https://doi.org/10.2352/ISSN.2470-1173.2018.09.IRIACV-126

Robust Pose Estimation with a Stereoscopic Camera in Harsh Environments

Longchuan Niu**, **Sergey Smirnov*, **Jouni Mattila***, **Atanas Gotchev****, **Emilio Ruiz*****
**** Laboratory of Automation and Hydraulics Engineering, Tampere University of Technology, Tampere, Finland***
***** Laboratory of Signal Processing, Tampere University of Technology, Tampere, Finland***
****** Fusion for Energy, Barcelona, Spain***

Abstract

Remote teleoperation of robotic manipulators requires a robust machine vision system in order to perform accurate movements in the navigated environment. Even though a 3D CAD model is available, the dimensions and poses of its components are subject to change due to extreme conditions. Integration of a stereoscopic camera into the control chain enables more precise object detection, pose-estimation, and tracking. However, the conventional stereoscopic pose-estimation methods still lack robustness and accuracy in the presence of harsh environmental conditions, such as high levels of radiation, deficient illumination, shiny metallic surfaces, etc. In this paper we investigate the ability of a specifically tuned iterative closest point (ICP) algorithm to operate in the aforementioned environments and suggest algorithmic improvements. We demonstrate that the proposed algorithm outperforms current state-of-the-art methods in both robustness and accuracy. The experiments are performed with a real robotic manipulator prototype and a stereoscopic machine vision system.

Introduction

Computer Aided Teleoperation (CAT) usually implies several different aspects or tools within the robotic operation chain. The main goal of the teleoperation in our application is to perform maintenance and tool manipulations with several kinds of objects inside a radioactive fusion reactor, where human presence is prohibited.[1]

In order to perform operations during the reactor maintenance break, a robotic manipulator must insert different tools inside several mounting holes for the different pre-defined reactor components. Even though the 3D CAD models of the components to be manipulated are known with high accuracy in advance, these elements are subject to small drifts in their poses, which have six degrees of freedom, and material deformation due to extreme heat and magnetic loads during machine operation. For precise and reliable teleoperation, the environment dimensions and poses have to be estimated accurately and converted into the robot's world coordinates [1]. Once that relation, that is, the rigid-body transformation is found, operations such as tool pickup, insertion, turning, retraction, and putting down can be made semi-automatic.

The problem of pose estimation, however, remains challenging, due to the harsh environment within the chamber. *Radiation tolerant* cameras are the only sensors capable of working in the chamber, and no stationary equipment is allowed. Apart from the low resolution and grayscale output of these cameras, other limitations connected to the environment are also present, including a high level of image noise due to the radiation; deficient illumination of the scene due to constraints on available light sources; non-Lambertian reflectance of shiny metallic surfaces and objects, etc. All these are difficulties that make any vision-based object detection and pose-estimation system problematic.

A previous study on pose-estimation CAT systems based on the 3D template matching algorithms showed significant limitations of the monocular approach [2]. In our application [3], we use a stereoscopic camera mounted to the last joint of a robot manipulator as a sensing tool to perform vision tasks, object detection, and pose-estimation. The same camera system can also be used by the operator, for instance when inspecting objects or the robot itself.

A stereoscopic camera system can reconstruct the geometry of a 3D scene based on stereo correspondences. Subsequently, it generates a depth map in the form of a grayscale image describing the geometry. We utilize this property in order to recover a *3D point cloud* representation of a scene, then try various *iterative closest point* (ICP) alignment approaches [4, 5] in order to detect and finally recover the pose of a target object.

Problems and Limitations

Current ICP methods are limited by the use scenario. Depth maps and point clouds generated by a stereoscopic camera system are significantly degraded due to various factors of the operating environment, and thus only a small portion of points can be trusted. For instance, the depth of shiny surfaces usually cannot be well estimated due to violation of the Lambertian reflectance model. High levels of noise can also result in false matches within textureless areas, and low-resolution grayscale imagery significantly limits the discriminative power of the stereo-matching algorithms. All these difficulties result in systematically erroneous depth values (outliers), which significantly disorient conventional general-purpose ICP methods.

Strong luminance gradients are the only features in the stereo images that can be trusted for their error-free behaviour. In the textureless and smooth scenes, strong image gradients usually correspond to object boundaries or significant changes in the surface (e.g., slope). In contrast to other robust image features, such as scale-invariant features (SIFT) [6] or speeded-up robust features (SURF) [7], image gradients are much denser and tolerate image noise.

Nevertheless, using the object boundaries as matching primitives can also limit the selection of the underlying ICP method.

[1] The nuclear-fusion reactor, constructed within the ITER project (`http://www.iter.org`).

As the surface normals generally cannot be estimated at borders and object edges, only *point-to-point* minimization is possible. More advanced *point-to-plane* [8] or generalized *plane-to-plane* [9] minimization approaches cannot be utilized.

The recently proposed *edge-point ICP* method [4] is capable of operating within this type of constraints. The method successfully works when the estimated point cloud contains few outliers and when a good initialization point is provided. From the algorithmic point of view, outliers are not only wrongly estimated depth values, but also points that have no corresponding points in the target (model) point cloud, or vice-versa.

Another substantial property of depth-from-stereo methods is the generation of content-dependent occlusion artifacts in their output. Occlusion hole-prediction methods exist, but they all rely on high-quality depth of the neighboring zones and use some guessing mechanisms, which is not allowed in precise alignment tasks. During the preparation of reference point clouds, based on the supplied CAD models, such artifacts are usually not taken into account, as it is not possible to predict from which viewpoint the object will be captured. Thus, large numbers of reference points may become outliers, with no corresponding point in the estimated cloud. Depending on the number of mismatched points, performance of the ICP alignment can be seriously degraded.

Contributions

In this paper we propose an efficient method to increase the robustness and the accuracy of the ICP alignment in which target point clouds are estimated using stereoscopic capture in the harsh industrial environments. We use the *approximate planarity* assumption in order to recover good initialization points for the ICP algorithm and illustrate its suitability for successful convergence. In contrast to conventional methods, we also use dynamically sampled reference point clouds, especially targeted to each particular stereo-observation. We model artifacts appearing in the depth-from-stereo methods in order to minimize the number of outliers in the reference clouds and thus increase final alignment accuracy.

Prior Art

Depth-from-Stereo

Estimation of the scene geometry from a binocular camera setup is usually called *stereo-matching* or the *depth-from-stereo* problem. Even though this field is already well developed, and many advanced techniques are available, in our problem we not only required estimating the depth but also correctly manipulating the depth values, projecting them back to the 3D space with real-world coordinates. Therefore, conventional stereo-matching methods, based on stereo-image rectification [10], might underperform due to the introduction of artificial camera transforms and excessive image interpolation steps. Moreover, a deviation from geometrically parallel camera configuration is possible (e.g., the camera optical axes might be crossed), thus introducing substantial image deformation in rectification-based methods.

Instead, *plane-sweeping depth estimation* methods, using calibrated camera parameters, allow direct processing of the captured imagery [11]. Figure 1 illustrates the depth-estimation method, based on the plane-sweeping principle. In this method, the entire observable scene is divided into a number of fronto-parallel planes (hypothesizes), where stereo correspondences might be found. Such hypothesizes can be selected for example by selecting the possible depth range (i.e., minimum and maximum possible depth values) and number of layers, which controls the trade-off between fidelity and computational complexity of the method.

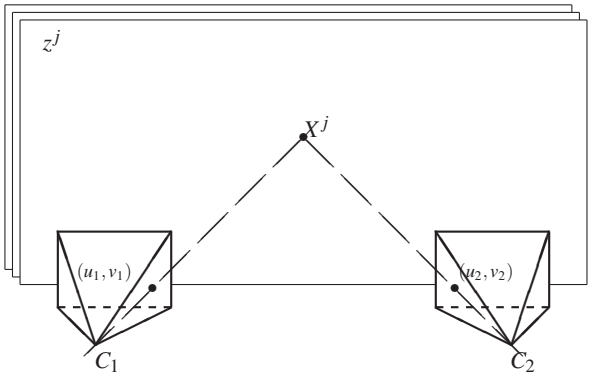


Figure 1. *Illustration of the plane-sweeping principle of the depth-from-stereo estimation methods*

For every hypothetical depth z_j, one can project a pixel (u_1, v_1) from a reference camera to a 3D space, using pre-calibrated camera matrix C_1:

$$\mathbf{X}^j = C_1^{-1}\dot{\mathbf{x}}_1, \tag{1}$$

where $\dot{\mathbf{x}}_1$ is the homogeneous projective coordinate of a current pixel $\dot{\mathbf{x}}_1 = (u_1 \cdot z_j, v_1 \cdot z_j, z_j, 1)^T$, $\mathbf{X}^j$ is the resulting point coordinate in a 3D space; and $j = 1, .., N$ where N is the selected number of layers.

Every obtained 3D point $\mathbf{X}^j$ can be further projected onto the sensor plate of a second camera using a similar equation:

$$\dot{\mathbf{x}}_2 = C_2\mathbf{X}^j \tag{2}$$

where $\dot{\mathbf{x}}_2$ is a projective pixel position in a second camera image plane, and the actual pixel coordinates can be recovered as:

$$u_2 = \frac{\dot{\mathbf{x}}_2.x}{\dot{\mathbf{x}}_2.z} \quad , \quad v_2 = \frac{\dot{\mathbf{x}}_2.y}{\dot{\mathbf{x}}_2.z} \tag{3}$$

Similarly to conventional rectification-based methods [10], one can construct a 3D cost volume, in which pixel dissimilarities are calculated between the original pixel in the reference camera and the corresponding pixel in the second one:

$$C(u, v, j) = \|I_1(u_1, v_1) - I_2(u_2, v_2)\|, \tag{4}$$

where I_1 and I_2 denote the first and second images, respectively, and because the (u_2, v_2) coordinates are not necessarily integers, the corresponding sampling should be performed for instance with bilinear interpolation.

After appropriate cost aggregation [11], the depth map can be recovered by using the so-called *winner-takes-all* approach:

$$Z_1(u, v) = z_{\hat{j}}, \hat{j} = arg\min_j \tilde{C}(u, v, j), \tag{5}$$

where $\tilde{C}(\cdot)$ denotes the aggregated cost volume.

The coordinates of the point cloud in the reference camera can now be reconstructed using the same equation as in (1), replacing $z_{\hat{j}}$ with the estimated value.

ICP Methods

Since the first invention of the ICP method [12], many updates have been proposed [5, 4, 13]. One of the directions for improvements has been reducing the influence of outliers on the global error. Thus, many widely accepted techniques remove too many point correspondences while calculating global error [5]. A number of linearized methods were suggested using SVD [14], quaternions [15], and dual quaternions [16] for minimizing the error metric with a closed-form solution. High quality of the sensed (input) point cloud is an essential requirement for conventional ICP algorithms. A relatively moderate fraction of outlying points in the input cloud can significantly degrade performance of the method, thus preventing its usage for real-world applications. This is an important aspect for point clouds estimated via stereoscopic camera in harsh environments. As the passive vision systems (including depth-from-stereo methods) usually fail in the presence of textureless or shiny (i.e., non-Lambertian) surfaces, their depth maps become corrupted with a high number of false estimates. Consequently, input point clouds could be contaminated with outliers, thus preventing use of the technique for pose estimation tasks.

Edge-point ICP [4] uses an additional type of filtering step, where points not connected to a strong image gradient are removed from the point cloud. Even though this operation can significantly reduce the number of available points in the cloud, their discriminative power significantly improves, thus resulting in better performance, especially in cases when textureless areas dominate the scenes.

HandEye Calibration and World Coordinates

The object pose in terms of camera coordinates has to be transformed into robot world coordinates, for which hand-eye calibration [17] is needed. Figure 2 indicates the relationship between the robot end-effector, the camera, and the object in the world coordinates with the formula:

$$P = R \cdot X \cdot A \tag{6}$$

where P is the required pose of an object, A is the estimated alignment in the camera coordinate space, X is the eye-hand transformation matrix, and R is the current position of the robot hand/wrist.

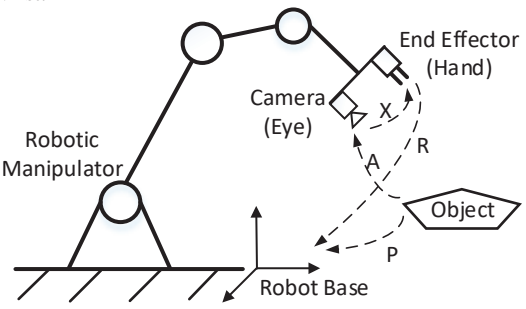


Figure 2. *Hand-eye calibration and world coordinates*

Sampling of CAD Models

Sampling of CAD models is usually done once during algorithm development and all estimated points in the point cloud are matched against this reference cloud.

An example of sampling of CAD models is provided in Figure 3, which shows a sensed point cloud before alignment.

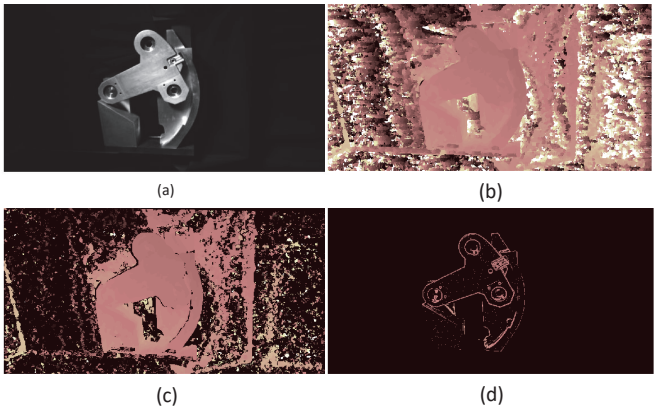


Figure 3. *Sampling of CAD model: (a) given image, (b) raw depth, (c) depth after L2R, and (d) depth after sampling*

Proposed Method

Typical industrial environments, which are also considered in our application, usually contain many planar surfaces. Such surfaces are easier to manufacture and they are more convenient when constructing large-scale structures. Target objects can also be considered as having at least one major planar surface, facing the stereoscopic sensor. Even though a strict planarity constraint may not be fully satisfied due to obstacles and other features on the object surface, often we can still rely on the *approximate planarity of the surfaces*. In our method, we propose imposing such constraints in order to estimate a good initialization point for the alignment algorithm and to avoid point mismatches due to occlusion artifacts.

The point cloud estimated from a scene can be analyzed for the presence of plane structures. This can be done, for instance, using the random sample consensus (RANSAC) [18] plane-fitting method. General plane-fitting methods in 3D point clouds usually utilize a generalized plane equation:

$$ax + by + cz + b = \mathbf{a}^T \hat{\mathbf{x}} = 0, \tag{7}$$

where $\mathbf{a} = [a,b,c,d]^T$ is the vector of plane parameters to estimate, and $\hat{\mathbf{x}} = [x,y,z,1]^T$ is the homogenous point coordinate from the cloud.

A conventional way to perform the analysis is to select three random points from the cloud, fit the plane parameters and estimate the number of other points that belong to the same plane with some kind of tolerance. The process is repeated multiple times, and the plane equation containing the largest number of inliers is considered the largest plane found in the scene.

As the point cloud estimated with the stereo-camera setup usually does not capture highly slanted or parallel-to-the-optical axis planes, we can utilize a relaxed plane equation:

$$z = ax + by + c = \mathbf{a}_s^T \hat{\mathbf{x}}_s. \tag{8}$$

Following a similar RANSAC methodology, the matrix of three selected points X and the vector of corresponding depth values $\mathbf{z}$ can be utilized to recover the plane parameters using the

Moore-Penrose pseudo-inverse:

$$X = \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{pmatrix}, \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \tag{9}$$

$$\mathbf{a}_s = \mathbf{z} \cdot X^T (XX^T)^{-1} \tag{10}$$

Here, $n = 3$ for the initial plane estimation and can be arbitrary during the plane refinement stage, when plane parameters are estimated using all the found inliers. Inliers can be selected using pre-defined threshold value θ, as points whose distance to plane is lower than a threshold $|ax_i + by_i + c - z_i| < \theta$.

The parameter θ can also control the expected proximity of an object surface to a plane model. For objects with dominating planarity, θ can be reduced to account only for possible depth estimation errors, while for objects containing many bumps or cavities, larger values of θ can be beneficial.

When the CAD model is aligned with its major plane (i.e., model origin and X-Y coordinates belong to it), the obtained plane parameters can directly be used to estimate good initialization of the rotation matrix. Two of the Euler angles can be estimated as:

$$\beta_x = tan^{-1} b, \tag{11}$$

$$\beta_y = -tan^{-1} a, \tag{12}$$

where β_x and β_y are Euler angles around X and Y axes, respectively.

Rotation around the Z axis cannot be estimated by such a coarse method; however, the generic assumption of vertical camera orientation can still be used to provide meaningful initialization. As a guess for an initial translation, we use the median-centroid of a point cloud. This assumption may introduce certain limitations of the method, particularly when a significant part of the surrounding scene is also visible to the stereo camera setup.

Advanced CAD Model Sampling

Apart from the transform matrix, we also propose a method to reduce the number of mismatches in the point cloud estimated by using the depth-from-stereo method. As the rotational component in the true underlying transformation can be arbitrarily large, projective distortions appearing in the sensed images may be significant. We use dynamic CAD model re-sampling as a mechanism to reduce possible outliers in the model point cloud, hence improving the accuracy of the final alignment.

In conventional ICP methods, the model point cloud is usually statically defined and re-used every time a new observation is made. In practical cases, however, excessive numbers of mismatched points prevents this use.

We use heuristics in order to remove possible outliers from the reference cloud. For instance, a left-to-right correspondence check rendering is done with the transform found in the initial alignment step. We render images for both the reference and secondary camera (with the same configuration as in the stereoscopic setup). This allows us to apply the same left-to-right correspondence check as in the estimated depth. We use rendered images of a CAD model to find strong edges in the scene and prepare a point cloud according to the same process as for the source point cloud. Applying these heuristics, the reference cloud contains the same amount of occlusion and similar results with regard to the edge properties as the source cloud.

For efficient processing, we propose the following scheme. Figure 4 shows the procedure per frame.

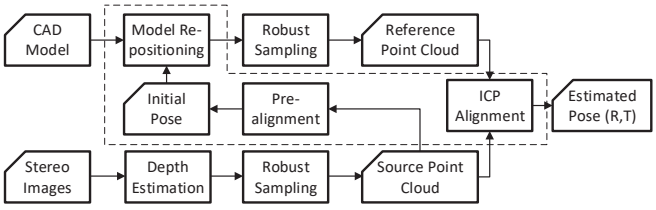


Figure 4. *Flowchart of proposed ICP implementation*

Standard edge-point ICP initializes its model point cloud by sampling only once, which is not robust in the case of a stereoscopic camera. Thus, in our proposed ICP (new blocks within the dashline), we render our CAD model such that it shows approximately what the camera is seeing. The model point cloud is estimated every time using the pre-alignment, and we sample the CAD model relative to our initial estimated alignment.

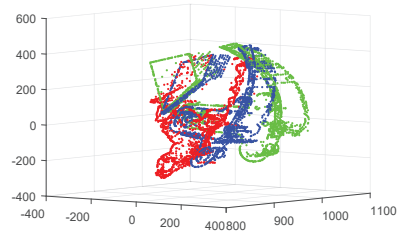


Figure 5. *Comparison of sampled point cloud; Red, sensed; Green, standard ICP; Blue, proposed ICP*

As we can see in Figure 5, the blue point cloud has a better initialization point than the green one, which helps to overcome the issues with local minima.

Experiments

In order to validate the proposed method, we run two sets of experiments, based on photographs of two target objects, namely a CLS mockup and a knuckle, illustrated in Figure 6. The CLS mockup is made of steel, from laser-cut sheet material welded together with high precision, while knuckle was mainly 3D printed with a fused deposition modeling (FDM) printer, sanded, and then painted with shiny metallic paint; the manufacturing accuracy is thus lower for the knuckle. Surrounding elements in the knuckle were made with precise steel-cutting approaches. Overall, both target objects well represent the expected reflectivity and textureless properties of the application, as well as corresponding to an underlying CAD model with tolerances up to 0.2 mm.

In order to obtain a comprehensive set of experimental data, we gathered a significant number of stereo-images using different camera offsets, orientations and different illumination conditions. Overall, 31 stereo-pairs per target object were acquired using the calibrated stereo camera setup. Camera positions were selected such that for the closest (to the object) camera position, the target object barely fit in the camera view, while for the position further away from the target objects, it occupies just a small fraction of

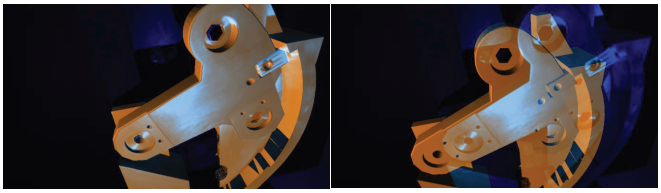
the image, representing a wide range of distances. Figure 6 shows two of the acquired images.

(a) (b)

Figure 6. *Sample images of (a) "CLS-mockup" and (b) "knuckle" objects used in the experiments.*

The main goal of our experiment was to estimate the robustness, reliability, and accuracy of the proposed method as well as to compare it with competing approaches. In order to estimate robustness, for every acquired stereo-pair we independently ran the alignment algorithm 30 times and measured the number of false pose estimates, that is, when the aligned object completely disagrees with the acquired data.

This can be done in semi-automatic mode, in which the software asks the operator to confirm whether current alignment was successful. Figure 7 shows two alignment results, where the CAD model was projected to the camera space and rendered according to the estimated object pose. Two images, the acquired and the rendered one, are combined together in different color channels and presented as a single RGB image, which we refer to as the "augmented" image. Such representation can easily be evaluated by the operator for correctness of alignment and thus be selected as correct or not.

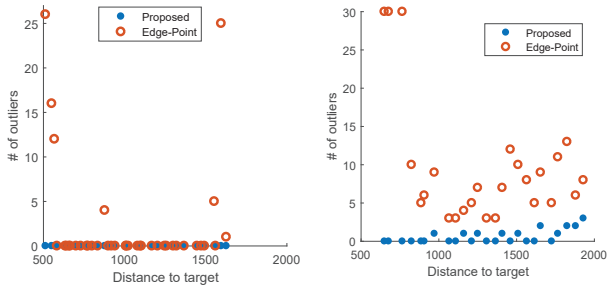


(a) (b)

Figure 7. *Example of ICP alignment with augmented images: (a) successful, (b) unsuccessful*

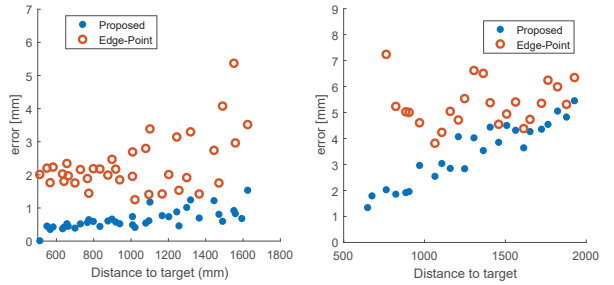
The 31 observations of every stereo-pair provided a number of pose estimates, including the relative rotation and the translation between the camera and the CLS mockup or knuckle. We used semi-manually estimated positions as the threshold for selection of correct estimates, or inliers. All inliers from these observations are averaged together in order to obtain the centroid of the estimated points. Now, by measuring the Euclidean distance between the centroid and every other estimate, one can obtain the average displacement (deviation) for this particular stereo-pair. While taking an estimated Z (depth) value as a reference variable, one can plot a figure in which the horizontal axis represents depth (Z-distance between camera and the object) and the vertical axis shows the respective deviation value. Figure 8 and Figure 9 show these graphs for a few different experiments.

As we can see, when the target object is too close to the camera, it can no longer observe all the distinctive edges. This indicates a general lower limit of the pose-estimation system where too-close observations are not reliable. In addition, the images show that both the CLS mockup and the knuckle achieve good accuracy and stability within the middle range. This can be explained as fairly consistent behavior within the expected operational range. With the increase in distance, the repeatability error grows but also becomes unstable, which could suggest the existence of an upper limit for the system. This limit, however, was not reached during these sets of experiments.



(a). *"CLS-Mockup"* ***(b).*** *"Knuckle"*

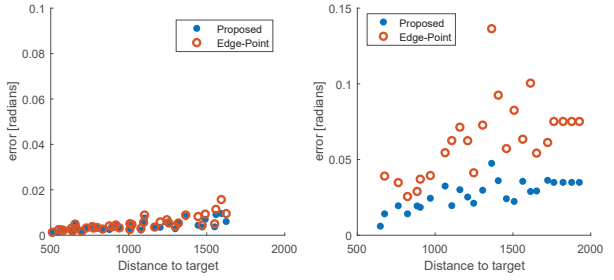
Figure 8. *Number of outliers for CLS mockup and knuckle datasets.*



(a). *"CLS-Mockup"* ***(b).*** *"Knuckle"*

Figure 9. *Position stability (repeatability) for CLS mockup and knuckle datasets.*

A similar procedure can be done for the rotational part of the found transforms. We extract the rotation matrix from each estimated transform and convert them to a vector of Euler angles. Then, the mean Euler angle value for one image is chosen as the correct rotation, and the error in the rotations is expressed as the difference between the mean Euler vector and the rest of the vectors. In order to obtain a single variable out of all the observations, we convert the angular error vectors to a list of combined errors, taking the L2 norm of each vector. Then, the mean value of all combined errors is taken to represent the integral error metric for one particular image. The process is repeated for every image in the dataset in order to obtain a closed curve.



(a). *"CLS-Mockup"* ***(b).*** *"Knuckle"*

Figure 10. *Angular stability (repeatability) for CLS mockup and knuckle datasets.*

Figure 10 exhibits similar performance of the method as in Figure 9. The optimal range of the system is reached in the range

of 65 to 100 cm, and the the overall integral angular error is on the order of 0.02 to 0.05 rad.

Conclusion

The measurement of repeatability error demonstrates fairly consistent behavior, even though the target object was imaged from different perspectives. Overall, our proposed method has shown more robustness and accuracy than the standard edge-point ICP method in terms of the number of outliers and precision of pose estimation. The results verify the effectiveness of the proposed method.

Acknowledgment

The work leading to this publication has been funded in part by Fusion for Energy and TEKES under Grant F4E-GRT-0689. This publication reflects the views only of the authors, and Fusion for Energy or TEKES cannot be held responsible for any use which may be made of the information contained herein.

References

[1] Pritam Prakash Shete, Abhishek Jaju, Surojit Kumar Bose, and Prabir Pal, "Stereo vision guided telerobotics system for autonomous pick and place operations," in *Proceedings of the 2015 Conference on Advances In Robotics*. ACM, 2015, p. 41.

[2] Z Ziaei, A Hahto, J Mattila, M Siuko, and L Semeraro, "Real-time markerless augmented reality for remote handling system in bad viewing conditions," *Fusion Engineering and Design*, vol. 86, no. 9, pp. 2033–2038, 2011.

[3] L. Niu, O. Suominen, M.M. Aref, J. Mattila, E. Ruiz, and S. Esque, "Eye-in-hand manipulation for remote handling: Experimental setup," in *International Conference on Robotics and Mechatronics*. IOP Conference Series, Hong Kong, 2017.

[4] Masahiro Tomono, "Robust 3d slam with a stereo camera based on an edge-point icp algorithm," in *IEEE International Conference on Robotics and Automation*, Kobe, Japan, May 12-17 2009.

[5] Dmitry Chetverikov, Dmitry Stepanov, and Pavel Krsek, "Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm," *Image and Vision Computing*, vol. 23, no. 3, pp. 299–309, 2005.

[6] David G Low, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision*, 1999, vol. 2, pp. 1150–1157.

[7] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," *Computer vision–ECCV 2006*, pp. 404–417, 2006.

[8] François Pomerleau, Francis Colas, Roland Siegwart, and Stéphane Magnenat, "Comparing icp variants on real-world data sets," *Autonomous Robots*, vol. 34, no. 3, pp. 133–148, 2013.

[9] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun, "Generalized-icp.," in *Robotics: science and systems*, 2009, vol. 2, p. 435.

[10] Daniel Scharstein and Richard Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[11] S. Smirnov, A. Gotchev, and M. Georgiev, "Comparison of cost aggregation techniques for free-viewpoint image interpolation based on plane sweeping," in *Ninth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM*, 2015.

[12] Paul J Besl, Neil D McKay, et al., "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[13] Roberto Marani, Vito Reno, Massimiliano Nitti, Tiziana D'Orazio, and Ettore Stella, "A modified iterative closest point algorithm for 3d point cloud registration," *Computer-Aided Civil and Infrastructure Engineering*, vol. 31, no. 7, pp. 515–534, 2016.

[14] K Somani Arun, Thomas S Huang, and Steven D Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on pattern analysis and machine intelligence*, , no. 5, pp. 698–700, 1987.

[15] Berthold KP Horn, Hugh M Hilden, and Shahriar Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *JOSA A*, vol. 5, no. 7, pp. 1127–1135, 1988.

[16] Michael W Walker, Lejun Shao, and Richard A Volz, "Estimating 3-d location parameters using dual number quaternions," *CVGIP: image understanding*, vol. 54, no. 3, pp. 358–367, 1991.

[17] Roger Y Tsai and Reimar K Lenz, "Real time versatile robotics hand/eye calibration using 3d machine vision," in *Robotics and Automation, 1988. Proceedings., 1988 IEEE International Conference on*. IEEE, 1988, pp. 554–561.

[18] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

Author Biography

Longchuan Niu is a PhD student at TUT, Tampere, Finland, in the field of robotics and computer vision. He received his M.Sc. with distinction at TUT in 2000. After that he worked at Nokia R&D Finland as a senior software engineer until joining TUT in 2016.

Sergey Smirnov received his B.Sc. degree from Yaroslavl Demidov State University, Russia (2003), and M.Sc. degree from TUT, Finland (2010). His research interests include depth image-based rendering (DIBR), image analysis, and 3D reconstruction and visualization.

Jouni Mattila received his M.Sc. and Dr. Tech. degrees in 1995 and 2000, respectively, both from TUT, Tampere, Finland. He is currently a Professor in Machine Automation in the Laboratory of Automation and Hydraulics, TUT. His research interests include machine automation, developing nonlinear model-based control systems for robotic mobile manipulators and off-highway machinery, etc. He is currently a Technical Editor of the IEEE/ASME Transactions. on Mechatronics.

Atanas Gotchev is a professor of the 3D Media Group at the Laboratory of Signal Processing at TUT and serves as Director of the nationalwide research facility Centre for Immersive Visual Technologies (CIViT). He has broad competence of 3D imaging gathered as Chair of Research Exchange Committee of FP6 3DTV Network of Excellence, Scientific Coordinator of the FP7 project Mobile3DTV, and Project Manager of the FP7 Marie Curie IAPP Action PROLIGHT. His research expertise is in the areas of sampling and reconstruction of multi-dimensional signals; multi-sensor 3D scene sensing and reconstruction, and signal processing for ultra-realistic displays. He has co-authored about 170 scientific publications and has eight invention disclosures.

Emilio Ruiz Morales received his M.Sc. degree in 1990 from ULB, Brussels. He is currently a Senior Engineer and Technical Responsible Officer of several R&D projects at the Remote Handling Project Team/ITER delivery of the Fusion For Energy agency. His background expertise and research work are in control systems and surgical, nuclear and remote handling robotics.