiElectro Optics Technology with a frame rate of 30 frames per second was also used to collect data. Tables 2 and 3 contain more detailed information about the technical specifications of the thermal infrared camera and UAV used in the paper.

Parameters	Values	
Focal length	25 mm	
Image dimensions	640 × 480	
Sensor dimensions	17 µm	
Thermal sensitivity	0.03 – 30 °C	

Table 2. Technical specifications of the thermal infrared camera.

Parameters	Values	
Flight duration	30-40 min	
Flight altitude	300 m	
Maximum weight	6.5 kg	
Dimensions	0.9 m	

Table 3. UAV technical specifications.

3.1 Camera Calibration

In this paper, interior orientation parameters and lens distortions are estimated as input to the point cloud generation algorithm to improve the geometric accuracy of dense point cloud generated using SfM and MVS algorithms. Following the extraction of the focal center of ellipse targets in the image's two-dimensional space using the Hough transform algorithm, the connection between the two-dimensional and three-dimensional space is established using collinearity condition equations, and the interior orientation parameters are calculated. Figure 5 depicts the results of ellipse target focal center extraction. The lens distortions are also estimated iteratively using Brown's equations.

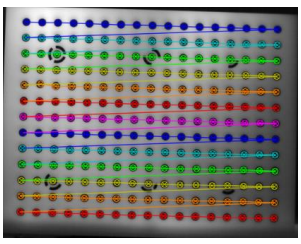


Figure 5. Extracting the focal center of ellipse targets.

Figure 5 depicts the results of extracting the focal center of 221 ellipse targets using the Hough transform algorithm. Table 4 also contains the numerical results of the interior orientation parameters and lens distortions estimated using collinearity condition equations of geometric calibration based on pixels for the thermal infrared camera used in this

paper. Table 4 displays the values and average standard deviation of each calibration parameter for 13 images captured using the proposed method from the calibration pattern based on the pixel. The c_x and c_y parameters in Table 4 represent the focal length along the x and y axes, the x_p and y_p parameters are the principal point coordinates, the k_1, k_2, and k_3 parameters are the lens's radial distortions, and the p_1 and p_2 parameters are the tangential distortion coefficients

C_x		C_y	
Values (pixel)	**SD[1] (pixel)**	**Values (pixel)**	**SD (pixel)**
1508.67	89.863736	1511.01	90.447885
x_p		y_p	
342.023	27.392121	198.959	23.885371
P_1		P_2	
- 0.00156352	5.796580 **e**-03	0.00465375	3.240426 **e**-03

K_1		K_2		K_3	
Value (pixel)	**SD (pixel)**	**Value (pixel)**	**SD (pixel)**	**Value (pixel)**	**SD (pixel)**
- 0.136702	0.145714	1.15309	5.101987	- 29.8807	51.482982

Table 4. Interior orientation parameters and thermal infrared camera lens distortions (pixels).

of the non-metric thermal infrared camera used in this paper.

3.2 Keyframes Extraction and Generating Point Cloud

The purpose of this paper is to investigate the performance evaluation of keyframes extraction in the generation of the thermal infrared dense point cloud to increase density and reduce reprojection error of the 3D point cloud. These criteria are effective in triangulation calculations to 3D reconstruction because keyframes have accurate geometry and high radiometric quality. The proposed algorithm was developed in Python and used the OpenCV library. The primary frames of thermal infrared video are extracted in the first step as input data for the proposed method. Then the blur frames are removed from the video frame dataset using the FFT metric by selecting the optimal threshold. Figure 6 depicts the results of extracting blur and non-blur frames.

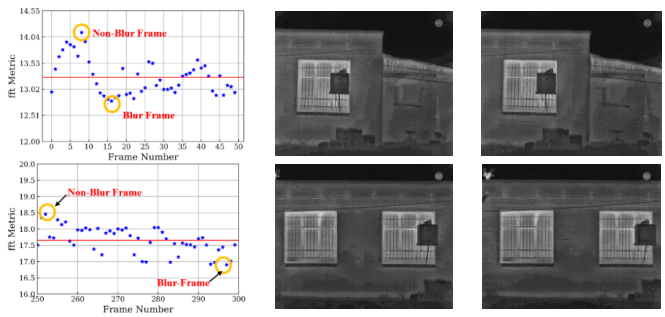


Figure 6. The results of the radiometric quality of the frames.

Figure 6 depicts the results of evaluating the radiometric quality of the frames. Figures (6. A) Show the high and low frequencies of the image of the magnitude spectrum of the frames, Figures (6. B) Display a non-blur frame, and Figures (6. C) Illustrate a blur frame. Video frames are considered in intervals of 50 frames in this paper to select and remove blur frames in specific intervals. As previously stated, the optimal threshold value is determined based on high-frequency values. To identify and remove blur frames in the desired intervals, the average value of the magnitude spectrum image is calculated for each frame. The average values of the magnitude spectrum of frames with minimum values in the desired range are then extracted as blur frames, while frames with maximum values are extracted as non-blur frames.

Following that, blur frames are removed from the keyframe extraction process to improve the geometric accuracy and radiometric quality of the thermal infrared dense point cloud generation. After assessing the radiometric quality of the frames, 853 blur frames were determined and removed from the test dataset, which contained 1957 primary frames, and 214 blur frames were identified and removed from the evaluation dataset, which had 907 primary frames.

Two upper and lower thresholds were used in this paper to evaluate the ratio coefficient of the corresponding key points between pairs of frames. In other words, determining the threshold involves calculating the percentage of corresponding key points between two frames. The upper and lower threshold values in this paper are 0.6 and 0.8, respectively. If the ratio calculated between two frames is greater than the upper threshold value, the baseline between the two frames is short; if the numerical value of the ratio is less than the lower threshold, the baseline between the two frames is long.

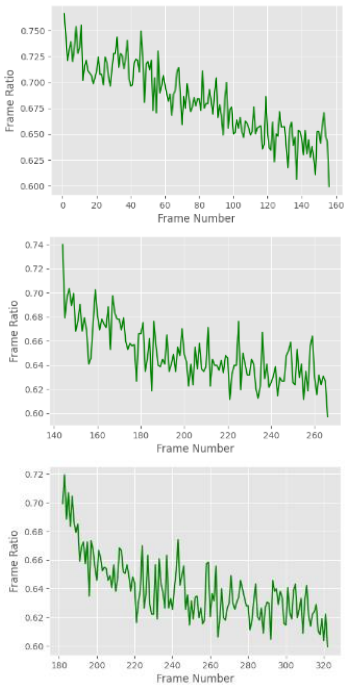
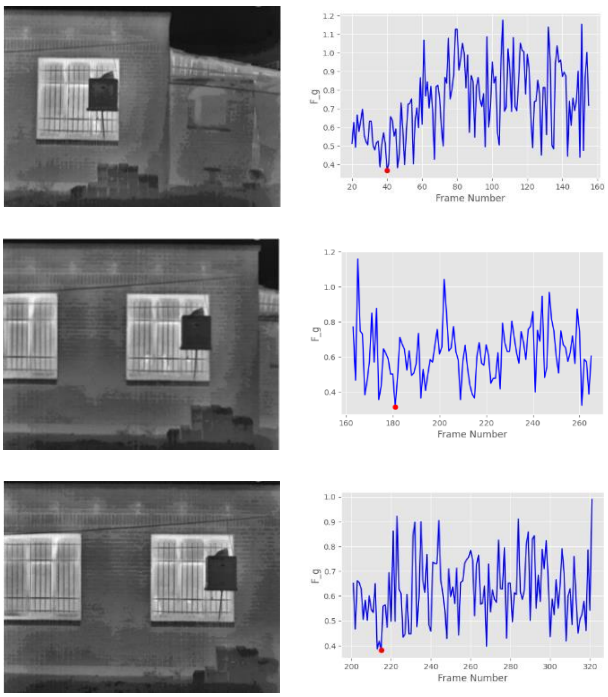


Figure 7. The calculation of the corresponding ratio coefficient between the frames.

[1] Standard deviation

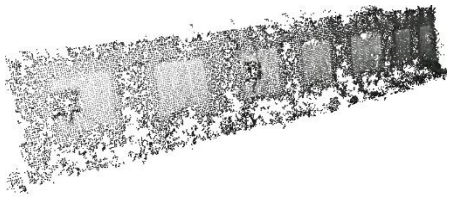
Figure 7 shows an example of calculating the corresponding ratio coefficient between the frames. Figure 7 shows the corresponding ratio coefficient between the frames in the various intervals of the frames between the upper and lower threshold limits, which are 0.6 and 0.8, respectively. Finally, the frames with the standard overlapping baseline in the range between the upper and lower thresholds proceed to the step of evaluating the degeneracy conditions with the optimal selection criterion of GRIC. The degeneracy conditions between pairs of frames are estimated using the GRIC in this step. The numerical value of GRIC is then estimated for the fundamental and homography matrices using equation (8). In the following step, using equation (10), a keyframe is extracted where the fundamental matrix model between two frames has the lowest value compared to the homography matrix model. Figure 8 depicts the GRIC criteria and keyframes results for several extracted keyframes.



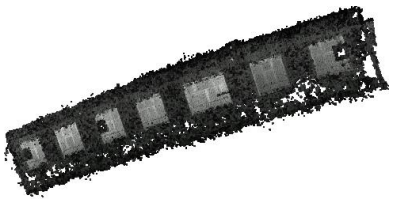
A B

Figure 8. The results of GRIC criteria and keyframes.

In Figure 8, column (8. A) Shows the keyframe extraction based on the comparison of GRIC criteria between the fundamental and homography matrices and selecting the minimum value of the fundamental matrix as the keyframe, and column (8. B) Illustrates the extracted keyframe. The thermal infrared dense point cloud was then generated using SfM and MVS algorithms for keyframes extracted from both test and evaluation datasets. Figure 9 depicts the output results of the thermal infrared 3D dense point cloud for using video data and keyframes from the test dataset.



A. Video



B. Keyframe

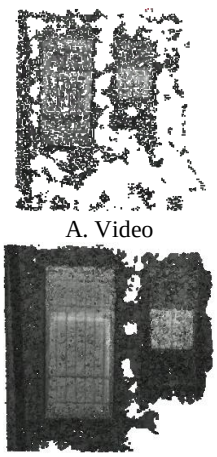
Figure 9. The output of the generation of the 3D thermal dense point cloud of the test dataset .

Figure (9. A) Depicts the output of the thermal infrared 3D dense point cloud of the test dataset for video mode, and figure (9. B) Illustrates the output of the 3D dense point cloud for keyframe mode. In this regard, the visual results show that using keyframes increases the density of the output point cloud. In addition, the results of the numerical evaluation of the density of dense point cloud generated and the amount of reprojection error for the modes of using video data and the test dataset keyframes are presented in Table 5.

Data Type (Video / Keyframe)	Point cloud density (points per square meter)	Reprojection error (pixels)	GSD (cm)
Video	800106	0.83	0.75
Keyframe	1779067	0.41	0.75

Table 5. The output results of the thermal infrared dense point cloud generation for the test dataset.

The numerical evaluation of the density of dense point cloud generated and the amount of reprojection error for the modes of using video data and the keyframes of the used test dataset are presented in Table 5. In this regard, the results show an increase in density and a decrease in reprojection error of the test dataset's point cloud generated using keyframes. Also, Figure 10 shows the output results of generating a 3D dense thermal point cloud using video data and the keyframes of the evaluation dataset.



A. Video

B. Keyframe

Figure 10. The output of the generation of the 3D thermal dense point cloud of the evaluation dataset.

Figure (10. A) Depicts the output of the thermal infrared 3D dense point cloud of the evaluation dataset in video mode, and figure (10. B) Illustrates the output of the 3D dense point cloud in keyframe mode. In this regard, the visual results show that using keyframes increases the density of the output point cloud. In addition, the results of the numerical

evaluation of the density of dense point cloud generated and the amount of reprojection error for the video data use modes, as well as the keyframes of the used evaluation dataset, are presented in Table 6. The results of the numerical evaluation of the density of dense point cloud generated and the amount of reprojection error for the use modes of video data, as well as the keyframes of the used evaluation dataset, are presented in Table 6. In this regard, the results show an increase in density and a decrease in reprojection error of the point cloud generated using keyframes for the evaluation dataset. Based on the results of tables 5 and 6, the use of keyframes increases the density by about 0.03% to 0.10% of points per square meter and reduces the reprojection error by about 0.005% of pixels (2 times) for the thermal infrared dense point cloud are tested and evaluated datasets. In the future, the proposed method in this paper will be quantitatively and qualitatively compared to a competitive process of keyframe extraction to generate the thermal dense point cloud. The conclusion and suggestions for future research are presented next.

Data Type (Video / Keyframe)	Point cloud density (points per square meter)	Reprojection error (pixels)	GSD (cm)
Video	16461	1.31	0.75
Keyframe	178524	0.66	0.75

Table 6. The output results of the thermal infrared dense point cloud generation for the evaluation dataset.

4. CONCLUSION

The primary goal of this paper is to investigate the effect of keyframe extraction from a thermal infrared video sequence on the geometrical accuracy of a dense thermal point cloud. Therefore, a method for evaluating the effect of extracting keyframes from a sequence of thermal infrared images to generate a dense thermal point cloud has been presented. Based on the findings, extracting keyframes from an image sequence improves the geometric accuracy of the point cloud, increases the speed, and decreases the volume of triangulation calculations. Based on the paper's results, keyframe extraction increases the density of the thermal infrared point cloud by about 0.03% to 0.10% points per square meter. It reduces the reprojection error by about 0.005% pixels (2 times). Among the paper's challenges are the limitations of the thermal infrared camera, such as low contrast and spatial resolution. These constraints reduce the texture of the images and limit the number of features that can be extracted from them to assess the baseline of the overlapping, match the features, and generate points in three-dimensional space. Another limitation and an effective step for identifying and extracting blur frames are selecting the optimal threshold to evaluate the radiometric quality of the frames. In this regard, it is possible to mention improving the contrast of thermal infrared images to increase the details and number of features that can be extracted from the thermal infrared images, as well as improving matching methods, which will be investigated in future studies. In this regard, it is possible to mention enhancing the contrast of thermal infrared images to increase the details and the number of features that can be extracted from thermal infrared images, as well as enhancing the matching, to check and solve the mentioned limitations. In future research, an attempt will be made to compare the limitations above and the proposed method of this paper to a competitive method from both a quantitative and qualitative standpoint.

REFERENCES

Abdel-Qader, I., Abudayyeh, O., Kelly Michael, E., 2003. Analysis of Edge-Detection Techniques for Crack Identification in Bridges. Journal of Computing in Civil Engineering, 17(4), 255-263. doi: 10.1061/ (ASCE) 0887-3801(2003)17:4(255).

Ahmed, M. T., Dailey, M. N., Landabaso, J. L., Herrero, N., 2010. Robust Key Frame Extraction for 3D Reconstruction from Video Streams. Paper presented at the VISAPP (1).

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., Wu, A. Y., 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. Journal of the ACM (JACM), 45(6), 891-923.

Azimi, A., Hosseininaveh, A., Remondino, F., 2022. A NOVEL GEOMETRIC KEY-FRAME SELECTION METHOD FOR VISUAL-INERTIAL SLAM AND ODOMETRY SYSTEMS. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 9-14.

Bakogiannis, E., 2020. Using Unmanned Aerial Vehicles (UAVs) to analyze the urban environment. European Journal of Engineering and Formal Sciences, 3(2), 10-18.

Brown, D. C., 1971. Close-range camera calibration, Photogrammetric Engineering. Engineering and Remote Sensing, 37(8), 855-866.

Caciora, T., Herman, G. V., Ilieș, A., Baias, Ș., Ilieș, D. C., Josan, I., Hodor, N., 2021. The use of virtual reality to promote sustainable tourism: A case study of wooden churches historical monuments from Romania. Remote Sensing, 13(9), 1758.

Cai, J.-F., Ji, H., Liu, C., Shen, Z., 2009. Blind motion deblurring using multiple images. Journal of Computational Physics, 228(14), 5057-5071. doi:https://doi.org/10.1016/j.jcp.2009.04.022.

Chia, A. Y. S., Leung, M. K., Eng, H.-L., Rahardja, S., 2007. Ellipse detection with hough transform in one dimensional parametric space. Paper presented at the 2007 IEEE International Conference on Image Processing.

Choi, J., Kwon, S., Son, K., Yoo, J., 2016. Fast key-frame extraction for 3D reconstruction from a handheld video. International journal of advanced smart convergence, 5(4), 1-9.

Chowdhury, T., Rahnemoonfar, M., Murphy, R., Fernandes, O., 2020. Comprehensive semantic segmentation on high resolution uav imagery for natural disaster damage assessment. Paper presented at the 2020 IEEE International Conference on Big Data (Big Data).

Crete, F., Dolmiere, T., Ladret, P., Nicolas, M., 2007. The blur effect: perception and estimation with a new no-reference perceptual blur metric. Paper presented at the Human vision and electronic imaging XII.

Dadras Javan, F., Savadkouhi, M., 2019. THERMAL 3D MODELS ENHANCEMENT BASED ON INTEGRATION WITH VISIBLE IMAGERY. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Dahaghin, M., Samadzadegan, F., Dadrass Javan, F., 2021. Precise 3D extraction of building roofs by fusion of UAV-based thermal and visible images. International journal of remote sensing, 42(18), 7002-7030.

Datta, A., Kim, J.-S., Kanade, T., 2009. Accurate camera calibration using iterative refinement of control points. Paper presented at the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops.

De, K., Masilamani, V., 2013. Image sharpness measure for blurred images in frequency domain. Procedia Engineering, 64, 149-158.

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.

Frederic, P. M., Dufaux, F., Winkler, S., Ebrahimi, T., Sa, G., 2002. A no-reference perceptual blur metric. Paper presented at the IEEE 2002 International Conference on Image Processing.

Furukawa, Y., Curless, B., Seitz, S. M., Szeliski, R., 2010, 13-18 June 2010. Towards Internet-scale multi-view stereo. Paper presented at the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Gonzalez, R. C., Woods, R. E., 2002. Digital image processing. In: Prentice hall Upper Saddle River, NJ.

Han, D., Lee, S. B., Song, M., Cho, J. S., 2021. Change detection in unmanned aerial vehicle images for progress monitoring of road construction. Buildings, 11(4), 150.

Hossein Pour, H. R., Samazadegan, F., F, D. J., 2016. Keyframe Selection from Video Stream for 3D Reconstruction. The 1st National Conference on Geospatial Information Technology.

Jarzabek-Rychard, M., Karpina, M., 2016. QUALITY ANALYSIS ON 3D BUIDLING MODELS RECONSTRUCTED FROM UAV IMAGERY. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41.

Koch, T., Körner, M., Fraundorfer, F., 2019. Automatic and semantically-aware 3D UAV flight planning for image-based 3D reconstruction. Remote Sensing, 11(13), 1550.

Kumar, G., Reddy, V., Srinivas Kumar, S., 2018. Video shot boundary detection and key frame extraction for video retrieval. Paper presented at the Proceedings of the Second International Conference on Computational Intelligence and Informatics.

Kylili, A., Fokaides, P. A., Christou, P., Kalogirou, S. A., 2014. Infrared thermography (IRT) applications for building diagnostics: A review. Applied Energy, 134, 531-549. doi:https://doi.org/10.1016/j.apenergy.2014.08.005.

Liu, F., Hu, P., Zheng, B., Duan, T., Zhu, B., Guo, Y., 2021. A field-based high-throughput method for acquiring canopy architecture using unmanned aerial vehicle images. Agricultural and Forest Meteorology, 296, 108231.

Longuet-Higgins, H. C., 1981. A computer algorithm for reconstructing a scene from two projections. Nature, 293(5828), 133-135.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91-110.

Marziliano, P., Dufaux, F., Winkler, S., Ebrahimi, T., 2002. A no-reference perceptual blur metric. Paper presented at the Proceedings. International Conference on Image Processing.

Ming-Chao, C., Boult, T. E., 1997. Local blur estimation and super-resolution. Paper presented at the Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Motayyeb, S., Fakhri, S. A., Varshosaz, M., Pirasteh, S., 2022. ENHANCING CONTRAST OF IMAGES TO IMPROVE GEOMETRIC ACCURACY OF A UAV PHOTOGRAMMETRY PROJECT. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 389-398.

Ong, E., Lin, W., Lu, Z., Yang, X., Yao, S., Pan, F, Moschetti, F., 2003. A no-reference quality metric for measuring image blur. Paper presented at the Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.

Pagaduan, R. A., Aragon, M. C. R., Medina, R. P., 2021. iBlurDetect: Image Blur Detection Techniques Assessment and Evaluation Study.

Paris, S., Kornprobst, P., Tumblin, J., Durand, F., 2009. Bilateral filtering: Theory and applications: Now Publishers Inc.

Peleshko, D., Rak, T., Noennig, J. R., Lytvyn, V., Vysotska, V., 2020. Drone Monitoring System DROMOS of Urban Environmental Dynamics. Paper presented at the ITPM.

Poux, F., Valembois, Q., Mattes, C., Kobbelt, L., Billen, R., 2020. Initial user-centered design of a virtual reality heritage system: Applications for digital tourism. Remote Sensing, 12(16), 2583.

Rashidi, A., Dai, F., Brilakis, I., Vela, P., 2013. Optimized selection of key frames for monocular videogrammetric surveying of civil infrastructure. Advanced Engineering Informatics, 27(2), 270-282. doi:https://doi.org/10.1016/j.aei.2013.01.002

Seo, J. K., Kim, S. H., Jho, C. W., Hong, H. K., 2003. 3D estimation and key-frame selection for match move. Paper presented at the ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications.

Seo, Y.-H., Kim, S.-H., Doo, K.-S., Choi, J.-S., 2008. Optimal keyframe selection algorithm for three-dimensional reconstruction in uncalibrated multiple images. Optical Engineering, 47(5), 053201.

Sieberth, T., Wackrow, R., Chandler, J. H., 2016. Automatic detection of blurred images in UAV image sets. ISPRS Journal of Photogrammetry and Remote Sensing, 122, 1-16. doi:https://doi.org/10.1016/j.isprsjprs.2016.09.010

Snavely, N., Seitz, S. M., Szeliski, R., 2008. Modeling the world from internet photo collections. International Journal of Computer Vision, 80(2), 189-210.

Suhr, J. K., 2009. Kanade-lucas-tomasi (klt) feature tracker. Computer Vision (EEE6503), 9-18.

Tomasi, C., Manduchi, R., 1998. Bilateral filtering for gray and color images. Paper presented at the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271).

Torr, P. H., 1998. Geometric motion segmentation and model selection. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 356(1740), 1321-1340.

Torr, P. H. S., Fitzgibbon, A. W., Zisserman, A., 1999. The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences. International Journal of Computer Vision, 32(1), 27-44. doi:10.1023/A:1008140928553

Ullman, S., 1979. The interpretation of structure from motion. Proceedings of the Royal Society of London. Series B. Biological Sciences, 203(1153), 405-426.

Usamentiaga, R., Garcia, D., Ibarra-Castanedo, C.,& Maldague, X., 2017. Highly accurate geometric calibration for infrared cameras using inexpensive calibration targets. Measurement, 112, 105-116.

Varadarajan, S., Karam, L. J., 2008. An improved perception-based no-reference objective image sharpness metric using iterative edge refinement. Paper presented at the 2008 15th IEEE International Conference on Image Processing.

Wang, J., Zeng, C., Wang, Z., Jiang, K., 2022. An improved smart key frame extraction algorithm for vehicle target recognition. Computers & Electrical Engineering, 97, 107540.

Xie, Z., Wan, F., Bu, Q., Zhou, X., Zhang, J., Chen, S., 2015. Aerial sequential frame decimation for scene reconstruction. Paper presented at the 2015 IEEE International Conference on Information and Automation.

Yun-Chung, C., Jung-Ming, W., Bailey, R. R., Sei-Wang, C., Shyang-Lih, C., 2004. A non-parametric blur measure based on edge analysis for image processing applications. Paper presented at the IEEE Conference on Cybernetics and Intelligent Systems, 2004.

Zhang, C., Wang, H., Li, H., Liu, J., 2017. A fast key frame extraction algorithm and an accurate feature matching method for 3D reconstruction from aerial video. Paper presented at the 2017 29th Chinese Control and Decision Conference (CCDC).