SIViP (2016) 10:835–842
DOI 10.1007/s11760-015-0823-z

ORIGINAL PAPER

A recursive robust filtering approach for 3D registration

Abdenour Amamra[1] · **Nabil Aouf**[1] · **Dowling Stuart**[1] · **Mark Richardson**[1]

Received: 16 February 2015 / Revised: 10 September 2015 / Accepted: 19 September 2015 / Published online: 5 October 2015

Abstract This work presents a new recursive robust filtering approach for feature-based 3D registration. Unlike the common state-of-the-art alignment algorithms, the proposed method has four advantages that have not yet occurred altogether in any previous solution. For instance, it is able to deal with inherent noise contaminating sensory data; it is robust to uncertainties caused by noisy feature localisation; it also combines the advantages of both L_∞ and L_2 norms for a higher performance and a more prospective prevention of local minima. The result is an accurate and stable rigid body transformation. The latter enables a thorough control over the convergence regarding the alignment as well as a correct assessment of the quality of registration. The mathematical rationale behind the proposed approach is explained, and the results are validated on physical and synthetic data.

Keywords Time-varying 3D registration · Recursive least squares · Kalman filter · Robust $\boldsymbol{H}_\infty$ filter

✉ Abdenour Amamra
amamra.abdenour@gmail.com

Nabil Aouf
n.aouf@cranfield.ac.uk

Dowling Stuart
s.dowling@cranfield.ac.uk

Mark Richardson
m.a.richardson@cranfield.ac.uk

[1] Centre for Electronic Warfare, Cranfield University, Cranfield, United Kingdom

1 Introduction

The widespread abundance of affordable 3D sensing devices has encouraged many enthusiasts to contribute new solutions for 3D reconstruction [1]. The latter require data alignment tools that enable the recovery of the 6DOF regarding viewpoints where the different scans had been captured. Theoretically, each viewpoint has a different coordinate system. Knowledge of the transformation that maps a given 3D point from one frame to another, therefore, becomes necessary.

In practice, the alignment requires some keypoints from a *Source* and a *Target* point cloud. Hence, alignment problem amounts to the determination of the mapping between the source and the target frames. To this end, we assume the keypoints being available and we focus on 3D registration. Generally, the determination of the best transformation is based on L_2 norm minimisation. However, L_2 optimisers assume a prior availability of the entire datasets before processing takes place. From a practical point of view, such an assumption is too optimistic due to sizeable noisy data streamed at relatively high frame rates that one encounters in practice. For this reason, our novel 3D registration solution delivers the 6DOF pose between viewpoints recursively and is capable of handling 3D points' noise and uncertainty for a more efficient estimation.

The remainder of this paper is organised as follows: In the first section, the related works about 3D registration are discussed and different alignment solutions that had been proposed so far are analysed. In the following section, 3D registration problem is formulated in a Least Squares (LS) form. In the next section the link between 3D registration and RLS is settled and fitted into Kalman filter's (KF) equations [2]. The parametric uncertainty of the 3D feature points is afterwards determined to be later used in the Robust H_∞

(RF) modelling for sparse alignment. Our contribution is validated on both synthetic and real datasets. Lastly, the paper is concluded and potential future works are recommended.

2 Related works

Since its invention by Besl and McKay [3], the Iterative Closest Point algorithm (ICP) has been considered as a reference in point cloud alignment literature. However, a good initial guess or some feature correspondences are necessary to avoid local minima. Newer variants of that algorithm have been proposed to deal with its limitations such as EM-ICP [4] and Softassign [5]. Unlike the original implementation that assigns to every point in the source its closest correspondent in the target, subsequent variants allow each point to be checked against the entire target dataset. To this end, weighting coefficients are associated with the elements to discard the describe their quality [5]. Other variants inspired by the original algorithm (ICP) were further proposed such as nonlinear ICP [6], generalised ICP [7] as well as non-rigid ICP [8]. Larusso et al. [9] showed that all closed-form solutions are computationally similar. However, performance can significantly differ from one solution to another. Thus, no single algorithm is exclusively optimal for all scenarios. Umeyama [10] states in his work that Horn and Arun's algorithms fail when the datasets become highly corrupted with noise. He further proposed an alternative solution that utilises Lagrange Multipliers [11].

A solution for the recursive estimation of rigid body transformations with the Extended Kalman Filter (EKF) was first proposed by Pennec and Thirion [12]. Ma and Ellis [13] followed the same strategy in order to align datasets contaminated with isotropic Gaussian noise using the Unscented Particle Filter (UPF) [14]. This algorithm can accurately estimate the parameters for very small datasets (less than one hundred elements). An Unscented Kalman Filter (UKF) algorithm was also adapted by Julier and Uhlmann [15] to align two datasets following a sequential estimation. All these recursive algorithms minimise $\boldsymbol{L_2}$ norm but consider the parameters being accurately determined beforehand. Nevertheless, it is impossible to assert the certainty of parameters in real scenarios. In our solution, however, we consider them (parameters) being uncertain, and we confine estimation error to a small range by optimising $\boldsymbol{L_\infty}$ norm instead of $\boldsymbol{L_2}$.

Micusik and Pflugfelder [16] used a second-order cone programming (SOCP) to minimise the $\boldsymbol{L_\infty}$ norm for non-overlapping cameras. They have shown a good performance with a fairly small error magnitude. Lee et al. [17] further claimed that by using $\boldsymbol{L_\infty}$ a number of computer vision problems such as homography estimation can be formulated and solved using Bisection method.

In the light of this background, our work takes advantage of the mature recursive estimation framework in order to compute a robust and optimal solution for 3D registration problem by means of $\boldsymbol{L_\infty}$ norm minimisation.

3 Problem statement

Given two sets of source and target 3D point clouds $\boldsymbol{Q} = \{\boldsymbol{q_1}, \ldots, \boldsymbol{q_n}\}, \boldsymbol{P} = \{\boldsymbol{p_1}, \ldots, \boldsymbol{p_n}\}$ respectively. Each of the elements $\boldsymbol{p_i}, \boldsymbol{q_i}$ within the sets of points has three components $\boldsymbol{p_i} = (\boldsymbol{x_p}, \boldsymbol{y_p}, \boldsymbol{z_p})_i$ and $\boldsymbol{q_i} = (\boldsymbol{x_q}, \boldsymbol{y_q}, \boldsymbol{z_q})_i$. The $\boldsymbol{k}$th point q_k in the source point cloud has been matched a priori with the $\boldsymbol{k}$th point in the target point cloud $\boldsymbol{p_k}$. The purpose of 3D registration is to find a rigid body transformation ($\boldsymbol{R}$: rotation, $\boldsymbol{t}$: translation) that maps the source $\boldsymbol{Q}$ onto the target $\boldsymbol{P}$. The determination of such a mapping can be modelled as an optimisation problem [18]. Nevertheless, due to noisy outputs streamed by the sensor, an exact solution is very unlikely to determine. Thus, a realistic model must take into account alignment error $\boldsymbol{e_i}$ as follows:

$$p_i = Rq_i + t + e_i \tag{1}$$

The rigid body transformation $[\boldsymbol{R}, \boldsymbol{t}]$ is optimal when the sum of the squares of errors $(\boldsymbol{e_i})$ becomes minimal:

$$e^2 = \underset{R,t}{\operatorname{argmin}} \sum_{i=1}^{n} \| p_i - (Rq_i + t)^2 \| \tag{2}$$

where

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}; \quad t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{3}$$

It is possible to simplify the problem of Eq. (2) by decoupling the translation vector $\boldsymbol{t}$ and eliminating scale difference as follows:

$$\hat{t} = \bar{p} - \hat{R}\bar{q}; \quad \hat{s} = \sum_{i=1}^{n} \left(\frac{\bar{p}_i \hat{R} \bar{q}_i}{\| \bar{q}_i \|^2} \right) \tag{4}$$

$$\tilde{t} = \bar{p} - \bar{q}; \quad \tilde{s} = \sum_{i=1}^{n} \left(\frac{\bar{p}_i \bar{q}_i}{\| \bar{q}_i \|^2} \right) \tag{5}$$

As claimed by Horn et al. [19], $\bar{\boldsymbol{q}}$ and $\bar{\boldsymbol{p}}$ are the centroids respective to the source and the target point clouds; $\hat{\boldsymbol{t}}, \hat{\boldsymbol{s}}$ are the optimal translation and scale between the two dataset. Whereas $\tilde{\boldsymbol{t}}, \tilde{\boldsymbol{s}}$ are their respective initial guesses when the initial rotation is assumed to be $\boldsymbol{R_0} = \boldsymbol{I_3}$. As a result of this simplification, the problem of pose estimation in Eq. (2) is now reduced to:

$$e^2 = \underset{R}{\operatorname{argmin}} \sum_{i=1}^{n} \| \bar{p}_i - R\bar{q}_i^2 \| \tag{6}$$

Once the optimal rotation $\hat{\boldsymbol{R}}$ computed, $\hat{\boldsymbol{t}}$ and $\hat{\boldsymbol{s}}$ can be deduced using Eq. (4). On the other hand, the optimal rotation $\hat{R}$ can be obtained by minimising $\sum_{i=1}^{n} \| \bar{p}_i - R\bar{q}_i \|^2$ using a LS optimiser. The resulting estimation is sufficient for most applications as long as robustness is not a determining factor. However, if the inputs become significantly contaminated with noise, the result becomes unstable (i.e., very sensitive to perturbations in the data) and more likely to drift away from the optimal solution.

4 3D registration with RLS

Despite the performance of time-varying filters, 3D registration has profited very poorly from their assets even after closed-form methods were proven weak in various practical situations. Moreover, the authors of a number of recent image registration surveys did not even allude to the possibility of solving 3D alignment with recursive filtering tools [20]. The power of the recursive solutions can be appreciated due to what has been claimed earlier and to the possibility of cooperation between different registration instances working together. The latter can share their most updated estimates instantaneously. As a result, they can benefit from each other's contributions, which in turn reduces the probability of falling into a local minimum.

4.1 Recursive modelling of 3D registration

In order to express the cost function of Eq. (6) in a recursive fashion, the original problem should be rewritten as shown in Eqs. (7)–(10). Such a transformation allows us to fit 3D registration problem in a recursive least squares framework.

$$p = Rq + e \tag{7}$$

$$\begin{cases} p_x = r_{11}q_x + r_{12}q_y + r_{13}q_z + e_x \\ p_y = r_{21}q_x + r_{22}q_y + r_{23}q_z + e_y \\ p_z = r_{31}q_x + r_{32}q_y + r_{33}q_z + e_z \end{cases} \tag{8}$$

By analogy, the state variable $\boldsymbol{x_k}$ now represents the rotation matrix $\boldsymbol{R}$ of Eq. (7). The optimiser uses pairs of corresponding points in order to refine the entries of the state vector now containing the entries of rotation matrix $\mathcal{R}_9$. For instance, at every time-step $\boldsymbol{k}$ we have:

$$v = \begin{bmatrix} q_x & q_y & q_z \end{bmatrix}$$

$$\begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} v & & \\ & v & \\ & & v \end{bmatrix} \begin{bmatrix} r_{11} \\ r_{12} \\ r_{13} \\ r_{21} \\ r_{22} \\ r_{23} \\ r_{31} \\ r_{32} \\ r_{33} \end{bmatrix} + \begin{bmatrix} e_x \\ e_y \\ e_z \end{bmatrix} \tag{9}$$

$$p = H_R \mathcal{R}_9 + e \tag{10}$$

$\boldsymbol{x_k} = \begin{bmatrix} \boldsymbol{r_{11}} & \boldsymbol{r_{12}} & \boldsymbol{r_{13}} & \boldsymbol{r_{21}} & \boldsymbol{r_{22}} & \boldsymbol{r_{23}} & \boldsymbol{r_{31}} & \boldsymbol{r_{32}} & \boldsymbol{r_{33}} \end{bmatrix}^{\boldsymbol{T}} \in \Re^9$; $\boldsymbol{A_k} = \boldsymbol{I_9}$; $\boldsymbol{B_k} = \boldsymbol{0_9}$ as no control variable is required. $\boldsymbol{w_k} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{0}, \boldsymbol{Q_k})$ is a random variable representing process noise for which $\boldsymbol{Q_k} = \boldsymbol{\sigma_k^2 I_9}$; $\boldsymbol{\sigma_k} > \boldsymbol{0}$ should be small because the process is accurately determined. $\boldsymbol{z_k} \in \Re^3$ is the actual noisy measurement vector whose elements are the coordinates of the target feature point. $\boldsymbol{y_k} \in \Re^3$ is the predicted observation vector that contains the 3D position of the target feature point. $\boldsymbol{v_k}$ is a random variable for which $\boldsymbol{R_k} = \begin{bmatrix} \boldsymbol{\sigma_x} & \boldsymbol{\sigma_y} & \boldsymbol{\sigma_z} \end{bmatrix} \boldsymbol{I_3}$, it represents noise process contaminating target feature point localisation. The complete scheme of KF-based registration is explained in Algorithm 1. The latter works as follows: (1) Initialise the state vector (rotation matrix) with the entries of $\boldsymbol{I_9}$. If available, an initial guess would be preferable. (2) Iterate over feature points; acquire a new target feature $\boldsymbol{z_k}$ and build $\boldsymbol{H_k}$. (3) KF prediction. (4) KF correction where the estimate $\boldsymbol{x_k}$ and the covariance of error in estimation $\boldsymbol{P_k}$ are corrected with $\boldsymbol{K_k}$.

Algorithm 1 KF-based registration

Source and target point clouds
$\boldsymbol{P}, \boldsymbol{Q}$: *3D feature points;*
$[\boldsymbol{p_i}, \boldsymbol{q_i}] = \emptyset$: *Correspondences list;*
$[\boldsymbol{p_i}, \boldsymbol{q_i}]$ = *FindCorrespondences(P,Q);*
$\widehat{\boldsymbol{x}}_{\boldsymbol{0}} = [1.0\ 0.0\ 0.0\ 0.0\ 1.0\ 0.0\ 0.0\ 0.0\ 1.0]^T$
$\boldsymbol{Q} = \sigma I_9$
$\widehat{\boldsymbol{P}}_{\boldsymbol{0}} = Q$
$\boldsymbol{R_0} = [\sigma_x\ \sigma_y\ \sigma_z] I_3$
for *each pair of correspondences* $(\boldsymbol{k} = \boldsymbol{1}, \boldsymbol{n})$
$z_k = [\, p(k).x\ \ p(k).y\ \ p(k).z]$
$H_k = H_R(p(k))$
Prediction

$$x_k = \hat{x}_{k-1} \tag{11}$$
$$y_k = H_k x_k \tag{12}$$
$$P_k = \hat{P}_{k-1} + Q \tag{13}$$

correction

$$K_k = P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} \tag{14}$$
$$\hat{x}_k = x_k + K_k(z_k - y_k) \tag{15}$$
$$\hat{P}_k = (I - K_k H_k) P_k \tag{16}$$

end

The computational complexity of KF registration is proportional to $\boldsymbol{O(n \times 9^3)}$ in the worst case, where $\boldsymbol{n}$ is the number of keypoints used to compute the optimal registration and **9** is the size of the state vector. On the other hand, the best complexity regarding alternative registration algorithms such as ICP, EMICP and WICP is proportional to $\boldsymbol{O(n^2 \times 3^{2.37})}$. KF 3D registration can be easily expanded to include the three components of translation vector in $\boldsymbol{H_k}$.

5 Robust H_∞ registration

5.1 3D points uncertainty

In order to handle instability in parameters estimation, the uncertainties should be confined into a small range. To this end, the behaviour of the noisy inputs must be thoroughly studied. Uncertainties are modelled empirically by looking at how 3D points are distributed, and how do 3D sensors sense the real world.

5.2 z-Resolution of RGBD cameras

The authors have already shown in a previous research [21] that the points within a 3D image lie on parallel clusters that were named "**Z-Levels**". Such a structure allows us to quantify correctly the amount of uncertainty in every feature point.

5.3 Depth noise statistics

RGBD sensors' measurement-noise has a Gaussian distribution with varying standard deviations. These standard deviations rely on the range between the sensor and the scene. The standard deviation $\boldsymbol{\sigma_{z_k}}$ of a given Z-level $\boldsymbol{z_k}$ is defined by the length of the interval where $\boldsymbol{z_k}$ is expected to vary as shown below:

$$\sigma_{z_k} = (Z_{k+i} - Z_{k-i})/2 \tag{17}$$

Here, $\boldsymbol{\sigma_{z_k}}$ represents the average distance separating the two Z-levels $\boldsymbol{Z_{k+i}}$ and $\boldsymbol{Z_{k-i}}$ and the central one $\boldsymbol{z_k}$. Empirically, the best estimation of the standard deviation regarding noise affecting the 3D points lying on $\boldsymbol{z_k}$ is obtained when $\boldsymbol{i = 3}$. That is, the true depth $\hat{\boldsymbol{z}}_{\boldsymbol{k}}$ taken by a given Z-level is expected to be equal to $\boldsymbol{z_k \pm ((Z_{k+3} - Z_{k-3})/2)}$. The standard deviations concerning the remaining two coordinates $(\boldsymbol{x_k}, \boldsymbol{y_k})$ are deduced from the intrinsic parameters of the camera $(\boldsymbol{f_x}, \boldsymbol{f_y}, \boldsymbol{c_x}, \boldsymbol{c_y})$ and $\boldsymbol{\sigma_{z_k}}$ as follows:

$$\begin{cases} u_i = (f_x/z_i)x_i + c_x \\ v_i = (f_y/z_i)y_i + c_y \end{cases} \tag{18}$$

$$\begin{cases} x_i = (z_i/f_x)\,(u_i - c_x) \\ y_i = (z_i/f_y)\,(v_i - c_y) \end{cases} \tag{19}$$

$$\begin{cases} \sigma_{z_k} = 0.5(Z_{k+i} - Z_{k-i}) \\ \sigma_{x_k} = (\sigma_{z_k}/f_x)\,(u_k - c_x) \\ \sigma_{y_k} = (\sigma_{z_k}/f_y)\,(v_k - c_y) \end{cases} \tag{20}$$

Every point is, therefore, affected by certain amount of noise characterised by the standard deviations $\boldsymbol{\sigma_{x_k}}, \boldsymbol{\sigma_{y_k}}, \boldsymbol{\sigma_{z_k}}$ towards the directions of the axes $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$, respectively. Hence, the covariance matrix attributed to each point $\boldsymbol{p(x, y, z)}$ is described as:

$$C\,(x, y, z) = \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y & \sigma_x\sigma_z \\ \sigma_y\sigma_x & \sigma_y^2 & \sigma_y\sigma_z \\ \sigma_z\sigma_x & \sigma_z\sigma_y & \sigma_z^2 \end{bmatrix} \tag{21}$$

$\boldsymbol{C}$ represents the spread of uncertainty around the point $\boldsymbol{p(x, y, z)}$. As can be seen in Fig. 1 a, the projection of covariance ellipsoids of a given 3D point on the planes $\boldsymbol{zx, zy, yx}$ yields three ellipses. The more accurately a feature point is captured, the smaller the norm of its covariance matrix (blue point in Fig. 1a). On the other hand, the less accurate the capture of a given feature is, the larger the norm of its covariance matrix (red point in Fig. 1a).

Kanazawa and Kanatani [22] claimed that the incorporation of feature uncertainty does not contribute any further improvements to the estimation. On the other hand, Brooks et al. [23] as well as us in a previous work [24], both noticed a reduced error in estimation after considering uncertainty. Based on the conducted experiments with registration algorithms and the fact that Weighted-ICP (WICP takes into account data uncertainty) outperforms ICP, as will be shown in the results, it is obvious that the incorporation of feature-location uncertainty improves pose estimation remarkably.

6 Robust H_∞ (RF) filter for 3D registration

In this section, we propose a time-varying registration algorithm that incorporates modelling and measurement uncertainties as follows:

$$x_k = (A_k + \Delta A_k)\,\hat{x}_{k-1} + B_k u_k + w_k \tag{22}$$

$$y_k = (H_k + \Delta H_k)x_k + v_k \tag{23}$$

$\boldsymbol{\Delta H_k}$ represents the uncertainty in observation model, whereas $\boldsymbol{\Delta A_k}$ is the uncertainty in process model. In our case, the two matrices take the values:

$$\Delta A_k = \sigma_A I_9$$

$$\sigma_A = \left[\sigma_{r_{11}}\sigma_{r_{12}}\sigma_{r_{13}}\sigma_{r_{21}}\sigma_{r_{22}}\sigma_{r_{23}}\sigma_{r_{31}}\sigma_{r_{32}}\sigma_{r_{33}}\right]$$

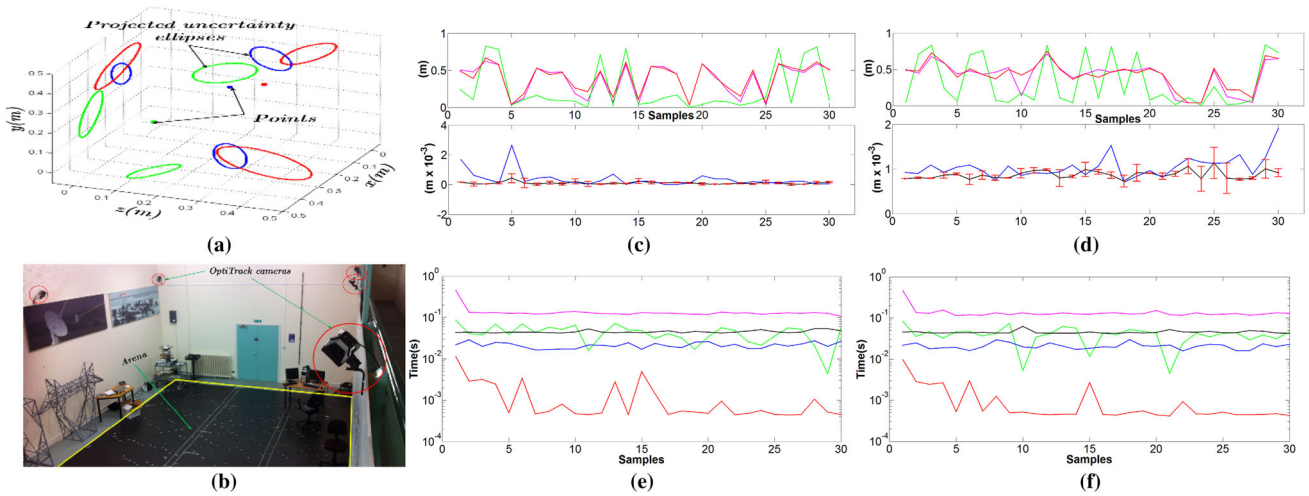

Fig. 1 **a** 3D points uncertainty ellipses, **b** ground truth (OptiTrack) and real data acquisition with Kinect in an indoor scene, **c**, **d** 3D registration RMSE (m) of the new and the old Kinect, respectively; **e**, **f** time elapsed during registration for the new and the old Kinect, respectively. EMICP (*pink*), WICP (*green*), Horn (*red*), RF (*black*), KF (*blue*) (colour figure online)

$$V_k = \begin{bmatrix} q_x + \sigma_x & q_y + \sigma_y & q_z + \sigma_z \end{bmatrix} \tag{24}$$

$$\Delta H_k = \begin{bmatrix} V_k & & \\ & V_k & \\ & & V_k \end{bmatrix} \tag{25}$$

If these matrices cannot be determined, RF would still be able to control the instability disturbing its parameters [25] by assuming it being of the form:

$$\begin{bmatrix} \Delta A_k \\ \Delta H_k \end{bmatrix} = \begin{bmatrix} M_{1k} \\ M_{2k} \end{bmatrix} \Gamma_k N_k \tag{26}$$

$\boldsymbol{M_{1k}}, \boldsymbol{M_{2k}}$ and $\boldsymbol{N_k}$ are known matrices, $\boldsymbol{\Gamma_k}$ is unknown but it should satisfy the bound:

$$\Gamma_k^T \Gamma_k \leq I \tag{27}$$

Our purpose is to design a state estimator of the form:

$$x_{k+1} = \tilde{A}_k x_k + \tilde{K}_k t_k \tag{28}$$

The latter should be stable (the eigenvalues of $\tilde{\boldsymbol{A}}_{\boldsymbol{k}}$ must be less than one in magnitude). The determination of the parameters of the filter can be done through the procedure described in our previous work [24].

The adaptation of RF is proven to be flexible and capable of delivering accurate state estimations, however uncertain system's parameters are. Estimation error compared to the ground truth measurements will show the effectiveness of RF 3D registration against alternative non-robust methods such as KF and the more established algorithms available in the literature. In real scenarios, the exact model is very unlikely to determine [26]. Yet the non-robust tools do not consider uncertainties in their parameters. Hence, if by chance the parameters are accurate, these tools perform as well as RF. On the other hand, when the system is not precisely characterised, they become significantly unstable. For instance, RF registration combines the robustness of $\tilde{\boldsymbol{H}}_{\infty}$ (it is less affected by the accuracy of system's parameters) and the optimality of KF on linear systems to produce an accurate and stable estimate. Such a quality guaranties a high precision of estimation and more stability towards inputs' perturbations.

7 Results and discussion

In this section, the results regarding KF and RF registration are validated with tests on real and synthetic 3D data. Our test benchmark includes: WICP [27]; Expectation Maximisation ICP algorithm (EMICP) [28] and Horn's closed-form solution based on quaternions (HORN) [29].

Here, accuracy is measured by the distance separating the target and the source point clouds after the registration. In order to fairly assess every algorithm, processing time elapsed to find the best pose is also recorded. Throughout experiments, it is noticeable that the plotted metrics (processing time and RMSE $= \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\hat{\boldsymbol{R}}\boldsymbol{Q_i} + \hat{\boldsymbol{t}} - \boldsymbol{P_i}\|^2}$) are not homogeneous. For this reason, a logarithmic scale was used to cope with the difference of scale within the same plot.

The number of keypoints extracted from every point cloud is about 400 points. In practice, an average-sized point cloud in a single frame contains up to 400 useful key points. Computation time has been calculated for the five algorithms running on an **i7-2670QM** working at **2.2 GHz**, with **12.0 GB** of memory. A sample is a set of 400 pairs of correspond-

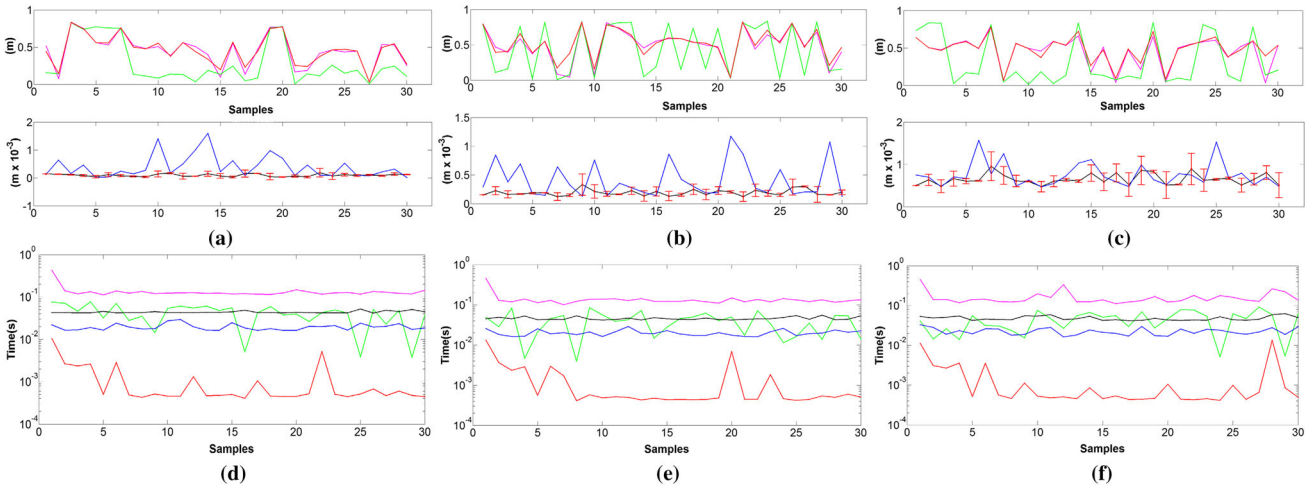

Fig. 2 **a–c** 3D registration RMSE of the small, average and large noise, respectively; **d–f** time elapsed during registration for small, average and large noise, respectively. EMICP (*pink*), WICP (*green*), Horn (*red*), RF (*black*), KF (*blue*) (colour figure online)

ing $<source, target>$ keypoints. 30 samples were tested in each of the following five scenario (two with real and three with synthetic data).

7.1 Real data

In this experiment, image data are delivered by two versions of Kinect[1] sensor (Kinect 1 is based on structured light principle; whereas, Kinect 2 is a time-of-flight camera). In addition, SIFT3D extractor and CSHOT [21] descriptor were used to obtain feature points from the real data.

In order to collect real 3D point clouds, the camera was carried and moved around in an infinity-shaped (∞) trajectory within the arena of our autonomous navigation laboratory. Simultaneously, a high-quality tracking system (OptiTrack[2]) was used as a ground truth reference, Fig. 1b. One hundred and twenty different pairs of overlapping point clouds were captured by each of the two Kinects. RGBD image data acquisition runs simultaneously as the robot moves around. At each time-step, we acquire a single pair of colour and depth images (both constitute a single point cloud) for the indoor scene. Hence, a total of 120 pairs of point clouds are aligned in a pairwise manner between $(\boldsymbol{C_i}, \boldsymbol{C_{i+1}})$. The last sample $\boldsymbol{C_{120}}$ is registered against both $\boldsymbol{C_{119}}$ and $\boldsymbol{C_1}$ to test the loop closure.

Scenario 1: New Kinect
RMSE The average RMSE for the five algorithms (see Fig. 1c) was as follows: 0.27 m for EMICP (pink), 0.13 m for WICP (green), 0.28 m for Horn (red), 0.15 mm for RF (black) and 0.7 mm for KF (blue).

[1] http://www.microsoft.com/en-us/kinectforwindows/. 2015.

[2] http://www.naturalpoint.com/optitrack/. 2015.

Scenario 2: Old Kinect
RMSE was 0.28 m for EMICP, 0.22 m for WICP, 0.3 m for Horn, 0.95 mm for RF and 1.13 mm for KF (see Fig. 1d).
Average processing time for both scenarios was 114.3 ms for EMICP, 26.7 ms for WICP, 1.05 ms for Horn, 23.1 ms for RF and 11.64 ms for KF (see Fig. 1e, f).

7.2 Synthetic data

In this experiment, we consider only artificial 3D keypoints, where, $\boldsymbol{Q_i}$ (source keypoints) as well as a random 3D transformation $[\boldsymbol{R_i}, \boldsymbol{t_i}]$ had been generated randomly. The target 3D keypoints are built using the equation, $\boldsymbol{P_i} = \boldsymbol{R_i}\,\boldsymbol{Q_i} + \boldsymbol{t_i}$. To realistically simulate physical data, a normally distributed anisotropic white noise was added to the clean datasets. The latter had different magnitudes $\boldsymbol{\sigma_i}$: *large* (20 mm $\leq \sigma_i \leq$ 80 mm), *average* (10 mm $\leq \sigma_i \leq$ 20 mm) and *small* (0.1 mm $\leq \sigma_i \leq$ 10 mm). For each, is generated 1000 point clouds, results were as follows:

Scenario 1: Small noise magnitude
RMSE was 0.42 m for EMICP, 0.18 m for WICP, 0.46 m for Horn, 0.18 mm for RF and 0.44 mm for KF (see Fig. 2a).

Scenario 2: Average noise magnitude
RMSE was 0.54 m for EMICP, 0.48 m for WICP, 0.56 m for Horn, 0.22 mm for RF and finally, 0.43 mm for KF (see Fig. 2b).

Scenario 3: Large noise magnitude
RMSE was 0.49 m for EMICP, 0.35 m for WICP, 0.51 m for Horn, 0.63 mm for RF and 0.89 mm for KF (see Fig. 2c).
Average processing time for all three scenarios was 115.1 ms for EMICP, 27.03 ms for WICP, 1.08 ms for Horn, 22.8 ms for RF and 10.43 ms for KF (see Fig. 2d–f).

As illustrated in Table 1, one can obviously notice how significantly poorly EMICP and Horn perform. This draw-

Table 1 RMSE (mm) for the whole sets of samples: 1000 for each simulation scenario and 120 for every version of Kinects

Noise	EMICP	WICP	Horn	RF	KF
Kinect					
New	274	152	246	0.72	1.62
Old	310	162	302	1.03	2.03
Small	298	193	285	0.55	1.52
Average	323	235	315	0.91	1.78
Large	332	260	343	0.96	2.10

back often occurs when the shapes present some symmetry. On the other hand, WICP is better endowed to cope with such drawbacks since it leverages knowledge about the quality of features, which helps it in discarding noisy elements. More importantly, KF and RF are both comparably superior in term of accuracy, but RF is more precise due to the control of uncertainty in parameters.

8 Conclusion and future works

A novel approach for robust 3D point cloud registration was presented. This contribution is based on a recursive optimal state estimation. After establishing the link between WLS and its original counterpart (LS), 3D point cloud registration problem was fitted to KF scheme. However, since KF parameters for 3D registration (state and projection matrices) are built from noisy data, a non-negligible estimation instability was noticed. Consequently, we modelled the uncertainty and overcame it with an RF-based solution.

The accuracy of the proposed solution was tested on many synthetic as well as real 3D samples delivered by Kinect. Precision, on the other hand, can be seen on the relatively small difference in accuracy among comparably noisy samples (red error bars in Figs. 1c, d and 2a–c on the black line).

The proposed solution requires some feature points to be extracted from the source and the target point clouds before the alignment is carried out. The number of keypoints is relatively small compared to the size of point clouds. In addition, our solution can be extended to any dimension for data that can be point clouds, meshes as well as surfaces, given that some distinctive features are available.

As a future work, we intend to investigate alternative applications of recursive filtering algorithms in the field of computer vision. It would be also interesting to implement RF registration in the graphic processor to reach higher frame rates. In addition, in a multiview registration scenario (many sensors streaming images concurrently), data fusion algorithms open a new perspective for the users to reconstruct 3D scenes and to track moving objects cooperatively. This new horizon is convenient for the technologies of virtual and augmented reality.

References

1. Cho, S.-Y.: 3D ear shape reconstruction and recognition for biometric applications. Signal Image Video Process. **7**(4), 609–618 (2013)
2. Ali, A., Jalil, A., Ahmed, J., Iftikhar, M.A., Hussain, M.: Correlation, Kalman filter and adaptive fast mean shift based heuristic approach for robust visual tracking. Signal Image Video Process. **9**, 1–19 (2014)
3. Besl, P.J., McKay, H.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
4. Tamaki, T., Abe, M., Raytchev, B., Kaneda, K.: Softassign and EM-ICP on GPU. In: 2010 First International Conference on Networking and Computing, pp. 179–183 (2010)
5. Gold, S., Rangarajan, A., Lu, C.-P., Pappu, S., Mjolsness, E.: New algorithms for 2D and 3D point matching. Pattern Recognit. **31**(8), 1019–1031 (1998)
6. Fantoni, S., Castellani, U., Fusiello, A.: Accurate and automatic alignment of range surfaces. In: 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, pp. 73–80 (2012)
7. Servos, J., Waslander, S.L.: Multi channel generalized-ICP. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 3644–3649 (2014)
8. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging **18**(8), 712–721 (1999)
9. Larusso, A., Eggert, D., Fisher, R.: A Comparison of four algorithms for estimating 3-D rigid transformations. In: Procedings of the British Machine Vision Conference 1995, pp. 24.1–24.10 (1995)
10. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. **13**(4), 376–380 (1991)
11. Nielsen, L.: Least-squares estimation using Lagrange multipliers. Metrologia **37**(2), 183–183 (2000)
12. Pennec, X., Thirion, J.-P.: A framework for uncertainty and validation of 3-D registration methods based on points and frames. Int. J. Comput. Vis. **25**(3), 203–229 (1997)
13. Ma, B., Ellis, R.E.: Surface-Based Registration with a Particle Filter, vol. 3216. Springer, Heidelberg (2004)
14. Wan, E.A., van der Merwe, R.: Kalman Filtering and Neural Networks. Wiley, New York (2001)
15. Julier, S.J., Uhlmann, J.K.: Unscented filtering and nonlinear estimation. Proc. IEEE **92**(3), 401–422 (2004)
16. Micusik, B., Pflugfelder, R.: Localizing non-overlapping surveillance cameras under the L-Infinity norm. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2895–2901 (2010)
17. Lee, H., Seo, Y., Hartley, R.: Homography estimation with L∞ norm minimisation method. 14th Korea–Japan Joint Workshop FCV, pp. 87–91 (2008)
18. Schönemann, P.H.: A generalized solution of the orthogonal procrustes problem. Psychometrika **31**(1), 1–10 (1966)
19. Horn, B.K.P., Hilden, H.M., Negahdaripour, S.: Closed-form solution of absolute orientation using orthonormal matrices. J. Opt. Soc. Am. A **5**(7), 1127 (1988)

20. Tam, G.K.L., Cheng, Z., Lai, Y., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X., Rosin, P.L.: Registration of 3D point clouds and meshes: a survey from rigid to non-rigid. IEEE Trans. Vis. Comput. Graph. **19**(7), 1–20 (2013)
21. Amamra, A., Aouf, N.: Robust and sparse RGBD data registration of scene views. In: 2013 17th International Conference on Information Visualisation, pp. 488–493 (2013)
22. Kanazawa, Y., Kanatani, K.: Do we really have to consider covariance matrices for image features?. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, vol. 2, pp. 301–306 (2001)
23. Brooks, M.J., Chojnacki, W., Gawley, D., van den Hengel, A.: What value covariance information in estimating vision parameters?. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, vol. 1, pp. 302–308 (2001)
24. Amamra, A., Aouf, N.: Real-time multiview data fusion for object tracking with RGBD sensors. Robotica, pp. 1–25 (2014). doi:10.1017/S026357471400263X
25. Simon, D.: Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. Wiley-Interscience, New York (2006)
26. Nguyen-Tuong, D., Peters, J.: Model learning for robot control: a survey. Cogn. Process. **12**(4), 319–340 (2011)
27. Marinov, A., Zlateva, N., Dimov, D., Marinov, D.: Weighted ICP algorithm for alignment of stars from scanned astronomical photographic plates. Serdica J. Comput. **6**(1), 101–110 (2012)
28. Hermans, J., Smeets, D., Vandermeulen, D., Suetens, P.: Robust point set registration using EM-ICP with information-theoretically optimal outlier handling. In: CVPR 2011, pp. 2465–2472 (2011)
29. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A **4**(4), 629 (1987)