ROBUST CLASSIFICATION AND SEGMENTATION OF PLANAR AND LINEAR FEATURES FOR CONSTRUCTION SITE PROGRESS MONITORING AND STRUCTURAL DIMENSION COMPLIANCE CONTROL

R. Maalek [a, *], D. D. Lichti [b], J. Ruwanpura [a]

[a] Dept. of Civil Engineering, University of Calgary, 2500 University Dr. NW, Calgary, AB, Canada - (rmaalek, janaka)@ucalgary.ca
[b] Dept. of Geomatics Engineering, University of Calgary, 2500 University Dr. NW, Calgary, AB, Canada - ddlichti@ucalgary.ca

Commission V, WG V/3

KEY WORDS: Point Cloud Segmentation, Construction Site Progress Monitoring, Robust Statistics, Deterministic Minimum Covariance Determinant, Complete Linkage

ABSTRACT:

The application of terrestrial laser scanners (TLSs) on construction sites for automating construction progress monitoring and controlling structural dimension compliance is growing markedly. However, current research in construction management relies on the planned building information model (BIM) to assign the accumulated point clouds to their corresponding structural elements, which may not be reliable in cases where the dimensions of the as-built structure differ from those of the planned model and/or the planned model is not available with sufficient detail. In addition outliers exist in construction site datasets due to data artefacts caused by moving objects, occlusions and dust. In order to overcome the aforementioned limitations, a novel method for robust classification and segmentation of planar and linear features is proposed to reduce the effects of outliers present in the LiDAR data collected from construction sites. First, coplanar and collinear points are classified through a robust principal components analysis procedure. The classified points are then grouped using a robust clustering method. A method is also proposed to robustly extract the points belonging to the flat-slab floors and/or ceilings without performing the aforementioned stages in order to preserve computational efficiency. The applicability of the proposed method is investigated in two scenarios, namely, a laboratory with 30 million points and an actual construction site with over 150 million points. The results obtained by the two experiments validate the suitability of the proposed method for robust segmentation of planar and linear features in contaminated datasets, such as those collected from construction sites.

1. INTRODUCTION

Construction project progress monitoring and deviation control are essential to allow decision makers to identify discrepancies between the planned and the as-built states of a project in order to take timely measures where required (Maalek and Sadeghpour, 2012). In practice, monitoring is performed manually, a time consuming, error-prone and labour-intensive task particularly on large scale projects (Golparvar-Fard et al. 2009). To reduce the time and cost associated with such manual approaches, a limited (and/or frequency) of onsite data are collected, which diminishes the ability of the project manager to identify the causes of delays and cost overruns on time.

In addition, the reliable determination of project performance is highly dependent on the accuracy of the data collected during the monitoring process (Saadat and Cretin, 2002). Currently, site supervisory personnel spend 30-50% of their time manually inspecting and controlling the quality of the manually accumulated onsite data (Golparvar-Fard et al. 2009). Reduction of this time by means of a novel approach to onsite data collection and analysis suggests that more time can be allocated towards improving vital construction related concerns such as safety, as well as workforce productivity and communications. In order to help overcome the aforementioned limitations of current manual practices, automating the

monitoring and control processes on construction sites has been proposed in recent years.

2. LITERATURE REVIEW

2.1 State-of-the-Art in Construction Management

In current practices, the time of completion of an activity is recorded in order to measure the potential deviations between the planned and the actual states of the project (Cox et al. 2003, Golparvar-Fard et al. 2015). However, this metric does not provide sufficient information to determine: i) the compliance of the dimensions of the as-built structures to those of the planned; and ii) the schedule delays throughout the progression of an activity (Maalek et al. 2014). In order to help improve these limitations, the "scope of work performed" should be determined by means of a remote sensing technology (Maalek et al. 2014). Terrestrial laser scanners (TLS) are widely used to measure the 3D coordinates of the structural elements.

Current research in construction management is devoted to the automatic extraction the "scope of the work performed" for each structural element from the accumulated TLS point clouds. However, most object-based recognition models use the planned 4D model as a priori knowledge to assign the collected 3D point clouds to their corresponding structural elements (Golparvar-Fard et al. 2009, 2015; Bosché et al. 2015). This

---

* Corresponding author

ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W5, 2015
ISPRS Geospatial Week 2015, 28 Sep – 03 Oct 2015, La Grande Motte, France

approach may not be reliable in cases where the location of the as-built structure differs from that of the planned (Shahi et al. 2013) or the issued-for-construction (IFC) plan with sufficient detail is not readily available.

In order to reduce this dependency on the planned model, it is proposed to generate the 3D/4D as-built model using only the geometric primitives of the accumulated points. Since the most generic building elements as well as most man-made objects are constructed from the intersection of planar (columns, beams) and linear (reinforcement bar) features (Nunnally, 2010; Vosselman et al. 2004), the classification and segmentation of planar and linear features are the major focus of this study.

2.2 Point Cloud Classification and Segmentation

As mentioned, the automatic detection of planar surfaces from TLS point clouds is the initial step to identify the most important structural elements. In order to extract features from point clouds, the initial step is devoted to labelling and grouping of the point clouds with similar physical attributes, also known as the classification and segmentation processes respectively (Rabbani et al. 2006).

2.2.1 PCA-based Point Cloud Classification:
There are two commonly-used methods to classify point clouds into planar surfaces, namely, 3D Hough transform and principal components analysis (PCA). Vosselman et al. (2004) use the 3D Hough transform to define every point in space with a plane in the parameter space, which allows the determination of planar surfaces without the estimation of the normal vectors. However, the use of Hough transformation for planar classification is computationally expensive and the results are highly affected by outliers (Lari, 2014). Therefore, special consideration is given to the use of PCA for the classification of point cloud.

PCA is the eigenvalue decomposition of the covariance matrix of a multivariate data set. It is used to summarize the variation of the data set in independent (orthogonal) axes (Johnson and Wichern, 2007). In the case of a three-dimensional point cloud, three orthogonal axes can be determined. Many researchers have used PCA for the classification of planar surfaces (Tovari and Pfeifer, 2005; Rottensteiner et al., 2005; Rabbani et al. 2006; Pu and Vosselman, 2006; Belton and Lichti, 2006; Filin and Pfeifer, 2006; Kim et al. 2007; Bremer et al. 2013; Lari, 2014). First, for each point cloud, a neighbourhood is defined. The PCA is performed on the pre-defined neighbourhood of each point. For coplanar points, the variation of a noise-free dataset in the direction of the surface normal is equal to zero. If the pattern of the neighbourhood of the desired point forms a planar surface, the point is classified as a plane.

Currently, there are methods available to classify points to planar/linear surfaces for datasets with no data contamination (i.e. no outliers). However, the classification of a dataset affected by outliers[1] using the classical PCA method is highly affected by the presence of outlying points (Serneels and Verdonck, 2008, Hubert et al. 2012). In order to improve the classification results for contaminated data sets, Nurunnabi et al. (2012a, b) proposed the use of robust PCA, which incorporates a robust estimate of the covariance matrix called the fast minimum covariance determinant (Fast-MCD) proposed by Rousseeuw and Driessen (1999). Their proposed robust PCA method for planar classification and segmentation showed

significant improvement in contaminated data sets. Their comparison to the random sample consensus (RANSAC) method indicated that the robust PCA is better able to detect more outliers (Nurunnabi et al., 2013, 2014). In order to determine the most efficient robust covariance matrix estimate, a review of the current state of robust dispersion (covariance) estimates is given in the following sub-section.

2.2.2 Robust Dispersion Estimates:
Robust statistics are methods of estimating models of contaminated data by reducing the effect of the outliers (Maronna et al. 2006). The *breakdown value* is the measure of robustness of an estimator with respect to the outlying observations (Hampel, 1971). It indicates the smallest fraction of contaminants in a sample that causes the estimator to break down (i.e. to take on values that are arbitrarily meaningless). An estimate with a breakdown point of 50% is ideal since it is able to detect the pattern of the majority of the uncontaminated data with up to 50% data contamination. There are currently two well-known multivariate dispersion estimates with high breakdown values (i.e. 50%), namely, the minimum volume ellipsoid (MVE) and the MCD.

The MVE is the smallest ellipsoid that covers a subset of *h* data points out of a set of *n* observations. The *(n-h)* points left are the outliers of the dataset. The MCD is concerned with selecting *h* points out of *n* for which the covariance matrix has the lowest determinant. The MCD has the same breakdown point as the MVE except that it is asymptotically normal (Butler et al. 1993) and has a higher convergence rate (Davies, 1992). In the study conducted by Jensen et al. (2007), it was concluded that the MCD is more suitable for larger sample sizes with a large percentage of data contamination. Therefore, an estimator of the MCD is preferred for the processing of point clouds in highly occluded areas such as a construction site.

There are currently two well-known MCD estimators namely, the fast-MCD (Rousseeuw and Driessen, 1999) and deterministic-MCD (Det-MCD; Hubert et al. 2012). Compared to the fast-MCD, deterministic Det-MCD is permutation invariant (i.e. the outcome of the estimator is not a function of the order of the observations). This is of great importance since the reordering of the point cloud samples does not affect the result of the robust covariance estimation subset. In addition, the computation time of the Det-MCD is much lower than that of Fast-MCD (Hubert et al. 2012). Therefore, in this study, the Det-MCD proposed by Hubert et al. (2012) is used to improve the classification of point clouds.

2.2.3 Point Cloud Segmentation:
Two methods are generally used to segment the classified planar/linear point clouds, namely, region growing and clustering. Region growing methods are widely implemented (Tovari and Pfeifer, 2005; Rottensteiner et al., 2005; Rabbani et al. 2006; Pu and Vosselman, 2006; Belton and Lichti, 2006; Belton, 2008; Nurunnabi et al. 2012a; 2012b; 2013; 2014) due to their computational efficiency. However, since the result of the segmentation is a function of the selected seed point/region (i.e. not permutation invariant), it is not considered as a robust method (Wang and Shan, 2009). Therefore, particular interest is given to segmentation procedures using cluster analysis.

In cluster analysis, an n-dimensional array of attributes is first defined. The points sharing similar attributes are then segmented into the same cluster. In the research carried out by Song and Feng (2008) and Shi et al. (2011), the k-means

---

[1] Which is the case on construction sites.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
Editors: S. Oude Elberink, A. Velizhev, R. Lindenbergh, S. Kaasalainen, and F. Pirotti
doi:10.5194/isprsannals-II-3-W5-129-2015
130

clustering algorithm was used to group point clouds with similar attributes. However, a k-means clustering approach requires a priori knowledge of the number of clusters and hence is not suitable for applications when this is unknown. In the work of Filin and Pfeifer (2006), clustering of the point clouds was carried out by seeking the mode of the histogram of the frequency of the attributes. However, the correct identification of the mode may be challenging in multivariate attribute cases (Haralick and Sahpiro, 1992). In the work of Lari and Habib (2014), a two-step segmentation method is proposed. First a region growing method is used to identify planar patches. These planar patches are then grouped/clustered in order to complete the segmentation. However, the choice of threshold used to cluster the attributes is currently subjective, which may result in over or under segmentation depending on the dataset.

As explained, the attributes in this study are robustly estimated during the classification process. Therefore, compact clusters are expected to be formed. In the research carried out by Bayne et al. (1980), Golden and Meehl (1980), Hartigan (1985), and Everitt et al. (2011) the complete linkage method was shown to be efficient for identifying compact clusters. This method does not require a priori knowledge about the number of clusters. In addition, it is not highly affected by outliers. However, it can break large clusters (Steinbach et al. 2003), resulting in over-segmentation. Here, an iterative robust complete linkage algorithm is proposed to reduce over-segmentation.

3. OBJECTIVE AND METHODOLOGY

The overall goal of this research is to automatically summarize acquired point clouds of construction sites into a set of vertices (i.e. automatic generation of the as-built model) using only the geometric primitives. To that end, a novel method is proposed to robustly segment coplanar and collinear points as a means of extracting the most common structural elements (beams, columns, slabs and reinforcement bars). Initially, the points are classified into planes and lines through a robust PCA, which uses the Det-MCD proposed by Hubert et al. (2012) to robustly estimate the covariance matrix. The coplanar and collinear points with similar attributes are then grouped together using a novel clustering approach. The modified convex-hull algorithm is used to detect the boundaries of each segment. The closest segments are then intersected in order to generate the 3D as-built model. The detailed explanation of the aforementioned stages is given in the following.

3.1 Robust Planar and Linear Classification

In order to classify point clouds into planes and lines, a neighbourhood is defined around each point. The 50 mm neighbourhood size is chosen based on the dimensions of the smallest structural elements that are required to be extracted[2]. Robust PCA is performed to determine the pattern of the variation within each neighbourhood. For coplanar points, the variation of the data in the direction of the surface normal is zero. For collinear points, all of the variation is summarized in one direction. This is illustrated in Figure 1.
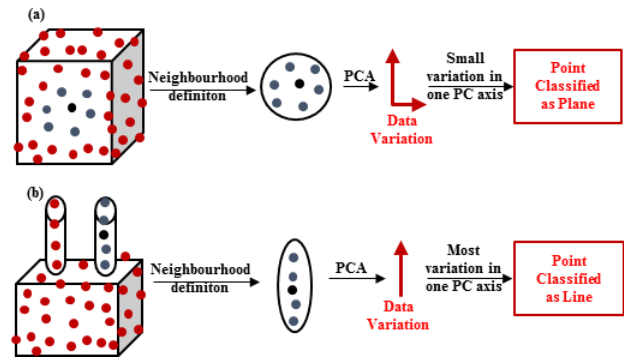

Figure 1. Classification of accumulated point clouds into a) planar surface; b) linear features

In order to illustrate the benefits of using robust PCA over classical PCA, in particular for the identification of mixed pixels, a point cloud comprising four adjacent planes scanned from a single instrument location was simulated. Random errors were added to the data using the specifications of the Leica HDS6100 TLS[3], the instrument used to collect real data for this research. Mixed pixel artefacts were added between two of the planes using the following equation:

$$SM = \frac{(X1 * S1 + X2 * S2)}{X1 + X2} \qquad (1)$$

where X_1, X_2, S_1, S_2 and S_M are shown in Figure 2a. The simulated point clouds are shown in Figure 2b.
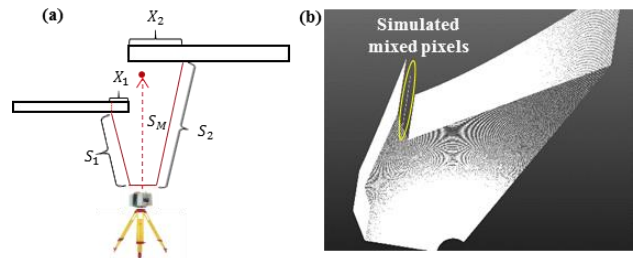

Figure 2. a) Schematic representation of the mixed pixel phenomena; b) simulated point clouds of the planar walls

Figure 3 illustrates the results of the classification of the data depicted in Figure 2b. Figure 3a represents the percentage of misclassified mixed pixels with respect to the threshold used for the percentage of variance, explained by the largest eigenvalue (the neighbourhood size was fixed at 100). It can be seen that the planar classification results using the robust PCA includes fewer type II errors than the classical PCA. Figure 3b shows the relative percentage of improvement in the number of misclassified mixed pixels with respect to the neighbourhood size (the threshold of the maximum normalized eigenvalue was fixed to 55%). It can be inferred that the percentage of improvement in the misclassified points within the planar classification is more evident as the neighbourhood size increases. The results shown in Figure 3 indicate that the proposed robust PCA is less dependent on the thresholds used (i.e. more robust) and the choice of initial neighbourhood size.

---

[2] In the work of Belton and Lichti (2006) and Weinmann et al. (2014), efforts were made to optimize the neighborhood size while performing the classical PCA. As will be proven in the following, the robust PCA is able to detect the outliers present within the predefined neighborhood, which reduces the dependency of the classification results on the initially defined neighborhood size.

[3] The manufacturer suggests a random error with Gaussian distribution of mean zero and standard deviation of 2 mm and 125 µrad for range and angular measurements respectively. The beam width is 3 mm at exit with angular divergence of 110 µrad on each side.
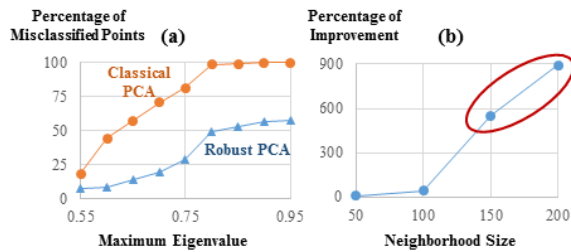
Figure 3. a) Percentage of misclassified points with respect to the threshold; b) percentage of improvement of the misclassification with respect to the neighbourhood size

3.2 Robust Planar Segmentation

From the robust PCA, points belonging to planar and linear features are identified. For each planar point, the four planar attributes, the robustly estimated surface normal vector and location (robust mean of the neighbourhood), are used to cluster points with similar attributes. As expressed in Section 2.2.3, the complete linkage algorithm is used to cluster coplanar points. According to the complete linkage algorithm, initially, a cluster is assigned to each point. The two clusters (say *U* and *V*) with the most similarity are merged together to form cluster *UV*. The distance between the similarity attribute of cluster (*UV*) and any cluster *W* is then calculated as follows:

$$d_{(UV)W} = \max \{d_{UW}, d_{VW}\} \qquad (2)$$

The cluster with the minimum distance to cluster *UV* is merged into *UV*, say point *W*, and the process is continued for cluster *UVW*. The grouping is finalized when the distance measured by Equation (2) is greater than a predefined threshold. The process is then repeated for the remaining clusters. However, the choice of the similarity threshold is subjective, which reduces the robustness of the method[4]. In order to reduce the dependence of the segmentation on the specific value of the threshold, a new iterative process is proposed (Figure 4).
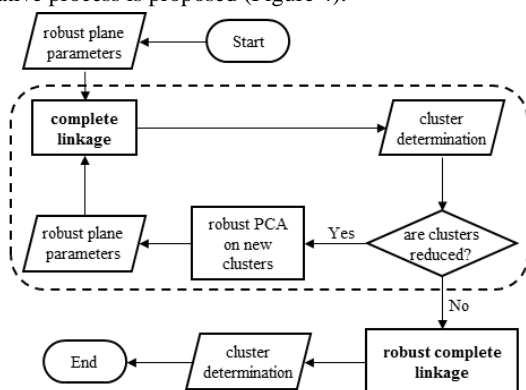

Figure 4. Iterative complete linkage algorithm for robust clustering of planar surfaces

Initially, the complete linkage algorithm is performed on the robustly-estimated plane parameters to group the coplanar points with similar attributes. The threshold is chosen so as to prevent under-segmentation[5]. For each cluster, the plane parameters are then estimated from the eigenvalue

---

[4] A large threshold may result in under-segmentation, whereas a small value may result in over-segmentation.

[5] In this study, a difference of ±1% of the magnitude of the attribute is used to accept similarity.

decomposition of the covariance matrix robustly estimated by DetMCD. The complete linkage algorithm is then carried out for the new plane parameters. The process is continued until the number of clusters remain constant.

For the identified cluster, a robust complete linkage is implemented to help reduce the dependency on the initial threshold (i.e. minimize over segmentation). First, the closest clusters are identified, say clusters I and J with sizes $N_I \leq N_J$. A random set of observations from cluster I is added to cluster J (no more than 25% of N_J)[6]. For the newly developed cluster, the DetMCD is performed to identify the outliers. The two clusters are merged if and only if less than half of the determined outliers are from cluster I. The process continues until no more clusters can be added to cluster IJ. The process is then repeated for the remaining clusters. In order to improve the computation efficiency, clusters with attributes that are farther than a certain threshold are not examined.

3.3 Robust Extraction of Flat Slab Floor and Ceiling

A new method is proposed to identify and extract the points on planar slab floors and ceilings before performing the proposed robust PCA using only the histogram of point elevation. This is particularly beneficial to help reduce the calculation time of the proposed segmentation procedure. A similar idea was introduced in (Arastounia and Lichti, 2013) to reduce the points on the ground in an electrical substation dataset. Here, a robust floor and ceiling extraction method is proposed to minimize the dependency on the thresholds used.

The typical histogram of point elevation for a room or a construction site with flat slab ceiling and floor is schematically shown in Figure 5. As illustrated, the histogram of elevation consists of two major peaks, representing the points of the floor and the ceiling. To determine the location of these two modes, the median-shift algorithm proposed by Shapira et al. (2009) is used. The two modes are regarded as points P_f and P_c in Figure 5. In order to robustly identify the points on the ceiling and the floor using the identified modes (peaks), first, all points within a predefined radius (r), here 5cm, from the modes P_f and P_c are identified. The Det-MCD algorithm is then applied on the specified points in order to identify the floor and ceiling.
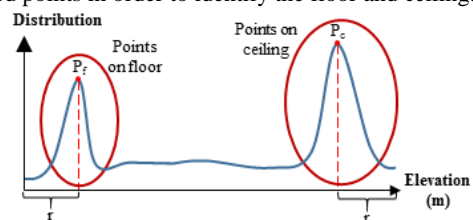

Figure 5. Expected distribution of the elevation of the points

3.4 Linear Segmentation

Every line in space can be uniquely defined by the intersection of two non-parallel planes. This concept is used to segment collinear points. After performing the robust PCA, each linearly classified point is defined by the robust directional vector and the robust location (mean of the neighbourhood). The cross product of the directional vector and the location vector results in a normal vector of a plane that passes through the line of interest and the origin. Initially, this metric is used within the

---

[6] The DetMCD algorithm is most efficient with 25% or less outlier contamination (Hubert et al. 2012).

complete linkage method to segment points with similar normal vectors. For each planar segment, the origin is then moved to an arbitrary location outside of the plane. The normal vector for each point in the cluster is again estimated using the robust directional vector and the new location[7]. The complete linkage algorithm is again performed to determine the final segments.

3.5 Boundary Detection and Robust Surface Fitting

Using the clustering methods proposed in Sections 3.2 and 3.4, spatially discontinuous surfaces with similar attributes are also grouped together. In order to enforce surface continuity, outer boundary points are determined using the modified convex hull algorithm proposed by Sampath and Shan (2007) and inner boundary points are defined using the method proposed by Lari (2014). Therefore, discontinuous surfaces are separated into different clusters.

The plane and line parameters for each identified cluster are robustly estimated using DetMCD. The closest planes and lines are then intersected to determine the vertices of the structural elements.

4. EXPERIMENTS

Two sets of LiDAR data were collected using a Leica HDS6100 TLS. The first set of experiments was for the as-built modelling of a laboratory at the University of Calgary. The second set of data was collected from an actual construction site and the planar and linear features are robustly segmented.

4.1 Experiment 1: Mechanics of Materials Laboratory

The first set of data was collected from the Mechanics of Materials laboratory at the University of Calgary (Figure 6). As illustrated in Figure 6a, the laboratory consists of many metallic tables, which may result in data contamination due to multipath reflections. Therefore, it can be considered as a fair representation of an actual indoor construction site. Approximately 30 million 3D points of the interior surfaces were recorded from three different scan-stations. Figure 6b shows the plan view of the planned model. As illustrated, the lab consists of 26 different walls. The elevation of the ceiling relative to the floor is 2.7 m. The planned model suggests that the roof, floor and the surrounding walls are planar surfaces. The objective of this experiment is to robustly extract the planes representing the walls, floor and ceiling in order to control dimension compliance.
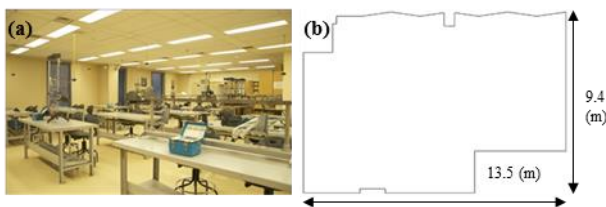

Figure 6. a) "Mechanics of Material" laboratory; b) plan view of the laboratory

4.1.1 Robust Extraction of Floor and Flat Slab Ceiling:
First, the points of the flat slab floor and ceiling are extracted using the method presented in Section 3.3. The histogram of the elevation is shown in Figure 7, which complies with the hypothesis presented in Figure 5. The smaller peak, shown in blue, represents the metallic tables. The precision, recall and accuracy (Olsen and Denlen, 2008) of the extracted points are 91.5%, 100% and 92% for the floor and 92.4%, 100% and 93.4% for the ceiling respectively[8]. As illustrated, no Type II errors were detected during the planar feature extraction, which indicates the robustness of the proposed method. In addition, the extracted points accounted for approximately half of the total accumulated points, which suggests a significant reduction in the time of data classification and segmentation.
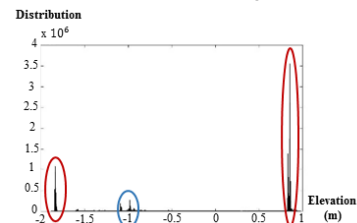

Figure 7. Histogram of elevation of the actual point cloud

4.1.2 Segmentation and As-built Model:
Using the methods presented in Section 3.1, the robust PCA was performed on the remaining points. The planar parameters were then clustered using the method described in Section 3.2. The results of the segmentation are shown in Figure 8b. Approximately 94.7% of the points were segmented correctly. Figure 8c shows the as-built 3D model of the laboratory. The vertices were determined by intersecting the nearest planar clusters using the method described in Section 3.5.
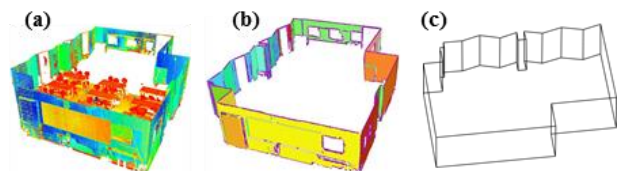

Figure 8. a) LiDAR point cloud; b) results of the robust segmentation (obstacles are removed for clarity) - purple represents segment boundaries; c) as-built 3D CAD model

4.2 Experiment 2: Graduate Student Residence Hall Construction Site

The second dataset was collected from the Graduate Student Hall of Residence construction site at the University of Calgary (Figure 9a). Approximately, 150 million points were collected from four scan locations with the Leica HDS6100, shown in Figure 9b. The building is a concrete structure with box-shaped columns. The goal was to robustly segment the planar surfaces (floor slab and column facets) and linear features (reinforcement bar) using the methods proposed in Section 3.
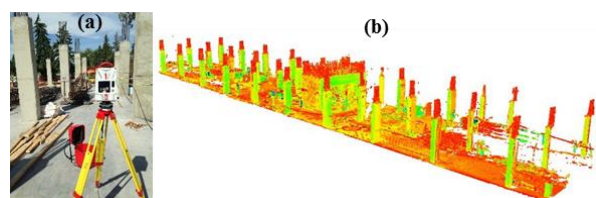

Figure 9. a) Construction site; b) point clouds of the site

---

[7] Since the DetMCD covariance estimate is very close to affine equivariant, the translation of the origin will not impact the segmentation results.

[8] The actual values are derived manually

4.2.1   Robust Floor Extraction: The points on the planar floor slab were extracted using the method proposed in Section 3.3. Figure 10 shows the histogram of point elevation of the acquired data. As illustrated, the shape of the histogram of the points on the floor complies with that proposed in Figure 5. Approximately 65 million points were removed using the proposed method, which led to a great reduction in the calculation time for the planar and linear segmentation of the remaining points. The precision, recall and accuracy rates are 90.3%, 100% and 94.6% respectively[9].
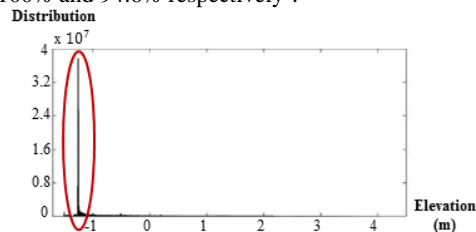

Figure 10. Histogram of point elevation

4.2.2   Robust Classification: The robust PCA proposed in Section 3.1 was performed on the remaining point cloud to identify the planar and linear features. The results of the classification are presented in Figure 11. Figure 11a represents the point cloud after the removal of the points on the floor. Figure 11b illustrates the points classified as lying on planar surfaces. As illustrated in Figure 11b, the proposed robust classification and floor extraction methods are able to correctly distinguish planar plates with a thickness of 5 cm from the points on floors. Figure 11c shows the remaining points after removing planar surfaces. The points classified as linear are shown in Figure 11d. The precision, recall and accuracy for the planar classification are 93.2%, 92.4% and 91.6% respectively. For linear classification, the precision, recall and accuracy are 91.8%, 89.6% and 92.8% respectively.
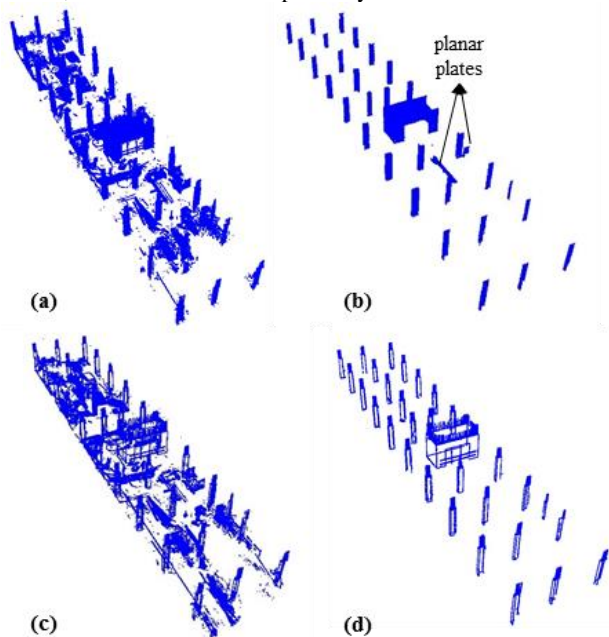

Figure 11. Robust planar and linear classification: a) after removing the floor; b) points classified as planar surfaces; c) points after removing the points classified as planes; d) points classified as linear

4.2.3   Robust Segmentation: The results of the robust segmentation of the classified point cloud are shown in Figure 12. Figures 12a through 12c[10] show the improvement of the planar segmentation results after each stage of the method proposed in Section 3.2. Figure 12a represents the segmentation of planar surfaces after the first iteration, in which 185 clusters were identified and over-segmentation is apparent. Figure 12b shows the planar segmentation after the last iteration (the third iteration). The number of clusters has been reduced to 132. Figure 12c illustrates the results after the robust complete linkage algorithm has been applied. The number of clusters was further reduced to 87. After this stage, approximately 95.2% of points were segmented correctly.

Figure 12d shows the linear segmentation results. The point density has been reduced for clarity. The reinforcement bar on the top of the elevator shaft has also been magnified to better represent the linear segmentation results. Approximately, 91.4% of the reinforcement bars were clustered correctly. For the remaining linearly classified points, about 86.9% of points were clustered correctly. It may be possible to improve the linear segmentation by means of a better choice for the location of the origins in the method proposed in Section 3.4.
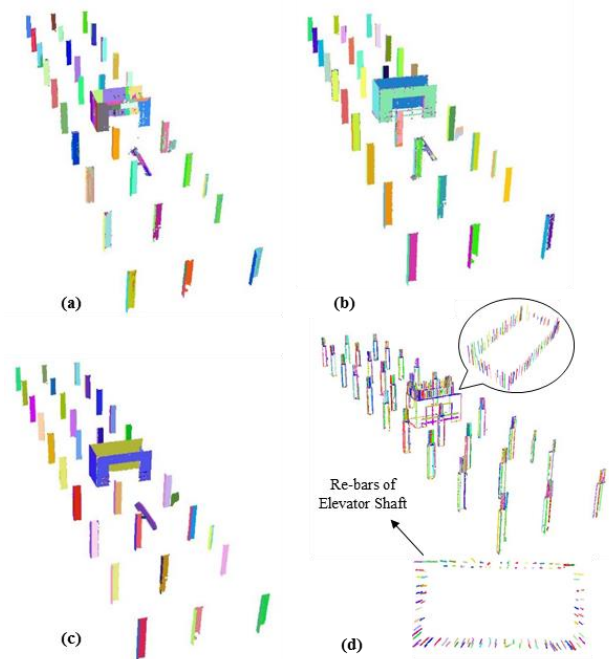

Figure 12. Planar segmentation: a) first iteration, 185 clusters; b) last iteration, 132 clusters; c) after robust complete linkage, 87 clusters. d) Linear segmentation results, 347 clusters

5.   CONCLUSTION

The use of LiDAR for construction site progress monitoring and structural dimension compliance control is evolving markedly. However, the point clouds collected in a dynamic environment such as a construction site are expected to be contaminated with outliers. Here, a robust method for the classification and segmentation of planar and linear features in LiDAR data collected from construction sites has been introduced. The classification method uses a robust PCA to reduce the effects of

---

[9] Approximate values since the actual points are determined manually

[10] The boundary detection has been carried out to differentiate discontinuous surfaces

outliers on the pattern of the data. It was also shown that the results of the classification are less affected by the choice of the size of neighbourhood. However, a robust optimum neighbourhood search method is required to further enhance the classification results.

A novel method for robust planar segmentation was proposed using an iterative complete linkage clustering method and the DetMCD covariance estimator. The method is particularly beneficial since its performance is not a function of a subjectively pre-defined threshold.

A robust method for extraction of planar floors and ceilings has been developed. This method has shown to be very efficient in extracting the points on floors and ceilings as well as reducing the calculation time for the classification and segmentation of the remaining points.

A new two-step method for linear segmentation was also introduced. Currently, the choice of the second origin after the initial segmentation is arbitrary and subjective and hence more investigation is required to find the optimum location of the origins to improve the linear segmentation results.

The applicability of the proposed planar and linear segmentation methods have been investigated in two datasets. The results indicate promise for the robust segmentation and classification of planar and linear features in contaminated datasets.

In future studies, the applicability of the proposed methods will be examined on two construction sites located at the University of Calgary as construction progresses. The inconsistencies between the planned 4D BIM model and the automatically generated as-built model will be investigated through a novel change detection algorithm. The robust segmentation and classification of NURB surfaces and the use of alpha-shapes in detecting the boundaries of these types of segments will also be studied.

ACKNOWLEDGEMENTS

REFERENCES

Arastounia, M., and Lichti, D.D. 2013. Automatic extraction of insulators from 3D LiDAR data of an electrical substation. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. II-5/W2, pp. 19-24.

Bayne, C.K., Beauchamp, J.J., Begovich, C.L., Kane, V.E. 1980. Monte carlo comparisons of selected clustering procedures. *Pattern Recognition,* 12, pp. 51-62

Belton, D. and Lichti, D. 2006. Classification and segmentation of terrestrial lasers canner point clouds using local variance information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (Part 5), Dresden, Germany, pp. 44-49.

Bosché, F., Ahmed, M.,Turkan, Y., Haas, C. T., Haas, R. 2015. The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Automation in Construction,* 44, pp. 212-226.

Bremer, M. Wichmann, V. and Rutzinger, M. 2013. Eigenvalue and graph-based object extraction from mobile laser scanning point clouds ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. II-5/W2 (2013), pp. 55-60.

Butler, R.W., Davies, P.L., and Jhun, M. 1993. Asymptotics for the Minimum Covariance Determinant Estimator. *The Annals of Statistics* 21, pp. 1385-1400.

Cox, R. F., Issa, R. R. A., and Ahrens, D. 2003. Management's perception of Key Performance Indicators for construction. *J. Constr. Eng. Manage.,* 129(2), pp. 142-151.

Davies, P. L. 1992. The asymptotics of Rousseeuw's minimum volume ellipsoid. *Ann. Statist.*, 20, pp. 1828-1843.

Everitt, B. S., Landau, S., Leese, M., and Stahl, D. 2011. *Cluster Analysis.* 5th Edition, John Wiley & Sons, UK, pp 43-110.

Filin, S., Pfeifer, N. 2006. Segmentation of airborne laser scanning data using a slope adaptive neighborhood. *ISPRS J. of Photogrammetry and Remote Sensing*, 60 (2), pp. 71-80.

Golden, R. R., & Meehl, P. E. 1980. Detection of biological sex: An empirical test of cluster methods. *Multivariate Behavioral Research*, 15, pp. 475-496.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. 2015. Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. *J. Comput. Civ. Eng.*, 29(1).

Golparvar-Fard, M., Peña-Mora, F., Arboleda, C. A., and Lee, S. H. 2009. Visualization of construction progress monitoring with 4D simulation model overlaid on time-lapsed photographs, *J. Comput. Civ. Eng*. 23, pp. 391-404.

Hampel, F. R. 1971. A general qualitative definition of robustness. *Ann. Math. Statist.*, 42, pp. 1887-1896.

Haralick, R. M., Shapiro, L. G. 1992. *Computer and Robot Vision*. Addison-Wesley, New York.

Hartigan, P. M. (1985). Statistical algorithms: algorithm AS 217. Computation of the dip statistic to test for unimodality. *Appl Stat*, 34, pp. 320-325.

Hubert, M., Rousseeuw, P. J., Verdonck, T. 2012. A deterministic algorithm for robust location and scatter. *J. Comput. Graph Stat*, 21, pp. 618-637.

Jensen, W. A., Birch, J. B. and Woodall, W. H. 2008. Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology*, 40, pp. 167-183.

Kim C., Habib A., Mrstik P. 2007. New Approach for Planar Patch Segmentation using Airborne Laser Data. Proceedings of the ASPRS, Tampa, Florida.

Lari, Z. 2014. Adaptive Processing of Laser Scanning Data and Texturing of the Segmentation Outcome, Ph.D. Thesis University of Calgary.

Maalek, R., and Sadeghpour, F. 2012. Reliability assessment of Ultra-Wide Band for indoor tracking of static resources on construction sites. CSCE, Edmonton, Canada.

Maalek, R., Ruwanpura, J., and Ranaweera, K. 2014. Evaluation of the State-of-the-Art Automated Construction Progress Monitoring and Control Systems. CRC Conference, Atlanta, USA.

Nunnally, S. W. 2010. *Construction Methods and Management*. eighth ed., Pearson Education, USA, pp. 517-518.

Nurunnabi, A., Belton, D., West, G. 2012a. Diagnostic-robust statistical analysis for local surface fitting in 3d point cloud data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1(3), Melbourne, Australia, pp. 269-274.

Nurunnabi, A., Belton, D., West, G. 2012b. Robust segmentation for multiple planar surface extraction in laser scanning 3D point cloud data. In: Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba Science City, Japan, 11–15 November, pp. 1367-1370.

Nurunnabi, A., Belton, D., West, G. 2013. Robust Outlier Detection and Saliency Features Estimation in Point Cloud Data. In: 10th Canadian Conference on Computer and Robot Vision, May, Regina, Canada, pp. 28-31.

Nurunnabi, A., Belton, D., West, G. 2014. Robust statistical approaches for local planar surface fitting in 3D laser scanning data. *ISPRS J. of Photogramm. Remote Sens.*, 96, pp. 106-122.

Olsen, D. L., Denlen, D. 2008. *Advanced Data Mining Techniques*. Springer, New York, NY, pp. 138.

Pu, S., Vosselman, G. 2006. Automatic extraction of building features from terrestrial laser scanning. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (Part 5), Dresden, Germany, (on CD-ROM).

R. A. Maronna, R. A., Martin, D. R., and Yohai, V. J. 2006. *Robust Statistics: Theory and Methods*. Wiley.

Rabbani, T., van den Heuvel, F., Vosselmann, G. 2006. Segmentation of point clouds using smoothness constraint. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (Part 5), Dresden, Germany, pp. 248-253.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K. 2005. Using the Dempster Shafer method for the fusion of LIDAR data and multispectral images for building detection. *Information Fusion*, 6 (4), pp. 283-300.

Rousseeuw, P. J., and Van Driessen, K. 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41, pp. 212-223.

Saadat, M., and Cretin, L. 2002. Measurement systems for large aerospace components. *Sensor Review*, 22(3), pp. 199-206.

Serneels, S., & Verdonck, T. 2008. Principal component analysis for data containing outliers and missing elements. *Computational Statistics and Data Analysis*, 52(3), pp. 1712-1727.

Shahi, A., West, J. S., Haas, C. T. 2013. On site 3D marking for construction activity tracking. *Automation in Construction* 30, pp. 136-143.

Shapira, L., Avidan, S., Shamir, A. 2009. Mode-detection via median-shift. In: IEEE Computer Vision and Pattern Recognition (CVPR), pp. 1909-1916.

Shi, B. Q., Liang, J., and Liu, Q. 2011. Adaptive simplification of point cloud using k-means clustering. *Comput. Aid. Des*. 43(8), pp. 910–922.

Song, H., Feng, H. Y. 2008. A global clustering approach to point cloud simplification with a specified data reduction ratio. *Comput-Aided Des*, 40 (3), pp. 281–292

Stanimirova, I., Daszykowski, M., Walczak, B. 2007. Dealing with missing values and outliers in principal component analysis. *Talanta*, 72, pp. 172-178.

Steinbach, M., Ertoz, L., Kumar, V. 2003. Challenges of clustering in high dimensional data", *University of Minnesota Supercomputing Institute Research Report*, 213, pp. 1–33.

Tóvári, D. and Pfeifer, N. 2005. Segmentation based robust interpolation − a new approach to laser filtering. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(3/W19), pp. 79–84.

Vosselman, G., Gorte, B., Sithole, G., Rabbani, T. 2004. Recognizing structure in laser scanner point clouds. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 46 (Part 8/W2), pp. 33-38.

Wang, J., Shan, J. 2009. Segmentation of LiDAR point clouds for building extraction. Proceedings of ASPRS Annual Conference, Baltimore, USA.

Weinmann, M., Jutzi, B. and Mallet, C., 2014. Semantic 3D scene interpretation: a framework combining optimal neighborhood size selection with relevant features. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-3, pp. 181–188.