Korean Journal of Remote Sensing, Vol.26, No.6, 2010, pp.629~644

Speeding up the KLT Tracker for Real-time Image Georeferencing using GPS/INS Data

Supannee Tanathong and Impyeong Lee †

Laboratory for Sensor and Modeling, Department of Geoinformatics,
The University of Seoul, 90 Jeonnong-dong, Dongdaemun-gu, Seoul, 130-743 Korea

Abstract : A real-time image georeferencing system requires all inputs to be determined in real-time. The intrinsic camera parameters can be identified in advance from a camera calibration process while other control information can be derived instantaneously from real-time GPS/INS data. The bottleneck process is tie point acquisition since manual operations will be definitely obstacles for real-time system while the existing extraction methods are not fast enough. In this paper, we present a fast-and-automated image matching technique based on the KLT tracker to obtain a set of tie-points in real-time. The proposed work accelerates the KLT tracker by supplying the initial guessed tie-points computed using the GPS/INS data. Originally, the KLT only works effectively when the displacement between tie-points is small. To drive an automated solution, this paper suggests an appropriate number of depth levels for multi-resolution tracking under large displacement using the knowledge of uncertainties the GPS/INS data measurements. The experimental results show that our suggested depth levels is promising and the proposed work can obtain tie-points faster than the ordinary KLT by 13% with no less accuracy. This promising result suggests that our proposed algorithm can be effectively integrated into the real-time image georeferencing for further developing a real-time surveillance application.

Key Words : KLT Algorithm, Image Georeferencing, Image Matching, Exterior Orientation, Error Propagation.

1. Introduction

The statistical report of worldwide disasters observed during the past three decades reveals an exponential increase in hazard occurrence, as illustrated as Fig. 1 (Rodriguez *et al*, 2009). Over the recent few years, the world has experienced a huge number of high-impact disasters. In 2008, cyclone Nargis killed more than hundred thousand people in Myanmar, and the Sichuan earthquake in China affected greater than eighty thousand people (Rodgriguez *et al*, 2009). The recent 2010 earthquake in Haiti caused the death of more than two hundred thousand people (CBC News, 2010 and USGS, 2010). Disasters do not only cause the loss of lives but also damage properties and natural environment,

Received October 19, 2010; Accepted December 20, 2010.
†Corresponding Author: Impyeong Lee (iplee@uos.ac.kr)

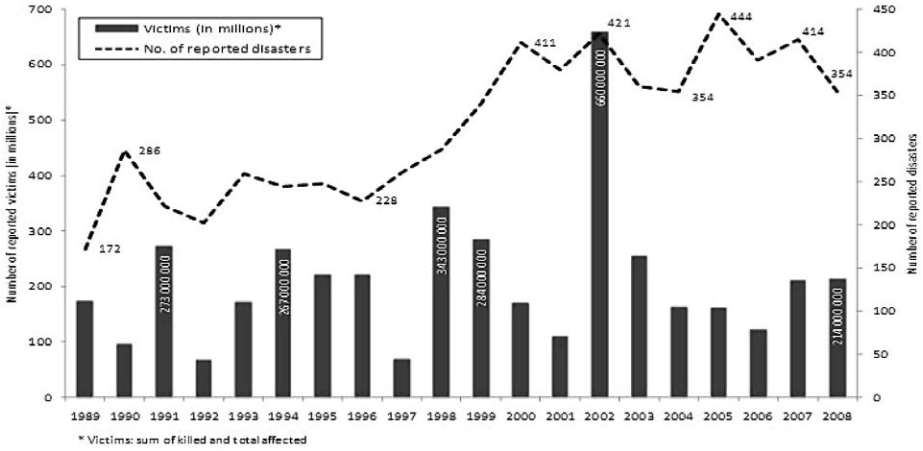


Fig. 1. Disaster trends in occurrence and numbers of victims.
(Rodriguez *et al*, 2009)

and account for economic losses. These consequences of hazardous events significantly lead to the increasing demand for surveillance applications to strengthen mitigation and rehabilitation activities.

To support disaster response by immediately acquiring the on-going emergency status, many surveillance applications have been launched with airborne real-time sensors based on a UAV platform. The mounted sensors include a digital camera, a laser scanner, a Global Positioning System (GPS) and an Inertial Navigation System (INS), which can acquire aerial images, ranging data, positions and attitudes of the platform tagged with the GPS time. These sensory data are then transmitted and processed in the ground for decision makers to address the target areas and establish a proper mitigation plan. The real-time captured data may be used for analysis in conjunction with existing spatial data such as maps. However, these images are usually acquired at individually different locations and attitudes with being partially overlapped. For the immediate uses of these images, they must be rectified with the same coordinate system as the existing spatial data principally in an absolute ground coordinate system. This rectification, referred to as an image georeferencing process, requires the accurate position and attitude of the camera at the time of exposure of each image, called the extrinsic camera parameters or exterior orientation parameters, aka EO (Schenk, 1999). Throughout this paper, these two words are used interchangeably.

The camera parameters provided by GPS/INS sensors are known as not so accurate unless high quality but expensive equipment is employed. Although the accuracy is one of the most important factors that we aim to achieve, due to a limited budget, the inexpensive implementation should be concerned. Therefore, an economical technique is instead considered to acquire sufficiently accurate exterior orientation parameters. A potential method to adjust the exterior orientation parameters of all images simultaneously is the aerial triangulation with a bundle block adjustment model (Schenk, 1999 and, Wolf and Dewitt, 2000). This technique requires the knowledge about intrinsic camera parameters, tie points between overlapped images, in conjunction with the control information about the ground features plus initial approximations for extrinsic

camera parameters and ground points (GP). The control information can be provided with either ground control points (GCP) requiring intensive human operations or direct measurements of the exterior orientation parameters. The use of GCP may produce highly accurate output but prevent automation and hence real-time processing. The direct measurements can be efficiently achieved using GPS/INS sensors with reasonable accuracy without any human intervention and hence more adequate for real-time processing. The input and output of the aerial triangulation with bundle block adjustment are summarized as Fig. 2.

To establish real-time image georeferencing, in principle, all inputs must be determined in real-time. The intrinsic camera parameters can be determined in a priori from the camera calibration process before flight. The control information from the direct measurements is more promising, for a real-time requirement, than the indirect measurement method using GCPs since the exterior orientation parameters can be obtained from the GPS/INS sensors boarded with a UAV in real-time. The initial approximations for EOs are simply derived from the GPS/INS data, and the initial approximations for ground points are determined using the initial EOs and tie points. To completely establish a real-time system, the tie points between adjacent images must also be acquired from a computational process instantaneously. For a system to meet a real-time requirement, the computational time of any process in the system should be less than the image acquisition process (Choi and Lee, 2010). For each new image acquired from the airborne sensor, the computational time used to obtain tie points between this image and its adjacent prior-acquired image must be sufficiently small. This process is required to complete before a new image is captured.

Typically, tie points can be determined through an image matching (or feature tracking) technique in computer vision. From our literatures on feature detection and matching, the Kanade-Lucas-Tomasi (later be referred to as KLT throughout this paper) is one of the promising techniques that can be employed for tie point acquisition (Lucas and Kanade, 1981, and, Shi and Tomasi, 1994). According to its lightweight computational characteristic for feature selection (candidates for tie points, in this case), in this study, KLT is utilized. Due to the designed period of camera's exposure time, in conjunction with the planned trajectory plus the effect of wind, the displacement between two image frames or, in a specific meaning, the distance between corresponding points in a pair of adjacent images is not always small. On the contrary, the KLT only works effectively well when the displacement is not so large. To reduce the expensive computation and ambiguity, when the distance is large, the search space must be limited. The hierarchical multi-resolution searching strategy is a powerful tool to reduce the range of searching (Moallem *et al*, 2002). The KLT also has the implementation of pyramidal

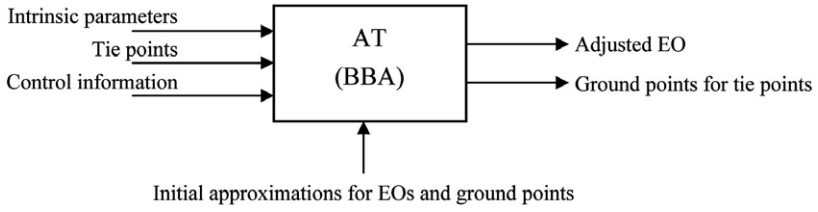


Fig. 2. Aerial triangulation with bundle block adjustment.

multi-resolution not only to reduce the search space for faster processing but also to allow the algorithm to track under the large displacement (Bouguet, 2000). The following question is that at which finest resolution level the searching process should start. Starting from too coarse resolution can increase the computational time for the searching process while less coarse resolution may not lead to the correct solution. In this work, we determine the appropriate number of resolution levels through the general law of propagation of variances using the knowledge of uncertainties associated with the measured GPS/INS data (Ghilani and Wolf, 2006).

This paper presents an attempt to develop an automated image matching based on the KLT tracker to obtain a set of tie points for real-time image georeferencing. The goal of this proposed work is to achieve the real-time requirement by utilizing the GPS/INS data and the knowledge about their uncertainty. The remainder of this paper is organized as follows: Section 2 reviews and discusses feature detection and matching techniques. Section 3 explains our proposed methodology to develop an automated real-time image matching. The experimental results are discussed in Section 4. Finally, the discussion and conclusion are presented in the last section.

2. Review of Feature Detection and Matching

Aerial triangulation is a promising technique to adjust together the exterior orientation parameters of each image by employing intrinsic camera parameters, tie points between overlapping images, the control information about the ground features plus some initial approximations for extrinsic camera parameters and ground points. To achieve real-time image georeferencing, theoretically, all input parameters must be obtained in near real-time or, at least, before a new consecutive image is acquired. The camera parameters, including the camera position and orientation, are typically obtained instantaneously from the GPS/INS sensors boarded with a UAV. Tie points are usually used to identify and establish the relationships among images that possess certain similarities. In analytical photogrammetry, tie points may be measured manually through a photogrammetric plotter. This technique is, however, computationally inefficient and does prohibit the real-time requirement. To develop a real-time system, tie points must be computed within a short period of computational time. Therefore, many researchers employ image matching technique in computer vision to determine tie points (Gruen, 1985).

In computer vision, there exist a number of powerful techniques for image matching. The most widely used methods are KLT (Kanade-Lucas-Tomasi) and SIFT (Scale-Invariant Feature Transform). KLT was developed by Shi and Tomasi (1994) based on the original work presented by Lucas and Kanade (1981). The algorithm locates significant features to track by examining the minimum eigenvalues of the autocorrelation matrix defined on the derivative image. SIFT was first published by Lowe (2004). It is a technique to extract distinctive invariant features from images that encourages reliable image matching. The excellent characteristics of SIFT features that are invariant to image scale and rotation cause the algorithm to be computationally expensive (Hubner and Pajarola, 2009). SIFT features are considered to be more robust than KLT. The KLT algorithm is more sensitive to noises but require less computational time in comparison with SIFT (Nakaguro *et al*, 2007). According to its lightweight computation, KLT is more suitable for a real-time system and thus utilized in this study.

KLT is one of the known tracking techniques that has gained much interest in motion tracking according to its effectiveness in selecting good features to track and its tracking algorithm. In the recent few decades, KLT has been widely used in a variety of applications ranging from eye tracking to military surveillance activities. This technique is classified as a sparse tracker in which only a subset of points in an images are identified and tracked along subsequent frames (Bradski and Kaehler, 2008). KLT refers to those points as *good features to track* which can be, intuitively, either corners or isolated pixels since they are likely to be unique and contain some significant information to pinpoint their corresponding feature in the consecutive frames. Therefore, they are quite appropriate to be used as tie points for an image georeferencing process.

As summarized in (Bouguet, 2000, and Bradski and Kaehler, 2008), a point is considered to be a good feature when the minimum eigenvalue, λ_m, of its *autocorrelation matrix G* defined on the derivative images centered at its position is larger than a predefined threshold. Such a matrix is defined as (1).

$$G = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (1)$$

where (p_x, p_y) and (w_x, w_y) denote the position of an image point and the dimension of the local window $(2w_x+1, 2w_y+1)$, respectively. I_x and I_y are defined as the image derivatives at the point (x, y) in x and y direction. Once the good features in the first image are determined, their corresponding points, or equivalently tie points, in the consecutive image may be found if there exists an optical flow vector $d = [d_x \; d_y]^T$ that minimizes the residual function ϵ defined as Equation (2), between the first image A and the subsequence image B.

$$\epsilon(d) = \epsilon(d_x, d_y) = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} (A(x,y) - B(x+d_x, y+d_y))^2 \quad (2)$$

From the equation above, it is noticeable that the residual function will return the minimal value only if $d_x < w_x$ and $d_y < w_y$. As to maintain the spatial coherence among neighboring points, a small local window is preferred. Thus the KLT algorithm is only applied for tracking when the displacements are small (Zivkovic, 2004). This conclusion corresponds to our preliminary experimental result performed on a sequence of thirty-seven aerial images with approximately 75% overlapping ratio. We developed an ordinary KLT tracker without pyramidal implementation to measure its performance. Due to the characteristics of trajectory and overlapping ratio of our system, the life span of a feature point is quite short. Therefore, we measured the tracking performance based on pairs of images and its succession instead of the whole sequence. As presented as Fig. 3, the tracking performance is quite impressive when the displacements between image pairs are small (the distance is less than 15% of the image dimension). However, when the displacements are large, the results are poor since the tracker could not match even a single correct conjugate point.

Due to the designed period of camera's exposure time, in conjunction with the planned trajectory plus the effect of wind, the displacement between two image frames or, in a specific meaning, the distance

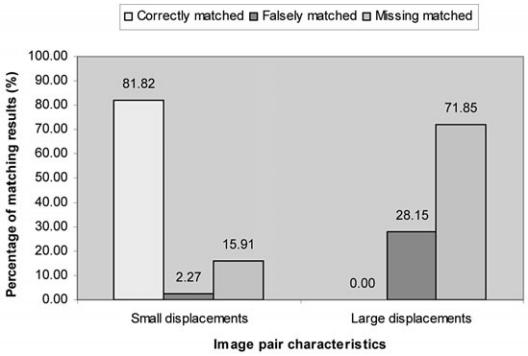


Fig. 3. Performance of the ordinary KLT tracker with no multi-resolution implementation.

between corresponding points in a pair of adjacent images is not always small. On the contrary, the KLT only works effectively well when the displacement is not so large. In order to handle large motions, larger than the local window size, Bouguet (2000) proposed a hierarchical multi-resolution implementation of KLT. With the pyramidal implementation that limits the range of search space, not only does the ambiguity decrease but the computational time does also. Furthermore, this technique enables the original KLT algorithm to track under large displacement. At each level of KLT pyramid, the method preserves the optical flow vector d to be smaller than the window size while the overall displacement vectors can be as large as $(2^{L_m+1} - 1)$ times of the window size, where L_m denotes the depth of the pyramid. With the implementation of pyramid iteration, equation (3) can be redefined as follows.

$$\epsilon^L(d^L) = \epsilon^L(d_x^L, d_y^L) = \sum_{x=p_x^L-w_x}^{p_x^L+w_x} \sum_{x=p_y^L-w_y}^{p_y^L+w_y} (A^L(x, y) - B^L(x + g_x^L + d_x^L, y + g_y^L + d_y^L))^2 \quad (3)$$

where $g^L = (g_x^L, g_y^L)$ means the initial guess for optical flow at the pyramidal level L.

The pyramidal KLT tracker runs iteratively from the coarsest resolution ($L = L_m$) to the finest resolution, which is the actual image resolution ($L = 0$). The procedure is summarized from Bouguet (2000) as (1) the optical flow vector is initially computed at the deepest pyramidal level L_m, (2) the computational result is then propagated to the upper level L_m-1 in a form of an initial guess for optical flow vector at the current level L_m-1, (3) given that initial guess, the refined optical flow vector is then computed for level L_m-2, (4) step are repeated until the highest pyramidal level is reached at the original image resolution ($L = 0$).

It is not unusual that the image sequences acquired from our boarded UAV have large displacements between subsequent images. To extract tie points from these pairs of images, we unquestionably require the KLT tracker with multi-resolution implementation. Although the pyramidal implementation enables the KLT tracker to run under the large displacement case, it is not yet a complete solution to real-time tie point extraction. If the displacement between a pair of images is considerably large, the KLT algorithm is supposed to start tracking at a very coarse resolution. In a similar meaning, the number of pyramidal depth-level for multi-resolution tracking may be high to ensure the distance between corresponding points in a pair of images is entirely tracked. However, a higher number of pyramidal levels causes the increasing computational time. Therefore, we need to keep the number of pyramidal depth levels to be promisingly small to reduce the tracking time but it must be sufficiently deep to maintain the same accuracy.

The potential number of depth levels is determined by considering the distance between corresponding points in image pairs. For each pair of corresponding points, one point is determined through the KLT good feature extraction while another point in the second image is through the KLT tracking process. The positions of corresponding points are yet not known in advance. In effect, the distances between tie points for further determining the number of pyramidal levels also remain obscure. To reduce the processing time, we need to find solutions to shorten the distance between the corresponding points such that it will effectively decrease the number of depth levels, and to compute the promising pyramidal depth levels. According to the problems discussed so far, in this paper, we attempt to find the answers to these open questions by using the acquired GPS/INS data. The background and proposed solutions are discussed in the next section.

3. Proposed Methodology

As presented as Fig. 3, the KLT tracker works perfectly well when the displacements between pairs of images are small. However, according to various conditions of the image acquisition system, these distances are not always small. Therefore, to acquire tie points under this situation, the KLT algorithm must be implemented in multi-resolution scheme. The larger the displacement is, the more number of pyramidal depth levels is required to ensure the tracking coverage, and also the more computational time uses. In this study, we attempted to shorten this displacement by employing the GPS/INS data to effectively lead to a fewer number of pyramidal depth levels. The pyramidal KLT is implemented based on Equation (3). The algorithm, typically, approximates the initial guess for the optical flow vector at the coarsest level to zero (no initial guess) in which the displacements remain unshortened. This work suggests promising initial guessed positions through the collinearity equation, which in effect will decrease the displacement as much as the accuracy of the GPS/INS data. Although the distance is believed to be considerably shortened, its quantitative value is still obscure. In this paper, we approximated this figure given the uncertainty of the GPS/INS equipments by employing the general laws of variance propagation (Mikhail, 1976). The potential number of pyramidal level can be further derived from this result. The detailed description of each proposed approach is discussed thoroughly in this section.

1) Initial Guessed Position for the Optical Flow Vector

The final outcome of this work is a set of tie points that will be used together with the other input parameters to establish the image georeferencing process. Tie points are the corresponding positions in multiple images that relate them with their certain similarities. The multi-resolution implementation of the KLT tracker runs iteratively from the coarsest resolution to the finest resolution in which the optical flow vector at a lower level is propagated to the upper level in a form of initial guess. As mentioned before in Section 2, the initial guess is offered to enable the algorithm to track under the large displacement case while still preserves the optical flow vector to be smaller than the local window size. In this section, we suggest an approach to obtain promising initial guessed positions by incorporating the GPS/INS data.

Being similar to many other non-linear problems, estimated solutions are usually significantly sensitive to the initial approximations of the unknowns. Therefore, one critical question is how to obtain suitable approximations. If they are not sufficiently close to the true values, it is possible that the solution may not be converged to the true values (Schenk, 1999). A simple but efficient approach to lead the tracker to the final solution is to provide initial tracking positions that are close to the corresponding tie points. Typically, the pyramidal KLT approximates the initial guess for optical flow at the coarsest level to zero (no initial guess). The novel innovation is that, in this work, we propose to supply KLT with the initial guessed positions of the conjugate points using an exterior orientation method. With our initial positions that are adequately close to the corresponding points, the KLT pyramidal process is expected to converge relatively fast at a few depth levels. This will significantly reduce the computational time of the overall tracking process and that yield the appropriate solution for the real-time requirements. The overview of the solution is presented as a simple flowchart as Fig. 4.

With the location of the perspective center (*PC*) and the orientation of the camera (`w`, `j`, `k`) bundled

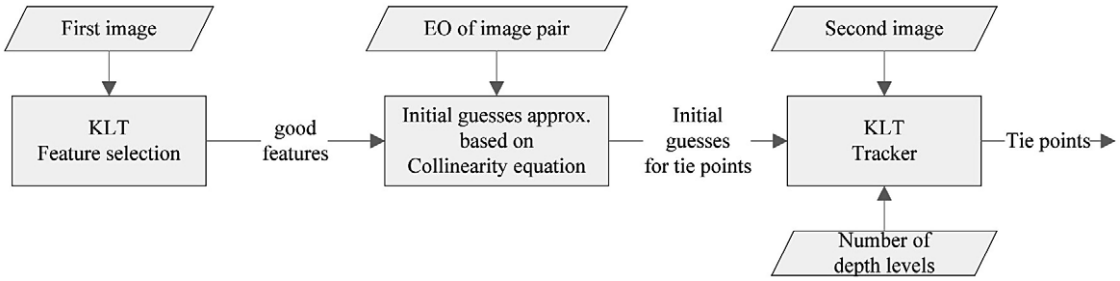


Fig. 4. The overview of our approach to obtain the initial guessed positions.

with the images, we can determine the approximated positions of the ground points corresponding to the image points or vice versa using the collinearity equations. For the sake of simplicity, in this work, it is assumed that the terrain is a virtual horizontal plane with an average elevation. The ground points corresponding to the good features in the first image are obtained using the rearranged prevalent collinearity equation (4).

$$x = x_p - c \cdot \frac{(X_p - X_c)r_{11} + (Y_p - Y_c)r_{12} + (Z_p - Z_c)r_{13}}{(X_p - X_c)r_{31} + (Y_p - Y_c)r_{32} + (Z_p - Z_c)r_{33}}$$
$$y = y_p - c \cdot \frac{(X_p - X_c)r_{21} + (Y_p - Y_c)r_{22} + (Z_p - Z_c)r_{23}}{(X_p - X_c)r_{31} + (Y_p - Y_c)r_{32} + (Z_p - Z_c)r_{33}} \quad (4)$$

where (X_p, Y_p) and c, are the interior orientation parameters that correspond to the image principle point and focal length. (X_c, Y_c, Z_c) and r_{ij}, that are referred to as exterior orientation parameters, denote the perspective center and rotational matrix of w, j, and k, respectively.

Those ground points are then projected to their conjugate point in the subsequent images using the given equations. Finally, the corresponding points will be input as the initial guess for the optical flow. The initial guess, as expected, contains some errors, which can be computed from the propagation of the errors associated with the exterior orientation of the pair of images and the ground elevation. With our fine initial guessed positions, the pyramidal tracking process can be started at a few depth levels (higher resolution). The suggested number of pyramidal depth level is discussed in Section 3.2. Fig. 5 presents our approach to estimate the initial guessed positions for the optical flow vector.

Although this work assumes a flat terrain, it is sufficient for initial guesses as the pyramidal process

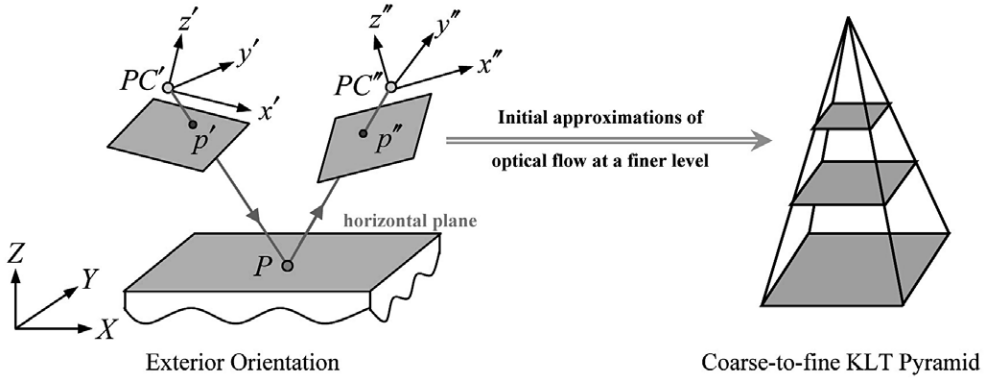


Fig. 5. The approach to estimate the initial guessed position for the optical flow vector.

will further shape them closer to the final optical flow vector. For those who expect high accurate initial approximations, DSM may be an alternative as suggested by a recent work from Hong *et al*. (2009) that proposed an image matching method using the LIDAR-derived DSM.

2) A Promising Number of Pyramidal Depth Level

The level of resolution significantly affects the processing time of the KLT tracker. As discussed earlier, the number of resolutions, or specifically the number of pyramidal depth levels, is a computational result of the displacement between pair of subsequent images. Section 3.1 discusses a solution to approximate the initial guessed position to the optical flow vector which can lead to decrease this displacement. Although the distance is believed to be considerably shortened, its quantity in term of numerical is still obscure. This section presents an approach to employ the knowledge about uncertainty of related measurements to estimate this figure.

Let define the point $p_1(x_1, y_1)$ in the first image as a feature of interest to be tracked and the point $p_2(x_2, y_2)$ in the second image be its corresponding point obtained through the collinearity equation to form a pair of tie point. The coordinates of the corresponding points are not competent to be used as tie points for the image georeferencing process since the initial guessed positions, inevitably, contains some errors. Due to its potential to guide to the right tie points, therefore, this work referred to those corresponding points as the initial guessed positions for tie points. The errors of the initial guessed positions are the effects of the errors in the following measurements:

(a) The coordinates of the interest points in the first image to be tracked for its tie point.
(b) The initial approximations of the exterior orientation parameters in both images obtained via the GPS/ INS.
(c) The estimation of the average terrain elevation.

Generally, the unknown values are often determined indirectly by making direct measurement of other quantities which are functionally related to the desired unknowns (Ghilani and Wolf, 2006). The obscure distance between the unknown correct positions of tie points in the second image and the positions of the corresponding points from the collinearity equations may be referred to as the error of the initial guessed positions of tie points. Fig. 6 illustrates this unknown distance. The error can be approximated through the general law of variance propagation given that the errors of those three related measurements above are known. The final tie point position in the second image can be effectively derived from the collinearity equation only if the given exterior orientation parameters of both images are accurate. The corresponding ground coordinate is first determined based on the first image's camera parameters, then this coordinate is propagated to derive the corresponding image point in the second image or, in this case, tie point using the second

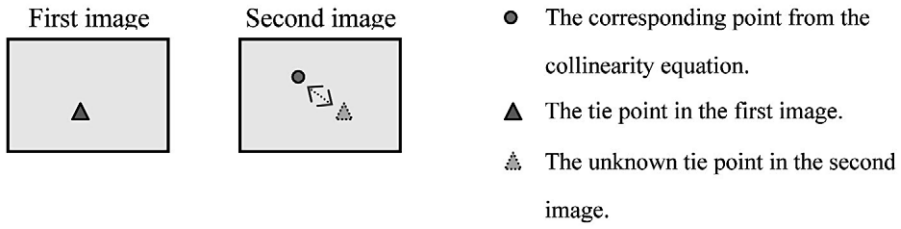


Fig. 6. The unknown distance between the initial guessed point and the correct tie point.

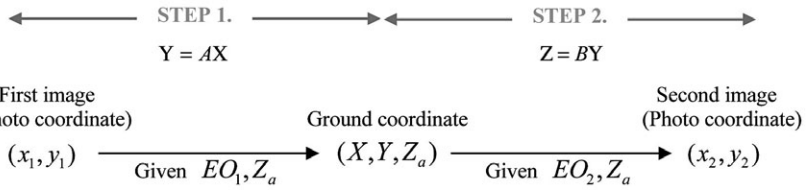


Fig. 7. The error propagation diagram.

image's parameters. Fig. 7 illustrates the diagram of this derivation.

These two-step diagram involves primarily the collinearity equation but in the invert order. In the first step, the computation involves the feature positions in the first image and its camera parameters in order to derive the corresponding ground coordinates. The indirect observation equation for the ground coordinate $P(X, Y, Z_a)$ written in another form of the collinearity equation is placed below for simply illustrating the solution.

$$\begin{aligned} X &= X_1^c + (Z_a - Z_1^c)\frac{r_{11}x_1 + r_{21}y_1 + r_{31}(-c)}{r_{13}x_1 + r_{23}y_1 + r_{33}(-c)} \\ Y &= Y_1^c + (Z_a - Z_1^c)\frac{r_{12}x_1 + r_{22}y_1 + r_{32}(-c)}{r_{13}x_1 + r_{23}y_1 + r_{33}(-c)} \end{aligned} \quad (5)$$

where the notation is used in the general mean as usual collinearity equation.

Every measurement has some uncertainty associated with it. The GPS/INS data, which are used for exterior orientation parameters, incorporated in this calculation are also subject to some uncertainty. The uncertainty from GPS corresponds to error at perspective center position: $s^2_{X_1^C}$, $s^2_{Y_1^C}$ and $s^2_{Z_1^C}$. Similarly, the uncertainty from INS corresponds to error at orientation of the camera: $s^2_{w_1}$, $s^2_{f_1}$ and $s^2_{k_1}$.

Since the collinearity equation is nonlinear, the Taylor series approximation is then applied to linearize it by taking the partial derivative of the equation (5), with respect to the unknown X_1^c, Y_1^c, Z_1^c, w_1, f_1, k_1 and Z_a. The design matrix A is given as the notation below for further reference.

$$A = \begin{bmatrix} \frac{\partial X}{\partial X_1^c} & \frac{\partial X}{\partial Y_1^c} & \frac{\partial X}{\partial Z_1^c} & \frac{\partial X}{\partial w_1} & \frac{\partial X}{\partial f_1} & \frac{\partial X}{\partial k_1} & \frac{\partial X}{\partial Z_a} \\ \frac{\partial Y}{\partial X_1^c} & \frac{\partial Y}{\partial Y_1^c} & \frac{\partial Y}{\partial Z_1^c} & \frac{\partial Y}{\partial w_1} & \frac{\partial Y}{\partial f_1} & \frac{\partial Y}{\partial k_1} & \frac{\partial Y}{\partial Z_a} \end{bmatrix}$$

The dispersion matrix of the parameters X_1^c, Y_1^c, Z_1^c, w_1, f_1, k_1 and Z_a to construct the propagation equation $Y = AX$ is formed as a 7×7 matrix containing the variance of the uncertainty of all parameters in respective order. Applying the general law of variance propagation (Ghilani and Wolf, 2006), the uncertainty of the ground coordinate (X, Y) is determined as

$$D\{Y\} = AD\{X\}A^T \quad (6)$$

The resulted uncertainty will be used further for deriving the uncertainty of the tie point in the second image as presented in the previous Fig. 7. For simply referring to related equation, we repeat the collinearity equation (4) and re-labeled the equation as (7). At this point, the indirect measurement of the photo coordinate of the tie point in the second image is defined by the collinearity equation below.

$$\begin{aligned} x_2 &= -c \cdot \frac{r_{11} \cdot (X - X_2^c) + r_{12}(X - Y_2^c) + r_{13}(Z_a - X_2^c)}{r_{31} \cdot (X - X_2^c) + r_{32}(X - Y_2^c) + r_{33}(Z_a - X_2^c)} \\ y_2 &= -c \cdot \frac{r_{21} \cdot (X - X_2^c) + r_{22}(X - Y_2^c) + r_{23}(Z_a - X_2^c)}{r_{31} \cdot (X - X_2^c) + r_{32}(X - Y_2^c) + r_{33}(Z_a - X_2^c)} \end{aligned} \quad (7)$$

where the notation is used in the general mean as usual collinearity equation.

The unknown is defined in a similar way as previous equation plus two additional unknowns for

the ground coordinates (X, Y) which are the outcome of the former derivation. The design matrix B is defined from the partial derivative of equation (7) with respect to the unknown X_2^c, Y_2^c, Z_2^c, ω_2, f_2, k_2 ground coordinate X, ground coordinate Y and Z_a.

$$B = \begin{matrix} \frac{\partial x_2}{\partial X_2^c} & \frac{\partial x_2}{\partial Y_2^c} & \frac{\partial x_2}{\partial Z_2^c} & \frac{\partial x_2}{\partial \omega_2} & \frac{\partial x_2}{\partial f_2} & \frac{\partial x_2}{\partial k_2} & \frac{\partial x_2}{\partial X} & \frac{\partial x_2}{\partial Y} & \frac{\partial x_2}{\partial Z_a} \\ \frac{\partial y_2}{\partial X_2^c} & \frac{\partial y_2}{\partial Y_2^c} & \frac{\partial y_2}{\partial Z_2^c} & \frac{\partial y_2}{\partial \omega_2} & \frac{\partial y_2}{\partial f_2} & \frac{\partial y_2}{\partial k_2} & \frac{\partial y_2}{\partial X} & \frac{\partial y_2}{\partial Y} & \frac{\partial y_2}{\partial Z_a} \end{matrix}$$

The dispersion matrix Y to construct the error propagation equation Z = BY is formed as a 9 ×9 matrix containing the variance of the uncertainty of X_2^c, Y_2^c, Z_2^c, ω_2, f_2, k_2 ground coordinate X, ground coordinate Y and Z_a. Based on the general law of variance propagation, the uncertainty of the tie point in the second image can be determined as

$$D\{Z\} = BD\{Y\}B^T \tag{8}$$

Finally, the candidates for tie-points residing in the second image is theoretically defined in the range of the below definition.

$$\begin{aligned} x &= x_2 \pm \sigma_x \\ y &= y_2 \pm \sigma_y \end{aligned} \tag{9}$$

where (σ_x, σ_y) is the uncertainty of the tie point as a result of equation (8).

With the assumption that the window size in x and y coordinate is equal, the number of the pyramidal depth level L can be approximated through the propagated error. Using the framework from Bouguet (2000), the ordinary point u is defined as $u^L = u/2^L$ on the pyramidal image I^L. The feasible number of the pyramidal depth level L is determined by the inequation below:

$$L > \log_2\left(\frac{\max\{\sigma_x, \sigma_y\}}{w}\right) \tag{10}$$

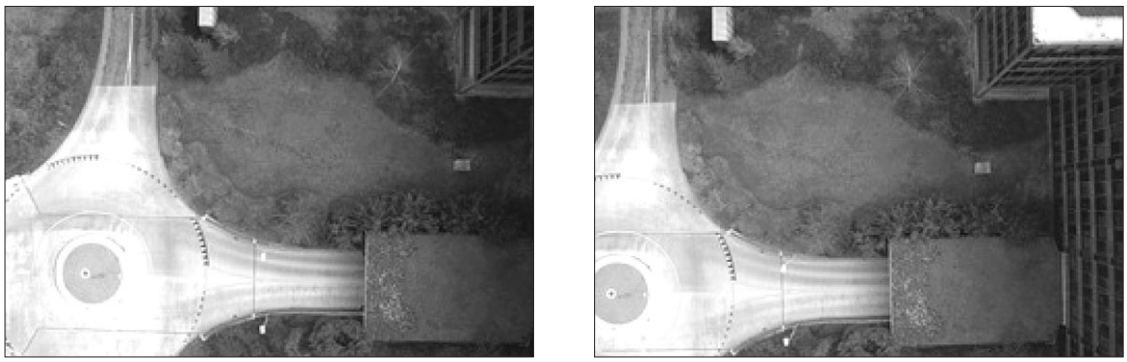
Based on the inequality plus the requirement for the number of resolutions to be integer, the quantity L must be rounded up to its nearest larger integer.

4. Experimental Results and Analysis

1) Requisite for Experiments

In this study, two experiments were conducted to evaluate the proposed solutions that can lead to a real-time image georeferencing. The first experiment presents an attempt to measure the processing time when the initial guessed positions to the optical flow vector are incorporated. The second experiment is to assess how efficient the proposed solution to estimate the suggested number of multi-resolution is. Experiments in this study were conducted on a pair of aerial images which was acquired from a UAV boarded approximately 70 m over the ground, as presented as Fig. 8. These two images are subsequent images retrieved from a sequence of aerial images in conjunction with the GPS/INS data. The sequential images are captured with an approximately 75% overlapping area.

Prior to the experiments, we employed the KLT feature extraction algorithm to obtain a set of significant features in the first image to be used as tie points, and they must be well distributed over the image. To complement the image georeferencing process, feature points were extracted based on the 3 ×3 pattern of the overlap configuration as presented as Fig. 9. Since the objective of the experiments is to measure the computational time of the solution, therefore, in this experiment, the KLT algorithm was implemented to extract at most 10 tie points per block. However, in practical, one tie point per block is sufficient. Due to the characteristic of the input image (Fig. 8(a)) and the quality of the extracted point, we obtained 57 significant points with quite well distributed.



(a) First image (b) Second image

Fig. 8. An input pair of consecutive aerial images used in the experiment.

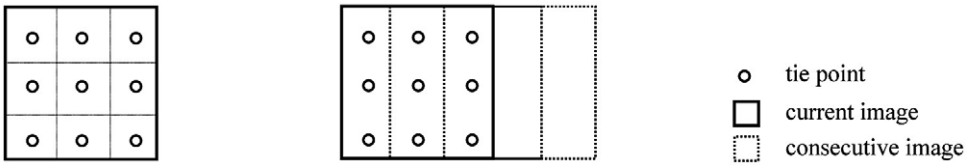


Fig. 9. A tie point pattern.

2) Effectiveness of Initial Guess on Computational Time

This experiment was conducted to measure the accuracy and the computational time of (a) the ordinary KLT tracker and (b) the improved KLT tracker which has been supplied initial guessed positions to the optical flow vector, in a comparison scheme. Prior to the computational time measurement, we started the experiment by measuring the displacements between two image frames for both (a) and (b) case. To determine the distances, we manually marked the positions in the second image that corresponded to those extracted points in the first image. The KLT tracker is considered to be accurate if the tracking results are, by far, less than 3 pixels from these manually extracted tie points. The 3-pixel threshold is, literally, large in term of accuracy. However, since the dimension of the input image is quite large (4288×2848), this threshold is quite reasonable and acceptable.

The displacement for case (a) is obtained as the distance between the good features extracted in the first image and the manually extracted corresponding point. For case (b), we first calculated the initial guessed positions for the corresponding points in the second image based on the collinearity equation given the exterior orientation parameters of both image. The distance is then measured from the position of the initial guessed corresponding point to the manually extracted point. The illustration of these two distances is presented as Fig. 10. Table 1 presents the displacements between tie points for case (a) and (b).

Originally, the displacement for case (a) measured between the positions of the good features and the manually extracted tie points is quite large in which its maximum Euclidean distance is up to 617 pixels. Using the collinearity equation given the camera parameters from the GPS/INS sensors, the distance measured between the initial guessed tie points and the manually extracted points is shortened in which its Euclidean distance is dramatically reduced to 210

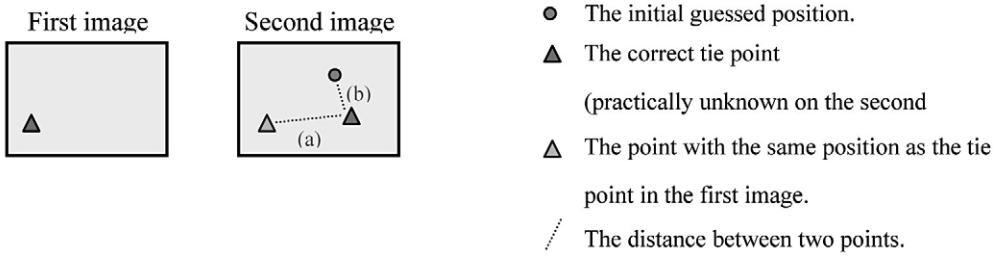


Fig. 10. The illustration of displacement measurement.

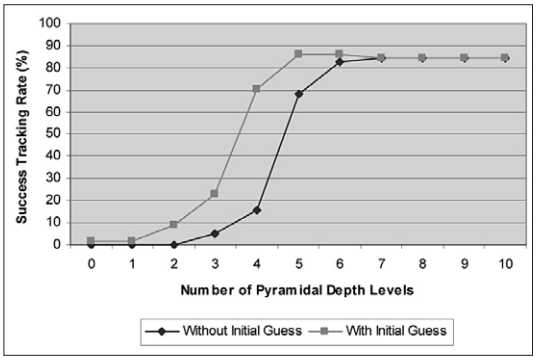
Table 1. Displacements measured between tie points

Condition	(a) Original displacement		(b) Shorten displacement	
Parameter	Average	Maximum	Average	Maximum
X	527.4	605.1	39.2	70.7
Y	105.6	171.7	141.1	197.7
Euclidean {X,Y}	537.1	617.4	146.1	210.0

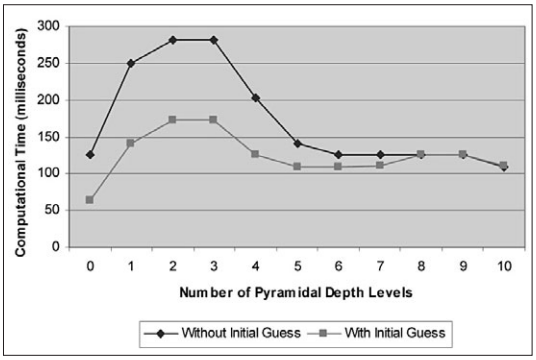
pixels. The presenting result strengthens our proposed solution to tremendously decrease the displacement between the input pair of subsequent images by employing the GPS/INS data.

The following experiment is to measure and compare the computational time consumed by the original KLT tracker and the speeded up KLT tracker. Since the displacements between the image frames are larger than the size of the local window, 10 in this study, both KLT trackers are required to implement in multi-resolution scheme. The number of pyramidal depth level is not yet determined in advance. Therefore, we deliberately performed the tracking process from 0 (at the original image size) to 10 depth levels. The tracking accuracy and computational time for both original KLT tracker and the proposed implementation at each depth level are measured and recorded for further comparison. The result is presented as Fig. 11.

As illustrated as Fig. 11(a), the original KLT tracker without the initial guessed positions for tie points can achieve its maximum success tracking rate at 84.2105% at the 7th pyramidal depth level. Employing the initial guessed positions based on the GPS/INS data, our proposed KLT tracker achieves the equivalent success rate using only 5 pyramidal depth levels. This measurement underscores the highest tracking capability of the KLT tracker at 84.2105%. Even with initial guesses, due to the ambiguity of the features over dimmed surfaces in conjunction with an oblique angle of cameras, the



(a) Measurement of success tracking rate



(b) Measurement of computational time

Fig. 11. Comparison of the KLT tracking measurements with and without guesses.

KLT tracker could not improve its matching capability. The experiment also suggests the appropriate number of resolutions for the ordinary KLT as 7, while with initial guesses the right number for the proposed KLT tracker is 5. The quantity is not dramatically reduced, however. Fig. 11(b) reveals the coming along computational time spent on the tracking process. At the maximum tracking rate, the original KLT running thorough 7 resolutions consumes 125 milliseconds for processing, while the proposed KLT tracker with initial guesses, tracking through 5 resolutions, consumes slightly smaller at 109 milliseconds or 12.8% speed improvement. The experimental results demonstrate our valid assumption that the initial guessed positions for tie points can effectively accelerate the tracking execution time. The processing speed, however, is not dramatically improved. This is according to the low accuracy of the GPS/INS sensors we employed that lead to the derivation of not sufficiently good quality of the initial guessed positions.

3) Accuracy of Suggested Number of Pyramidal Multi-resolutions

This experiment was conducted to suggest a promising number of pyramidal depth levels prior to the tracking process. As discussed in section 3.2, in advance of performing a real-time tie point extraction, the number of pyramidal depth levels must be determined. Consider the collinearity equation (4), all parameters involved usually associate with some uncertainties. In this study, only the uncertainties of the high impacted parameters are considered. The errors from the GPS sensor affected the accuracy of the camera's position: X^c, Y^c and Z^c, and the errors from the INS sensor affected the camera's orientation: `w`, `f` and `k`, as well as, the errors from estimating the average terrain elevation account for calculating the displacement between tie point based on the general law of variance propagation. Referring to the general specification of the GPS/INS sensors, the uncertainties involved are summarized as follows:

(a) The uncertainty of the GPS sensor is defined within 30 cm.

(b) The uncertainty of the INS sensor is defined within 0.5 degree.

(c) The measurement error from estimating the average terrain elevation is defined within 10 m.

The computation is performed in a point-by-point basis and this experiment picks 5 random points for this purpose. The propagated errors for each point are calculated and presented as Table 2. The table shows the errors propagated to the ground coordinate as the result of Step 1 calculation and the final error propagated to the tie points in the second image as the result of Step 2 calculation.

With the local window size set equally to 10 in both x and y axis, using equation (10) on the calculated results in Table 2, the computed results are

Table 2. Result of the propagated error calculation

Input	Estimated uncertainties					Output
Image coord (First image)	Step 1. Variance in ground coord (m^2)			Step 2. Error in second image coord (pixels)		No. Depth Levels
	$Var\{X\}$	$Var\{Y\}$	$Cov\{XY\}$	X	y	
(2446.94,397.67)	1.9045	5.1068	2.5983	82.7690	238.8860	4.5783
(1554.20,169.45)	0.5999	13.2078	-1.2092	240.8711	281.1260	4.8131
(2107.74,396.72)	0.7444	6.7926	1.3456	131.8865	237.0524	4.5671
(1809.80,797.92)	0.6772	4.5655	-0.9597	183.7800	154.6393	4.1999
(1783.13,169.01)	0.4915	11.5021	0.4033	195.0516	282.4609	4.8199

rounded up to the largest integer and the promising number of pyramidal depth level is suggested to 5. This Figure is corresponding to the result presented in Section 4.2 in which the proposed KLT tracker achieved its maximum tracking rate at 84% at its smallest depth level 5 and 109 milliseconds were consumed.

5. Discussion and Concluding Remarks

To establish a real-time image georeferencing, all input parameters must be determined in real-time. The intrinsic camera parameters can be obtained in advance from the camera calibration process while the control information can be derived instantaneously from the real-time GPS/INS data. Throughout the system, the possible bottleneck is the tie point acquisition process. In this paper, we presented a combine solution to develop a real-time image matching technique to obtain a set of tie points in less computational time. Due to its lightweight computational characteristics, the KLT tracker is utilized. The proposed work speeded up the ordinary KLT tracker by supplying the initial guessed tie-points obtained through the collinearity equation plus the GPS/INS data to shorten the tracking distance. The original KLT only works effectively well when the displacement between tie-point is small. Due to the period of exposure time plus the effect of wind, this distance is not always small. To drive an automated solution, this paper suggests the appropriate number of depth levels for multi-resolution tracking under large displacement using the knowledge of uncertainties in the measured GPS/INS data. The experimental results show that our suggested depth levels is promising and the proposed work could obtain tie-points faster than the ordinary KLT by 13% with equal or no less accuracy. The experimental results are promising enough to support that our improved KLT tracker can be effectively integrated into the real-time image georeferencing for further developing a real-time surveillance application.

Acknowledgment

This research was supported by a grant (07KLSGC03) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded by the Ministry of Land, Transport and Maritime Affairs. We also would like to acknowledge Dr. Armin Gruen and Dr. Henri Eisenbeiss from the Institute of Geodesy and Photogrammetry, ETH Zurich for the permissions to use their aerial images in our experiments

References

Bouguet, J. Y., 2000. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm. Technical Report, Intel Corporation, Microsoft Research Labs.

Bradski, G. and Kaehler, A., 2008. Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly, California, pp. 316-367.

CBC News, 2010. The world's worst natural disasters. [online] http://www.cbc.ca/world/story/2008/05/08/f-natural-disasters-history.html

Choi, K. and Lee, I., 2009. A Sequential AT Algorithm based on Combined Adjustment. Korean Journal of Geomatics, 27(6): 669-678.

Ghilani, C. D. and Wolf, P. R., 2006. Adjustment

Computations : Spatial Data Analysis. John Wiley & Sons, New Jersey, pp. 84-95.

Gruen, A. W., 1985. Adaptive Least Squares Correlation: A Powerful Image Matching Technique. South Africa Journal of Photogrammetry, Remote Sensing and Cartography, pp. 175-187.

Hong, J., Lee, I., Oh, T. and Choi, K., 2009. Data Fusion of LIDAR and Image Data for Generation of a High-quality Urban DSM. Proceedings of the 2009 Urban Remote Sensing Joint Event, Shanghai, May 2009.

Hubner, T., and Pajarola, R., 2009, Real-time Feature Acquisition and Integration for Vision-based Mobile Robots. ISVC09, USA, pp. 189-200.

Lucas, B. and Kanade, T., 1981. An Iterative Image Registration Technique with An Application to Stereo Vision. Proceedings of the International Joint Conference on Artificial Intelligence, pp. 674-679.

Lowe, D., 2004. Distinctive Image Features from Scale Invariant Keypoints. Int. Journal of Computer Vision, 60(2): 91-110.

Mikhail, E. M., 1976. Observations and Least Squares. University Press of America, New York.

Moallem, P., Faez, K., and Haddadnia, J., 2002. Fast Edge-based Stereo Matching Algorithms through Search Space Reduction. Proc. IEICE Transaction on Information and System, pp. 1850-1871.

Nakaguro, Y., Dailey, M., and Makhanov, S., 2007. SLAM with KLT Point Features. Proc. International Workshop on Advanced Image Technology, Thailand, pp. 262-267.

Rodriguez, J., Vos, F., Below, R., and Guha-Sapir, D., 2009. Annual Disaster Statistical Review 2008 - The Numbers and 23Trends, Centre for Research on the Epidemiology of Disasters (CRED). [online] http://www.cred.be/sites/default/files/ADSR_2008.pdf

Schenk, T., 1999. Digital Photogrammetry. TerraScience, Ohio, pp. 225-255, 381-405.

Shi, J. and Tomasi, C., 1994. Good Features to Track. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 593-600.

USGS, 2010. Magnitude 7.0 - Haiti Region. [online] http://earthquake.usgs.gov/earthquakes/recenteqsww/Quakes/us2010rja6.php

Wolf, P. R. and Dewitt B., 1999. Elements of Photogrammetry with Applications in GIS. McGraw-Hill, pp. 366-403.

Zivkovic, Z., 2004. Improving The Selection of Feature Points for Tracking. Pattern Analysis and Applications, 7(2): 144-150.