



**HAL**  
open science

## Direct matching for improving image-based registration

Fernando Israel Ireta Muñoz, Andrew I. Comport

► **To cite this version:**

Fernando Israel Ireta Muñoz, Andrew I. Comport. Direct matching for improving image-based registration. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015), Late Breaking Results, Sep 2015, Hamburg, Germany. hal-02061947

**HAL Id: hal-02061947**

**<https://hal.science/hal-02061947>**

Submitted on 8 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# DIRECT MATCHING FOR IMPROVING IMAGE-BASED REGISTRATION



Fernando I. Ireta Muñoz & Andrew I. Comport

ireta@i3s.unice.fr comport@i3s.unice.fr

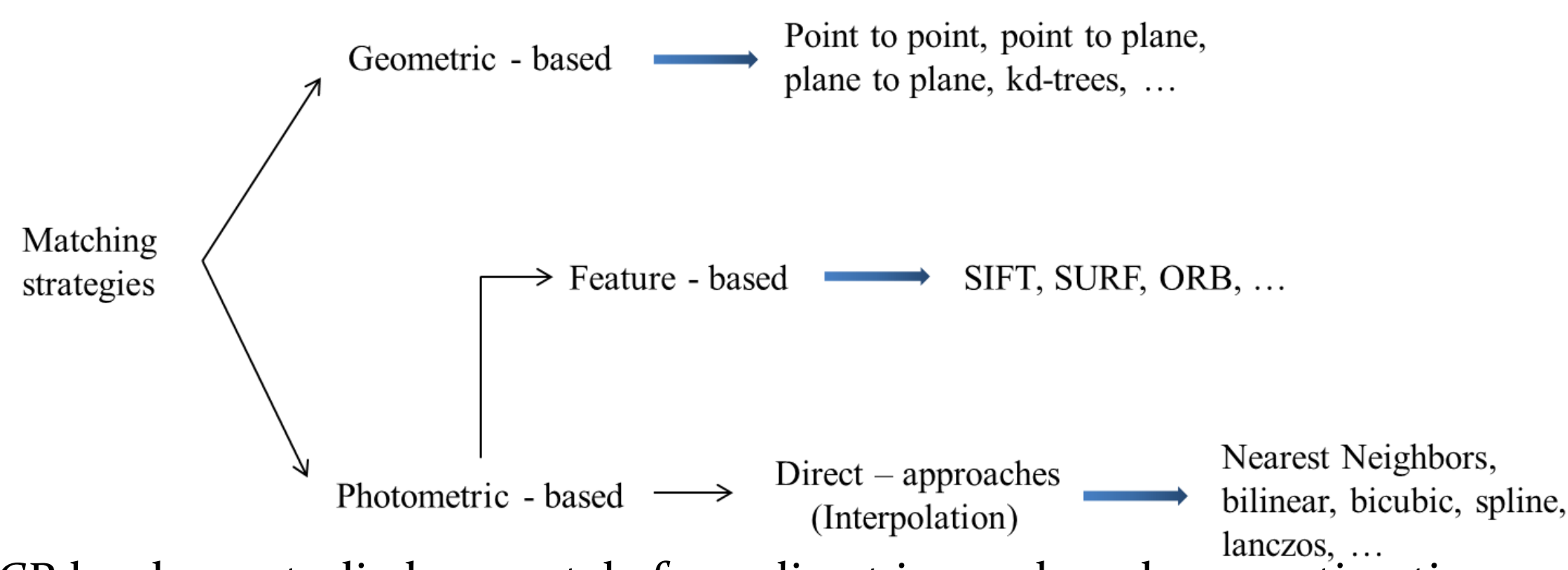


## OBJECTIVE

- Improve direct image-based methods to solve the registration problem that provides the alignment between two measurements acquired at different poses, hence a **matching strategy is proposed based only on the intensity of the pixels and image formation without feature extraction**, which allows to **increase the local convergence domain and speed up the alignment whilst maintaining the robust and accurate properties of direct approaches**.
- Extend the proposed method to simultaneously register both color and depth data and further improve results when RGBD-sensors are available.

## INTRODUCTION

- The registration problem has been widely studied in the field of computer vision, direct methods solve the problem by minimizing the photometric and the geometric error directly in sensor space as opposed to ICP (Iterative closest point) or feature-based image approaches which require feature extraction and matching.
- Many optimizations have been applied in one or more stages of the ICP method, but searching the closest point is the most expensive stage in computational cost. Diverse methods have been proposed to improve searching for the closest points and they can be classified as follows:



- Whilst ICP has been studied separately from direct image-based pose estimation approaches, these matching techniques share much similarity in obtaining the estimation of the pose in different applications (Figure 1).

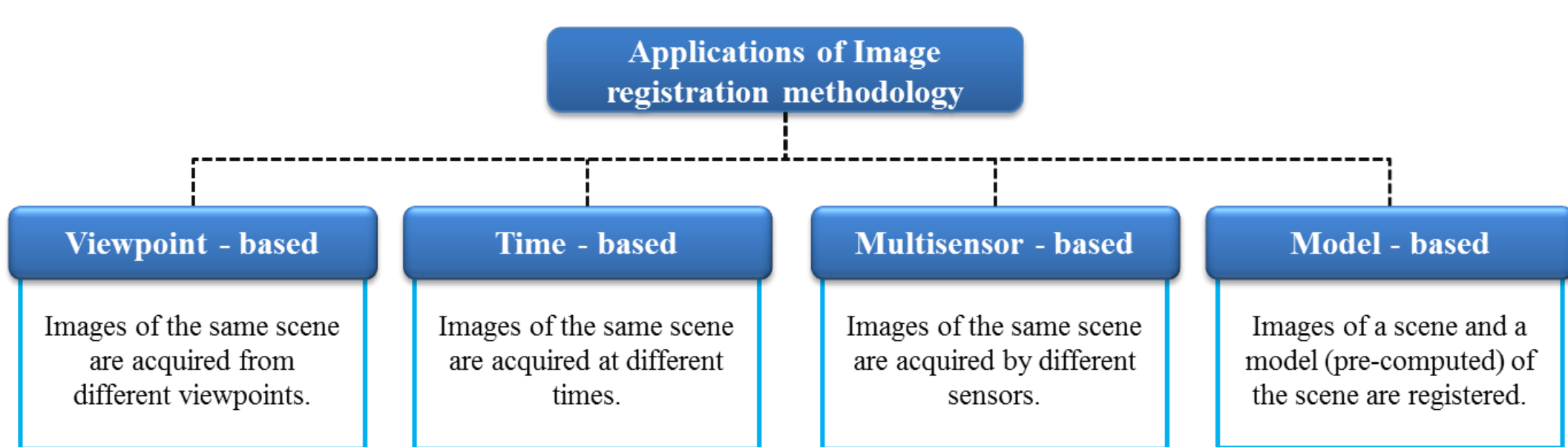
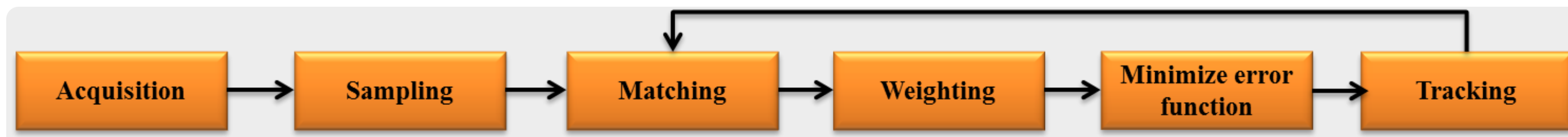


Fig 1. Applications of Image-based registration [1]

## GENERAL OVERVIEW OF POSE ESTIMATION

The pose estimation strategy common to many classic techniques involves the following stages:



If the aforementioned stages are developed, first a measurement vector  $\mathcal{M}^* = [M_1^*, M_2^*, \dots, M_n^*]$  is obtained in a reference position and its corresponding match is found in the following position, where a new measurement vector  $\mathcal{M} = [M_1, M_2, \dots, M_n]$  is obtained.

Based on the point pairs between two datasets, an error metric is defined as:

$$e_i = M_i - f(M_i^*, x)$$

where the unknown motion parameter  $x$  is computed such that:

$$x = \begin{bmatrix} v \\ \omega \end{bmatrix} = -(J^T W J)^{-1} J^T W e_i \in \mathbb{R}^{6 \times 1}$$

Finally, the pose estimate is computed at each iteration using an homogeneous update  $\hat{T}$  until convergence as  $\hat{T} \leftarrow \hat{T}T(x)$ , where the homogeneous pose matrix  $T(x)$  is related to the angular  $\omega$  and linear  $v$  velocity by the exponential map  $T(x) = e^{[x]_\wedge}$ , with the operator  $[\cdot]_\wedge$  as:

$$[x]_\wedge = \begin{bmatrix} \omega_x & v \\ 0 & 1 \end{bmatrix}$$

where  $[\cdot]_\wedge$  is the skew symmetric matrix operator.

$J \rightarrow$  Jacobian computed by derivation of the error function  $W \rightarrow$  Weighted matrix

## RELATED WORK

- When a RGB-D sensor is available, two images of dimensions  $m \times n$  taken at different views of the same scene can be considered to estimate the unknown pose between them. Each image is associated with an intensity function and a depth function as  $\mathcal{I} = (I(p), Z(p))$ , where the parameter  $p = [u \ v]^T$ , correspond to the pixel coordinates of  $\mathcal{I}$ .
- Three main strategies can be considered to compute the unknown pose. Since the techniques shares much similarity in the iterative estimation of  $T(x)$ , the main differences are shown in the following table.

GEOMETRIC - BASED	PHOTOMETRIC - BASED	HYBRID - BASED[4]
Measurement: 3D Euclidean points $M_i = K^{-1} \bar{p}_i Z_i = [X_i \ Y_i \ Z_i]^T$	Measurement: Brightness at pixels coordinates $M_i = I(p_i)$	Measurement: 3D + brightness $M_i = \begin{bmatrix} K^{-1} \bar{p}_i Z_i \\ I(p_i) \end{bmatrix} = [X_i \ Y_i \ Z_i \ I(p_i)]^T$
Minimization of Error function: Point to plane. $e_{gi} = \bar{R}R(x)N_i^T(M_i - \Pi_3 \hat{T}T(x)M_i^*)$	Minimization of Error function: Bilinear interpolation $e_{pi} = I(\omega(\hat{T}T(x), K^{-1} \bar{p}_i Z_i)) - I^*(p^*)$	Minimization of error function: Bi-objective $e_i = \begin{bmatrix} \lambda e_{gi} \\ e_{pi} \end{bmatrix}$

Without the matching step in the pose estimation algorithms, the distance is computed for every set of coordinates  $M_n$  and its performance depends on the dimension  $n$  of the measurements vector.

$K \rightarrow$  Calibration Matrix  $N^* \rightarrow$  Normals of the reference points  $\omega(\hat{T}T(x), p) \rightarrow$  warping function [4]  
 $R(x) \rightarrow$  Rotation Matrix  $\bar{R} \rightarrow$  Homogeneous update for rotation  $\Pi_3 = [1, 0, 0] \in \mathbb{R}^{3 \times 4} \rightarrow$  Projects  $T(x)$  onto the  $3 \times 4$  space  $\lambda \rightarrow$  constant that scales the relative error distributions  $p = [X \ Y \ Z]^T \rightarrow$  3D Euclidean point

## References

- Barbara Zitov and Jan Flusser. Image registration methods: a survey. Image and Vision Computing, 21:977-1000, 2003.
- L. Morency and T. Darrell. Stereo tracking using ICP and normal flow constrain. In Pattern recognition, 2002. Proceedings, 16th International Conference on, volume 4, pages 367-372.col.4, 2002.
- Mariusz Muja and David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In International Conference on Computer Vision Theory and Application VISSAPP '09, pages 331-340. INSTICC Press, 2009.
- M. Meilland and A.I. Comport. On unifying key-frame and voxel-based dense visual SLAM at large scales. In International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3-8 November 2013. IEEE/RSJ
- J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d systems. In Proc. Of the International Conference on Intelligent Robot Systems (IROS), Oct. 2012.

## METHODS FOR IMPROVING DIRECT IMAGE - BASED REGISTRATION

### Image-based minimization

- Similarly to several ICP strategies that accelerate the alignment, a creation of a balanced kd-tree for the reference image is proposed, which is done only once before the iterative loop [2].

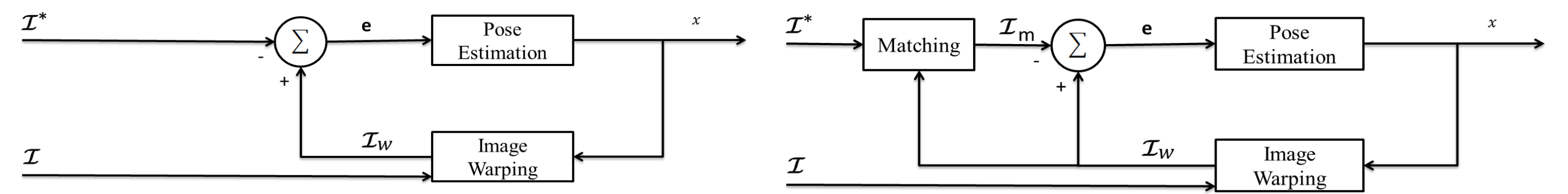


Fig 2. Direct image-based registration

Fig 3. Direct matching image-based registration

- The kd-tree is defined such:  $k^* = k(\mathcal{M}^*)$ , where each measurement of the vector  $\mathcal{M}^* = [M_1^*, M_2^*, \dots, M_n^*]$  involves the pixel coordinates with their respective intensity value as:

$$M_i^* = [p_i^* \ I^*(p_i^*)]^T$$

- A new set  $\mathcal{M} = [M_1, M_2, \dots, M_n]$  is defined as the warped measurements vector, where

$$M_i = [p_i^w \ I(\omega(\hat{T}T(x), K^{-1} \bar{p}_i^* Z_i^*))]^T$$

- The matching points are estimated with the search function of the FLANN library[3]. The correspondences are identified by an index as:  $m_i = match(k^*, \mathcal{M})$ , which gives the  $i$ -position where the nearest neighbor for  $M_i$  was found in  $\mathcal{M}^*$ . Therefore, the new photometric error is computed such as:

$$e_{pi} = M_i - M_i^m$$

where  $M_i^m = \mathcal{M}^*(m_i)$  is a vector with the matched points.

### ALTERNATIVE APPROACH

- If the match is computed at each iteration, then the metric error generated between the warped and matched pixel coordinates could be minimized. The error function is computed as:

$$e_{UVI} = M_i - M_i^m = \begin{bmatrix} \lambda e_{UV} \\ e_I \end{bmatrix}$$

- In order to find correspondences using information from the reference data measurements, the projected warped points could be projected-back into the reference image trough the inverse of the transformation matrix as:

$$M_i = \left[ \omega \left( K \Pi_3 (\hat{T}T(x))^{-1}, P_{w_i} \right) \ I(\omega(\hat{T}T(x), P_i^*)) \right]^T$$

### Bi-objective minimization

- The measurements vectors are created in 4 dimensions (3D + intensity) as follows:

$$\mathcal{M}^* = \begin{bmatrix} P^* \\ \mathcal{I} \end{bmatrix} \quad \mathcal{M} = \begin{bmatrix} K^{-1} \bar{p}^w Z^* \\ \mathcal{I}(\omega(\hat{T}T(x), P^*)) \end{bmatrix}^T \in \mathbb{R}^{4 \times mn}$$

- The matched points are used in the minimization of the two error functions point to plane and bilinear interpolation such that:

$$e_{XVZI} = \left[ \lambda \bar{R}R(x)N^{*T}(P_i^m - \Pi_3 \hat{T}T(x)P_i^*) \quad I(\omega(\hat{T}T(x), P_i^*)) - I_i^m \right]^T$$

$p^w = [u^w \ v^w]^T \rightarrow$  warped pixel coordinates  $m \rightarrow$  vector that contains the index values where the closest point was found  $P_i^w \rightarrow$  Warped 3D points  $I_i^m = I^*(p_i^m) \rightarrow$  matched intensities

## RESULTS

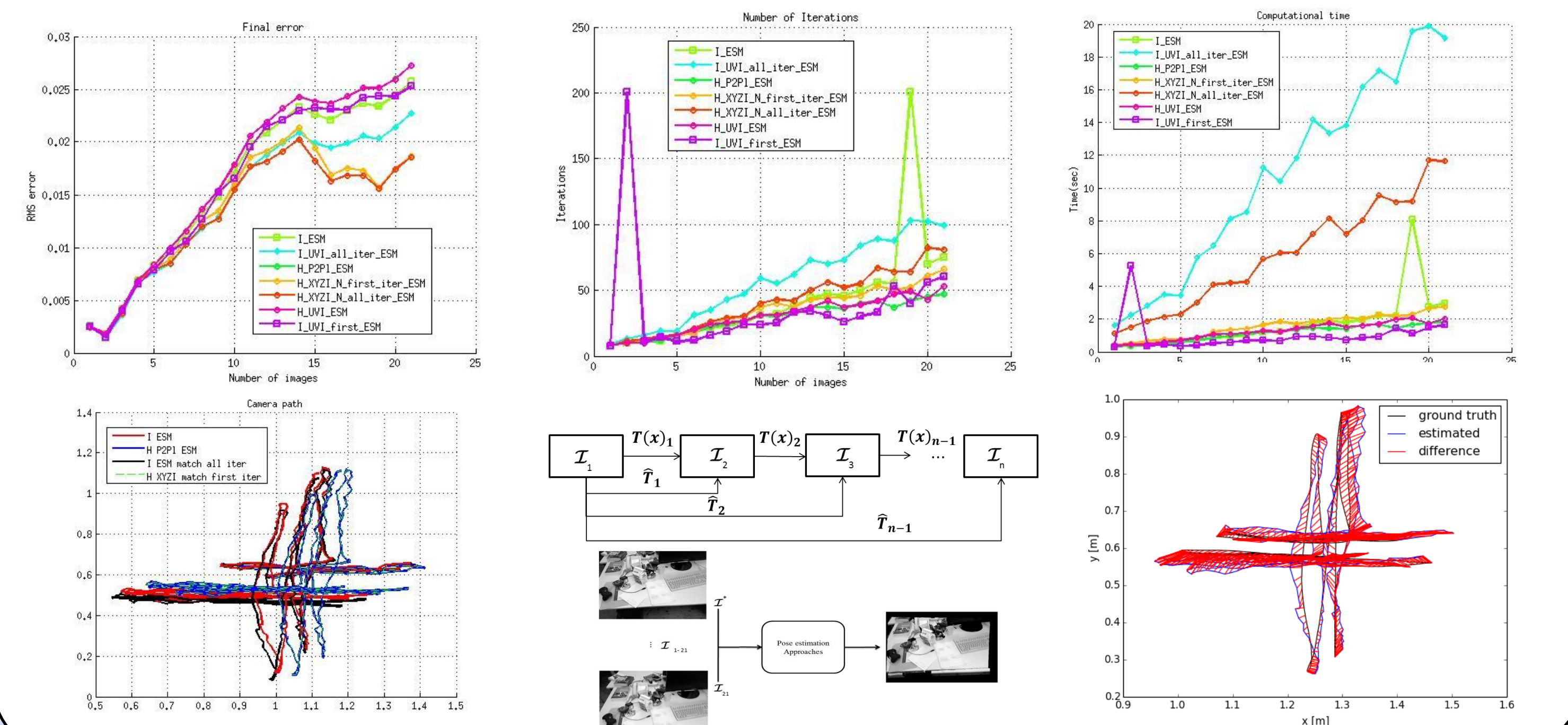
### - Simulated images

50 synthesized images with Gaussian noise and random initial pose were generated for the simulations, where a multiresolution pyramid was used to improve the computational efficiency. The computational time estimated during the iterative loop and the number of iterations until achieve convergence are computed for each image generated. The average of these results is shown in the following table:

Method	Iterations (mean)	Time (sec) (mean)
Bilinear interpolation (I_ESM)	61.40	0.6515
Bilinear interpolation + matching at first iteration (I_UV1_first_ESM)	57.28	0.4809
Hybrid method (H_P2P1_ESM)	14.88	0.3784
Hybrid + matching at first iteration (H_XYZLN_first_iter)	10.86	0.3559
Hybrid + matching in all iteration (H_XYZLN_all_iter)	10.26	0.4808

### - Real images

Part of real RGB-D images in a real sequence [5] were used and resized to carry-out the new image-based algorithms. The first image was considered as the groundtruth and the correspondences are found for each image that follows, where each one generates a bigger error when the camera moves gradually its position. Finally, the full trajectory was estimated (sequence freiburg1\_xyz).



## CONTRIBUTIONS

- A direct image-based registration method that benefits from a "direct" matching step which avoids any feature extraction yet benefits from a larger convergence domain and faster minimization.
- Whilst the novelty lies mainly in the direct matching approach, the technique readily extends for RGB-D sensor registration by minimizing the bi-objective error function.
- The improvements are demonstrated for both, simulated and real images and it has been shown that better estimations can be achieved using FLANN, which provide faster convergence when the matching is performed in the first iteration only.
- In conclusion, the bi-objective strategy is the most computationally efficient, accurate and robust when the matching method is integrated in the minimization of the error function.

