

# FUSION OF OPTICAL AND TERRESTRIAL LASER SCANNER DATA

Julien Li-Chee-Ming, Costas Armenakis

Department of Earth and Space Science and Engineering  
Geomatics Engineering, GeoICT Lab  
York University, Toronto, Canada  
{julienli} {armenc} @yorku.ca

Commission I, WG I/5

**KEY WORDS:** Terrestrial Laser Scanning, Photogrammetry, Sensor Registration, Point Texture Mapping

## ABSTRACT :

Optical imagery and range data can be registered to create photo-realistic scene models via texture mapping. Presented in this paper is an alternative approach where true colour (RGB) point clouds are generated by automatically fusing a close-range optical (RGB) image acquired with an uncalibrated digital camera with the corresponding high-density 3D lidar point cloud collected with a terrestrial laser scanner (TLS). The alignment of optical pixel colour values and lidar point cloud is obtained by estimating the position and orientation of the camera with respect to the lidar point cloud reference system. To perform this sensor co-registration, an automated corner feature extraction algorithm, followed by area-based image matching is applied between the optical data and the lidar intensity image to establish point correspondence. The matching process is solely based on point matches and does not use external control or calibration patterns. The 3D lidar points of the corresponding lidar intensity image corner points are then extracted from the point cloud. As these 3D lidar points correspond to the extracted optical image corner points, a bundle self-calibration adjustment with additional parameters is applied using the extended collinearity equations to estimate the interior and exterior orientation of the camera. The RANSAC robust estimator is used to reduce the influence of outliers in the estimation of the camera parameters. Having established the mathematical relationship between image space and lidar points a photo-realistic 3D model is generated. Through reverse mapping, each point in the lidar point cloud is assigned the RGB value of the image pixel upon which it is projected. Experiments are performed observing typical urban scenes, particularly building facades. The feasibility and potential of estimating the co-registration parameters using a TLS is evaluated in terms of accuracy of the results. The true calibration parameters, provided by the TLS manufacturer, are used in the validation of the registration parameters. The technique has reliably aligned a camera with the TLS geometry for the simultaneous generation of point based photo-realistic 3D models.

## 1. INTRODUCTION

Geometrically and photo-realistically accurate three dimensional (3D) building models have many applications in geomatics and other disciplines. Examples include aids in mapping and navigation, city planning, city and building evacuation planning, real-estate advertising, and film and video game development. One way of creating these 3D models is by generating prismatic primitives through processing range points collected by light detection and ranging (lidar) instruments followed by photo-realistic texture mapping from optical images covering the building. Finding the mathematical mapping by computing the relative orientation between sensors is referred to sensor registration. (Zitoza et. al., 2003) present a review of registration methods, classified according to their nature (area-based and feature-based) and according to four basic steps of the image registration procedure: feature detection, feature matching, mapping function design and image transformation and resampling. Most registration methods fall into two broad categories. First, photogrammetric methods register two or more sets of optical images, where most methods require specially fabricated and positioned targets. Secondly, range-based methods register two or more sets of range images or point clouds. Over recent years, several approaches have been developed that register 2D imagery with 3D range data. The most relevant to this study is the work done by (Aguilera et al., 2009) in automatic co-registration of

terrestrial laser scanner (TLS) and uncalibrated digital camera data. This method is simple and flexible while overcoming limitations of other methods. Specifically, it is solely based on simple point matches, it does not use calibration patterns and it does not require stereoscopic setup or physical attachment of the two sensors. This paper presents a similar method to automatically register optical (RGB) imagery from a digital camera with range imagery from a TLS in order to verify the achievable accuracy and evaluate possible alternatives in the work flow.

The method utilizes well-known algorithms to ensure datasets are registered with precision and reliability, and integrates robust error checking to account for the heterogeneity of the data. The structure of the paper is as follows. Section 2 and 3 describe feature detection and matching, respectively. Section 4 explains the mapping function design and image resampling approach. Section 5 presents the experimental results. Finally, Section 6 outlines the conclusions and future work.

## 2. FEATURE DETECTION

Feature detection and matching can be done interactively by the user. This is time consuming and introduces human error, especially if many images are processed. An alternative is to automatically extract then match distinct features in each image.

This is known as feature-based image matching. Many alternatives exist, where in any case features have specific properties. Specifically, a feature is a salient point in the image and the feature is highly correlated with the corresponding feature in an overlapping image. The particular criteria for this experiment are: 1) the variance of the range measurements surrounding the feature is small. Features such as tree leaves and building edges are noisy and cannot be accurately matched, hence they do not serve as good correspondence points; and 2) all the features are evenly distributed throughout the image, and they do not lie in the same plane. This ensures the estimated transformation is accurate.

### 2.1 Generation of 2D Intensity Images

Automatically extracting features from 3D objects is difficult and computationally intensive. The proposed solution to efficient feature extraction involves constraining the problem to 2D. This is practical for point clouds collected from a static TLS because each scan can be considered as 2.5D, implying the intensity values can be projected to a plane without distortion.

In the first step, the irregular point cloud is resampled to a grid by equation (1). Each pixel is populated with an intensity value using a nearest neighbour (NN) interpolation. The intensity value  $i(X, Y)$  represents the strength of the return signal as it is recorded by the TLS at position  $(X, Y)$  in the TLS coordinate system. The horizontal and vertical dimensions of each pixel ( $D_x$  and  $D_y$ , respectively) are set such that the intensity image has the same resolution of its corresponding optical image. If this is satisfied and the optical image has the same object in its field of view as the intensity image, the image scales will be approximately equal. This facilitates matching corresponding features using an area-based approach. An orthographic projection is used, where the X-coordinate of each point is transformed to the horizontal axis of the grid (column  $c$ ). Similarly, the Y-coordinate of each point is transformed to the vertical axis of the grid (row  $r$ ).

$$c = \left\lfloor \frac{X - X_{\min}}{D_x} \right\rfloor$$

$$r = \left\lfloor \frac{Y - Y_{\min}}{D_y} \right\rfloor$$

$$I(r, c) = NN(i(X, Y)) \quad (1)$$

where  $\lfloor \cdot \rfloor$  represents rounding to the nearest integer and  $NN(\cdot)$  represents the nearest neighbour interpolation operator.

The size of the kernel depends on the resolution of the optical image and the point cloud's sampling interval. To eliminate 'salt and pepper noise', a median filter is used on the intensity image.

### 2.2 Corner Extraction

A wide variety of feature detection methods exist. The chosen method depends on the nature of the data. (Zitoza et. al., 2003) list the main area-based and feature-based methods, where edges and corners are commonly used features. The Scale Invariant Feature Transform (SIFT) is a popular approach developed by (Lowe, 1999). Features are extracted that are invariant to image translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection. However, difficulties were encountered during the matching process due to the degree of dissimilarity between the

datasets. This research is aimed towards modeling urban scenes, where parallelism and orthogonality exist naturally. Thus, corners serve as reliable features. To identify corners, histogram equalization enhances the edges and then a Sobel filter extracts horizontal and vertical binary edge maps from the gray scale of the optical image. The intersection of horizontal and vertical edge maps yields the corner features. The threshold of the edge detector is adjusted according to the image resolution and radiometric properties.

## 3. FEATURE MATCHING

Correspondence between the features detected on the lidar intensity image of the 3D object and those detected in the optical image is approximated using normalized cross-correlation. This area-based matching algorithm is most effective when the images have similar scales, intensities, vantage points, and orientations. To increase the correlation between the two images, histogram matching is used to manipulate the pixel value distribution of the intensity image to match the pixel distribution of the gray scale of the optical image. Notably, the optical image should only contain the details present in the gridded point cloud. Excluding sky and ground regions ensures the histogram matching will properly modify the intensity image.

To begin the matching process, a kernel is centered on an extracted corner in the optical image. The size of the kernel is specified by the user. The matching point on the intensity image is located where the normalized cross-correlation between the kernel and the intensity image is the greatest. If the second largest cross-correlation value is significantly lower than the greatest value, the match is likely correct. A kernel is then centered on the matched point in the intensity image and the variance of the range values within the kernel is calculated. The match is rejected if the variance is greater than a user defined threshold. High range variance suggests the point could be, for example, a noisy building corner or vegetation. In which case, the match is likely false or inaccurate.

In another effort to eliminate false matches, the 2D affine transformation relating the RGB image to the intensity image is estimated. Once more than 4 matches are obtained, a least-squares adjustment is performed and false matches are identified through the analysis of the observation residuals. However, the system should be highly redundant before matches are rejected to ensure only outliers are filtered. The match with the largest residuals is considered an outlier and it is replaced with a new pair of corresponding points. The process is repeated until the sum of the squares of the residuals is below a user defined threshold. An equal amount of features is selected from each quadrant of the image to ensure the solution is stable.

Once an accurate estimate of the affine transformation parameters is available, all of the corners extracted from the optical image are transformed to the intensity image. This densification of corresponding points ensures the camera calibration solution is accurately estimated through a RANSAC algorithm. Corresponding points do not match exactly in some areas of the image. To provide an estimate with sub-pixel accuracy, point positions are refined with least squares matching (LSM) (Grün, 1985).

#### 4. TRANSFORM MODEL ESTIMATION

The matched features are used to compute the registration parameters of the digital camera and the TLS. These parameters are then used to map textures from the optical image to the point cloud, yielding a photo-realistic 3D model. The problem of image to 3D model registration is closely related to the problem of camera calibration and pose determination, which estimates interior and exterior orientation of the camera. The camera model for a positive digital image is shown in Figure 1.

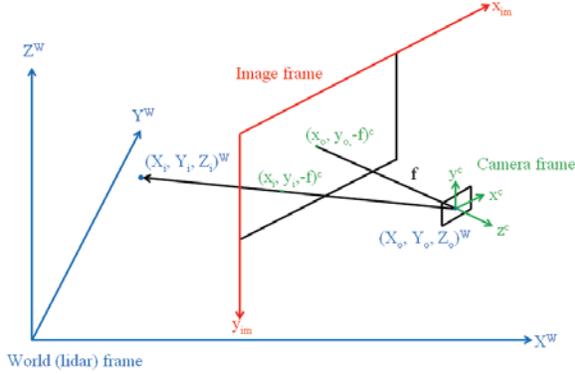


Figure 1. Camera-object geometry

Interior orientation refers to the parameters linking the pixel coordinates of an image point  $(x_{im}, y_{im})$ , with the corresponding coordinates in the camera reference frame  $(x, y, -f)$ . Specifically, the interior orientation parameters are the coordinates in pixel of the image center, or the principal point  $(x_o, y_o)$ , the focal length  $f$  and any parameters used to model lens distortion  $dx$ . Exterior orientation refers to the position  $(X_o, Y_o, Z_o)^W$  and orientation  $(\omega, \phi, \kappa)$  of the camera with respect to a world reference frame, in this case the TLS sensor frame. The orientation is described by  $r_{ij}$ , the elements in the 3D rotation matrix relating the 3D coordinates of a point in the TLS sensor frame to the camera coordinates of the corresponding point. The camera calibration and pose parameters are estimated by solving the collinearity equations. To increase the accuracy of the parameters, the collinearity equations are extended with corrections for the systematically distorted image coordinates. These modified collinearity equations are known as extended collinearity equations

$$\begin{aligned} x_{im}^i &= x_o - dx - f \frac{r_{11}(X_G^i - X_o) + r_{12}(Y_G^i - Y_o) + r_{13}(Z_G^i - Z_o)}{r_{31}(X_G^i - X_o) + r_{32}(Y_G^i - Y_o) + r_{33}(Z_G^i - Z_o)} \\ y_{im}^i &= y_o - dy - f \frac{r_{21}(X_G^i - X_o) + r_{22}(Y_G^i - Y_o) + r_{23}(Z_G^i - Z_o)}{r_{31}(X_G^i - X_o) + r_{32}(Y_G^i - Y_o) + r_{33}(Z_G^i - Z_o)} \end{aligned} \quad (2)$$

The most commonly used correction is for the radial lens distortion, expressed as

$$\begin{aligned} dx &= x_d(k_1 r^2 + k_2 r^4) \\ dy &= y_d(k_1 r^2 + k_2 r^4) \end{aligned} \quad (3)$$

where  $(x_d = x_{im} - x_o, y_d = y_{im} - y_o)$  are the coordinates of the distorted image points, the radial distance is  $r^2 = x_d^2 + y_d^2$  and  $k_1$  and  $k_2$  are the coefficients of the radial lens distortion polynomial. For most cameras, including the one used in this experiment, the effects of decentering distortion  $(p_1, p_2)$  are

negligible when compared to radial distortion, as seen in Table 2. This distortion is disregarded in this study.

Certain considerations must be taken into account when solving for these parameters, particularly when only one image is being used. 1) The distortion parameters can significantly affect the image position of a point, which causes the solution to converge at a false minimum. 2) A good estimate for the interior and exterior orientation parameters is required to reliably determine the distortion parameters. Based on these considerations, the calibration parameters are solved through a two-step approach. The first-step directly computes a closed-form solution for all the external parameters and internal parameters using Direct Linear Transformation (DLT) (Abdel-Aziz et al., 1971). The second step is a nonlinear optimization based on the extended collinearity equations that incorporates radial lens distortion.

##### 4.1 Solution by the DLT method

The first approximation of the registration parameters is obtained using the DLT method, expressed as Equation (4). Computing the perspective transformation matrix by solving a linear system of equations is computationally fast because it is a closed-form solution, requiring no iterations.

$$\begin{aligned} x_{im}^i &= \frac{m_{11}(X_G^i) + m_{12}(Y_G^i) + m_{13}(Z_G^i) + m_{14}}{m_{31}(X_G^i) + m_{32}(Y_G^i) + m_{33}(Z_G^i) + 1} \\ y_{im}^i &= \frac{m_{21}(X_G^i) + m_{22}(Y_G^i) + m_{23}(Z_G^i) + m_{24}}{m_{31}(X_G^i) + m_{32}(Y_G^i) + m_{33}(Z_G^i) + 1} \end{aligned} \quad (4)$$

Given  $N$  point matches, the linear system is:

$$Am = L \quad (5)$$

Where  $m$  is the vector of unknown DLT parameters,  $L$  is the vector of image observations and  $A$  is the Jacobian matrix of equation (4) with respect to  $m$ . The DLT parameters are obtained using the least squares method:

$$m = (A^T P A)^{-1} A^T P L, \quad P = I \quad (6)$$

The physical camera parameters are derived from the DLT parameters. This is a well-known algorithm and has been implemented and used by many groups. However, this model does not incorporate the nonlinear radial distortion. The distortion is absorbed in the residuals, thus the extracted camera parameters are biased.

In order to reliably estimate camera calibration parameters, RANSAC (RANDOM SAMPLING CONSENSUS) (Fischler and Bolles, 1981) is applied. It selects the optimum solution among several computed registrations. At least 6 matched points are required to obtain a solution for the projection matrix, but matching more points allows for a more accurate least squares solution and outlier detection. To begin the process, a user specified number matches is randomly chosen from the set extracted by the correspondence process. To ensure the features are well distributed, an equal amount of matches is selected from each quadrant of the image. The subset of points is kept if sum of the squares of the residuals from the least squares adjustment is the below a user defined threshold. Otherwise another subset of points is randomly chosen to compute the parameters.

Notably, solving the camera parameters in equation (6) falsely assumes the residuals consist of uncorrelated, zero-mean random noise with equal variance. To improve the solution, a non-linear optimal estimation procedure is undertaken, utilizing the extended collinearity equations and stochastic models for the observations and parameters.

#### 4.2 Solution by extended collinearity equations

The DLT solution is refined using an extended collinearity-based nonlinear minimization approach. Since the algorithm is iterative, the solution may converge incorrectly if an accurate initial estimate is not available. High correlation between the parameters may also cause divergence. For example, the focal length is highly correlated with the camera's Z position. A common solution uses convergent imagery to reduce this correlation. Also, rotating the camera along its optical axis will reduce the correlation between the principal point position and rotational parameters of exterior orientation (Clarke et al., 1998). However, this algorithm is based on processing a single image, thus an alternative is used. Firstly, the parameters provided by the DLT method serve as good approximations. Secondly, the parameters are constrained by  $P^o$ , the parameter weight matrix. This unified least squares solution with parameter observations is found by solving the system of equations given by (Mikhail, 1976)

$$(A^T P A + P^o) \Delta = A^T P w + P^o l^o \quad (7)$$

where A is the Jacobian of equation (2) with respect to the parameters, w represents the observation misclosure vector,  $l^o$  is the parameter misclosure vector, which is  $\hat{x} - x_o$ . P is the weight matrix of the observations, and  $\Delta$  is the parameter correction vector, expressed as

$$\Delta = [d\omega \ d\phi \ d\kappa \ dX_o \ dY_o \ dZ_o \ df \ dx_o \ dy_o \ dk_1 \ dk_2]^T \quad (8)$$

The observations and parameters are treated as uncorrelated due to the lack of knowledge of their precision. Therefore, P and  $P^o$  are diagonal matrices. Each iteration, weights are calculated using equation (9) and (10).

$$P(v) = \frac{1}{|v|} \quad (9)$$

$$P^o(\Delta) = \frac{1}{|\Delta|} \quad (10)$$

The weights are based on the residuals (v and  $\Delta$ ) from the previous iteration. This weighting scheme reduces the participation of outliers in the estimation of the calibration parameters.

The covariance matrix for the camera resection is given by

$$C_{\hat{x}} = \hat{\sigma}_o^2 (A^T P A + P^o)^{-1} \quad (11)$$

where  $\hat{\sigma}_o^2$  is the a posterior variance factor. Once the parameters of interior and exterior orientation of the optical image are estimated, a photo-realistic and accurate visualization of a 3D model is generated by reverse mapping the RGB image values to the corresponding 3D TLS points.

## 5. EXPERIMENTAL RESULTS

Experiments are performed using a dataset with known calibration parameters to assess the registration approach in terms of accuracy. A time of flight laser scanner, Optech's ILRIS-3D (Intelligent Laser Ranging and Imaging System) is used to obtain a range image. A consumer grade digital camera, the Nikon D50 is used to obtain an optical image. Calibration parameters are provided by Optech specifically for York University's ILRIS-3D scanner with the Nikon D50 camera and Nikon 20mm lens. These parameters, listed in Table 2, contribute to the validation of the results of this experiment.

The test site is situated in the north of Tennis Canada's Rexall Center, on the York University Campus. The field of view consists of many features that are not ideal to use as tie points, thus testing the filtering techniques. Such features include trees, which are physically unstable, and glass windows, which yield noisy and inaccurate range measurements. The input data are an optical image (3008 x 2000 pixels) (Figure 2) and a point cloud (1,342,674 points) (Figure 3).



Figure 2. Optical image of the test site

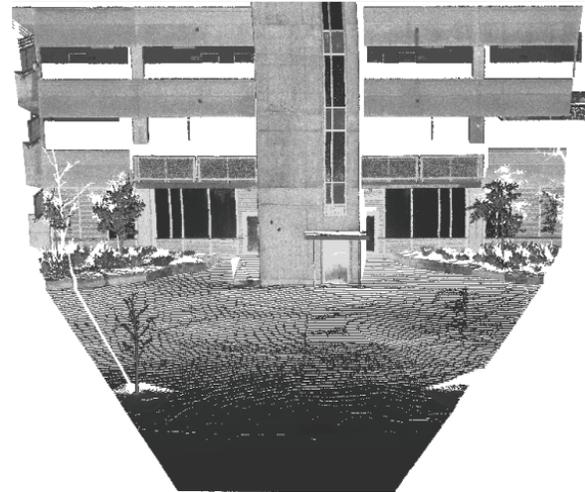


Figure 3. TLS data of the test site

The optical image is cropped by the user to match the field of view of the laser scanner. This allows the pixel size of the intensity image to be automatically set to match the pixel size of the optical image. Figure 4 shows the resulting intensity image

generated from the point cloud data using equation (1). The ground pixel dimensions of the intensity image are 1.52cm x 1.48cm. Figure 5 shows the cropped optical image. The green pixels represent the corner features that meet the criteria in the feature extraction process. Figure 6 shows the intensity image after histogram matching. The green pixels are the affine-transformed features shown in Figure 5. The size of the search kernel is 20 pixels. It is evident that corresponding points do not match exactly in some areas of the image. Point positions are refined using LSM to provide estimates with sub-pixel accuracy.

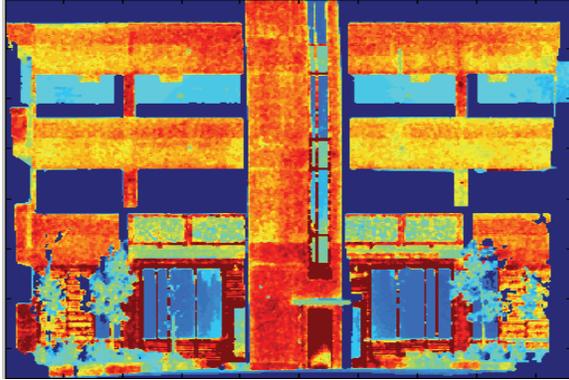


Figure 4. Intensity image of the TLS data



Figure 5. Detection of corner points in the cropped optical image

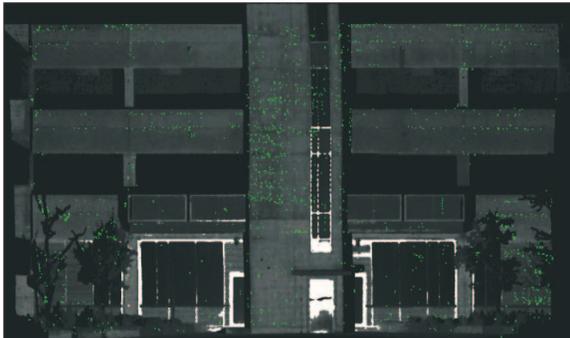


Figure 6. Matched corner point in the histogram matched intensity image

The interior and exterior orientation parameters approximated using the DLT method and RANSAC are provided in Table 1. It is evident that the results are biased from Optech's calibration

values given in Table 2. This bias is also evident through the residual analysis (Table 3) as the mean of the residuals is offset from zero.

In the second step, the parameters from the DLT method are used as an initial approximation in the collinearity-based least-squares adjustment. The calibration results are provided in Table 2. These values are more accurate and precise than the parameters derived from the DLT method because radial lens distortion is modelled and stochastic models for the observations and parameters are incorporated in the least-squares adjustment. However, the radial distortion parameters cannot be accurately estimated as the signal is masked by measurement noise. If estimates of the distortion parameters are necessary, the calibration procedure should replace the point cloud with a precise camera calibration test field, for example a traditional checkerboard pattern.

The accuracy of the offset in the Z direction ( $Z_o$ ) is low because many control points lie in approximately the same plane - the building façade is planar. A more stable solution is achievable with a larger variation in range measurements. Further, the accuracies are low for the rotational parameters of the exterior orientation and the principal point offsets because they are still highly correlated. As previously mentioned, shifts in the principal point can be compensated for with additional images captured by the camera in various orientations.

Parameter	Estimated Values	Standard Deviation
$\omega$ [degrees]	0.1172	0.258
$\phi$ [degrees]	0.0247	0.343
$\kappa$ [degrees]	-0.0074	0.665
$X_o$ [m]	-0.132	0.136
$Y_o$ [m]	0.214	0.244
$Z_o$ [m]	0.330	0.451
$f$ [mm]	20.261	0.254
$x_o$ [pixels]	1501.387	15.432
$y_o$ [pixels]	1012.407	9.432

Table 1. Calibration parameters: DLT+RANSAC

Parameter	Optech's Calibration Values	Estimated Values	Standard Deviation
$\omega$ [degrees]	-0.098	-0.124	0.253
$\phi$ [degrees]	0.136	0.049	0.274
$\kappa$ [degrees]	-0.402	-0.596	0.351
$X_o$ [m]	0.004	0.033	0.169
$Y_o$ [m]	0.038	0.073	0.068
$Z_o$ [m]	0.165	0.210	0.192
$f$ [mm]	20.345	20.353	0.052
$x_o$ [pixels]	1495.429	1497.518	3.122
$y_o$ [pixels]	1009.117	1010.550	4.18
$k_1$ [unitless]	2.622e-4	1.872e-3	3.322e-4
$k_2$ [unitless]	-4.494e-7	2.699e-7	2.058e-7
$p_1$ [unitless]	0.000	N/A	N/A
$p_2$ [unitless]	0.000	N/A	N/A

Table 2. Calibration parameters: Control and Collinearity

Statistic	DLT	Collinearity
Maximum[pixels]	13.8	7.3
Minimum[pixels]	-9.0	-5.4
Mean[pixels]	-5.2	2.3E-03
Standard deviation [pixels]	6.9	3.4
$\hat{\sigma}_o$ [pixels]	24.4	7.7

Table 3. Residuals analysis: DLT and Collinearity

Finally, once a complete camera model is computed, an automatic texture mapping algorithm is applied (Figure 6). It is evident that given correct and accurate correspondence, the camera is reliably calibrated.



Figure 6: RGB-based textured TLS point cloud

## 6. CONCLUSIONS AND FUTURE PERSPECTIVES

The presented paper has developed a procedure for automatic registration of a TLS and a digital camera, and simultaneous generation point based photo-realistic 3D models. The most relevant aspects of the proposed approach include automated feature extraction, matching and outlier detection. Automating these tasks drastically decreases the labour hours and human error attributed to the final product, especially if a large number of datasets is processed.

Several improvements could be considered. For instance, the user interaction to provide a first approximation of the area of interest could be automated. Also, linear or planar features matching instead of point features may reduce outliers and increase the registration accuracy. Future experiments will explore these possibilities and the alternative of registering an off-the-shelf camera with a laser scanner, also using lower resolutions and point densities.

### Acknowledgements

We would like to thank NSERC for the financial support of the research work. We would also like to thank Optech Inc for providing the calibration parameters and for their technical assistance. Finally, we thank Thomas Hanusch for his Matlab implementation of the least squares matching algorithm.

### References

- Abdel-Aziz, Y. I., and Karara, H. M., 1971. Direct linear transformation into object space coordinates in close-range photogrammetry. Proc. Symposium on Close-Range Photogrammetry, Urbana, Illinois: pp. 1-18.
- Aguilera, D., González, P., and Lahoz, J., 2009. An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. ISPRS Journal of Photogrammetry and Remote Sensing, 64 (3), pp. 308-316.
- Clarke, T., Wang, X., and Fryer, J., 1998. The Principal Point and CCD Cameras. Photogrammetric Record, 16(92), pp.293-312.
- Fischler, M. A., and Bolles, R. C., 1981. Random sample consensus. CM Communications.
- Grün, A., 1985. Adaptive least squares correlation: a powerful image matching technique. S. Afr. J. of Photogrammetry, Remote Sensing and Cartography, 14 (3), pp.175-187.
- Lawson, C.L., and Hanson, R.J., 1974. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, NJ.
- Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: Proceedings of International Conference on Computer Vision, Kerkyra, Greece, vol. 2, pp. 1150–1157.
- Mikhail, E.M., 1976. Observations and Least Squares. IEP-A Dun Donnelly, New York.
- Wolf, P., and Dewitt, B., 2000. Elements of Photogrammetry: with applications in GIS, McGraw-Hill, Boston.
- Zitoza, B., and Flusser, J., 2003. Image registration methods: a survey. Image and Vision Computing, 21, pp. 997-1000.