ISPRS Journal of Photogrammetry and Remote Sensing 134 (2017) 15–29

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Pairwise registration of TLS point clouds using covariance descriptors and a non-cooperative game

Dawei Zai [a], Jonathan Li [a,b,*], Yulan Guo [c,d], Ming Cheng [a], Pengdi Huang [a], Xiaofei Cao [a], Cheng Wang [a]

[a] Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, Xiamen, China
[b] Department of Geography and Environmental Management, University of Waterloo, Waterloo, Canada
[c] College of Electronic Science and Engineering, National University of Defense Technology, Changsha, China
[d] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 31 March 2017
Received in revised form 2 October 2017
Accepted 2 October 2017

Keywords:
Terrestrial laser scanning (TLS)
Registration
Covariance matrix descriptor
3D representation
Game theory

ABSTRACT

It is challenging to automatically register TLS point clouds with noise, outliers and varying overlap. In this paper, we propose a new method for pairwise registration of TLS point clouds. We first generate covariance matrix descriptors with an adaptive neighborhood size from point clouds to find candidate correspondences, we then construct a non-cooperative game to isolate mutual compatible correspondences, which are considered as true positives. The method was tested on three models acquired by two different TLS systems. Experimental results demonstrate that our proposed adaptive covariance (ACOV) descriptor is invariant to rigid transformation and robust to noise and varying resolutions. The average registration errors achieved on three models are 0.46 cm, 0.32 cm and 1.73 cm, respectively. The computational times cost on these models are about 288 s, 184 s and 903 s, respectively. Besides, our registration framework using ACOV descriptors and a game theoretic method is superior to the state-of-the-art methods in terms of both registration error and computational time. The experiment on a large outdoor scene further demonstrates the feasibility and effectiveness of our proposed pairwise registration framework.

1. Introduction

Terrestrial laser scanning (TLS) systems have been commonly used in many applications such as heritage preserving, 3D modeling, and manufacturing because a TLS system can rapidly acquire dense and accurate 3D point clouds with geometry, color and intensity information from surfaces. As a preprocessing step for many TLS point cloud related applications, registration aims to transform multiple scans in arbitrary initial positions and orientations into a common coordinate system to align their overlapping regions. Point cloud registration is indispensable for completing interest objects or scenes. This paper develops a method for automated pairwise registration of TLS point clouds.

Registration of TLS point clouds is a challenging task. First, the arbitrariness in initial positions of scans can affect the performance of point cloud registration methods and has to be considered in the operation. Second, the TLS data pose four major challenges as follows (Tam et al., 2013; Theiler et al., 2015; Yang et al., 2015): (1) **Data size**: A TLS system can rapidly acquire a large volume of point clouds (e.g., up to 300,000 pts/s). That means it needs to efficiently deal with a large amount of data. (2) **Noise and outliers**: Noise is presented as a form of random fluctuation of data or unwanted points close to a surface, and outliers are considered as those points far from the surface. Both noise and outliers are common and unavoidable. (3) **Uneven density**: Uneven distribution of point density is caused by the mechanism of the TLS system. That is, the closer the target is to the TLS system, the denser the acquired points are, and vice versa. (4) **Limited overlap**: Limited overlap is caused by different views in each scan. In particular, considering the efficiency of data acquisition, the number of scans is expected to be as low as possible, leading to insufficient common points between successive scans. All of these challenges can seriously affect the robustness of point cloud registration methods. Other problems may arise due to pattern repetitions in a scene or symmetries of a surface, which are inevitable and should be considered when designing a point cloud registration method.

In this paper, we propose a new pairwise registration method for TLS point clouds. Our method consists of two main stages: adaptive covariance (ACOV) descriptor generation and non-cooperative game construction. In the first stage, the method first

* Corresponding author at: Department of Geography and Environmental Management, University of Waterloo, Waterloo, Canada.
E-mail address: junli@xmu.edu.cn (J. Li).

https://doi.org/10.1016/j.isprsjprs.2017.10.001
0924-2716/

introduces the features to be fused in a covariance matrix and the keypoint extraction method based on covariance descriptors. Then, we propose a new ACOV feature descriptor. Finally, we introduce the distance metric and matching strategy used to find candidate correspondences between two point clouds. The second stage contains three steps: (1) the concept of game theory used for matching; (2) a non-cooperative matching game; and (3) an optimal solution for the matching game based on replicator dynamics. We tested our proposed method on three models and a large outdoor scene acquired by two different static TLS systems. Experimental results have demonstrated the effectiveness of our pairwise point cloud registration method. The contributions of this paper are as follows: (1) We propose a covariance-based descriptor to fuse various information (e.g., geometry, color, and intensity) acquired by a TLS system; (2) We generate covariance descriptors over adaptive neighborhood sizes, which are robust to noise, outliers, rigid transformations, and uneven density; and (3) We develop a game-theoretic matching method for pairwise registration of TLS point clouds. The remainder of the paper is organized as follows. Section 2 briefly reviews and discusses the representative works for pairwise point cloud registration. Section 3 introduces the pipeline of our registration method, including adaptive covariance (ACOV) descriptor generation and non-cooperative matching game construction. Section 4 presents the experimental results. Section 5 presents a detailed discussion of the derived results. Section 6 concludes the paper.

2. Related work on pairwise registration

Pairwise point cloud registration problem can be solved using a rigid transformation with six degrees of freedom by finding a set of correspondences (at least three) between two point clouds. A point cloud registration method usually composes of two steps: coarse registration and fine registration. Over the past decades, a number of pairwise point cloud registration methods have been proposed in different fields, such as computer vision, photogrammetry, and robotics. We will only review and discuss some representative works in this section. For a more complete overview, please refer to the evaluation papers (Salvi et al., 2007; Tam et al., 2013; Pomerleau et al., 2015; Weinmann, 2016).

The most commonly used fine registration methods include the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992; Chen and Medioni, 1992) and its variants (Bae and Lichti, 2008; Gressin et al., 2013; Yang et al., 2013). Bae and Lichti (2008) proposed a Geometric Primitive ICP with Random sample consensus (GPICPR). It uses the normal vector and geometric curvature of a local surface for matching and neighborhood search. The GPICPR algorithm provides an acceptable efficiency and accuracy in finding correspondences for registration. Gressin et al. (2013) demonstrated how the standard ICP algorithm can be improved by geometric features. It is noticed that these ICP methods are able to achieve registration results with high accuracy and efficiency, but they still require a coarse alignment to avoid a local minimum. Therefore, Yang et al. (2013) proposed a global optimal solution for ICP-type registration (Go-ICP) without a coarse alignment by integrating the ICP algorithm with a branch-and-bound (BnB) scheme. The Go-ICP algorithm works well on small-scale point clouds, but it still faces challenges on large-scale point clouds acquired by a TLS system.

Aiger et al. (2008) proposed a method named 4-Points Congruent Sets (4PCS) for coarse registration. This method extracts coplanar 4-points sets which are congruent as primitives. The 4PCS algorithm is demonstrated to be robust to varying degrees of noise, outliers, and overlaps, but it has a quadratic time complexity. Mellado et al. (2014) proposed the SUPER-4PCS algorithm with linear time complexity using a smart indexing for data organization. Theiler et al. (2014) used keypoints (DoG and Harris detectors) to down-sample the original point clouds and then adapted the 4PCS algorithm for registration. Their experiments demonstrated that K-4PCS obtained a sufficiently high registration accuracy for the subsequent ICP refinement with a linear time complexity. In addition, these feature descriptors evaluated (Guo et al., 2014) are commonly used as matching primitives to find correspondences between two point clouds in RANSAC-based methods. Barnea and Filin (2008) exploited features which are invariant to 3D rigid-body transformation as primitives. Rusu et al. (2009) proposed a Sample Consensus based method for initial alignment (SAC-IA) using a local feature called Fast Point Feature Histograms (FPFH) as primitives. Weinmann et al. (2011) proposed an efficient perspective-n-point (EPnP) algorithm by extracting characteristic 2D points based on SIFT features. Yang and Zang (2014) extracted crest lines as matching primitives and then proposed a deformation energy model to find correspondences, they obtained fine registration results with good accuracy in their experiments. Kelbe et al. (2016) generated feature descriptors and tested the RMSE of a blind view-invariant marker-free registration method for TLS data in forest environments. Yang et al. (2016) proposed a registration method based on semantic feature points extracted from large-scale urban scenes.

Albarelli et al. (2009) proposed a game-theoretic perspective for the matching problem. The matching problem was formulated as a non-cooperative game where the potential correspondences are analogous to strategies, while payoffs represent the degree of compatibility between any two correspondences. Further, Albarelli et al. (2010, 2015) cast the selection of correspondences between two point clouds in a game-theoretic framework and obtained a fine registration in a single step. Cirujeda et al. (2015) proposed a covariance descriptor to fuse color and shape information within several neighborhood radii, called multi-scale covariance (MCOV) descriptor. The MCOV descriptor was then combined with a game theoretic framework to find correspondences under global geometric constraints. These game-theoretic methods can obtain a fine registration between small-scale point clouds in a linear time without any initial alignment. However, they cannot handle large-scale TLS point clouds.

Compared to (Cirujeda et al., 2015), this paper proposes a new feature descriptor called ACOV to fuse various information (e.g., geometry, color, and intensity) over an adaptive neighborhood size and develops an improved game-theoretic matching method by adding a Laplacian term for pairwise registration of TLS point clouds.

3. The proposed registration method

This section introduces a new registration framework for TLS point clouds, which consists of two stages: adaptive covariance (ACOV) descriptor generation and non-cooperative game construction.

3.1. Covariance matrix descriptor

The proposed method is expected to fully use the information (e.g., geometry, color, and intensity) of data acquired by a TLS system, by means of the statistical concept of covariance matrix. The covariance matrix is generalized to multiple dimensions to combine a set of random variables and to describe statistically how these variables change in relation to each other. From the descriptor perspective, random variables combined in a covariance matrix must correspond to a set of features computed at a given point. In this section, the extracted features and the covariance matrix (which works as a descriptor) will be introduced firstly. Then,

keypoint extraction using covariance descriptors will be presented. Next, a new ACOV descriptor will be proposed. At last, the distance metric and matching strategy will be presented.

3.1.1. Feature extraction and covariance matrix

For a given point $\boldsymbol{p}$ and its neighborhood within radius r in the scene, a set of features are computed, including geometry (α, β, and γ) (Cirujeda et al., 2015), dimensionality (a_{1D}, a_{2D}, and a_{3D}) (Gressin et al., 2013; Lin et al., 2014), color (R, G, and B) and intensity (I) features.

As shown in Fig. 1, α is the angle between the normal vector at point $\boldsymbol{p}$ and the vector from point $\boldsymbol{p}$ to point $\boldsymbol{p_i}$, β is the angle between the normal vector at point $\boldsymbol{p_i}$ and the vector from point $\boldsymbol{p}$ to point $\boldsymbol{p_i}$, and γ is the angle between the normal vectors at points $\boldsymbol{p}$ and $\boldsymbol{p_i}$. These three angular measures are related to the local surface geometrical structure within the neighborhood of point $\boldsymbol{p}$.

The three dimensionality features a_{1D}, a_{2D} and a_{3D} are computed as follows:

$$a_{1D} = \frac{\lambda_1 - \lambda_2}{\lambda_1}, \quad a_{2D} = \frac{\lambda_2 - \lambda_3}{\lambda_1}, \quad a_{3D} = \frac{\lambda_3}{\lambda_1} \tag{1}$$

where λ_1, λ_2, and λ_3 ($\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$) are eigenvalues obtained by performing a Principal Component Analysis (PCA) on the neighborhood of point $\boldsymbol{p_i}$. If $\lambda_1 \gg \lambda_2$ and λ_3, a_{1D} is therefore larger than a_{2D} and a_{3D}, a linear structure is represented by the neighborhood of point $\boldsymbol{p_i}$. If $\lambda_1 \simeq \lambda_2 \gg \lambda_3, a_{2D}$ is the largest one, and a planar structure is represented by the neighborhood of point $\boldsymbol{p_i}$. At last, $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$ means that a_{3D} is larger than a_{1D} and a_{2D}, and a scatter structure is represented by the neighborhood of point $\boldsymbol{p_i}$.

The intensity entropy $H(I)$ is computed using the histogram of the intensity values of all points within the neighborhood of point $\boldsymbol{p_i}$:

$$H(I) = \sum_{g=0}^{255} p_g \cdot \log\ p_g \tag{2}$$

where p_g is the probability of a gray value $g \in [0, 255]$ obtained from the intensity histogram.

The extracted features contain geometry (α, β, and γ), dimensionality (a_{1D}, a_{2D}, and a_{3D}), color (R, G, and B) and intensity ($H(I)$) information. It should be noted that all of these selected features are invariant to a rigid transformation of the scene, which is important for point cloud registration. Finally, these extracted features are normalized to generalize a statistical distribution of random variables, such that all features can have an equal range. Specifically, α is normalized as $\alpha = (\alpha - \alpha_{min})/(\alpha_{max} - \alpha_{min})$, where α_{min} and α_{max} are the minimum and maximum values of $\{\alpha_{p_i}\}$, respectively. β, γ, and $H(I)$ are normalized following the same way as α. R, G, and B are divided by 255 for normalization.

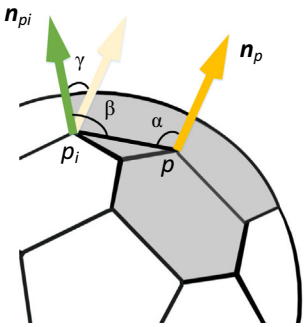


Fig. 1. The angular measures α, β, and γ for a point $\boldsymbol{p_i}$ within the neighborhood of point $\boldsymbol{p}$.

For a given point $\boldsymbol{p}$ and its neighborhood size r, a new patch π_p is obtained as follows:

$$\pi_p = \{\boldsymbol{f}_{p_i},\ \forall\ p_i\ s.t.\ \|\boldsymbol{p} - \boldsymbol{p_i}\| \leqslant r\} \tag{3}$$

where $\boldsymbol{f}_{p_i}$ is a 10-dimensional feature vector obtained at each point $\boldsymbol{p_i}$ within the neighborhood of point $\boldsymbol{p}$, i.e.,

$$\boldsymbol{f}_{p_i} = < \alpha_{p_i}, \beta_{p_i}, \gamma_{p_i}, a_{1D_{p_i}}, a_{2D_{p_i}}, a_{3D_{p_i}}, R_{p_i}, G_{p_i}, B_{p_i}, H(I)_{p_i} >, \tag{4}$$

where α, β, and γ are the angular measures to represent the surface geometric information, a_{1D}, a_{2D} and a_{3D} are three dimensionality features to describe the linear (a_{1D}), planar (a_{2D}) and scatter (a_{3D}) characteristics, R, G and B represent the color space values, and $H(I)$ is the intensity entropy.

When π_p is obtained, a covariance matrix is computed for the given point $\boldsymbol{p}$:

$$C_r(\pi_p) = \frac{1}{N_p - 1} \sum_{i=1}^{N_p} (\boldsymbol{f}_{p_i} - \boldsymbol{\mu})^{\mathrm{T}} (\boldsymbol{f}_{p_i} - \boldsymbol{\mu}) \tag{5}$$

where N_p is the cardinality of π_p and $\boldsymbol{\mu} = \frac{1}{N_p} \sum_{i=1}^{N_p} \boldsymbol{f}_{p_i}$. A covariance matrix descriptor combines these extracted features and describes their mutual relationship. As discussed in Kaiser et al. (2013), the variances on the diagonal of a covariance matrix represent the statistical variations of these features from their own average value. The variance treats all deviations from the mean equally regardless of their directions. As a result, the squared deviations cannot sum to zero. They give the appearance of no variability at all in the feature, which is already informative. Besides, the covariances between any two features on the off-diagonal entries provide further information, which leads to higher descriptiveness as compared to these individual features. Fig. 2 shows an example of covariance descriptors. The facial area of a statue is used to compute covariance descriptors, as presented in Fig. 2(b). Fig. 2(c)–(e) shows the same facial surface with down-sampling, Gaussian noise, and both down-sampling and Gaussian noise, respectively. The corresponding covariance descriptors with a scale of 10×10 are depicted in Fig. 2(f)–(i), respectively. It can be visually observed that these covariance descriptors are similar. More details and qualitative performances will be discussed in the following sections.

3.1.2. Covariance descriptor for keypoint extraction

The determinant $|C_r(\pi_p)|$ can be considered as a generalized variance, it is equal to the variance for the one-dimensional case. Similar to the variance which measures how far a set of numbers are spread out from their mean, $|C_r(\pi_p)|$ measures the variation extent of random variables (features) computed at point $\boldsymbol{p}$. That is, a larger $|C_r(\pi_p)|$ means that more random variables are dispersed in the feature space. It can be interpreted that $|C_r(\pi_p)|$ measures the degree of homogeneity of point $\boldsymbol{p}$ in the scene (Cirujeda et al., 2015). Namely, the points with larger determinant values can be considered as keypoints of the scene, because these points have more significant variations in shape, color and intensity.

Instead of computing covariance descriptors for each point in the scene, we use uniform sampling to obtain initial interest points for covariance descriptor computing. The uniform down-sampling process can improve the method in two aspects: (1) It can initially down-sample the point clouds and significantly increase the computing efficiency; (2) It roughly evens out the strong point distribution variation across TLS point clouds (Theiler et al., 2014).

When the initial keypoint set S_i is obtained after uniform sampling, the covariance descriptor and its corresponding determinant are computed for each $\boldsymbol{p}$ ($\boldsymbol{p} \in S_i$) over its neighborhood in the scene. When the determinant of covariance descriptor for each point $\boldsymbol{p}$ ($\boldsymbol{p} \in S_i$) is obtained, points with higher determinant are selected

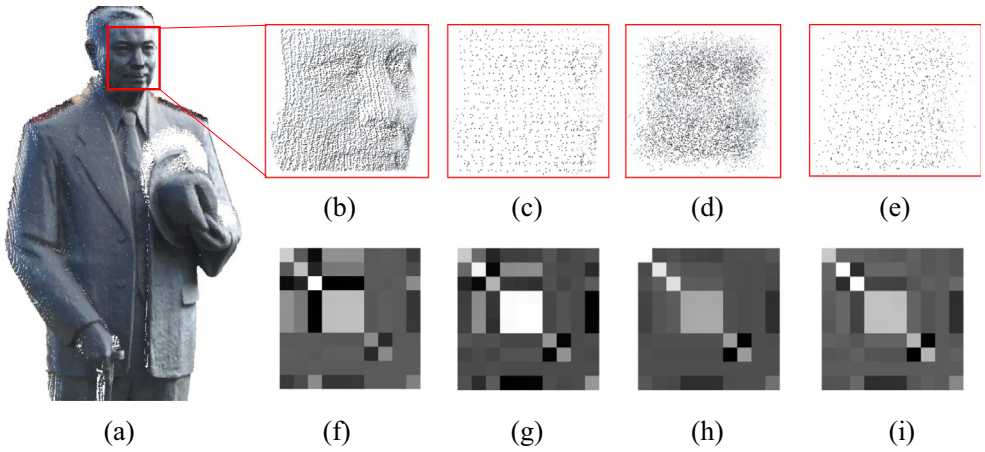


Fig. 2. An example of covariance descriptors. (a) Original point cloud of a statue. (b) Close-up view of the facial surface. (c) Point cloud after down-sampling (20% of the original resolution). (d) Point cloud with Gaussian noise ($\sigma = 0.1$ cm). (e) Point cloud with both down-sampling (20% of the original resolution) and Gaussian noise ($\sigma = 0.1$ cm). (f–i) The corresponding covariance descriptors with a scale of 10×10 for these point clouds shown in (b–e).

as keypoints as these points have a large variation in shape, color and intensity.

3.1.3. Adaptive covariance (ACOV) descriptor

For each keypoint, a covariance descriptor $C_t(\pi)$ can be computed easily over the neighborhood (t) of the keypoint in the original scene. The neighborhood size has an influence not only on the distinctiveness of the descriptor $C_t(\pi)$, but also on its robustness to occlusion, noise, and uneven density (Guo et al., 2013). An adaptive neighborhood size for keypoints can work better than a fixed size, especially when dealing with point clouds with uneven density. In Demantké et al. (2011) and Weinmann et al. (2015), the optimal neighborhood size is selected by computing the dimensionality features and minimizing a measure of unpredictability defined by Shannon entropy. Unnikrishnan and Hebert (2008) detected salient regions to capture shape variation at a point relative to its neighborhood using the Laplace-Beltrami scale-space (LBSS) theory. It is demonstrated that the selected salient neighborhood regions were repeatable and performed well for shape descriptor calculation. Inspired (Unnikrishnan and Hebert, 2008), we propose a method to adaptively select neighborhood sizes for the generation of covariance descriptors to capture variation in geometry, color and intensity.

As our covariance descriptor fuses geometry, color and intensity information, the corresponding determinant measures the degree of variation in geometry, color and intensity. We then define the corresponding determinants of covariance descriptors as the distinctiveness at a point relative to a selected neighborhood size t ($t \in \{t_m\}$). That is, the neighborhood region corresponding to the maximum determinant over a set of sizes $\{t_m\}$ is selected. The set of sizes $\{t_m\}$ for each keypoint is fixed as $t_m = t_0 \cdot b^m$ where t_0 is a basic size. The base b is set to $1.3, m$ is set to $\{0, 1, 2, 3, 4, 5\}$ and t_0 is set to 20 times the average resolution of the model in this paper, as analysed in Section 4.3.

3.1.4. Distance metric and matching strategy

A covariance matrix is symmetric positive definite and it lies on a Riemannian manifold rather than a Euclidean space. The covariance matrix can be used to define a region as it abstractly represents the geometrical location of shape and texture distributions within a region (Cirujeda et al., 2015). Although most of the distance measures (e.g., Euclidean distance) cannot be applied, the log-eigenvalue measure (Kaiser et al., 2013) can be used to measure the similarity of two arbitrary covariance matrices $\boldsymbol{C}_t(\pi_1)$ and $\boldsymbol{C}_t(\pi_2)$, it is defined as:

$$d(\boldsymbol{C}_t(\pi_1), \boldsymbol{C}_t(\pi_2)) = \sqrt{\sum_{i=1}^{D} \ln^2 \lambda_i(\boldsymbol{C}_t(\pi_1), \boldsymbol{C}_t(\pi_2))} \tag{6}$$

where $\lambda_i(\boldsymbol{C}_t(\pi_1), \boldsymbol{C}_t(\pi_2))$ is the set of generalized eigenvalues of $\boldsymbol{C}_t(\pi_1)$ and $\boldsymbol{C}_t(\pi_2)$ and D is the rank of $\boldsymbol{C}_t(\pi_1)$ or $\boldsymbol{C}_t(\pi_2)$ (D is 10 in our case). The generalized eigenvalues of $\boldsymbol{C}_t(\pi_1)$ and $\boldsymbol{C}_t(\pi_2)$ are defined by $\lambda_i \boldsymbol{C}_t(\pi_1)\boldsymbol{v}_i = \boldsymbol{C}_t(\pi_2)\boldsymbol{v}_i$ with $\boldsymbol{v}_i \neq bf0$. A generalized eigenvalue problem can be converted into a normal eigenvalue problem: $\boldsymbol{C}_t(\pi_1)^{-1}\boldsymbol{C}_t(\pi_2)\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$ (Kaiser et al., 2013).

Finally, the ACOV descriptor is used to match points between two point clouds P and Q. The keypoints extracted from P are denoted as P_k and the keypoints in Q for corresponding point searching are denoted as Q_k. We propose a new matching strategy to generate correspondence using the inclusive ratio strategy. That is, an arbitrary correspondence $\{\boldsymbol{p}, \boldsymbol{q}\}$ ($\boldsymbol{p} \in P_k$ and $\boldsymbol{q} \in Q_k$) is considered as a true positive candidate if and only if it satisfies:

$$\begin{cases} d(\boldsymbol{C}_t(\pi_p), \boldsymbol{C}_t(\pi_q)) \leqslant ratio \times \min_{i \in Q_k} d(\boldsymbol{C}_t(\pi_p), \boldsymbol{C}_t(\pi_i)) \\ d(\boldsymbol{C}_t(\pi_p), \boldsymbol{C}_t(\pi_q)) \leqslant ratio \times \min_{j \in P_k} d(\boldsymbol{C}_t(\pi_j), \boldsymbol{C}_t(\pi_q)) \end{cases} \tag{7}$$

where $d(\boldsymbol{C}_t(\pi_p), \boldsymbol{C}_t(\pi_q))$ is the log-eigenvalue distance measure described in Eq. (6). When the parameter *ratio* equals 1, the above strategy degenerates into a one-to-one matching. As *ratio* becomes larger, the rate of true positive candidates is increased, but false positives are also increased. Due to pattern repetitions and local symmetries in a scene, the one-to-one matching is not suitable. In contrast, the inclusive ratio strategy with a fixed *ratio* is more appropriate for a registration procedure as a point can correspond to several matches as long as they satisfy Eq. (7), at the cost of requiring a subsequent rejection method to deal with the large number of false positives.

3.2. Non-cooperative game

This section introduces a new rejection method to remove false correspondences using a non-cooperative game. The basic ideas for correspondence estimation using game theory is firstly introduced. Then, a payoff matrix is defined for a matching game. At last, an optimal solution to the matching game is evolved using the replicator dynamics.

3.2.1. Basic ideas

The concept of game theory used for matching purposes is first introduced by Albarelli et al. (2009), where correspondences are discarded in order to obtain the best combination of correspondences by imposing geometric constraints. This matching approach has been applied successfully to other pairwise registration related applications (Albarelli et al., 2010, 2012, 2015; Rodola et al., 2012; Cirujeda et al., 2015).

The underlying idea is to consider all correspondences (denoted as S) resulted from a descriptor matching stage, and isolate mutually compatible ones from the outliers by calculating payoffs (denoted as π) associated to each two selected correspondences. That is, when S is generated, our goal is to extract the subset including most correct correspondences from S by imposing constraints. Similar to a game G, each point cloud (P or Q) is a player and these correspondences S form the set of pure strategies. The game is summarized as a triplet $G = \{I, S, \pi\}$, where I is the player set $\{P, Q\}$, S is the pure-strategy set, and π is the combined payoff function. At equilibrium, only the mutually compatible correspondences are preserved and considered as inliers. Hence, the best set of strategies for the game (the registration between two point clouds) becomes the selection of mutually compatible correspondences. Previous research has demonstrated that by properly designing the payoff function π, these selected correspondences are globally optimal for the rigid transformation problem (Albarelli et al., 2010, 2015).

3.2.2. Definition of a payoff matrix

As a key step of the game theoretic matching framework, the definition of a payoff function π must consider pairwise interactions between correspondences S, which is denoted as $\pi: S \times S \rightarrow \mathbb{R}^+$. That is, the payoff function π is materialized in a symmetric payoff matrix Π to represent a degree of compatibility. Particularly, true positive correspondences conform to some property between each other while false correspondences and randomly paired correspondences should not comply with the same property. Considering that our problem conforms to a rigid transformation, it is natural to impose geometric constraints on those correspondences to discard false correspondences.

Let $s_i, s_j \in S$ be two arbitrary correspondences, and $s_i = \{\boldsymbol{p_a}, \boldsymbol{q_b}\}$, $s_j = \{\boldsymbol{p_c}, \boldsymbol{q_d}\}$, where $\boldsymbol{p_a}, \boldsymbol{p_c} \in P_k$ and $\boldsymbol{q_b}, \boldsymbol{q_d} \in Q_k$. Based on the above considerations, a distance measure δ between s_i and s_j is defined as:

$$\delta(s_i,\ s_j) = \frac{\min(d(\boldsymbol{p_a}, \boldsymbol{p_c}), d(\boldsymbol{q_b}, \boldsymbol{q_d}))}{\max(d(\boldsymbol{p_a}, \boldsymbol{p_c}), d(\boldsymbol{q_b}, \boldsymbol{q_d}))} e^{\frac{-|d(\boldsymbol{p_a}, \boldsymbol{p_c}) - d(\boldsymbol{q_b}, \boldsymbol{q_d})|}{\gamma}} \tag{8}$$

where $d(\boldsymbol{a}, \boldsymbol{b})$ represents the Euclidean distance between points $\boldsymbol{a}$ and $\boldsymbol{b}$. The min/max term was originally proposed in Albarelli et al. (2010) and the Laplacian term is proposed in this paper to enlarge the distribution range of payoff values. The parameter γ is used to control the strictness of the constraints. Note that, the Laplacian term plays a significant role in our game theoretic matching method. Unlike the previous works (Albarelli et al., 2010, 2012, 2015; Rodola et al., 2012; Cirujeda et al., 2015), our registration problem focuses on the applications of TLS point clouds. Since a compatibility is set between each pair of correspondences, the size of a payoff matrix is $|S| \times |S|$. Such a large memory requirement is intolerable for TLS point clouds. Fortunately, the payoff matrix can be simplified into a sparse symmetric matrix by setting those payoff values smaller than a given threshold (0.1 in our work) to 0. To achieve this, a wide range of payoff values is necessary and the Laplacian term is used to find a small number of functional entries in the payoff matrix.

Since the final transformation solution has to be one-to-one, an additional constraint is imposed on the payoffs: $\pi(s_i, s_j) = 0$ if $\boldsymbol{p_a} = \boldsymbol{p_c}$ or $\boldsymbol{q_b} = \boldsymbol{q_d}$. Thus, the payoff function π is defined as follows:

$$\pi(s_i,\ s_j) = \begin{cases} 0, & \text{if } \boldsymbol{p_a} = \boldsymbol{p_c} \text{ or } \boldsymbol{q_b} = \boldsymbol{q_d} \text{ or } \delta(s_i,\ s_j) < 0.1 \\ \delta(s_i,\ s_j), & \text{otherwise.} \end{cases} \tag{9}$$

An example of a payoff matrix is shown in Fig. 3. Each correspondence has a zero payoff with itself, correspondences sharing the same source or destination point also have a zero payoff (e.g., $\pi(s_1,\ s_2) = 0$). Correct correspondences have high payoff values (e.g., $\pi(s_1,\ s_3) = 1$), while less compatible correspondences have low payoff values (e.g., $\pi(s_1, s_4) = 0.2$) or even zero payoff (e.g., $\pi(s_2, s_3) = 0$), as shown in Fig. 3(b).

3.2.3. Evolution to an optimal solution

When the matching game G is generated, the optimal solutions for the rigid transformation problem correspond to Evolutionary Stable Strategies (ESS), a robust population-based generalization of Nash equilibrium (Albarelli et al., 2010). In the game, the support vector $\boldsymbol{x}$ is an ESS if it is a Nash equilibrium and satisfies that $\boldsymbol{x}^T \Pi \boldsymbol{y} > \boldsymbol{y}^T \Pi \boldsymbol{y},\ \forall \boldsymbol{y} \in \triangle$, where $\triangle = \{\boldsymbol{x} \in \mathbb{R}^{|S|} : \sum_{i=1}^{|S|} x_i = 1 \text{ and } x_i \geqslant 0\}$ is a probability distribution over the strategy set S. Our matching game starts by setting each strategy an initial probability $1/|S|$, which can be evolved iteratively by applying the replicator dynamics equation:

$$x_i(n+1) = x_i(n) \frac{(\Pi \boldsymbol{x}(n))_i}{\boldsymbol{x}(n)^T \Pi \boldsymbol{x}(n)} \tag{10}$$

where n is the number of iterations, and Π is the payoff matrix. When Eq. (10) converges, a correspondence $s_i \in S$ is considered as positive if $x_i > 0$. In contrast, if $x_j = 0$, the correspondence $s_j \in S$ is considered as negative. Once the plausible correspondences are obtained, the rigid transformation is computed to align two point clouds using Singular Value Decomposition (SVD).

4. Experimental results

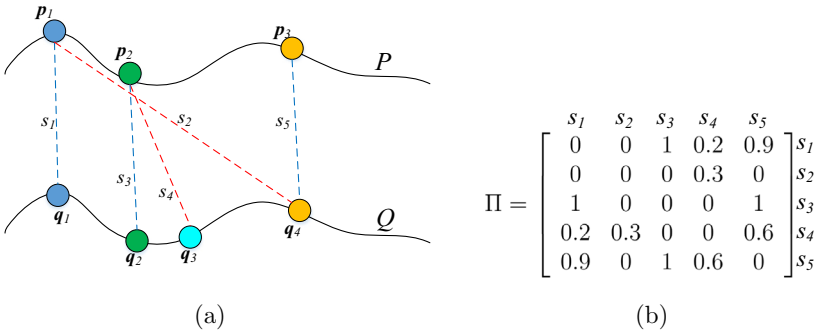
In this section, several experiments are conducted to demonstrate the superiority of the proposed ACOV descriptor and the point cloud registration framework.

4.1. Test data

Our proposed ACOV descriptor and registration method were evaluated on three models containing geometry, color and intensity information, which were acquired by two different TLS systems. For each model, ground truth for registration was obtained by first aligning all of these scans manually and then performing ICP fine registration. Fig. 4 shows the 3D models for test. The information of the tested models is summarized in Table 1, the average resolution $\bar{r}$ (the average distance between each two adjacent points on the model) of each model and the overlap between each two successive scans of a model are also given. Our methods were implemented in C ++ and the experiments were conducted on a PC with Windows 7, Intel Core(TM) i5-4460 3.2 GHz CPU and 48.0 GB RAM.

4.2. Keypoint extraction parameter setting

To obtain an appropriate neighborhood size for initial covariance descriptor calculation, we test the repeatability of keypoints over Gaussian noise and resolution changes. The keypoints were first extracted using the initial covariance descriptors with different neighborhood sizes from a model and its corresponding model with down-sampling (50% of the original resolution) and Gaussian noise ($\sigma = 5\bar{r}$). Then, the overlap was calculated between two corresponding keypoints, as shown in Fig. 5. On the three models, we can obtain the best overlap between corresponding keypoints



$$\Pi = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} \begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0.2 & 0.9 \\ 0 & 0 & 0 & 0.3 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0.2 & 0.3 & 0 & 0 & 0.6 \\ 0.9 & 0 & 1 & 0.6 & 0 \end{bmatrix} \end{array}$$

Fig. 3. An example of correspondences and the corresponding payoff matrix. (a) P and Q are two compared surfaces and five correspondences are selected. The blue and red dotted lines denote correct and wrong correspondences, respectively. (b) Matrix Π shows a sparse characteristic and expresses the compatibility between pairs of correspondences according to Eq. (8). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. An illustration of tested 3D models (shown in color). (a) Statue 1. (b) Statue 2. (c) Church. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Information of the tested models.

Model	Number of scans	Number of points	$\bar{r}$ (cm)	Overlap (%)
Statue 1[a]	4	864,451	0.18	50–70
Statue 2[a]	3	1,205,163	0.26	38–47
Church[b]	2	3,028,709	1.34	38

[a] The data is acquired by RIEGL VZ-1000 system.
[b] The data is downloaded from Semantic3D (Hackel et al., 2017).

when the neighborhood size is set to $25\bar{r}$. It is also observed that a neighborhood size within the range of $[25\bar{r},\ 35\bar{r}]$ can still produce promising results. That is, the initial covariance descriptor is robust to different neighborhood sizes. Note that, a larger neighborhood size requires more computational time. Therefore, in the following experiments, the neighborhood size for initial covariance descriptor calculation is set to $25\bar{r}$. Fig. 6 shows an example of keypoint extraction results on model "Church" and its down-sampled model.

4.3. ACOV generation parameter setting

As discussed in Section 3.1.3, our ACOV descriptor depends on three parameters: the basic size t_0, the scale m, and the base b. The performance of ACOV descriptor under different parameter settings was tested on three models using the Precision-Recall

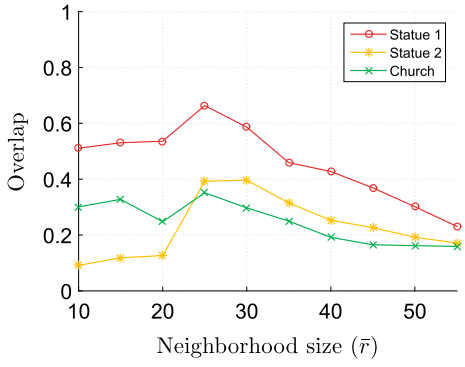


Fig. 5. The overlap of keypoints between one model and its variant with down-sampling (50% of the original resolution) and Gaussian noise ($\sigma = 5\bar{r}$) using different neighborhood sizes.

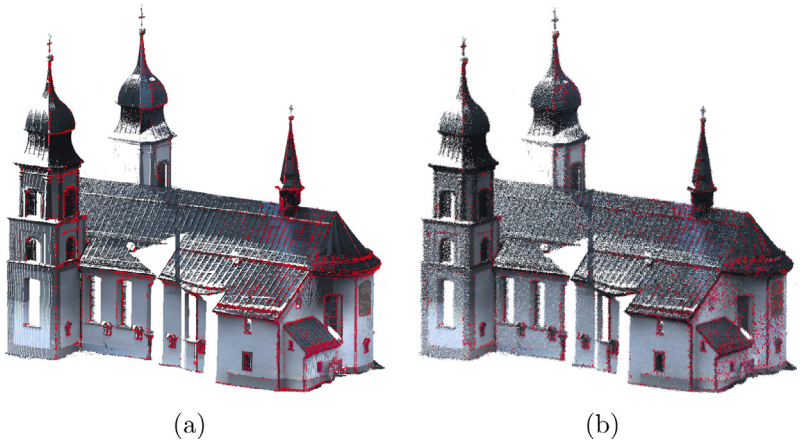


Fig. 6. An example of keypoint extraction using covariance descriptors with the neighborhood size $25\bar{r}$. Keypoints are shown in red. (a) Result of keypoint extraction from the original model "Church". (b) Result of keypoint extraction from the variant of model "Church". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(PR) Curve. Given a model and its variant with down-sampling (50% of the original resolution) and Gaussian noise ($\sigma = 5\bar{r}$), keypoints were first extracted from the model and its variant. Then, ACOV descriptors were computed for each keypoint. For each feature on the model, we computed its descriptor similarity against all features on the down-sampled model using the inclusive ratio (see Section 3.1.4), where the parameter *ratio* was set within a range of $[1, 5]$. Using these matching points, precision and recall can be computed as:

$$\begin{cases} precision = \frac{TP}{TP+FP} \\ recall = \frac{TP}{TP+FN} \end{cases} \tag{11}$$

where *TP* is the number of true positive correspondences, *FP* is the number of false positive correspondences, and *FN* is the number of false negative correspondences. The PR Curve can be generated by tuning the parameter *ratio*.

The basic size t_0 plays an important role in the generation of an ACOV feature descriptor. A larger value of t_0 enables the ACOV descriptor to encode more information of surface, but it also increases the sensitivity to occlusion and noise. We tested the performance of the ACOV descriptor with respect to varying basic sizes while keeping the other two parameters fixed ($b = 1.3$ and $m = 5$). Fig. 7(a)–(c) shows that the performance of the ACOV descriptor was improved steadily as the basic size was increased from $10\bar{r}$ to $20\bar{r}$. As the basic size was increased to $30\bar{r}$, the performance of the ACOV descriptor was further improved slightly on models "Statue 2" and "Church" (Fig. 7(b) and (c)), but the performance deteriorated sharply on model "Statue 1" (Fig. 7(a)). Therefore, the basic size t_0 is set to $20\bar{r}$ in this paper to maintain a robustness to occlusion and noise.

Similarly, we tested the performance of the ACOV descriptor with respect to varying base b as shown in Fig. 7(d)–(f). The other two parameters were set as $t_0 = 20\bar{r}$ and $m = 5$. The results show that the performance of ACOV descriptor was improved as the base was increased from 1.1 to 1.3. That is because the 6 neighborhoods are discriminating between each other. However, as the base was further increased, the performance deteriorated, as shown in Fig. 7(d)–(e). That is because a larger base increases the sensitivity to occlusion and noise. The performance of ACOV descriptor was increased slightly on model "Church" as the base was further increased, as shown in Fig. 7(f). That is because the scale of model "Church" is much larger than the other two models. Since a good overall performance can be achieved on these three models when b is 1.3, we set the base b to 1.3 in this paper.

We tested the performance of ACOV descriptor with respect to different scales while the other two parameters were set to $t_0 = 20\bar{r}$ and $b = 1.3$, the results are shown in Fig. 7(g)–(i). It can be found that there is no obvious difference in the performance of ACOV descriptor on model "Statue 1" for different scales, as shown in Fig. 7(g), although the performance was increased slightly as the scale was increased from 3 to 7. Moreover, increasing the scale will cost more computational and memory resources. Therefore, we set the scale to 5 in this paper.

4.4. Descriptor comparison

To demonstrate the superiority of our ACOV descriptor, we compared it with three state-of-the-art methods including FPFH (Rusu et al., 2009), SHOT (Tombari et al., 2010), and MCOV (Cirujeda et al., 2015). FPFH and SHOT descriptors are available in the open-source Point Cloud Library (PCL) (Rusu and Cousins, 2011). Therefore, the default parameters in PCL implementations for FPFH and SHOT are used, except for support radius. Further, FPFH and SHOT were tested with different support radii $(10\bar{r}, 15\bar{r}, 20\bar{r}, 25\bar{r}, 30\bar{r})$ on a model, the support radius with the best performance was used in this paper. The original implementation and default parameters was used for MCOV descriptor.

In this paper, we used Precision-Recall (PR) Curve to evaluate the distinctiveness of these descriptors over Gaussian noise and resolution changes. For fair comparison, the same inclusive ratio matching strategy in Section 3.1.4 was used to find candidate correspondences based on their associated descriptor similarity measures and parameter *ratio* was set within a range of $[1, 5]$. For each model, a set of points (we empirically set the number of points to 1000) were randomly selected to evaluate the descriptors. Then, for each selected point, we computed its descriptor similarity against the same set of points on the model with noise and down-sampling.

Note that, these three models provide different challenging scenarios for the comparison of descriptors in terms of color homogeneity, intensity homogeneity, repetitive patterns or identical local shapes. In order to test the proposed descriptor over Gaussian noise and varying resolutions, we down-sampled the models to 50% of their original resolution and added three different levels of Gaussian noise ($\sigma = 2\bar{r}, 5\bar{r}$ and $10\bar{r}$) to the models. Fig. 8 shows the variations of a model. The resulting PR Curves are presented in Fig. 9.

It was observed that on model "Statue 1", where its color is almost homogeneous, our proposed ACOV descriptor achieved the best performance. Specifically, the performance of ACOV was

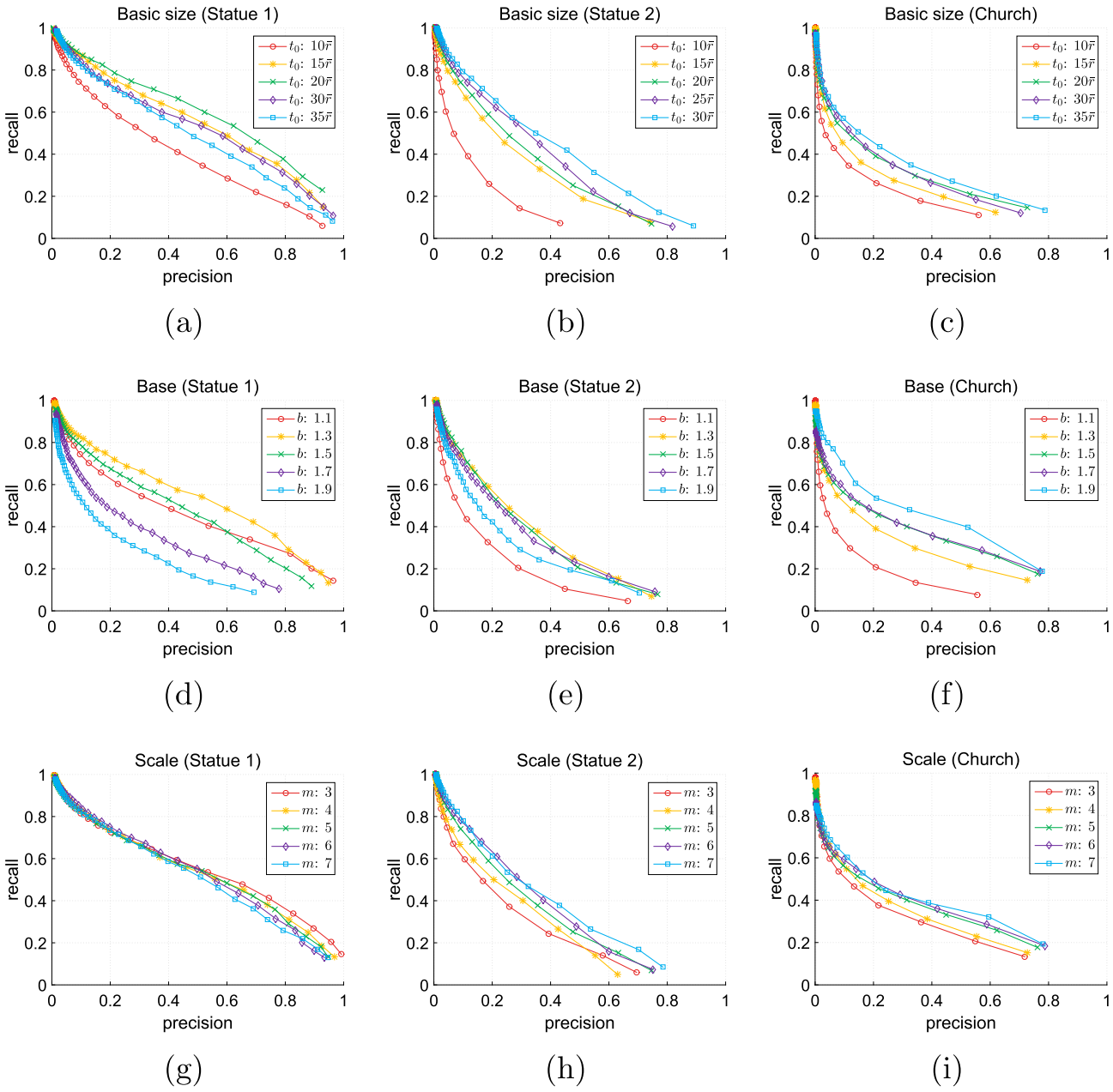


Fig. 7. Effect of the ACOV generation parameters. (a–c) Different basic sizes on three models, respectively. (d–f) Different bases on three models, respectively. (g–i) Different scales on three models, respectively. (We set $t_0 = 20\bar{r}, b = 1.3$ and $m = 5$ in this paper as a tradeoff between effectiveness and efficiency.)

significantly better than other descriptors under high levels of noise, e.g., with a noise deviation of $5\bar{r}$ or $10\bar{r}$, as shown in Fig. 9 (b) and (c). It can be inferred that ACOV is effective and robust to noise even in the case of a homogeneously colored model.

For model "Statue 2" where both the color feature and the intensity feature are homogeneous, our proposed ACOV descriptor achieved the best performance under Gaussian noise with a standard deviation of $2\bar{r}$ or $5\bar{r}$, as shown in Fig. 9(d) and (e). As the standard deviation of Gaussian noise increased to $10\bar{r}$, ACOV performed slightly worse than SHOT, but it was still better than the other two descriptors, as indicated in Fig. 9(f).

Model "Church" provides challenging scenarios such as repetitive patterns and local symmetries, our proposed ACOV descriptor outperformed all the other descriptors by a large margin under all levels of Gaussian noise, as indicated in Fig. 9(g)–(i). The strong robustness of ACOV to these challenging scenarios is due to at least two factors. First, because other methods work on local surface neighborhoods and rely on the variation of a local shape, their performance will be deteriorated on the data with identical local shapes. On the contrary, the ACOV descriptor offers a more comprehensive representation as it fuses geometry, color and intensity information. Second, compared to other descriptors using a fixed support radius, we use a more flexible approach to adaptively estimate the support radius for the calculation of ACOV descriptors, which further improves the robustness and distinctiveness of ACOV descriptor.

4.5. Registration parameter setting

Our proposed game theoretical registration method depends on the number of keypoints and the size of the payoff matrix Π. Note that, if the number of keypoints is too small, it is difficult to cover

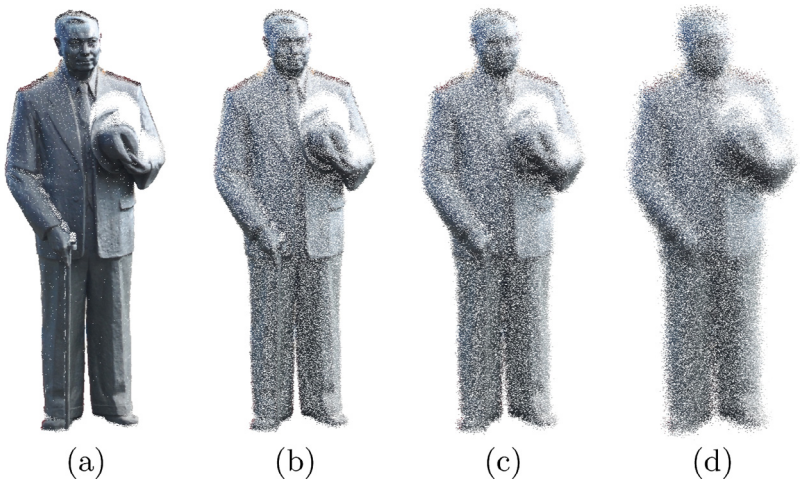


Fig. 8. An example of the variations of a model. (a) The original model. (b–d) The models with down-sampling (50% of the original resolution) and different levels of Gaussian noise ($\sigma = 2\bar{r}, 5\bar{r},$ and $10\bar{r}$).

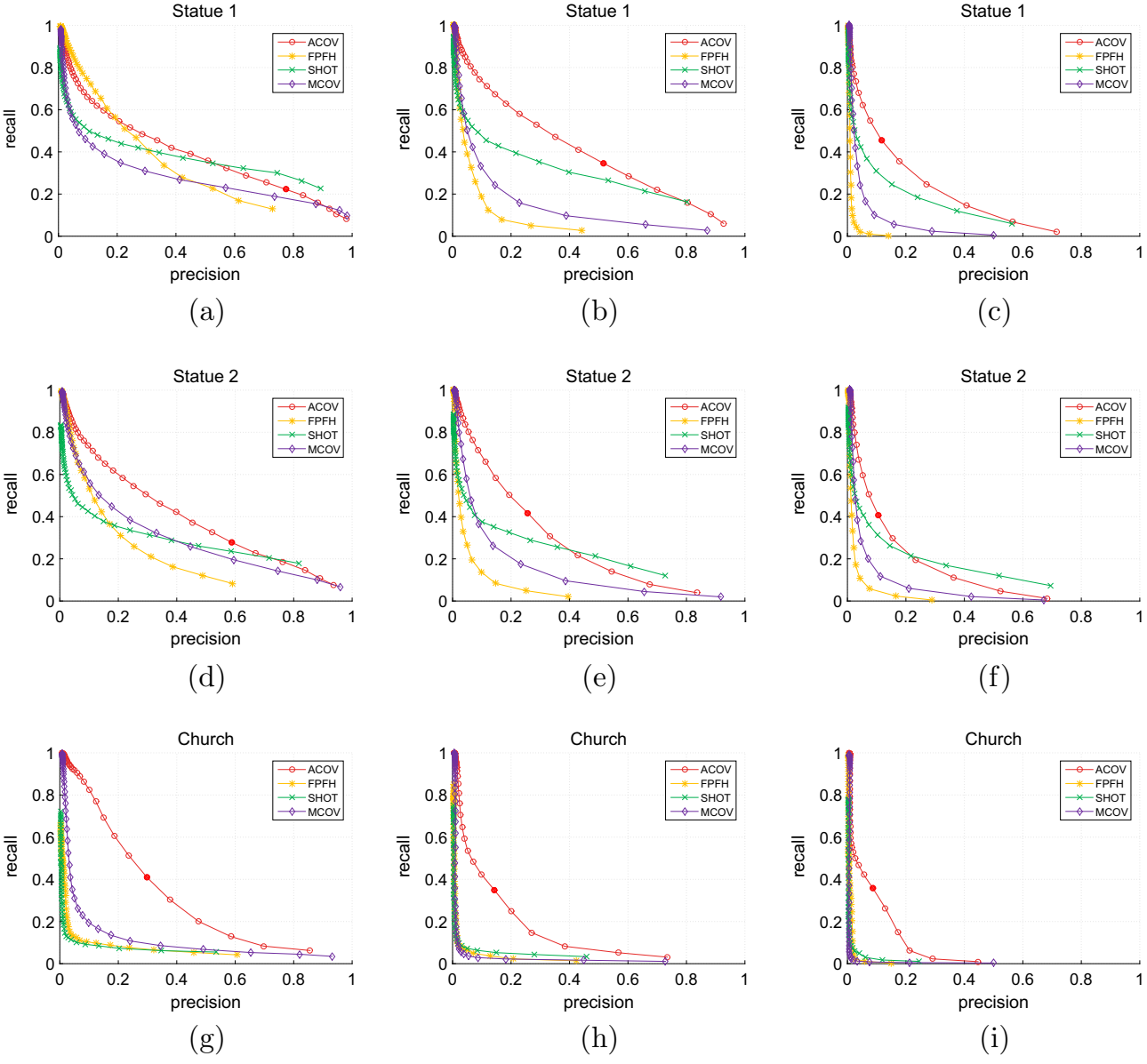


Fig. 9. PR Curves achieved by four 3D descriptors on three models. Different rows represent the results achieved on different models. Different columns represent the results on a particular model with down-sampling (50%) and different levels of Gaussian noise ($\sigma = 2\bar{r}, 5\bar{r}$ and $10\bar{r}$). The red solid point on ACOV PR Curve represents the result with the inclusive ratio $ratio = 1.5$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the overlapping region. In contrast, a large size introduces high time cost. The size of the payoff matrix Π is determined by two parameters *ratio* and γ. Parameter *ratio* determines the number of candidate correspondences, i.e., the dimension of the payoff matrix, while γ determines the number of nonzero elements in the matrix. In Section 4.4, we have discussed the performance of ACOV on three models where the parameter *ratio* is set within a range of $[1, 5]$. As shown in Fig. 9(a)–(l), a large *ratio* leads to a high recall but a low precision, which means that a large number of outliers are introduced. In contrast, a small *ratio* leads to a high precision but a low recall, which means that few true positive correspondences are obtained. Considering the tradeoff between precision and recall, we set *ratio* to 1.5. In our experiment, we can achieve a recall within the range of $(0.2,\ 0.5)$ and a precision larger than 0.1, as shown in Fig. 9.

Two successive scans with varying overlaps were selected from these three models to test the performance of the proposed registration method with different values of γ under different numbers of keypoints. Before keypoint extraction in these point clouds, uniform sampling was first used to obtain initial interest points. Then, three different numbers of keypoints with higher determinants (i.e., 5%, 10% and 15% with respect to the number of initial interest points) were selected, where the parameter γ was set within the range of $[0.1t_0, t_0]$, where t_0 is the basic size defined in Section 3.1.3.

The registration results with different parameters were shown in Fig. 10. It can be observed that with a small number of keypoints (5%), the proposed method failed to register two point clouds below 40% overlap, as shown in Fig. 10(a), (b), (d)–(f). However, with more keypoints and a suitable parameter γ, the proposed method achieved a good result. With a high overlap (e.g., 70%), our method achieved good result even with a small number of keypoints, as shown in Fig. 10(c). Fig. 10(a) shows with 25% overlap, a low γ lead to a poor result. However, with a large γ, our method still obtained comparable result as point clouds of model "Statue 1" with 40% or 70% overlap. Meanwhile, on the models "Statue 2" and "Church", our method achieved an acceptable result with sufficient keypoints. Therefore, it can be concluded that our proposed method is robust to parameter γ and the number of keypoints.

4.6. False correspondence rejection

Compared to commonly used methods for false correspondence rejection, e.g., geometric consistency (Chen and Bhanu, 2007; Yang et al., 2016), transformation parameters clustering (Zhong, 2009; Guo et al., 2013), and RANSAC (Rusu et al., 2009), the procedure of a game theoretic method converges to a limited set of correspondences by successively removing false correspondences.

Geometric consistency methods segment potential correspondences using geometric constraints and then select the largest geometrically consistent group to estimate the transformation. Transformation parameters clustering methods calculate all the plausible transformations using randomly selected correspondences and then group these transformations into several clusters in the parameter space. A confidence score is calculated for each cluster and the cluster with the highest score is selected for verification. These methods are not robust to large amounts of false correspondences (e.g., 99%). RANSAC methods usually find the optimal solution using brute-force search through an iterative evaluation. At each iteration, a subset of candidate correspondences are

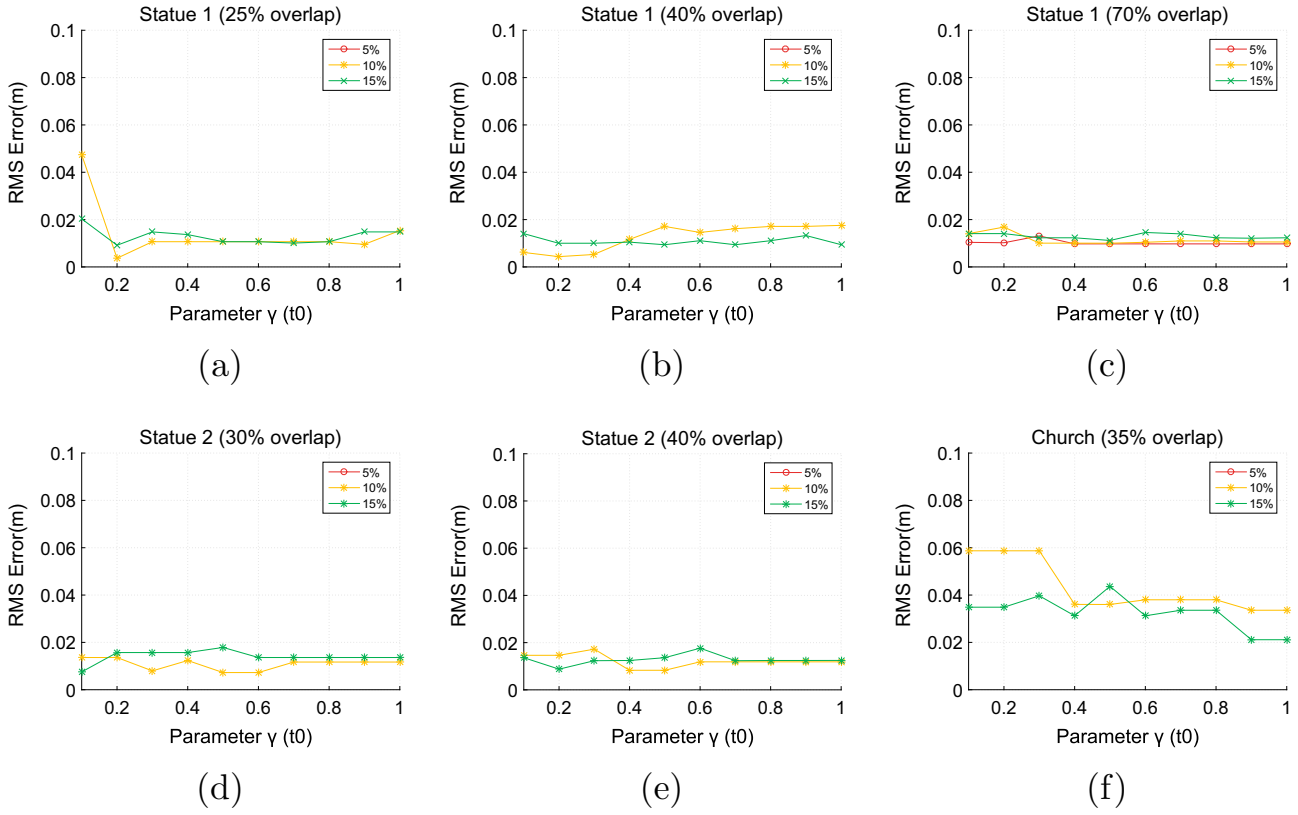


Fig. 10. The RMS Error between corresponding common points under three different numbers of keypoints. The numbers of keypoints are set to 5%, 10% and 15% of the number of initial interest points, respectively, are shown in red, yellow and green. With a small number keypoints (5%), the proposed method failed with an overlap less than 40%. (a) "Statue 1"(25% overlap). (b) "Statue 1"(40% overlap). (c) "Statue 1"(70% overlap). (d) "Statue 2"(30% overlap). (e) "Statue 2"(40% overlap). (f) "Church"(35% overlap). (Overlap is measured as $n_C/(n_P + n_Q)$, where n_c, n_P and n_Q are the numbers of common points, source point cloud and target point cloud, respectively.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Average time cost by two false correspondence rejection methods on three models.

	Time (s)					
Initial positive ratio	75%	50%	25%	10%	5%	1%
Game theory	0.01	0.01	0.02	0.05	0.08	0.11
RANSAC	0.43	0.72	6.61	102.59	346.91	–

randomly selected. That is, the ratio of inlier candidates changes randomly during these iterations. In contrast, a game theoretic method implicitly selects the elements in the payoff matrix, it can evolve to an optimal solution during iterations.

To test the performance of our game theoretic method, an experiment was conducted to compare our method with the RANSAC method. Specifically, the computational time needed to obtain positive correspondences was tested. Given two instances of a model, 500 candidate correspondences was established. Different numbers of negative correspondences (from 1% to 75%) were used to test the performance. The number of iterations in RANSAC was set to 10,000 in this experiment. The number of iterations for the game theoretic method was bounded by the number of initial candidates (500 in the current experiment) regardless of the initial positive ratio (Cirujeda et al., 2015). Starting from the initial set of 500 candidates, the average time for a method to converge to a sub-set of correspondences with 100% plausible correspondences on three models was measured, as shown in Table 2. It is observed that the proposed game theoretic method converged to the optimal solution within 1 s for any initial positive ratio. In contrast, when the initial positive ration was decreased to 25%, the RANSAC method took more time than our game theoretic method by two orders of magnitude. When the initial positive ratio was further decreased to 1%, the RANSAC method cannot converge within 10,000 iterations. For real scenes, a limited set of initial correspondences are always corrupted by outliers, the proposed game theoretic method is able to provide the optimal solution by isolating mutually compatible correspondences from many outliers.

4.7. Registration results

The proposed registration method was tested on a number of successive scans acquired from the models described in Table 1. These scans have various amount of noise, outliers, and overlap. Then, we aligned all scans of tested models using the proposed method. We compared our method with the state-of-the-art methods in terms of both RMS Error distance and efficiency (Aiger et al., 2008; Rusu et al., 2009; Mellado et al., 2014; Theiler et al., 2014).

Fig. 11 presents the results achieved by the proposed method under different levels of overlap. Our method achieved a good result without any ICP refinement even with 25% overlap. Fig. 12 presents the results achieved by the proposed method on the point clouds with varying levels of Gaussian noise. The proposed method successfully registered scans even with a high level of Gaussian noise. Fig. 13 shows the results achieved by the proposed method on point clouds corrupted with varying number of outliers. The successfully registered scans have demonstrated the robustness of our method with respect to outliers. In summary, the proposed method is robust to low overlap, Gaussian noise and outliers.

To further demonstrate the performance of our method, it is compared with the state-of-the-art methods (Aiger et al., 2008; Rusu et al., 2009; Mellado et al., 2014; Theiler et al., 2014). The performance of these registration methods are tested on three models. Each test was conducted without any assumption about initial positions or orientations and was run for 10 times over the set of scans (the initial positions and orientations were randomly

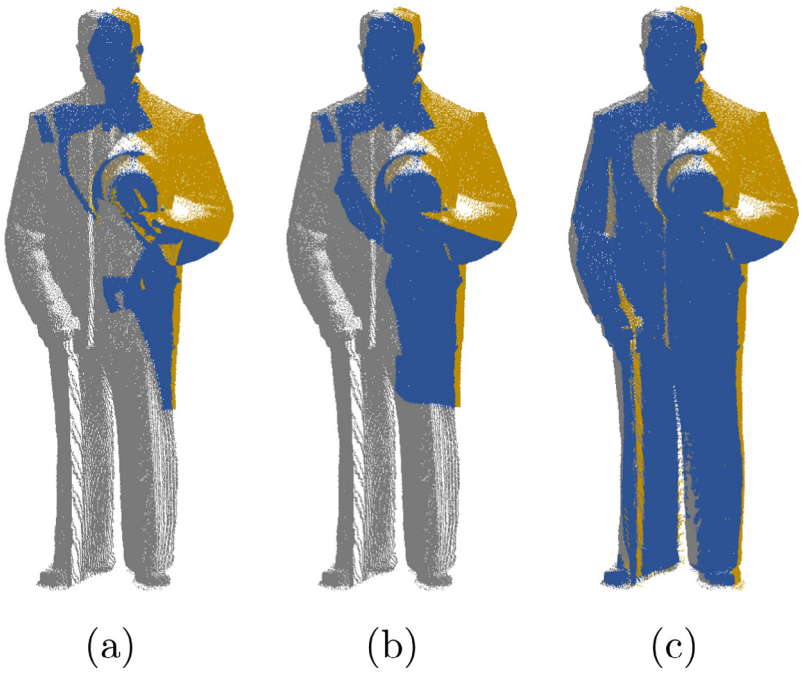


Fig. 11. The results achieved by the proposed method under different levels of overlap. The overlapping regions are shown in blue, the remaining parts of the input point clouds are shown in gray and golden. All results were obtained without any ICP refinement, or assumption about initial positions of the input data. (a) 25% overlap. (b) 40% overlap. (c) 70% overlap. (Overlap is measured as $n_C/(n_P + n_Q)$, where n_c, n_P and n_Q are the numbers of common points, source point cloud and target point cloud, respectively.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

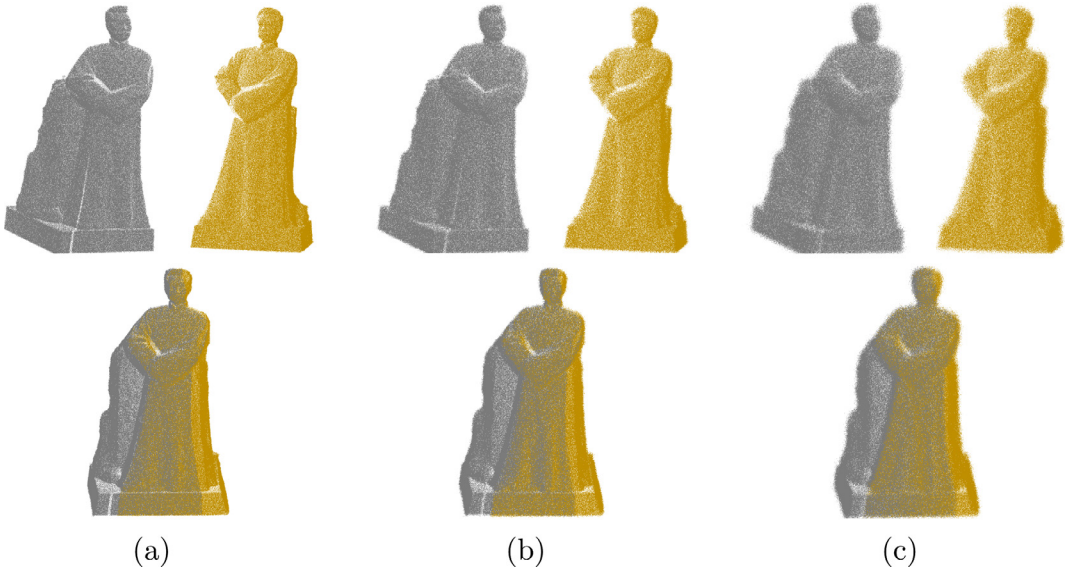


Fig. 12. The results achieved by the proposed method under varying levels of Gaussian noise. (a) $\sigma = 2\bar{r}$. (b) $\sigma = 5\bar{r}$. (c) $\sigma = 10\bar{r}$.

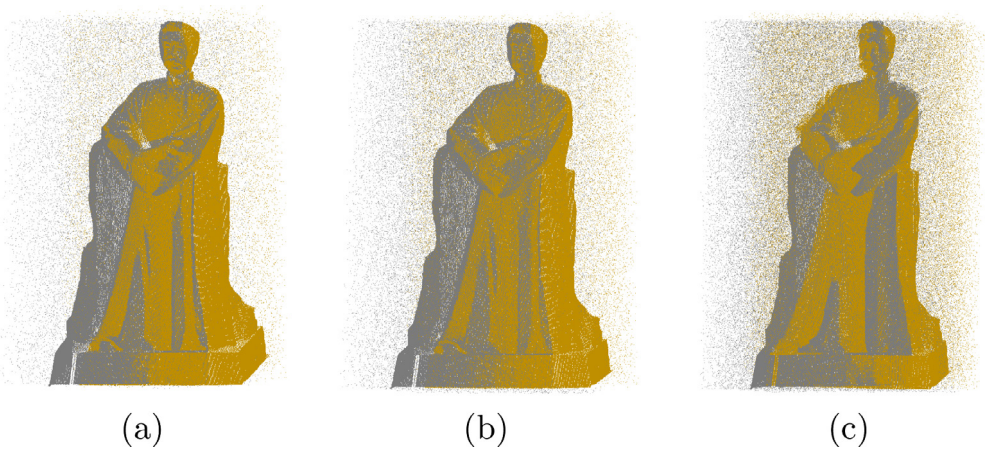


Fig. 13. The results achieved by the proposed method under varying amount of outliers. (a) 5% outliers. (b) 10% outliers. (c) 20% outliers. The ratio of outliers is calculated as the percentage of outliers related to the number of input points.

determined for each test). For each set of experiments, we computed the average RMS distance between common points of the two point clouds, and the average time for aligning all successive scans of each model. A pair of common points $\{\boldsymbol{p}, \boldsymbol{q}\}$ ($\forall \boldsymbol{p} \in P$ & $\forall \boldsymbol{q} \in Q$) is defined if the following two requirements are satisfied: (a) point $\boldsymbol{p}$ is the closest one in P to point $\boldsymbol{q}$ or vice versa; (b) the distance between point $\boldsymbol{p}$ and point $\boldsymbol{q}$ is smaller than a pre-defined threshold (0.1 m in this paper).

Table 3 shows the time cost by each step of our proposed registration framework. It is observed that the correspondence generation step took the majority of time. However, this step can be significantly accelerated by using a data structure (e.g., KD-Tree).

Table 4 shows the comparative performance of these methods in terms of RMS Error distance and computational time. It is observed that our method outperforms other methods in terms of both registration error and computational time. The method proposed by Rusu et al. (2009) fails to register the scans of model "Church" and the methods (Aiger et al., 2008; Mellado et al., 2014; Theiler et al., 2014) are inferior to our method because of the challenges on model "Church" (e.g., occlusion and noise), as shown in Fig. 14(a)–(d).

Table 3
Time cost by each step in our proposed registration framework.

Step	Time (s)		
	Statue 1	Statue 2	Church
Keypoint extraction	28.81	12.49	32.59
Covariance descriptor calculation	45.25	25.5	50.89
Correspondence generation	172.17	108.56	713.54
False correspondence rejection	34.34	32.01	84.16
Total time (s)	280.57	178.56	881.18

With the subsequent ICP refinement, our proposed method achieved ever better registration results, as shown in Table 4 and Fig. 14(e). The number of iterations for the convergence of ICP depends on the model and the initial position of point clouds under registration. In our experiments, the ICP algorithm usually converged after 20–40 iterations.

Table 4
A comparison of point cloud registration results.

Method	Average RMS Error (cm)			Total Time (s)		
	Statue 1	Statue 2	Church	Statue 1	Statue 2	Church
FPFH + RANSAC (Rusu et al., 2009)	1.39	1.48	–	438.30	910.05	–
4PCS (Aiger et al., 2008)	1.35	2.04	5.24	3502.24	898.32	9919.17
Super 4PCS (Mellado et al., 2014)	1.60	1.54	4.97	1050.80	302.73	3721.01
3DHarris + 4PCS (Theiler et al., 2014)	1.51	1.34	2.96	485.95	221.56	1847.56
Ours'	1.33	1.05	2.85	**280.57**	**178.56**	**881.18**
Ours' + ICP	**0.46**	**0.32**	**1.73**	288.87	184.38	902.68

4.8. Scene registration results

To further demonstrate the feasibility and effectiveness of our pairwise registration framework, a data set obtained from a square in Feldkirch, Austria by a TLS system was used (Hackel et al., 2017). The data set consists of two successive scans captured from different views, it covers a range of approximately 150 m by 15 m and includes about 46 million points. Note that, most objects in the scene are buildings.

When keypoints were extracted using initial covariance descriptors, we added a constraint for each point to reduce computational resources (memory and time): $a_{1D} > 0.5$. This constraint ensures that each extracted keypoint lies on a linear surface (Demantké et al., 2011). The numbers of keypoints extracted from these two scans were 5272 and 5755, respectively. 68,339 candidate correspondences were obtained successively using ACOV descriptors. Finally, 21 true correspondences were obtained using the proposed game theoretic method by setting parameter γ to 0.01, as shown in Fig. 15(a). Fig. 15(b) shows the registration result achieved by our proposed registration framework. The registration RMS Error distance was 0.055 m, the total computational time was 22,481 s. The correspondence generation step cost 98% of the time, as discussed in Section 4.7. It can be concluded that our proposed method can successfully register point clouds of a large scene.

4.9. Limitations

As shown in Fig. 16, our proposed registration method failed to register the two point clouds of a symmetric building. That is mainly because only the local information is used in our covariance descriptor and the game theoretic method depends on the position

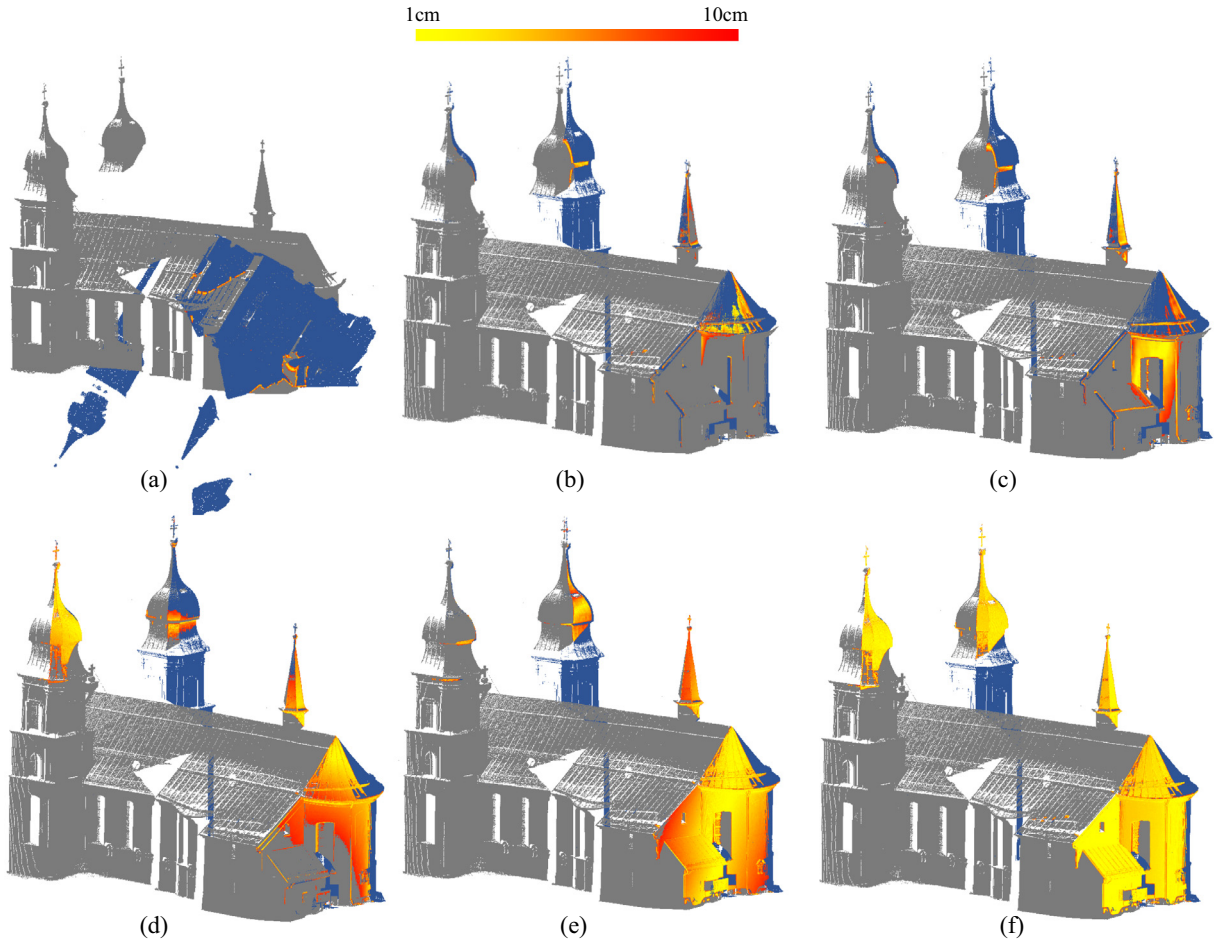


Fig. 14. Registration results on the point clouds of model "Church". (a) Achieved by FPFH + RANSAC. (b) Achieved by 4PCS. (c) Achieved by Super4PCS. (d) Achieved by 3DHarris + 4PCS. (e) Achieved by the proposed method without ICP refinement. (f) Achieved by the proposed method with ICP refinement.

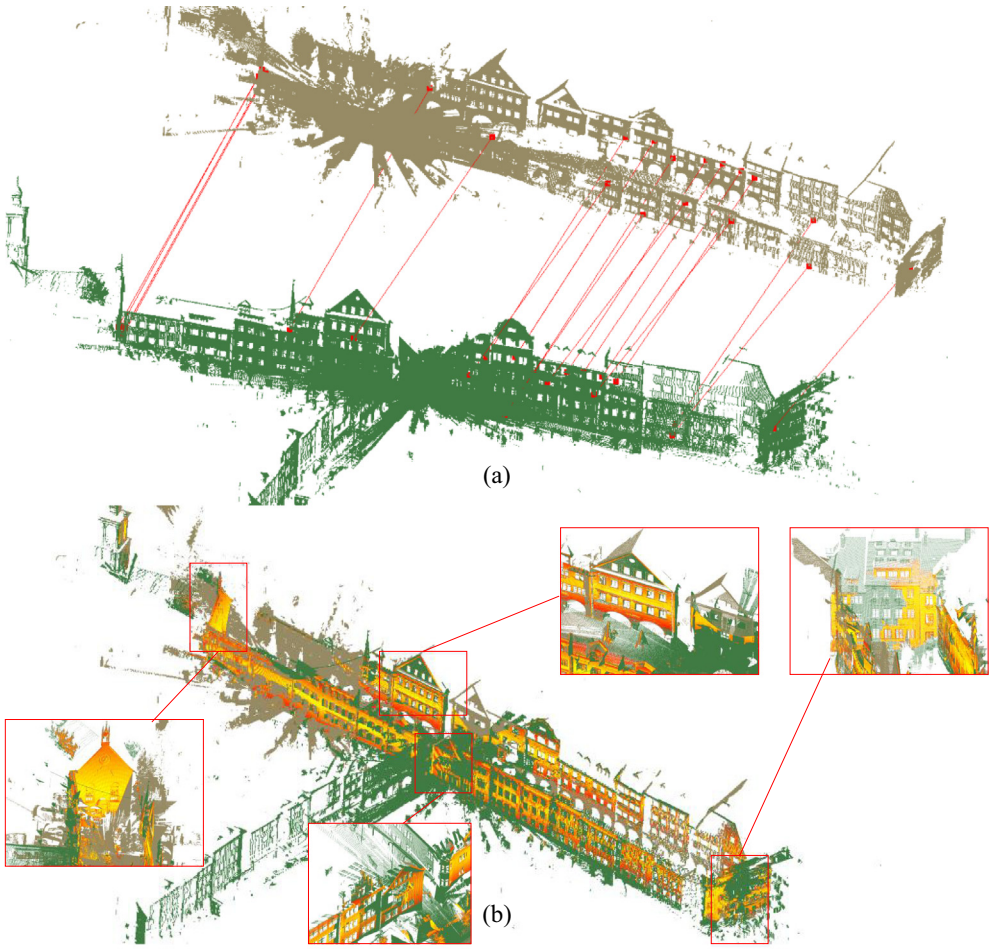


Fig. 15. Registration results of a real scene. (a) Correspondences obtained using ACOV descriptors and a game theoretic method. (b) Registration result.

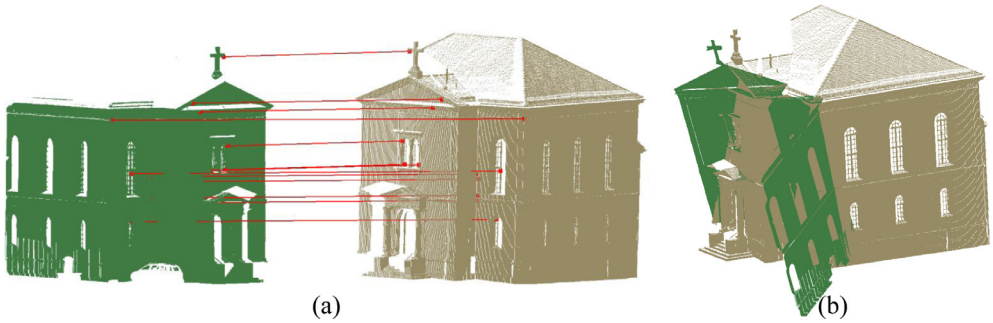


Fig. 16. An error case of the proposed registration framework. (a) Correspondences obtained using ACOV descriptors and a game theoretic method. (b) Registration result.

information of points. If more semantic or context information of the model is used, this problem is possible to be addressed.

5. Discussion

The proposed pairwise registration method benefits from two components: searching candidate correspondences using ACOV descriptor and rejecting false correspondences using a game theoretic method. For each component, we briefly discuss the main conclusions derived from the experiments.

Keypoint extraction plays an important role in candidate correspondence searching. The extracted keypoints should be repeatable. In this paper, keypoints are extracted through the distinctiveness of initial covariance descriptors. The neighborhood size for initial covariance descriptor determines the amount of surface that is encoded, it has a great impact on the keypoint extraction. Experiments on three models reveal that a neighborhood size of $25\bar{r}$ is appropriate to achieve high repeatability and efficiency (Fig. 5). For each detected keypoint, an ACOV descriptor is computed using an adaptive neighborhood size. Since the optimal neighborhood size may be different for these keypoints, the proposed method determines the neighborhood size by the highest salient variation in the feature space. The experiments demonstrate that the selected salient neighborhood size is repeatable and distinct (Fig. 9).

For the second component of false correspondences rejection, the proposed game theoretic method outperforms the RANSAC

method by a large margin in terms of robustness and efficiency (Table 2). Although a small value of parameter γ may lead to a poor registration result, our game theory method is insensitive to parameter γ (Fig. 10). Experiments on three models demonstrate that our registration method is robust to low overlap, noise and outliers, which are the major challenges posed by the TLS data (Figs. 11–13). For symmetric models, it is very challenging for our proposed method (Fig. 16). However, the surrounding environments of an object can be used to handle symmetric structures. The experiment on a real scene further demonstrates the feasibility and effectiveness of our registration method (Fig. 15).

In summary, the good performance of our pairwise registration framework is attributed to at least two factors: (1) The proposed ACOV descriptor is distinctive, repeatable and robust to noise and varying resolutions. (2) The game-theoretic technique can isolate mutually compatible correspondences from large outliers.

6. Conclusion

In this paper, we have presented a new method for pairwise registration of TLS point clouds. Our method was tested on three models and a large outdoor scene acquired by two different TLS systems. Our proposed method has several advantages over the state-of-the-art: (1) it performs point cloud registration accurately and efficiently without any initial position estimation; (2) it is robust to large noise, outliers and varying levels of overlap; (3) it can deal with very challenging scenes, such as pattern repetitions and local symmetries. Comparative experiments clearly demonstrate that our proposed method outperforms the other three methods by a large margin in terms of both registration error and computational time. It can be concluded that the proposed method can register TLS point clouds more accurately and efficiently.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants 41471379, 61471371, 61602499, in part by Fujian Collaborative Innovation Center for Big Data Applications in Governments, and in part by the National Postdoctoral Program for Innovative Talents under Grant BX201600172.

References

Aiger, D., Mitra, N.J., Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface registration. ACM Trans. Graphics 27 (3), 85.
Albarelli, A., Bulo, S.R., Torsello, A., Pelillo, M., 2009. Matching as a non-cooperative game. In: IEEE International Conference on Computer Vision. IEEE, Kyoto, Japan, pp. 1319–1326.
Albarelli, A., Rodola, E., Torsello, A., 2010. A game-theoretic approach to fine surface registration without initial motion estimation. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, San Francisco, CA, pp. 430–437.
Albarelli, A., Rodolà, E., Torsello, A., 2012. Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: a game-theoretic perspective. Int. J. Comput. Vision 97 (1), 36–53.
Albarelli, A., Rodolà, E., Torsello, A., 2015. Fast and accurate surface alignment through an isometry-enforcing game. Pattern Recogn. 48 (7), 2209–2226.
Bae, K.-H., Lichti, D.D., 2008. A method for automated registration of unorganised point clouds. ISPRS J. Photogramm. Remote Sens. 63 (1), 36–54.
Barnea, S., Filin, S., 2008. Keypoint based autonomous registration of terrestrial laser point-clouds. ISPRS J. Photogramm. Remote Sens. 63 (1), 19–35.
Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14 (2), 239–256.
Chen, H., Bhanu, B., 2007. 3D free-form object recognition in range images using local surface patches. Pattern Recogn. Lett. 28 (10), 1252–1262.
Chen, Y., Medioni, G., 1992. Object modelling by registration of multiple range images. Image Vis. Comput. 10 (3), 145–155.
Cirujeda, P., Cid, Y.D., Mateo, X., Binefa, X., 2015. A 3D scene registration method via covariance descriptors and an evolutionary stable strategy game theory solver. Int. J. Comput. Vision 115 (3), 306–329.
Demantké, J., Mallet, C., David, N., Vallet, B., 2011. Dimensionality based scale selection in 3D LiDAR point clouds. ISPRS-Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. 3812, 97–102.
Gressin, A., Mallet, C., Demantké, J., David, N., 2013. Towards 3D lidar point cloud registration improvement using optimal neighborhood knowledge. ISPRS J. Photogramm. Remote Sens. 79, 240–251.
Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J., 2014. 3D object recognition in cluttered scenes with local surface features: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 36 (11), 2270–2287.
Guo, Y., Sohel, F., Bennamoun, M., Lu, M., Wan, J., 2013. Rotational projection statistics for 3D local surface description and object recognition. Int. J. Comput. Vision 105 (1), 63–86.
Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. Semantic3D. net: A new large-scale point cloud classification benchmark. Available from: <1704.03847>.
Kaiser, M., Xu, X., Kwolek, B., Sural, S., Rigoll, G., 2013. Towards using covariance matrix pyramids as salient point descriptors in 3D point clouds. ISPRS J. Photogramm. Remote Sens. 120, 101–112.
Kelbe, D., van Aardt, J., Romanczyk, P., van Leeuwen, M., Cawse-Nicholson, K., 2016. Marker-free registration of forest terrestrial laser scanner data pairs with embedded confidence metrics. IEEE Trans. Geosci. Remote Sens. 54 (7), 4314–4330.
Lin, C.-H., Chen, J.-y., Su, P.-L., Chen, C.-H., 2014. Eigen-feature analysis of weighted covariance matrices for LiDAR point cloud classification. ISPRS J. Photogramm. Remote Sens. 94, 70–79.
Mellado, N., Aiger, D., Mitra, N.J., 2014. Super 4PCS Fast Global Pointcloud Registration via Smart Indexing. Wiley Online Library, pp. 205–215.
Pomerleau, F., Colas, F., Siegwart, R., 2015. A review of point cloud registration algorithms for mobile robotics. Found. Trends Robot. 4 (1), 1–104.
Rodola, E., Bronstein, A.M., Albarelli, A., Bergamasco, F., Torsello, A., 2012. A game-theoretic approach to deformable shape matching. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Providence, Rhode Island, pp. 182–189.
Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. In: IEEE International Conference on Robotics and Automation (ICRA). IEEE, Kobe, Japan, pp. 3212–3217.
Rusu, R.B., Cousins, S., 2011. 3d is here: Point Cloud Library (PCL). In: IEEE International Conference on Robotics and Automation (ICRA). IEEE, Karlsruhe, Germany, pp. 1–4.
Salvi, J., Matabosch, C., Fofi, D., Forest, J., 2007. A review of recent range image registration methods with accuracy evaluation. Image Vis. Comput. 25 (5), 578–596.
Tam, G.K., Cheng, Z.-Q., Lai, Y.-K., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X.-F., Rosin, P.L., 2013. Registration of 3D point clouds and meshes: a survey from rigid to nonrigid. IEEE Trans. Visualiz. Comput. Graphics 19 (7), 1199–1217.
Theiler, P.W., Wegner, J.D., Schindler, K., 2014. Keypoint-based 4-points congruent sets–automated marker-less registration of laser scans. ISPRS J. Photogramm. Remote Sens. 96, 149–163.
Theiler, P.W., Wegner, J.D., Schindler, K., 2015. Globally consistent registration of terrestrial laser scans via graph optimization. ISPRS J. Photogramm. Remote Sens. 109, 126–138.
Tombari, F., Salti, S., Di Stefano, L., 2010. Unique signatures of histograms for local surface description. In: European Conference on Computer Vision. Springer, Crete, Greece, pp. 356–369.
Unnikrishnan, R., Hebert, M., 2008. Multi-scale interest regions from unorganized point clouds. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, Anchorage, Alaska, USA, pp. 1–8.
Weinmann, M., 2016. Reconstruction and Analysis of 3D Scenes: From Irregularly Distributed 3D Points to Object Classes. Springer.
Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS J. Photogramm. Remote Sens. 105, 286–304.
Weinmann, M., Weinmann, M., Hinz, S., Jutzi, B., 2011. Fast and automatic image-based registration of TLS data. ISPRS J. Photogramm. Remote Sens. 66 (6), S62–S70.
Yang, B., Dong, Z., Liang, F., Liu, Y., 2016. Automatic registration of large-scale urban scene point clouds based on semantic feature points. ISPRS J. Photogramm. Remote Sens. 113, 43–58.
Yang, B., Zang, Y., 2014. Automated registration of dense terrestrial laser-scanning point clouds using curves. ISPRS J. Photogramm. Remote Sens. 95, 109–121.
Yang, B., Zang, Y., Dong, Z., Huang, R., 2015. An automated method to register airborne and terrestrial laser scanning point clouds. ISPRS J. Photogramm. Remote Sens. 109, 62–76.
Yang, J., Li, H., Jia, Y., 2013. Go-icp: Solving 3D registration efficiently and globally optimally. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Portland, Oregon, pp. 1457–1464.
Zhong, Y., 2009. Intrinsic shape signatures: A shape descriptor for 3D object recognition. In: IEEE 12th International Conference on Computer Vision Workshops. IEEE, Kyoto, Japan, pp. 689–696.