Journal of Engg. Research Vol. 7 No. (1) March 2019 pp. 1-19

A Low Cost Data Collection Approach to Pavement Mosaic Reconstruction

V. Khalifeh[1], A. Golroo[2], K. Ovaici[3] and T. Alipourfard[4]

(1) PhD Amirkabir University of Technology. Assistant Professor, Dept. of Civil and Environmental Engineering, Sirjan University of Technology, Kerman, Iran. Email: vahid.khalifeh@sirjantech.ac.ir

(2) Assistant Professor, Dept. of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran (Corresponding author). Email: agolroo@aut.ac.ir

(3) Assistant Professor, Dept. of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran. Email: kovaici@aut.ac.ir

(4) Ph.D. Student, Dept. of Geomatics Engineering, University of Tehran, Tehran, Iran. Email: tayebalipour@ut.ac.ir

ABSTRACT

Data collection is one of the most important and costly steps of pavement management systems. Traditional methods have been widely replaced with automated data collection vehicles due to their advantages such as safety, accuracy, precision, standardization, and repeatability. However, these vehicles are very expensive due to several high-cost sensors mounted on-board, which might not be financially efficient. The main goal of this paper is to propose a cost-effective data collection approach utilized to reconstruct the 3D model of a pavement surface, which can be utilized to evaluate pavement condition. For this purpose, an inexpensive sensor called Kinect V2 is applied including both cameras and infrared projector to capture depth data. Having calibrated the sensor and captured data, the color images were stitched together. Then, the depth data was added to the stitched images so that the 3D model of pavement was built. This approach makes a significant difference in terms of total cost of data collection for pavement distresses, in which their main feature is elevation such as roughness and rutting.

Keywords: Kinect V2, 3D Reconstruction, Scale invariant feature transform, Random sample consensus, Conformal coordinate transformation, Affine coordinate transformation.

INTRODUCTION

Roads always play a key role in the sustainable development of countries with regard to mobility of passengers and commodities. Roads should not stop operating due to some defects. Road maintenance is of significant importance to assure high quality services and continuous operability of roads. For this reason, an optimized maintenance plan should be developed to indicate appropriate time, treatment actions, and road section to be maintained over the life span of roads. With this regard, the concept of pavement management systems (PMSs) was developed (Shahin 2002).

The main concept of pavement management is to solve a multi-objective optimization problem, which should answer to three questions: Which road sections should be maintained? What type of treatment should be applied? When the road sections should be treated? This optimization should be conducted with regard to the increase in road overall condition and decrease in total life cycle costs (Golabi et al., 1982). The network level perspective as well as project level view is considered before any action is executed (Shahin, 2002).

Data collection is the main core of a PMS and is often the most expensive one. Pavement data can be collected using variety of methods. Generally speaking, there are two different methods used to collect pavement condition data including manual and automated (NCHRP, 2004; Findley et al., 2011). Each method has some advantages and disadvantages in terms of associated quality, time, and cost. So, there is a trade-off between collecting data at a high quality level and being cost effective.

Manual data collection methods employ an expert or a team of experts who can detect, recognize, and quantify distresses that exist on pavement (NCHRP, 2004; Findley et al., 2011). There is a certain amount of subjectivity and uncertainty in collected data due to expert judgments on distress types, severity, and density. Manual pavement data collection is time consuming and laborious, unsafe, unable to cover the entire road networks, and susceptible to lack of integrity.

Automated data collection methods usually deploy a vehicle with a series of sensors such as lasers, high resolution cameras, accelerometers, Global Positioning System (GPS), and radar to capture pavement condition. Automated data collection is classified into two categories, that is, semi-automated and fully automated (NCHRP, 2004; Findley et al., 2011). In order to differentiate between these two categories, two different steps in pavement condition assessment should be clearly defined: data collection and data processing. In the semi-automated approach, the data collection step is automated but the data processing including distress detection and definition in terms of type, severity, and density are manual, while, in the automated approach, both data collection and processing are automated (NCHRP, 2004; Findley et al., 2011). In summary, the automated data collection method is fast, accurate, precise, and repeatable. However, it is of high cost in terms of operation and maintenance.

Literature Review

Automated data collection vehicles have been widely employed by ministries of transportation and private companies around the globe. For instance, the Automated Road Analyzer (ARAN) is applied in North America (Tahberer, 2012), PAVUE is utilized in the European countries (Wang, 1999), Hawkeye is developed by the ARRB Engineering group in Australia (Novak, 1993), Komatsu is deployed in Japan (Fukuhara et al., 1990), and ROMDAS is used in New Zealand (Bennett, 1998). Most of the vehicles only capture and store the data on an on-board external drive. Data processing (e.g., image processing) is conducted in an office to acquire pavement condition information out of the collected data. Image processing is one of the most important techniques to assist pavement data processing. Pynn et al. developed an image processing method based on video images collected with a van to automatically detect cracks (Pynn et al., 1999). To detect and classify pavement cracks using captured images, also, the photogrammetry technique was proposed by Mustaffara et al. (2008).

The 3D surface reconstruction is an excellent approach in the measurement of pavement unevenness, e.g., roughness and rutting. It relies on 3D point clouds collected by laser scanners or stereo-vision algorithms using a pair of video cameras (Koch and Brilakis, 2011). Yu et al. proposed an integrated multi-sensor method for pavement 3D mapping (Yu et al., 2007). Li et al. developed a real-time 3D laser scanner to collect 3D pavement surface data (Li et al., 2010). An automated pavement data collection system has been introduced by Ouyang and Xu using a 3D camera and a structured laser light to obtain pavements transverse profile (Ouyang and Xu, 2013). Although the photometric techniques have been successfully deployed to reconstruct the 3D pavement surface, due to the immaturity or high cost they are not feasible (Grendy et al., 2011).

Microsoft Kinect is a notable inexpensive sensor able to generate real-time geometric feature, color, and audio data of the environment (Sanna et al., 2013). It consists of an infrared projector and infrared/color cameras that produce color and depth images (Microsoft Xbox Group, 2016). Although it is a multi-sensor device, it is inexpensive because of its mass production as being part of the Microsoft Xbox gaming console. Aiming at broadening the Xbox users beyond its usual gamer base, the company released the first generation of the Kinect in November 2010. Afterwards, in February 2012, they released Kinect V1 (for Windows). Then, the second generation (Kinect V2) was first released in 2014. Kinect V2 has a few advantages over Kinect V1. Table 1 summarizes the major differences between Kinect V1 and V2 (Gonzalez et al., 2015; Pagliari et al., 2014).

Kinect has been applied in various fields of study due to its mature techniques and affordable expenses. For instance, Kinect has been successfully employed for medical care studies especially in the field of rehabilitation. Lange et al. investigated and proved Kinect applicability in clinical use (Lange et al., 2012). Application of Kinect was studied in physical rehabilitation by Chang et al. The authors claimed that they can provide competitive motion tracking performance in comparison with other professional motion detection systems (Chang et al., 2012).

The application of Kinect as a navigation sensor for mobile robotics and depth data in indoor mapping was investigated (Khoshelham and Elberink, 2012; Oliver et al., 2012).

In the field of pavement condition data collection, Joubert et al. utilized a Microsoft Kinect and a high-speed USB camera as pothole detection tools (Joubert et al., 2011). Kamal et al. utilized Kinect as a sensor for pothole imaging and metrology (Kamal et al., 2016). Jahanshahi et al. applied Kinect as a data collection tool to detect potholes (Jahanshahi et al., 2013). They claimed that the automated data collection vehicles are very costly because of a series of expensive mounted sensors that may not be necessary to be applied at the network level. They declared that these vehicles usually

consist of sensors such as high-resolution digital cameras, which require advanced illumination systems to provide uniform lighting condition in the captured images. The digital camera and laser-illumination module, laser road-imaging system, on these vehicles cost about $150,000. Moreover, pavement surface profiler and laser sensors, which are commonly used for rutting-depth or surface-roughness measurement, cost about $130,000–$150,000. However, Kinect can minimize pavements data collection costs (Jahanshahi et al., 2013).

To sum up, a little attention has been paid by researchers to detect distresses on pavement using Kinect, only in terms of crack depth perception and pothole detection. However, reconstruction of pavement mosaic has not been conducted using Kinect because of its complexity. The complexity is related to recognizing detectable features on pavement surface in order to stitch adjacent depth images and build up continuous film of pavement surface to calculate distresses associated with height such as roughness and rutting.

Objective and Scope

This paper is aimed at deploying a cost-effective sensor with embedded camera and infrared projector called Kinect V2 to build up a 3D mosaic of pavement surface, which provides appropriate source of information to measure pavement distresses related to deformation such as roughness and rutting that can be ultimately employed for pavement management. The scope of this paper is limited to application of Kinect V2 mounted on a portable stand capturing data. The mosaic is developed based upon asphalt pavement.

METHODOLOGY

The first step was to calibrate color and depth cameras. The internal parameters of the cameras including focal length and principal point coordinates along with factors related to radial and tangential distortions were obtained to calculate real coordinate of each point of pavement surface. The second step was data acquisition. An optimum vertical distance of cameras from pavement surface was determined to maximize the accuracy of collected data. To achieve an adequate overlap between adjacent images considering the optimum vertical distance, a number of stations across and along pavement were determined. Deploying the optimum distance, color and depth images of pavement were collected by Kinect V2. The third step was to develop a 3D model of pavement. Due to lack of features of pavement surface, 3D modeling with conventional methods such as Iterative Closest Point (ICP) was not possible. Therefore, a novel approach was proposed and applied to develop a 3D model of pavement surface in this research. The flowchart of the research methodology is shown in Fig. 2.

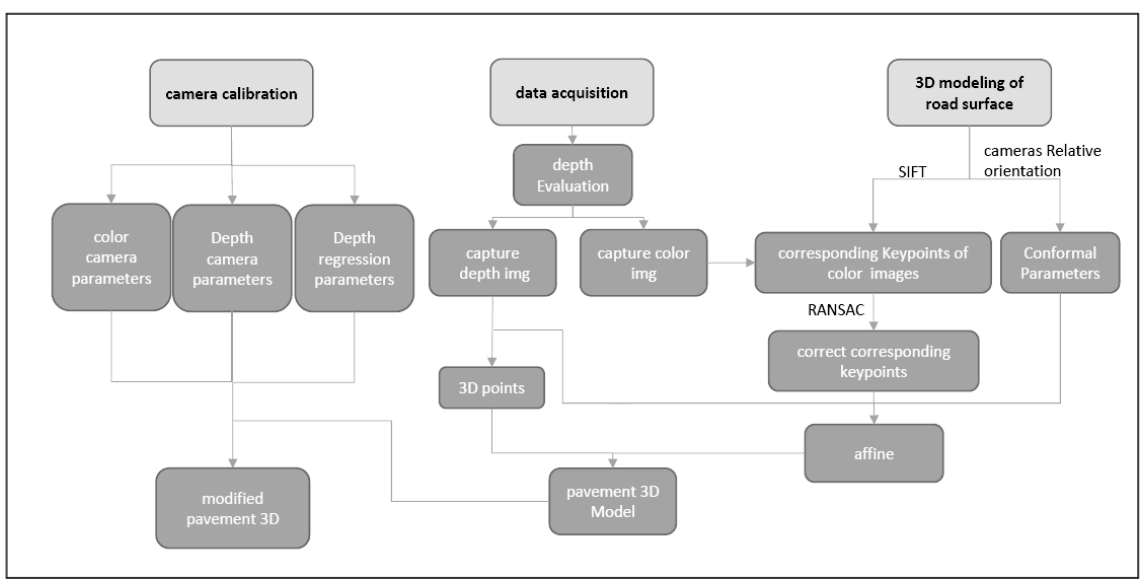

Fig. 1. Research Methodology.

Camera Calibration

Regardless of the quality/resolution of a camera and embedded lens, camera calibration was an essential step to measure internal parameters and interior orientation of the camera in order to adjust distortions caused by the lens. The prime parameters generally recognized include the following:

- *(f)*: camera focal length
- (x_P, y_P): principal point offset
- (k_1, k_2, k_3): radial lens distortion parameters
- (p_1, p_2): tangential distortion parameters

The camera focal length is a measure of how strong the system converges or diverges light. This length is a distance over which initially collimated rays are brought to a focus. A system with a shorter focal length has more optical power than one with a longer focal length. A line through camera center and perpendicular to a principal plain is the principal axis. A principal point is a perpendicular intersection point of a principal axis and an image plain, which is illustrated in Fig. 3 (Hartley and Zisserman, 2003).

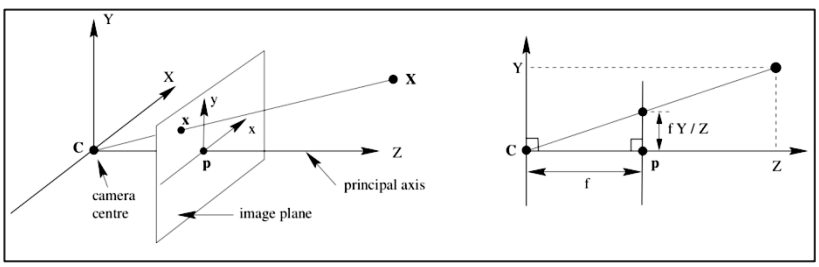

Fig. 2. Principle point and focal length (Hartley and Zisserman, 2003).

Lens distortions including radial and tangential are the main factors affecting camera calibration. Radial distortion causes image position to be distorted along a radial line from the optical axis, while tangential distortion/ decentering is due to imperfect centering of the lens components and other manufacturing defects. There are different methods applied for calibration. Self-calibration is the most practical one, in which camera interior parameters with lens distortion factors are simultaneously calculated (Fraser, 1997). The collinearity is the main mathematical term that be used in self-calibration. This mathematical term describes the relationship between any point in image space, the camera perspective center, and the same point in the object space. It assumes that the light ray is a straight line at the moment of exposure. That is, the exposure station, image point, and the object point must lie on a single straight line (ray). The mathematical form of the collinearity condition can be represented as below (Fryer and Brown, 1986):

$$x + \Delta x = x_p \quad f\frac{m_{11}(X - X_c) + m_{12}(Y - Y_c) + m_{13}(Z - Z_c)}{m_{31}(X - X_c) + m_{32}(Y - Y_c) + m_{33}(Z - Z_c)} \tag{1}$$

$$y + \Delta y = y_p - f\frac{m_{21}(X - X_c) + m_{..}(Y - Y_c) + m_{23}(Z - Z_c)}{m_{31}(X - X_c) + m_{32}(Y - Y_c) + m_{33}(Z - Z_c)} \tag{2}$$

$$\Delta x = \bar{x}(k_1 r^2 + k_2 r^4 + k_3 r^6 + ...) + p_1(r^2 + 2\bar{x}^2) + 2p_2\bar{x}\bar{y} \tag{3}$$

$$\Delta y = \bar{y}(k_1 r^2 + k_2 r^4 + k_3 r^6 + ...) + p_2(r^2 + 2\bar{y}^2) + 2p_1\bar{x}\bar{y} \tag{4}$$

$$\bar{x} = x - x_p \tag{5}$$

$$\bar{y} = y - y_p \tag{6}$$

where *(x, y, 0)* denote coordinates of an image point; *(x_p, y_p, f)* are interior orientation parameters (coordinates of a principal point and focal length); *(m11, ... , m33)* are elements of rotation matrix; *(X_c, Yc_c, Z_c)* denote coordinates of a perspective center; and *(X, Y, Z)* are coordinates of an object point. and are the image coordinate perturbation terms.

Calibration of color and depth cameras was conducted based on a standard method using two separate test fields. Regarding the color camera, 20 images from different angles and at different distances were taken from a planar chessboard as shown in Fig. 4. From each image, 77 corresponding points were extracted (intersection of grids) to implement collinearity condition accompanied with radial and tangential distortions.

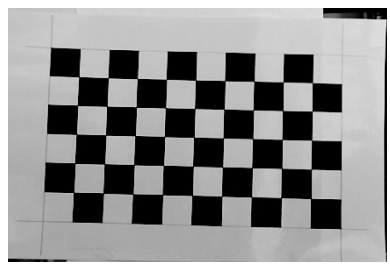

Fig. 3. Test field for color camera calibration.

However, in order to calibrate the depth camera, it was necessary that a plain with elevated targets was considered, which were clearly recognizable in the depth images. Therefore, three pairs of converged images were captured using a test field with 30 elevated targets to calibrate the depth camera as shown in Fig. 5. The distance, diameter, and elevation of targets were 20, 5, and 10 centimeters, respectively.

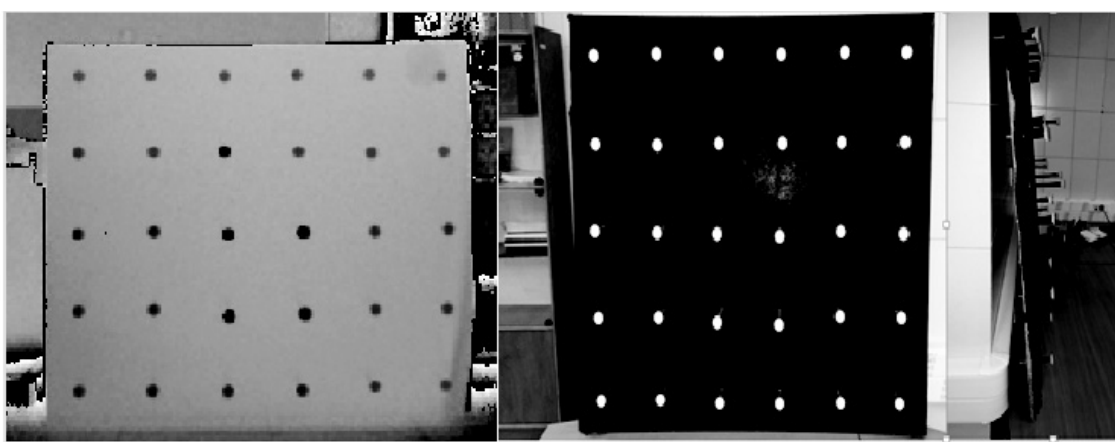

Fig. 4. Test-field for depth camera calibration (depth image (left), color image (middle), and targets (right)).

Using equation (7) depth camera calibration parameters can be computed.

$$R[x_i, y_i, f]^t + T = [X_i, Y_i, Z_i] \tag{7}$$

Equation (7) shows collinearity condition in which R and T are rotation and transition matrices, respectively. (x_i, y_i, f) are coordinates of points in an image space, (X_i, Y_i, Z_i) denote coordinates of points in an object space, and t is transpose. By substituting the corresponding points of two different images in Equation (7), the following equation is derived.

$$R_l[x_{il}, y_{il}, f]^t + T_l = R_r[x_i r, y_{ir}, f]^t + T_r \tag{8}$$

l and *r* represent the left and right images, respectively. Solving Equation (8) for all corresponding points of different images, cameras internal parameters and distortion were obtained in a bundle adjustment process.

Data Collection

Kinect V2 is employed to collect required data including depth and color images. The schematic hardware platform that was designed for data collection is illustrated in Fig 5. Accuracy of depth images is the most important stage in reconstruction of 3D modeling of pavement surface. Therefore, the depth images (i.e., from the camera to a reference plain) derived from Kinect V2 were compared to ground truth. The ground truth was measured applying a high accuracy laser rangefinder. This comparison was replicated ten times at the interval of 100 mm at the distance between 600 mm and 1500 mm. Using the Kinect, 20,000 depth points were extracted at each distance from one depth image. Evaluating ten images for each distance resulted in 200,000 depth points. Upper and lower thresholds (i.e., mean 3 standard deviation) were specified in each depth image to eliminate noises. Then, mean, standard deviation, median, mode, and error were determined for each distance (presented in Table 2) to indicate which distance has the highest accuracy in terms of estimating ground truth.

Fig. 5. Data collection hardware platform.

As clearly shown in Table 2, the optimum distance between Kinect V2 and the reference plain is equal to 1100 mm due to its lowest error. Based on this optimum distance, the scanned area was equal to 1400 *1150 mm. This optimum distance was applied for data collection to build a 3D model of pavement surface. The data collection was performed based upon 15 - 25% of image overlap. To achieve this overlap, the data collection stations were set up at the interval of one meter at both transverse and longitudinal directions.

Reconstruction of Road Surface

Due to lack of features of pavement surface, depth images cannot solely be registered. Therefore, a novel registering approach applying adjacent color images was proposed in this research. To register corresponding depth and color images, orientation of color and depth cameras was carried out. The 3D modeling approach consists of five steps, which are described below.

Align RGB and Depth Camera

The color and depth cameras in Kinect V2 are not coincident. Therefore, aligning two cameras is essential. To align color and depth cameras, the conformal coordinate transformation was used, which is the easiest method of transforming two coordinate systems (Bingqian, 2014). The conformal transformation has four parameters, two of which are related to the horizontal and vertical shifts, third one is related to rotation, and the fourth parameter is a scale factor. To compute the parameters, first, two color and depth images were taken from a same scene with both color and depth cameras. Then, the coordinates of the corresponding points in both images were specified. The conformal parameters were computed applying the following equations (Bingqian, 2014):

$$X_0 = m.x.cos(\alpha) - m.y.sin(\alpha) + \Delta x \quad (9)$$

$$Y_0 = m.x.sin(\alpha) + m.y.cos(\alpha) + \Delta y \quad (10)$$

where X_0 and Y_0 are the x and y transformed coordinates, m denotes the scale factor, Δx and Δy are translations (shifts), and α denotes rotation. Having assumed $a = m.cos\alpha$, $b = m.sin\alpha$, $c = \Delta x$, and $d = \Delta y$, Equations (9) and (10) are rewritten applying these assumptions as follows:

$$X_0 = ax + by + c \tag{11}$$

$$Y_0 = -bx + ay + d \tag{12}$$

The above mentioned equations can be represented in the matrix format as follows:

$$\begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} = \begin{bmatrix} x & y & 1 & 0 \\ y & -x & 0 & 1 \end{bmatrix} * \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \tag{13}$$

Regarding the fact that $tan(\alpha) = b/a$, the required angle of rotation was calculated. Also, the scale factor can be determined using Equation (14).

$$m^2 = a^2 + b^2 \tag{14}$$

To calculate the four aforementioned parameters *(i.e., a, b, c, and d),* coordinates of two corresponding points in the color and depth images were applied to write four equations. If the targets, hence the number of equations, are increased, the parameters can be found more accurately using the method of least squares. To specify the corresponding points in the color and depth images, it was necessary that a plain with elevated targets was considered so that they were clearly recognizable in the depth images. For this purpose, a plain shown in Fig. 5 was used to test the depth and color of images captured.

Registering Color Images

The first stage of registering two color images was feature extraction. Features utilized to match two images are too diverse. Matching methods are divided into two main categories: area based and feature based (Zitova and Flusser, 2003). In the area-based methods, features that are used in the matching process are the entire image pixels.

In the feature-based methods, sensible and specified features should be automatically detected and extracted. These features can be conventional (e.g., edges, lines, balance curves, and areas), prominent (e.g., corners and intersection of lines), and statistical (e.g., center of gravity). In feature-based methods, different descriptors are used, which should meet ideally the following requirements:

- Invariance: descriptors of corresponding features on the target image and the reference have to be the same.
- Uniqueness: two different features should have different descriptors.
- Stability: descriptors of a feature deformed slightly have to be close to the original descriptor of the feature.
- Independence: if the descriptor is a vector, its elements must be independent functions.

Although descriptors do not normally have these requirements at the same time, descriptors with most of these requirements should be used. A comprehensive assessment was conducted on performance of various descriptors (Mikolajczyk and Schmid, 2005). The authors proposed that the Scale Invariant Feature Transform (SIFT) descriptor expressed the best performance. The SIFT algorithm is a method to detect and extract independent and distinct features from the images that was developed by David Lowe in 1999. This algorithm is often utilized for applications such as object recognition, matching images, tracking and building 3D sceneries, object retrieval in multimedia databases, and autonomous robots. The SIFT algorithm is faster and more precise in terms of calculation than other algorithms. The main stages for revealing and extracting features based on the SIFT algorithm applied herein are as follows (Lowe, 2004):

Revealing extremums of scale space

For this purpose, the first target image was formed at different scales. Then, for each image scale called octave, several images with different standard deviations were constructed (Fig. 6). Images with different standard deviations were created by multiplying Gaussian kernel at the original image. After that, the difference between adjacent images was obtained per octave. The obtained results (using Equations 1517-) are called Difference of Gaussian images (DOG).

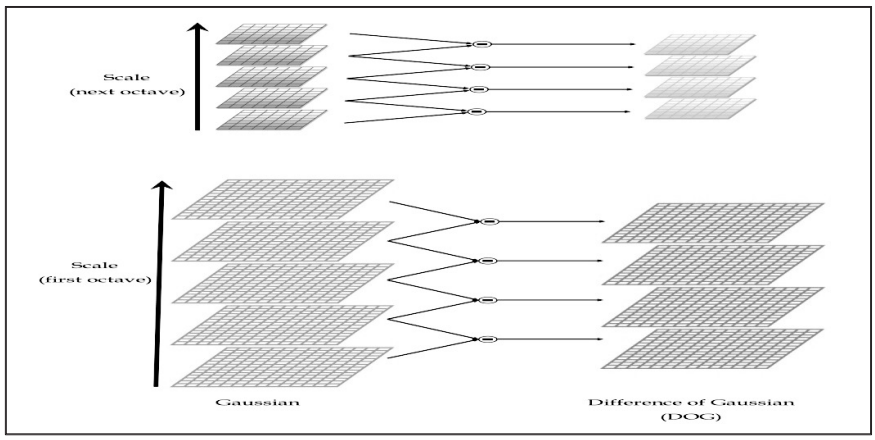

Fig. 6. Image pyramid formation using Gaussian function (Lowe, 2004).

$$\mathrm{L}(x, y, \sigma) = \mathrm{G}(x, y, \sigma) * \mathrm{I}(x, y) \tag{15}$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma\cdot}(e^{\frac{-(x^2+y^2)}{\cdot\,\sigma^2}}) \tag{16}$$

$$\mathrm{D}(x, y, \sigma) = \big(\mathrm{G}(x, y, k\sigma) - G(x, y, \sigma)\big) * \mathrm{I}(x, y) = \mathrm{L}(x, y, k\sigma) - \mathrm{L}(x, y, \sigma) \tag{17}$$

where I is the original image, G is the Gaussian kernel function, which produces image L with convolution multiplication, and k denotes a coefficient (i.e., integer number). By subtracting two resulting images per octave, image D is produced based on DOG.

Recognizing local extremums as key points:

The gray values of each pixel were compared with eight adjacent pixels as well as nine pixels in upper and lower adjacent images. The images with different σ were achieved by DOG as shown in Fig. 7. If the value of this pixel was more or less than all 26 neighbor pixels, it would be selected as a candidate/ key point.

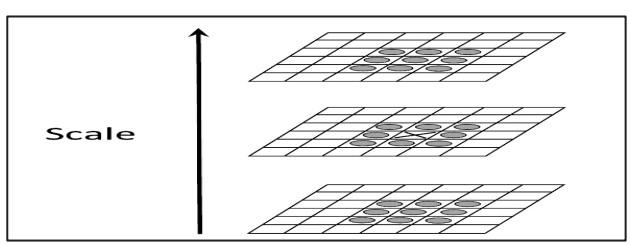

Fig. 7. Indicating key point (Lowe, 2004).

Determining the location of key points:

The next step after indicating the key points was matching them with adjacent data in terms of location, scale, and proportion of the original curves. Using this information, the key points with low contrast that were sensitive to noise were deleted. The key points located along the edge were also removed using the Hessian matrix (H) via Equation (18) (Lowe, 2004).

$$H = \begin{bmatrix} Dxx & Dxy \\ Dxy & Dyy \end{bmatrix} \tag{18}$$

Each derivative of Hessian matrix was obtained using the difference between neighbor points. Calculating the trace and determinant of the matrix H, the points that did not satisfy following condition were discarded from the set of the key points.

$$\frac{Tr(H)^2}{Det\,(H)} < \frac{(r+1)^2}{r} \tag{19}$$

where r is a result of division of maximum eigenvalue to minimum its values and D is the Difference-of-Gaussian function. The gradient value m(x, y) and also orientation θ(x, y) of each key point were determined using the following formulas (Lowe, 2004).

$$m(x,y) = \sqrt{\big(L(x+1,y) - L(x-1,y)\big)^2 + \big(L(x,y+1) - L(x,y-1)\big)^2} \tag{20}$$

$$\theta(x,y) = \arctan\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \tag{21}$$

As a result, an n*4 matrix was formed for all the key points.

$$\begin{bmatrix} x_1 & y_1 & m_1 & \theta_1 \\ x_2 & y_2 & m_2 & \theta_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & m_n & \theta_n \end{bmatrix}_{n*4} \tag{22}$$

Forming a descriptor of the key points:

To identify descriptors of the key points, first, windows with 16*16 pixels were selected around the key point locations. In the selected windows, gradient of each pixel was calculated using the adjacent pixels with size 4*4. Then, at the center of 16 pixels, the results of pixel gradients in four main directions and their bisectors were drown. The associated histograms were established in that each histogram includes eight branches. As a result, each descriptor included an array of four histograms. Therefore, each descriptor of SIFT was a vector with 8*4*4=128 elements. In Fig. 8, these operations are shown for the 8*8 window.

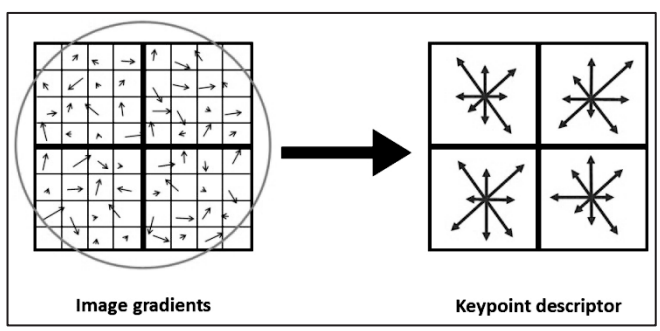

Fig. 8. Formation of descriptor for key points (Lowe, 2004).

Matching extracted features from two images:

A descriptor with 128 arrays of an original image was compared with all descriptor vectors of the key points in a target image using the inner product of two vectors. The concept of inner product according to the vector analysis is the same as the minimum Euclidean distance. The inner product of descriptor vectors from two images was a vector matrix of 1*n in which n is the number of the key points in a target image. Then, the matrix was sorted from the lowest values of an array to the highest values. For each descriptor in a target picture, the first and second close descriptors were selected as the descriptors for the main image. A pair of close descriptors were utilized to determine a pair of the corresponding key points. If the ratio of distance of the first minimum element to the second minimum element in inner product vector was lower than a threshold, the two key points in the two images as two corresponding points were selected. In this research, sensitivity analysis on the threshold was carried out and finally a value of 0.6 was selected.

Eliminating outliers with the Random Sample Consensus algorithm:

The Random Sample Consensus (RANSAC) algorithm is an iterative method to estimate parameters of a mathematical model from a set of observed data that contains outliers. This algorithm was introduced in 1981 by Fischler and Bolles. The RANSAC algorithm assumes that model parameters can be estimated with the correct data set in such a way that it optimally fits the data. The outliers occur because of noises, incorrect measurements, and false assumptions about the interpretation of data. The main difference between RANSAC and the method of least squares is related to the fitting approach. The least squares method fits an inaccurate line into all data including outliers. But, the RANSAC algorithm fits an accurate line on data after removing outliers.

Matching color images

After extracting the corresponding points from two adjacent color images, they were matched using the affine geometrical transformation, which was based on two transition parameters along x and y directions (Δx and Δy), one rotation parameter *(α)*, two scale factors along *x* and *y* directions *(mx and my),* and one parameter related to deviation from verticality (β) (Bingqian, 2014). These parameters were calculated using the following equation with six equations associated with three corresponding points:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} m_x & 0 \\ 0 & m_y \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin(\alpha+\beta) \\ -\sin\alpha & \cos(\alpha+\beta) \end{pmatrix} \times \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \tag{23}$$

(X, Y) is transferred coordinate of *(x, y)*. Having assumed $a = mx.\cos\alpha$, $b = -my.\sin(\alpha+\beta)$, $c = \Delta x$, $d = mx.\sin\alpha$, $e = my.\cos(\alpha+\beta)$ and $f = \Delta y$, Equation (23) was simplified and rewritten as follows:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{pmatrix} \times \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} \tag{24}$$

It is worth noting that the matching process was executed between only two images with an adequate overlap (i.e., more than 25% coverage). In order to register all color images, first the adjacent images were matched with each other across the pavement. After registering all color images in the transverse direction and combining them into a single image, these images were matched in the longitudinal direction to create a color image mosaic.

Depth data registration and reconstructing 3D model

After constructing the color image mosaic, each depth image was transferred to the space of corresponding color image using conformal parameters. Finally, registration of transferred depth images and formation of 3D point clouds were carried out based on affine transformation parameters achieved.

The depth data collected by Kinect V2 contained some noises. To remove the noises, two steps were taken. Firstly, the isolated data (single points) were removed. Secondly, having assumed a threshold (herein ±100 mm), data with less or more depth values than 1100 ± threshold were eliminated. Certainly, using different filters such as the median filter also resulted in smoothing of the depth data.

Results and Discussion

The proposed method was verified by taking some samples from pavement surfaces with different levels of roughness. The experiment was conducted in a portable static state condition using a single Kinect V2. Three main steps of the abovementioned approach including calibration process, color and depth data collection, and pavement 3D reconstruction were carried out on the samples as described below.

Calibration Process

The color camera was calibrated using a planar checkerboard with 20 pictures from different angles and distances, while the depth camera calibration was executed using elevated targets via taking three pairs of images in the field. In the calibration process, the internal orientation parameters (focal length and principle points) and distortion were simultaneously obtained based on collinearity condition. Table 3 shows the results of color and depth cameras calibration.

Applying calibration parameters on color and depth images resulted in calibrated images without any distortion effects. Calibration of depth data was also carried out through development of a correlation equation between depth data captured by Kinect V2 and the ground truth (measured via an accurate laser rangefinder). To develop this equation, the depth data was collected by Kinect V2 at ten vertical distances between Kinect V2 and pavement from 600 to 1500 mm (at an interval of 100 mm). The mean of 1,000 points randomly extracted from each image was defined as a calculated depth data and plotted against the ground truth. Several curves were fitted to the data. The best fitness was proposed by a quadratic polynomial function as shown in Fig. 9.

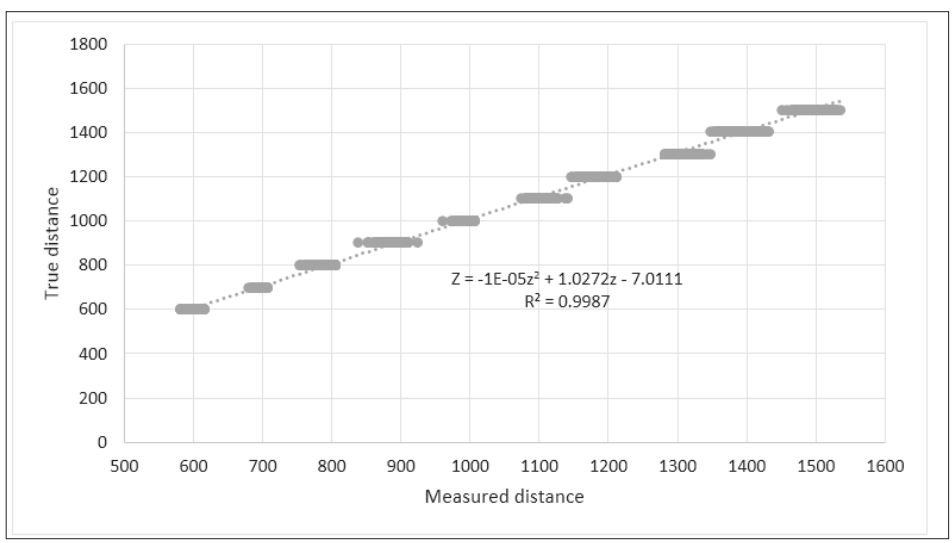

Fig. 9. Diagram of real depth value versus measured distance by Kinect V2.

As Fig. 9 shows, the relationship between the calculated depth data and ground truth is as follows:

$$Z = -0.00001z^2 + 1.0272z - 7.0111 \quad (25)$$

Z and z specify ground truth and calculated depth data, respectively. Using Equation 25, the calculated depth data was corrected to be applied in 3D modeling.

Data Collection

Design of experiment was carried out with regard to the fact that a variety of pavement roughness levels were captured to ensure the proposed methodology was valid in different roughness conditions. For this purpose, a pilot study was performed to indicate appropriate pavement sections. Then, numerous images were taken from the sections at the optimum vertical distance, i.e., 1100 mm and the horizontal interval of one meter along and across the pavement (25% overlap of adjacent images). At each station, three depth images were captured during daytime without direct sunlight. The mean depth was utilized for further calculation resulting in reducing noise and increasing accuracy of depth data. Fig. 10 shows the color and depth images taken from six adjacent stations across the pavement, which were registered.

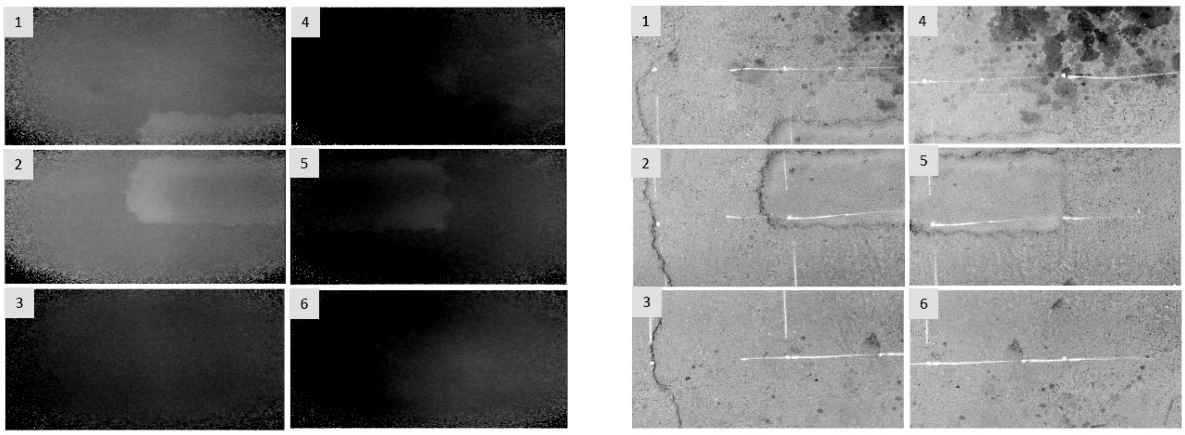

Fig. 10. Data samples collected by Kinect: (left) depth images; (right) corresponding color images.

Pavement 3D Reconstruction

Aligning the color and depth cameras was the first step in 3D reconstruction of pavement surface. As mentioned earlier, conformal geometry transformation was applied for orientation of cameras. This transformation was simple and provided higher accuracy than other transformations to align the cameras. To calculate the conformal parameters, it was necessary to determine the coordinates of at least two corresponding points in two color and depth corresponding images. If the targets, hence the number of equations, were increased, the parameters would be indicated more accurately using the method of least squares.

After the color and depth images were captured from the test field with elevated targets, the method of least squares was used to determine the unknown parameters. Points in the two corresponding images were identified through application of the Australis software (Photometrix Company 2015) presented in Table 4. The related conformal parameters are summarized in Table 5. Regarding the coordinate transformation of color and depth cameras, it was assumed that the color image was the reference space and the top left corner of color images was the original coordinate. Also, the x and y axes were from left to right and top to bottom directions, respectively.

The SIFT algorithm was employed for registration of color images. The SIFT first extracted all the key points in the images and then determined the corresponding points in two adjacent images as shown in Fig. 11. Having employed the RANSAC algorithm, the corresponding key point outliers were eliminated. Table 6 shows extracted key points by SIFT and effects of the RANSAC algorithm to discard outliers.

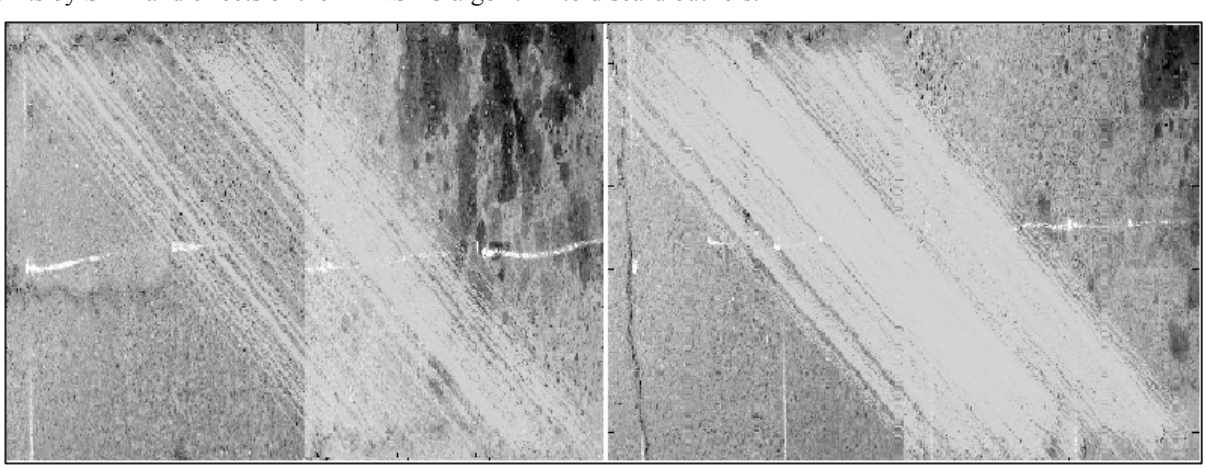

Fig. 11. Corresponding key points in two adjacent images. Images 1 and 2 (left) and images 4 and 5 (right).

As shown in Table 6, the number of key points in each image extracted by the SIFT algorithm is much more than the number of corresponding key points. Also, the RANSAC algorithm reduces the number of corresponding points by eliminating outliers. The number of key points in images 1, 2, and 3 is 22479, 30679, and 29365, respectively. The SIFT detected only 178 points as corresponding key points for registration of images 1 and 2. Likewise, 93 points were indicted as corresponding key points for registration of images 2 and 3, while 97 and 28 points were removed by RANSAC from 178 corresponding points in images 1 and 2, and 93 corresponding points in images 2 and 3, respectively. That is, in the first step of the registration of color images within the pavement sections, around 40 percent of points were discovered incorrectly as key points. But, by repeating the above process and increasing the number of pixels in registered images (12- and 23- integrated images), the number of corresponding key points was increased and the percentage of error was drastically decreased. This fact is also visible in registration of other images. To sum up, it was concluded that, in the beginning of the registration process of color images, correct key points were the least accurate.

In the registration process, only two adjacent color images were jointly registered at a time. Therefore, registration of all images required an iterative process. Fig. 12 illustrates this process.

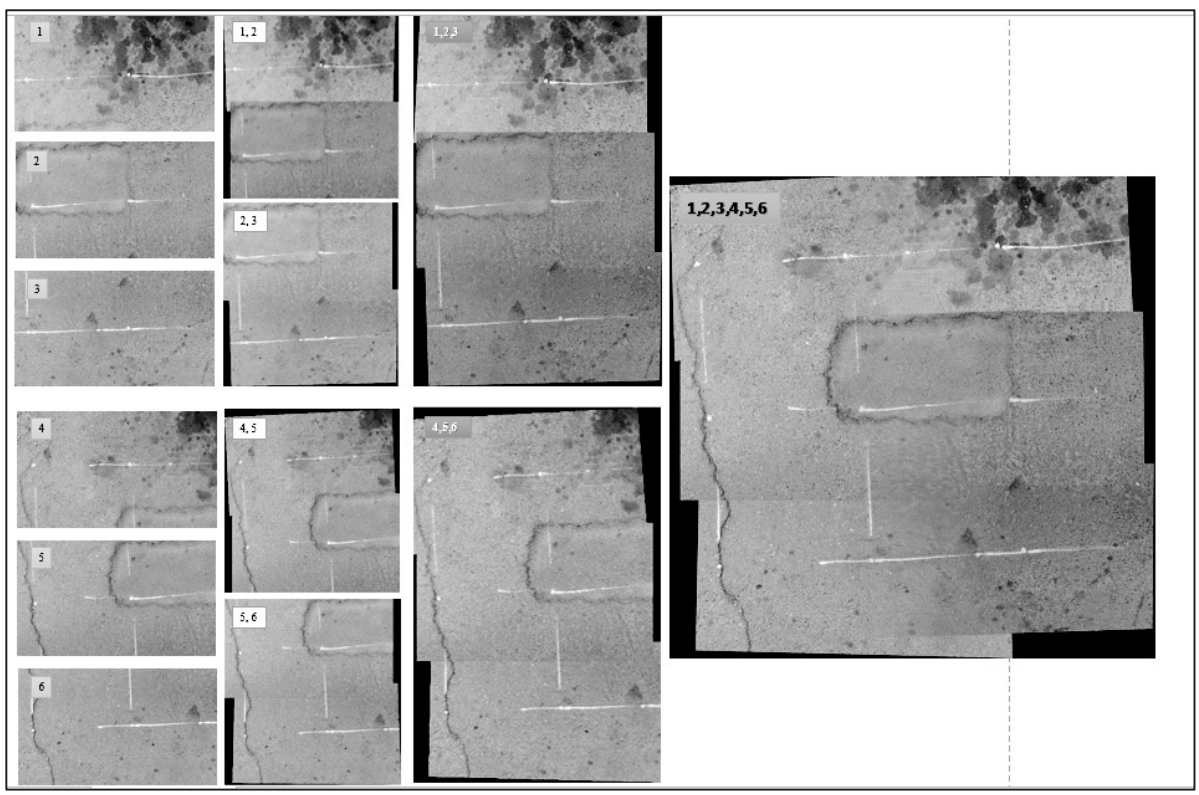

Fig. 12. Matching process of six color images using SIFT and RANSAC.

Determining of affine geometric transformation parameters was an output of color image registration that plays a crucial role in depth images matching. Correct corresponding key points were utilized to calculate the affine transformation parameters. Affine transformation parameters for some paired images are shown in Table 7.

Then, registration of depth images was carried out as depicted in Fig. 13. Finally, 3D reconstruction of pavement surface was accomplished. As illuminated in Fig. 13, the 3D model of pavement was made with high accuracy. Irregularities in the pavement surface including existing pothole are quite evident.

One of the most important applications of 3D models reconstruction based on registration of color images is to compute pavement roughness such as International Roughness Index (IRI), which is an ongoing research being conducted by the authors. Furthermore, these color images are of significant importance to recognize distresses such

as longitudinal and transverse cracks and measure their severity and density more accurate than 3D models. For instance, as shown in Figure 13, a transverse crack and a pothole were present on the pavement. Having utilized the color images of the pavement, length of crack and dimension of pothole were determined; i.e., crack length was 2500 mm and pothole dimensions were 1600 x 900 (length, width). But, the depth of the pothole should be calculated using 3D models. Therefore, more distresses can be identified on pavement surface using data derived from both 3D models and color images.

A single drawback of the proposed approach is to be time consuming due to a massive number of computations related to two described processes. The first one is color image registration, which is an iterative process. In each iteration, the number of pixels in the registered images is increased. Therefore, the SIFT algorithm spends significant amount of time to find the corresponding points. The second time-consuming process is the limitation of color image registration; i.e., only two images can be matched at a time so that the more pavement length/color images, the more computational time required. However, the time-consuming issue can be conquered through application of a computer with powerful and parallel processors.

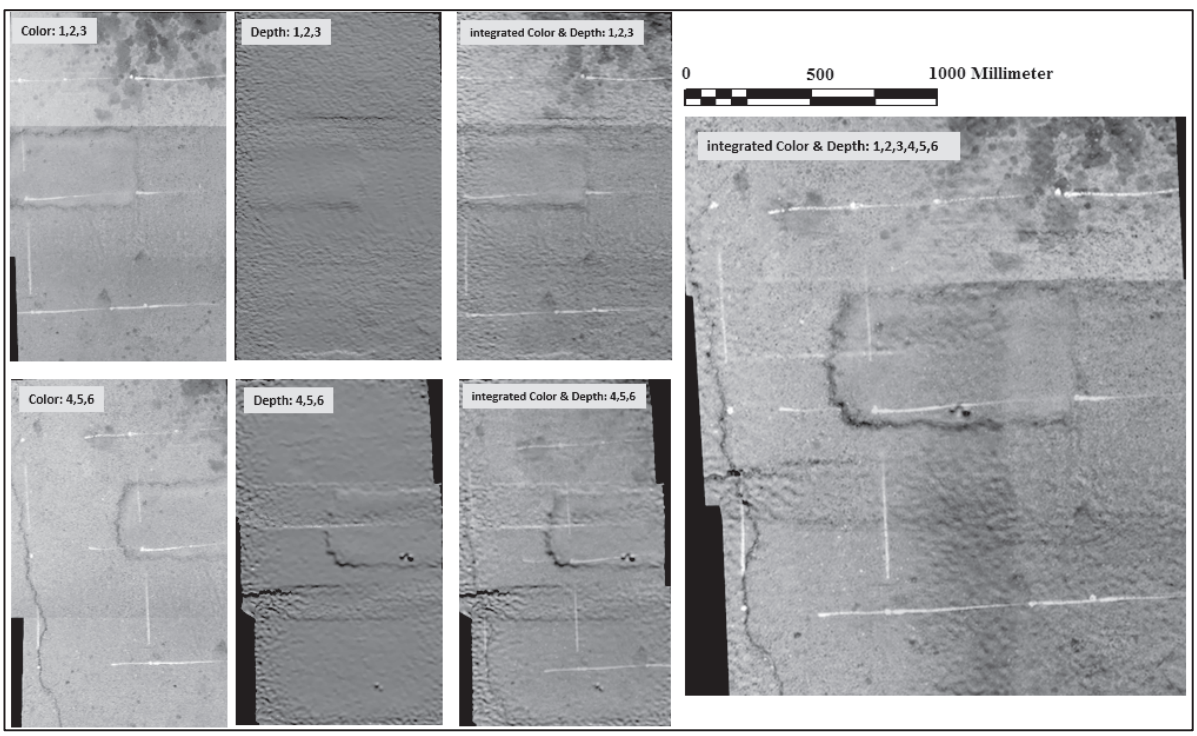

Fig. 13. (Left) color images matched, (middle) corresponding depth image matched, and (right) 3D pavement mosaic for 3 and 6 images.

CONCLUSION

Data acquisition is the core of pavement management systems, which is costly and time consuming. To date, several automated data collection methods have been studied, which reduce the duration of data collection but have not been very successful in decreasing costs. In this research, a cost-effective approach was presented for a 3D reconstruction of pavement surface. The proposed approach employed remarkably inexpensive Microsoft Kinect V2 to collect pavement surface data. Registering depth data and 3D modeling of pavement surface with conventional methods such as Iterative Closest Point (ICP) was not possible due to complexity of the pavement. Therefore, a novel approach was proposed based on color image registration. Since the color image and depth cameras did not conform, they were aligned using the conformal geometrical transformation. To register the color images, the SIFT algorithm was employed. The outliers in the color images were discarded using RANSAC algorithm. Having transferred the depth

data into the space of the color data through conformal parameters, the affine geometrical transformation resulted in the conformity of the depth data. This led to a 3D model of the pavement surface in a continuous film wherein the 3D coordinates of any image point were obtainable. Experimental field data showed that the proposed approach was capable of providing results with high accuracy. It is concluded that Kinect V2 has the potential to be deployed as a cost-effective tool in pavement surface modelling. This model can be easily utilized to calculate pavement roughness and other pavement distresses such as potholes.

FUTURE WORKS

The effects of each parameter in the registration of color images in the SIFT algorithm such as the number of octaves should be examined. And, the optimum value for each parameter must be indicated. In addition, pavement management indexes like International Roughness Index (IRI) based on proposed approach should be evaluated and validated by automated data collection vehicles.

ACKNOWLEDGMENT

The authors extend their appreciation to close range photogrammetry lab and its director, Mohammad Saadatseresht, and thank Ahmad Mahmoodzadeh and Saina Firoozi Yeganeh for data collection cooperation.

REFERENCES

Bennett, C. R. (1998). "Evaluation of a High Speed Transverse Profile Logger". *In 4th International Conference on Managing Pavements, South África.*

Bingqian, L. (2014). "Transformation Optics Methodology Review and ITS Application to Antenna Lens Designs". *master dissertation in Mathematics and Electrical Engineering, Pennsylvania State University.*

Findley, D. J., Cunningham, C. M. & Hummer, J. E. (2011). "Comparison of mobile and manual data collection for roadway components". *Transportation Research Part C: Emerging Technologies,* **19**(3), 521–540.

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.

Fraser, C. S. (1997). "Digital camera self-calibration". *ISPRS Journal of Photogrammetry and Remote sensing,* **52**(4), 149–159.

Fryer, J. G. & Brown, D. C. (1986). "Lens distortion for close-range photogrammetry". *Photogrammetric engineering and remote sensing,* **52**(1), 51–58.

Fukuhara, T., Terada, K., Nagao, M., Kasahara, A. & Ichihashi, S. (1990). "Automatic pavement-distress-survey system". *Journal of Transportation Engineering,* **116**(3), 280–286.

Geng, J. (2011). "Structured-light 3D surface imaging: a tutorial". *Advances in Optics and Photonics,* **3**(2), 128–160.

Golabi, K., Kulkarni, R. B. & Way, G. B. (1982). "A statewide pavement management system". *Interfaces,* **12**(6), 5–21.

Gonzalez-Jorge, H., Rodríguez-Gonzálvez, P., Martínez-Sánchez, J., González-Aguilera, D., Arias, P., Gesto, M. & Díaz-Vilariño, L. (2015). "Metrological comparison between Kinect I and Kinect II sensors". *Measurement,* 70, 21–26.

Hartley, R. & Zisserman, A. (2003). *"Multiple view geometry in computer vision"*. Cambridge university press.

Jahanshahi, M. R., Jazizadeh, F., Masri, S. F. & Becerik-Gerber, B. (2012). "Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor". *Journal of Computing in Civil Engineering,* **27**(6),743–754.

Lowe, D. G. (2004). "Distinctive image features from scale-invariant keypoints". *International journal of computer vision,* **60**(2), 91–110.

Microsoft Xbox Group. (2010). http://www.xbox.com/en-US/Kinect/default.htm (July 15 2015.

Mikolajczyk, K. & Schmid, C. (2005). "A performance evaluation of local descriptors". *IEEE transactions on pattern analysis and machine intelligence,* **27**(10), 1615–1630..

NCHRP Synthesis 334. (2004). "Automated Pavement Distress Collection Techniques". Transportation Research Board.

Novak, K. (1993). "Real-Time Mapping Technology". *International Archives of Photogrammetry and Remote Sensing,* 29, 569–569.

Pagliari, D., Menna, F., Roncella, R., Remondino, F. & Pinto, L. (2014). "Kinect Fusion improvement using depth camera calibration". *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* **40**(5), 479.

Photometrix Company Pty Ltd. <http://www.photometrix.com.au> (July 17 2015).

Shahin, M. Y. (2005). Pavement management for airports, roads, and parking lots (Vol. 501). New York: Springer.

Tahberer M (2012) "Supply of Automated Roadway Survey Vehicle". Fugro RoadWare Inc.

Wang, K. & Li, X. (1999). "Use of digital cameras for pavement surface distress survey". *Transportation Research Record: Journal of the Transportation Research Board,* (1675), 91–97.

Zitova, B. & Flusser, J. (2003). "Image registration methods: a survey". *Image and vision computing,* 21(11), 977–1000.

Table 1. Difference between Kinect V1 & V2 (Gonzalez et al., 2015; Pagliari et al., 2014).

Feature	Kinect V1	Kinect V2
Method of calculate depth of object in scene	Structured light	Time of flight
Dimension of Depth camera (pixel)	640*480	512*424
Dimension of Color camera (pixel)	640*480 @ 30 fps	1920*1080 @ 30 fps
Horizontal field of view	57 degrees	70 degrees
Vertical field of view	43 degrees	60 degrees
Tilt motor	Yes	No
Max depth distance	4 m	8 m
Min depth distance	80 cm	50 cm

Table 2. Descriptive statistics of distances measured by the Kinect.

Ground Truth (1)	Kinect				Error = ((1)-(2))/(1)
	Mean Observed Depth (2)	Standard Deviation	Mode	Median	
600	595.5	8.2	587	595	0.008
700	696.7	3.3	697	697	0.005
800	788.3	9.4	790	789	0.015
900	891.4	7.6	894	892	0.010
1000	994.5	6.2	998	995	0.006
1100	1098.8	7.0	1100	1098	0.002
1200	1182.3	11.2	1173	1182	0.015
1300	1296.5	8.6	1298	1296	0.003
1400	1387.8	13.4	1384	1387	0.009
1500	1492.3	12.6	1489	1491	0.005

Table 3. Parameters of color and depth camera calibration.

Type	Parameters	Color camera		Depth camera	
		Mean	Standard Deviation	Mean	Standard Deviation
Internal orientation	Focal length (x)	1047.92	4.54	363.90	1.52
	Focal length (y)	1048.46	4.73	365.60	1.83
	Principal point (x)	981.16	3.48	252.70	1.78
	Principal point (y)	538.19	2.96	203.10	2.12
Distortion	k1	0.01053	0.00835	0.08533	0.01437
	k2	0.05968	0.03519	-0.18742	0.06331
	k3	0.00000	0.00000	0.00000	0.00000
	p1	0.00255	0.00099	-0.00162	0.00242
	p2	0.00535	0.00127	-0.00268	0.00198

Table 4. Corresponding points in color and depth images from elevated test-field.

#	Color points		Depth points		#	Color points		Depth points	
	X	Y	X	y		X	Y	x	y
1	723.56	326.54	163.85	133.02	11	1488.77	471.00	430.45	184.91
2	873.45	323.63	216.38	132.37	12	723.76	637.35	162.96	241.82
3	1027.40	324.36	270.09	132.96	13	882.22	631.48	218.73	239.80
4	1179.36	321.14	323.07	132.01	14	1038.98	636.86	272.80	241.95
5	1330.49	319.48	375.90	131.76	15	1196.88	630.85	327.95	240.15
6	723.53	477.71	163.38	185.93	16	1343.97	628.39	379.96	239.56
7	880.09	475.45	218.49	185.38	17	725.96	789.17	163.31	295.01
8	1029.86	472.99	269.94	184.84	18	883.28	785.32	218.68	293.76
9	1184.00	472.94	324.37	185.06	19	1042.85	798.44	273.74	298.53
10	1335.87	474.41	377.55	185.81	20	1200.93	792.47	329.05	296.71

Table 5. Calculated Conformal parameters.

a	b	C	d
0.0153	2.8639	252.8349	-52.1486

Table 6. Key points of samples resulted from SIFT and effect of RANSAC.

image #	SIFT key points	SIFT corresponding key points	corresponding key points after running RANSAC
1	22479	178	81
2	30679		
2	30679	93	65
3	29365		
1,2	43863	24485	24482
2,3	57900		
4	33451	688	419
5	30359		
5	30359	108	58
6	29847		
4,5	53139	28017	28014
5,6	55257		
1,2,3	65375	1746	377
4,5,6	59415		

Table 7. Calculated affine parameters for some sample images.

image #	Affine Geometrical Transformation					
	A	b	C	d	e	f
1	0.9877	-0.0348	-10.0356	0.007321	0.978038	-943.241
2						
2	0.9916	-0.0012	75.9639	-0.01649	0.970309	981.246
3						
1,2	1.0000	3.98E-07	19.9993	-6.6E-07	1.0000	936.0015
2,3						
4	0.9886	0.0448	-83.6901	-0.03931	0.996217	-824.105
5						
5	0.9604	0.0047	107.0955	-0.00636	0.981134	926.3332
6						
4,5	1.0000	-5.5E-07	63.0000	3.09E-08	1	891.0011
5,6						
1,2,3	0.9918	-0.0429	-924.5230	0.025782	1.04447	-79.6267
4,5,6						

SUBMITTED: 02/06/2017
ACCEPTED : 08/01/2018

جمع البيانات لإعادة بناء حجارة الرصيف من الموازييك بواسطة نهج منخفض التكلفة

ف. خليفة[1]، أ. غولرو[2]، ك. أوفتشي[3]، تي. ألبورفارد[4]

(1) قسم الهندسة المدنية والبيئية، جامعة سيرجان للتكنولوجيا، كرمان، إيران

(2) قسم الهندسة المدنية والبيئية، جامعة أمير كابير للتكنولوجيا، طهران، إيران

(3) قسم الهندسة المدنية والبيئية، جامعة أمير كابير للتكنولوجيا، طهران، إيران

(4) قسم هندسة الجيوماتكس، جامعة طهران، طهران، إيران

الخـلاصة

جمع البيانات هو أحد الخطوات الأكثر أهمية والأعلى تكلفة في أنظمة إدارة عمليات الرصف. فقد تم استبدال الطرق التقليدية على نطاق واسع وذلك باستخدام مركبات لجمع البيانات اتوماتيكياً بسبب مزاياها المتعددة، مثل: السلامة والدقة والانضباط وتوحيد المقاييس وقابلية التكرار. ومع ذلك، فإن هذه المركبات باهظة التكاليف نظراً لوجود العديد من أجهزة الاستشعار عالية السعر والتي تكون مُثبتة على متن المركبة مما يجعلها غير عملية من الناحية المالية. إن الهدف الرئيسي من هذا البحث هو اقتراح نهج منخفض التكاليف لجمع البيانات المُستخدمة في إعادة بناء نموذج ثلاثي الأبعاد لسطح الرصيف والذي يمكن استخدامه لتقييم حالته. ولهذا الغرض، تم استخدام جهاز استشعار منخفض التكلفة يسمى Kinect V2 وهو يحتوي على الكاميرات وجهاز عرض يعمل بالأشعة تحت الحمراء لالتقاط بيانات العمق. وبعد معايرة جهاز الاستشعار والبيانات المُلتقطة، تم تجميع الصور الملونة معاً. وتمت إضافة بيانات العمق إلى الصور المُلتقطة؛ ومن ثم تم بناء نموذج ثلاثي الأبعاد للرصيف. وهذا النهج يُحدث فرقاً كبيراً من ناحية التكلفة الإجمالية لجمع البيانات عن عيوب الأرصفة والتي تتميز بشكل أساسي بالرفع ، مثل: الخشونة وتشوه الطبقات.