KYBERNETIKA — VOLUME *62* (2026), NUMBER 1, PAGES 133–146

AUTOMATIC POSE ESTIMATION FROM RIGID AND PARTIAL OBJECT IMAGERY

Mehmet Akif Alper

3D pose estimation algorithms have been the subject of widely studied research topic due to problems related to their reliability and precision in related applications. Despite numerous studies by researchers to attempt efficient solutions to application related problems, many proposed methods still not submit sufficiently recover estimates for practical, real-world scenarios in the field of Computer Vision. Therefore, we made extensive study and presented an innovative and practical method that enables a cheap and practical solution by integrating information from both depth and color cameras. Outlier points can impact the pose estimations. We additionally implemented outlier rejection method due to outliers coming from depth to color point projection. After applying and evaluating our proposed algorithm on a public dataset for pose estimation problem, we have shown that it significantly enhances the robustness and accuracy of pose estimation in six degrees of freedom (6-DoF).

Keywords: point cloud, transformation, projection, rotation, translation

Classification: 68T10, 68T45

1. INTRODUCTION

Accurate scene understanding necessitates the detection and estimation of rigid object poses, defined by the spatial positions and orientations of target objects within a reference coordinate system. Pose estimation represents a complex and extensively investigated research domain, addressing the challenges of determining object pose, motion, and trajectory. Recent advancements in digital technologies have facilitated the acquisition of three-dimensional point cloud data from the environment, providing a robust foundation for enhancing pose estimation methodologies. Such algorithms have become fundamental to a broad spectrum of computer vision applications, including robotic localization, virtual reality, camera calibration, biomedical analysis, and human–machine interaction. [2, 7, 20, 40, 42]. Publicly available datasets typically provide complete object measurement models; however, real-world applications often require solving relative pose and motion under conditions where objects are only partially visible and surrounded by clutter, thereby limiting their appearance. This challenge is particularly prevalent in scenarios involving non-static cameras, where target objects are in motion, resulting in incomplete representations of both the model and the observed target. Currently,

DOI: 10.14736/kyb-2026-1-0133

there is no standardized method for defining similarity between model and target images. Three-dimensional pose estimation can be achieved by exploiting point cloud data of both model and target imagery, in conjunction with pre-calibrated camera systems, to compute the pose in six degrees of freedom (6-DoF). Despite numerous algorithms proposed in the literature, the problem remains unresolved due to issues such as partial observations, noise sensitivity, and cluttered environments.

The accuracy of relative pose estimation is of critical importance in real-world applications. Unlike absolute pose estimation, relative pose estimation does not require a predefined coordinate system origin or orientation for the tracked objects, but rather focuses on the spatial relationship between them. Despite significant advancements in modern sensor technologies [12], achieving robust and precise relative pose estimation remains a considerable challenge, particularly in dynamic and unstructured environments where factors such as occlusion, noise, and limited field of view reduce overall reliability. Therefore, the accuracy of pose estimation is strongly influenced by the density and quality of the object's point cloud data. In space robotics, for instance, autonomous systems must perform tasks such as active debris removal and object grasp–release operations, where precise pose estimation is critical. Similarly, in biomedical applications, micro-robotic arms employed in surgical procedures demand extremely high accuracy and sensitivity due to the complexity and risk inherent in medical interventions. In the domain of autonomous driving, vehicles must continuously estimate the relative pose of surrounding vehicles and pedestrians to ensure safe navigation and decision-making. Although state-of-the-art methods have demonstrated promising results through the use of synthetic datasets for training, their performance often degrades when applied to real-world industrial settings involving unseen object models. To address these limitations, our proposed method is designed to achieve highly accurate relative pose estimation, making it particularly valuable for error-sensitive industrial applications.

We conducted a systematic evaluation of the proposed algorithm, proposed algorithm, using a diverse set of objects from the publicly available BigBird [34] and RGBD object [19] datasets. The BigBird dataset provides a large-scale collection of images of everyday objects, along with corresponding ground-truth pose information, making it a widely used benchmark for pose estimation research. This data set enables a comprehensive assessment in a variety of categories and conditions of objects, allowing us to rigorously validate the performance and robustness of our approach, and the BigBird data set can be downloaded on the internet [34]. Those objects are located on a plate, have poses from 0 to 180 degree with respect to center part of the plate. BigBird dataset also supplies a ground-truth pose that has been scanned with color and depth images, meshes, and calibration parameters. Additionally, BigBird also provides high quality color images that have been acquired from the Canon camera. RGBD object dataset gives 300 household objects with groundtruth pose measures, and scene videos. RGBD object samples are collected from Kinect. We have applied test experiments for the proposed algorithm on scenarios that are subject to lower resolution imagery.

Rigid object pose estimation can be formulated as the process of identifying fine-grained correspondences between a model and a target object, where the transformation between their 3D point representations is quantified by a rotation matrix R and a translation vector T across three axes. Unlike many existing approaches, our method does

not rely on labeled training data or pre-defined CAD templates to refine rigid object poses. The proposed algorithm operates by leveraging point clouds extracted from low-resolution depth data in conjunction with optical flow information from color imagery, without the need for explicit image feature extraction. We integrated our algorithm into the FilterReg pose estimation framework [13], enhancing it into a more accurate pose estimation method. The contributions of this work can be summarized in two key aspects. First, the proposed algorithm is capable of operating on partial object imagery, thereby enabling applicability to real-world scenarios without requiring idealized, noise-free object models. Second, the proposed algorithm achieves superior performance compared to state-of-the-art matching techniques on pose datasets.

In the rest of our paper, we have explained related researches in Section 2, we explained proposed algorithm method at Section 3. We submitted quantitative results and error analysis at Section 4 and summarized our findings and future development and contributions at Section 5.

2. RELATED WORK

Pose estimation algorithms can be defined with three main categories: Template based methods, feature based methods, and machine learning based methods. There are different approaches to solve pose estimation depends on application and data. Some algorithms utilize depth image and computes pose by searching point data matches locally as Iterative Closest Point algorithm, other algorithms utilize from stereo RGB imagery and solve pose from global optimization methods as correspondence matching.

Template-based pose estimation methods rely on constructing 3D shape models of objects, often through scanning, to serve as reference templates. Among these, ICP algorithm and its variants are among the most extensively studied approaches [5, 27]. The ICP algorithm iteratively minimizes point-to-point distances between model and target clouds, converging to a local minimum. When initialized with suitable parameters and small transformations, ICP can achieve highly accurate estimates of the rotation (R) and translation (T) that align the two point sets. Chen et al. [9] extended this approach with the point-to-plane ICP, in which correspondences are determined by minimizing a plane-based error metric. While these methods can produce accurate pose estimates, their performance degrades in scenarios involving planar surfaces, non-Gaussian noise, or significant object clutter.

To overcome these limitations, numerous refinements of ICP have been proposed. Iversen et al. [37] introduced shape descriptors to reduce computational costs during matching. Presnov et al. [31] enhanced ICP by integrating wearable sensor measurements through an Extended Kalman Filter (EKF), while Aghili et al. [1] proposed a similar fusion method using an Adaptive Kalman Filter (AKF) for pose estimation in space robotics. Myronenko and Song [27] introduced the Coherent Point Drift (CPD) algorithm, a probabilistic method that formulates registration as aligning a Gaussian Mixture Model (GMM) to the target point set by maximizing the posterior probability. Building on CPD, Delavari et al. [10] incorporated mesh construction and additional model parameters for biomedical applications, while Liu et al. [25] developed a likelihood field extension to handle sparse and large-motion point sets.

Another notable family of methods is based on the Normal Distributions Transform

(NDT). Biber et al. [3] modeled point clouds as sets of normal distributions and maximized the density-based similarity measure. Subsequent enhancements have sought to improve accuracy and robustness: Hong et al. [16] refined NDT by truncating and combining Gaussian components, while Liu et al. [26] proposed clustering-based improvements using Gaussian and k-means clustering. Despite these advances, NDT-based approaches remain prone to convergence issues. Several researchers have also explored template-based methods using LIDAR and CAD models. Opromolla et al. [28] proposed a centroid-based similarity measure for LIDAR data in space robotics, while Picos et al. [29] employed correlation filters to estimate object pose from point data. Philips et al. [30] developed a maximum evidence method for the estimation of the excavator pose with LIDAR, achieving robust 6 DoF inference. CAD models, in particular, provide idealized and noiseless object representations that can significantly improve the accuracy of pose estimation. He et al. [15] introduced a CAD-based method extracting key points to refine 6-DoF poses via error minimization, while Tsai et al. [38] combined template matching with Perspective-n-Point (PnP) methods for efficient Augmented Reality applications. Song et al. [35] further advanced CAD-based pose estimation through depth filtering to remove outliers and randomly pick out bins from color imagery.

Feature-based pose estimation represents another extensively studied area within computer vision. These methods mainly rely on extracting distinctive features and descriptors from both model and target frames, which are expected to remain robust under image deformations. Object poses are then refined by minimizing a defined error metric or by combining strategies. Feature-based approaches can be broadly categorized into local and global methods. Accurate pose estimation using these methods typically requires sufficient texture on both the model and target objects. Several notable contributions have been made in this domain. Chen et al. [8] leveraged optical flow measurements to handle large object motions, refining poses through template warping and SIFT feature correspondences. Liu et al. [23] introduced a novel feature descriptor, P2P-TL, which models target appearance to reduce both computational time and pose estimation error. Teng et al. [36] proposed a method for adjusting aircraft pose based on extraction of line features and matching of correspondence. Quan et al. [32] developed a voxel-based binary descriptor for 3D object geometry, enabling fine registration through feature matching. Liu et al. [22] estimated pose by matching edge features of the image, although the precision of their method is highly dependent on the geometry of the object and the information about the edge. Contour-based approaches have also been extensively explored. Leng et al. [21] proposed iterative pose refinement using model and target contours extracted from grayscale images. Schlobohm et al. [33] introduced projected contour features, improving pose recovery through global optimization. Zhang et al. [43] combined object shape and contour information, detecting inliers, rejecting outliers, and computing target pose. Similarly, Wang et al. [39] applied particle filtering to contour and edge features, achieving robust pose estimation in cluttered scenes.

Nowadays, machine learning algorithms rapidly developing and presenting novel approaches on particularly for robotics and computer vision problems. Machine learning algorithms comprised of convolutional filters and learning functions that can be trained with labeled images and can produce robust pose estimates for dynamic environments and can overcome the limitation of manual design features. Machine learning based

methods generally segment the target and estimates pose of the foreground object. Zeng et al. [41] proposed a Convolutional Neural Network (CNN) for robot manipulators that is capable of picking and releasing objects. Le et al. [4] presented a CNN network that formerly segments object and recover pose of the target for robots. Brachmann et al. [6] proposes an algorithm that uses random forest for classification from images and estimates pose of the object. Kendall et al. [18] applied fine tuning on the GoogleNet and proposed a CNN that refines pose from RGB images. Hua et al. [17] developed a hour glass neural networks by adding residual modules to neural networks that finds pose. Giefer et al. [14] proposed a cascade connected CNN networks for object localization and pose recovery. Machine learning based methods are dependent to pre-labeled training data and can be produced poor results for different geometric shapes, large variances on the object models, and computation.

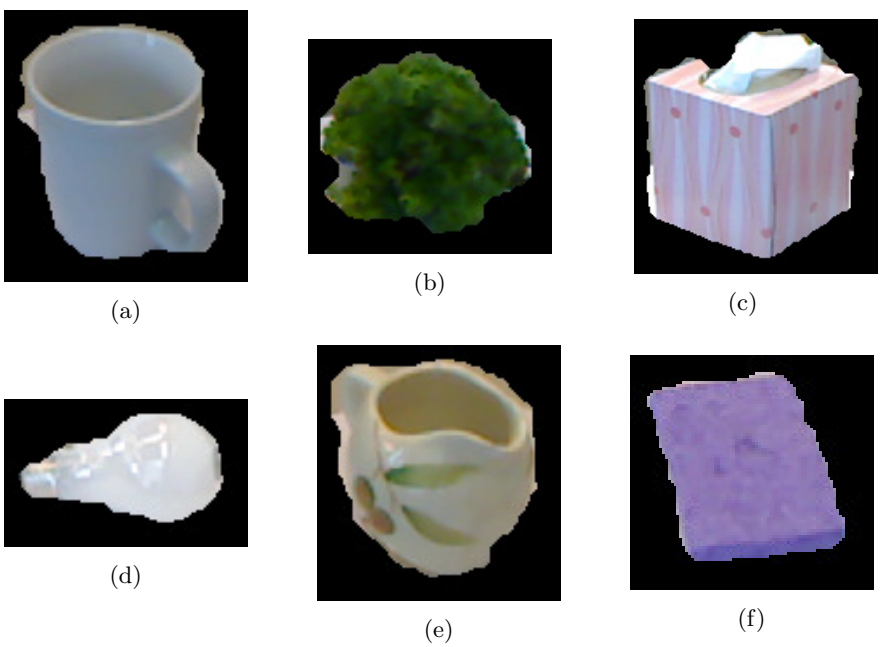

Fig. 1: Various images (mug, vegetable, tissue, light bulb, sticher, and sponge) are used from RGBD object dataset. Objects are segmented and sharp boundary found before computing the pose between model and target images.

3. METHOD

Our proposed algorithm operates in three main steps as detecting target object and finding pose with a series of processes.

3.1. Region of Interest (ROI) detection:

The proposed algorithm necessitates precise localization of the ROI corresponding to the target object, a task that can be addressed using neural network architectures. In this work, the Single Shot Video Object Detector (SSVD) [11] was used to identify ROI of objects, since the SSVD is a fast and automatic object detector that makes it applicable to the real-time object detector. The SSVD is comprised of three sub modules: pyramid network scales input image to different sizes, feature aggregator extracts features, and class/box subnet detects objects. The SSVD network was trained with test samples of the BigBird and RGB-D datasets, both of which provide objects captured under diverse poses with synchronized color and depth modalities as well as ground-truth pose annotations, see Figure 1. The BigBird dataset, publicly available, includes high-resolution 12-megapixel images, point clouds acquired via a Carmine depth sensor, and reconstructed meshes generated from fused point clouds, with ground-truth poses obtained through checkerboard calibration. Similarly, the RGB-D dataset, also publicly accessible, was collected using a Kinect camera, consisting of 640×480 pixel color images and depth frames at 30 Hz.

Although SSVD provides reliable detection of coarse object contours and boundary points, it often introduces outliers that decrease the accuracy of pose estimation. To mitigate this issue, the proposed algorithm integrated on optical flow measurements to refine boundary localization. Optical flow is estimated through a CNN-based method which uses warped image features [24]. By fusing optical flow with sparse depth information, the proposed algorith yields sharper object boundaries and substantially improves the precision of relative pose estimation.

3.2. Fuzzy Outlier Rejection

Object outliers are defined as point samples that deviate substantially from the overall spatial distribution of the point cloud of an object. These outliers typically emerge from sensor noise or inaccuracies in camera projection. Although object detectors generally delineate coarse object boundaries, such projections often introduce considerable noise, particularly around edges and corners, where depth values are expected to vary smoothly along the object surface. Consequently, the removal of outliers is essential for enhancing the reliability of pose estimation.

To mitigate this issue, we employed a fuzzy rule-based outlier detection framework for filtering anomalous points within the 3D point cloud. The fuzzy sets were constructed using inter-point distances as well as deviations from the mean and standard deviation of the distribution. In the first stage, points exhibiting pronounced differences relative to their neighbors were identified. Membership weights were then assigned according to their statistical deviation: points close to the main cluster were assigned low weights, points exceeding two standard deviations were assigned high weights, and intermediate cases received medium weights. Finally, a fuzzy clustering strategy was applied to classify outliers, wherein points located far from or outside the primary cluster were assigned a high membership degree for outliers.

3.3. Fusion and Pose Estimation

A rotation (R) is defined as the motion of an object about a fixed point, while a translation (T) denotes a displacement in which all points of the object are shifted by the same magnitude in the same direction, without any rotational component. The initial pose of a model is represented by (R_1, T_1), and the pose of the target object is expressed as (R_2, T_2), where the subscripts indicate their respective configurations. The relative pose of the target object, describing the transformation from the initial to the subsequent configuration, can thus be parameterized by (R_2, T_2).

Let the model point cloud be denoted as P_m and the target point cloud as P_t. These point sets are related through a relative transformation (R_1^2, T_1^2), see Eq. (1-2), and can be decomposed into rotational components—parameterized by Euler angles (yaw, pitch, and roll) and translational components along the Cartesian axes. Accordingly, the mapping of model points to target points is formulated in Eq. (3). Depth cameras provide 3D measurements (X, Y, Z), which can be projected into the 2D image plane (x, y) using a projection $Proj$, see (Eq. 4) where f_x and f_y denote focal length in x and y, c_x and c_y denote optical center in x and y for color camera. The optical flow between the model image (I_m) and the target image (I_t) is then computed using deep learning, where the matrix indices correspond to pixel coordinates in the optical flow field $I(x_f, y_f)$. 3D points and their 2D projections are available from Eq. (4) and optical flow computed from model and target imagery (x_f, y_f), so optical flow and depth correspondences can be associated using inverse projection $(Proj^{-1})$, thereby enabling the construction of a fused 3D point cloud for known depth by projecting the optical flow onto the depth data (Eq. 5).

A weighted average of the target points for depth is computed for depth (Eq. 5) and weights from depth points fused with optical flow correspondences, Eq. 6. Similarly, second order weighted average of the target points for depth is computed for depth (Eq. 7) and weights from depth points fused with optical flow correspondences, see Eq. 8. A normal distribution N is modeled over point correspondences (Eqs. 5-8).

Best alignment between model and target found by minimizing the objective function for depth, see Eq. 9. The proposed algorithm is employed to isolate the target object and to reconstruct 3D points with fusion of color and depth. Then, Optimization fuction (Eq. 9) turns to Eq. 10. Pose estimation is subsequently performed within an Expectation–Maximization (EM) framework. In the optimization step (Eq. 10), the motion parameters (θ) are estimated by minimizing Eq. 9 over θ where (X,Y,Z) fixed points, i denotes point cloud index and k denotes index of potential kinematic models. The algorithm aims to recover the Euler angles (roll, pitch, yaw) that align the model with the target pose. Finally, the accuracy of the estimated transformation is assessed using the root mean square (RMS) error between the predicted poses and the groundtruth measurements.

$$R_1^2 = R_2.R_1^{-1} \tag{1}$$

$$T_1^2 = T_2 - T_1 \tag{2}$$

	Algorithm	BigBird	RGBD Object
1	Proposed	**8.83**	15.93
2	FilterReg	**9.61**	17.87
3	CPD	**10.71**	19.33
4	Go-ICP	**9.37**	17.31

Tab. 1: RMS errors for pose estimations (angle) have been given for algorithms on BigBird and RGBD object datasets.

$$P_t = R_1^2.P_m + T_1^2 \tag{3}$$

$$Proj(X,Y,Z) = \left(\frac{X \cdot f_x}{Z} + c_x, \frac{Y \cdot f_y}{Z} + c_y\right) = (x,y) \tag{4}$$

$$M_{X_i}^0 = \textstyle\sum_{Y_k} N(X_i(\theta^{old}); Y_k \sum_{XYZ}) \tag{5}$$

$$\tilde{M}_{X_i}^0 = \alpha M_{X_i,\text{depth}}^0 + (1-\alpha) M_{X_i,\text{flow}}^0 \tag{6}$$

$$M_{X_i}^1 = \textstyle\sum_{Y_k} N(X_i(\theta^{old}); Y_k \sum_{XYZ}) Y_k \tag{7}$$

$$\tilde{M}_{X_i}^1 = \alpha M_{X_i,\text{depth}}^1 + (1-\alpha) M_{X_i,\text{flow}}^1 \tag{8}$$

$$J(\theta) = \sum_{X_i} \frac{M_{X_i}^0}{M_{X_i}^0 + c} \left(X_i(\theta) - \frac{M_{X_i}^1}{M_{X_i}^0}\right)^T \Sigma_{XYZ}^{-1} \left(X_i(\theta) - \frac{M_{X_i}^1}{M_{X_i}^0}\right) \tag{9}$$

$$J_{\text{fusion}}(\theta) = \sum_{X_i} \frac{\tilde{M}_{X_i}^0}{\tilde{M}_{X_i}^0 + c} \left(X_i(\theta) - \frac{\tilde{M}_{X_i}^1}{\tilde{M}_{X_i}^0}\right)^T \Sigma_{XYZ}^{-1} \left(X_i(\theta) - \frac{\tilde{M}_{X_i}^1}{\tilde{M}_{X_i}^0}\right) \tag{10}$$

The proposed algorithm addresses the challenge of object pose estimation by fusing point cloud data acquired from both depth sensing and optical flow estimation. Optical flow can be computed automatically from consecutive color image frames through image warping [24], enabling to detect small interframe displacements.

The proposed methodology establishes a correspondence between depth points and their associated RGB image coordinates. This association is made possible by the availability of intrinsic and extrinsic calibration parameters of depth and color cameras, which enable accurate projection of depth measurements into the color image domain. The mathematical formulation of these transformations is provided in Eq. (4)–(5).

Within the proposed framework, depth points are projected into a reference coordinate system defined by the imaging setup. To ensure robust pose estimation, these projected points must remain confined within the spatial extent of the object delineated in the RGB image. Leveraging camera calibration parameters, the algorithm aligns depth pixels with corresponding color image pixels, thereby facilitating consistent multimodal data fusion. This procedure is illustrated in Figure 2, with representative

Fig. 2: Sample box object is given, we need to project corresponding depth on RGB imagery. Interpolation helps to find unknown depth on RGB image.

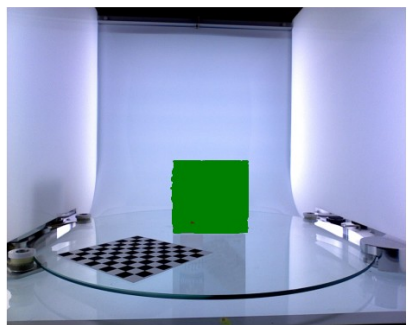

Fig. 3: Depth is projected and interpolated on RGB imagery, we eliminated edge depth points that can cause noise on point cloud.

dataset samples shown in Figure 3, and an example of depth points projected onto RGB imagery presented in Figure 4.

Comprehensive experiments were conducted to assess the effectiveness of the proposed approach. The evaluation results indicate that the proposed algorithm achieves superior pose estimation accuracy compared to existing state-of-the-art methods, as summarized in Figure 4.

3.4. Experiments

The proposed algorithm has been tested on public dataset termed BigBird and RGBD object dataset. After systematic tests on various objects and multiple pose variances, proposed has presented slightly higher pose refinement on partial objects that can be common on real world problems. Systematic tests has been completed by using 12 different objects where objects have different size, shape, and texture properties. Test objects are rotated from small to large pose changes (3-30 degrees). Since objects are rotated on single axes, we have reported results that quantified in terms of axis angle errors.

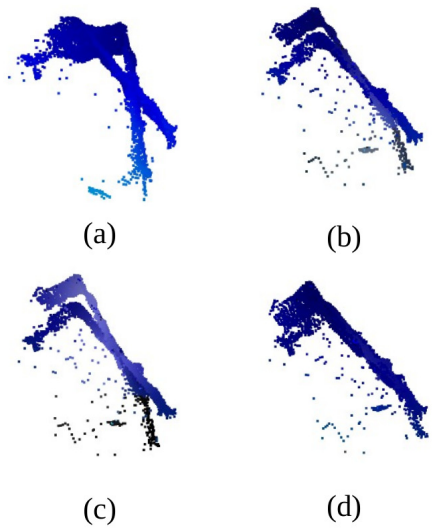

Fig. 4: Depth is projected and interpolated on RGB imagery, we eliminated edge depth points that can cause noise on point cloud. (a)initial point clouds, (b) results from FilterReg, (c) results from CPD, (d) results from proposed algorithm.

4. RESULTS

We have tested our proposed algorithm pose estimation algorithm to evaluate estimations by analyzing estimation results with ground truth poses from BigBird dataset. RMS results have been quantified for proposed algorithm, FilterReg, and CPD algorithms. It has been quantified and proposed algorithm computes pose of the rigid object in improved accuracy than FilterReg and CPD pose estimation. Instead of using plain depth or 2D color measurements of model and target, proposed algorithm enables higher accuracy pose estimations by combining of filtered depth measurements and corresponding optical flow from the CNN algorithm. Axis angle values has been utilized for error analysis. RMS values are computed and results have been reported in the next.

Overall, RMS values given in Table 1, proposed algorithm surpasses FilterReg, Go-ICP, and CPD respectively. Amplitude of RMS can be significantly affected with object size, shape, and texture properties of tracked objects. Object matching results have been given based on pose estimation results, see Figure 3.

5. CONCLUSION

Depth cameras generally provide limited resolution and are typically more expensive than conventional color cameras. As a result, pose estimation methods that rely exclusively on sparse depth measurements often suffer from reduced accuracy and robustness, which can be problematic in error-sensitive applications. By contrast, color cameras are comparatively inexpensive and offer higher spatial resolution, but they do not directly provide depth information. This motivates the fusion of sparse depth and color data to

enhance object detection and pose recovery.

Low-resolution depth cameras yield a sparse set of 3D points, and projection errors on tracked objects can introduce challenges in multimodal sensor fusion. In addition, the geometric shape and surface texture of an object strongly influence pose estimation accuracy, particularly in cases of small pose variations.

In this work, we propose a cost-effective and accurate pose estimation framework that integrates RGB and depth information. The algorithm employs pre-calibrated cameras with known extrinsic parameters and is designed to improve estimation accuracy for non-cylindrical 3D objects with textured surfaces. By fusing complementary RGB and depth measurements, the method rejects outliers, leading to robust and reliable pose estimation suitable for error-critical applications.

The proposed approach can be implemented on commercially available RGB-D sensors such as Microsoft Kinect, thereby enabling efficient and inexpensive pose estimation for indoor environments. Experimental evaluation using low-resolution depth and RGB data demonstrates that our proposed algorithm outperforms existing algorithms such as Coherent Point Drift (CPD) and FilterReg. Unlike CPD, which is prone to false correspondences due to local minima and sensitivity to outliers, the proposed algorithm achieves consistently lower pose estimation errors.

Furthermore, the approach is computationally efficient and applicable to real-time scenarios in dynamic environments. Potential use cases include robotic manipulation and mobile robotics, where accurate relative pose estimation is essential. Future work will extend this framework toward fully real-time implementations and deployment in mobile robot platforms for indoor navigation and manipulation tasks.

ACKNOWLEDGEMENT

This research partially supported by Cankiri Karatekin University.

(Received September 7, 2025)

REFERENCES

[1] F. Aghili and Ch.-Y. Su: Robust relative navigation by integration of ICP and adaptive Kalman filter using laser scanner and IMU. IEEE/ASME Trans. Mechatronics *2+* (2016), 2015–2026. DOI:10.1109/tmech.2016.2547905

[2] M. A. Alper, J. Goudreau, and M. Daniel: Pose and Optical Flow Fusion (POFF) for accurate tremor detection and quantification. Biocybernet. Biomedical Engrg. *40* (2020), 468–481. DOI:10.1016/j.bbe.2020.01.009

[3] P. Biber and W. Strasser: The normal distributions transform: a new approach to laser scan matching. In: Proc.2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453). DOI:10.1109/iros.2003.1249285

[4] T. Le, L. Hamilton and A. Torralba: Benchmarking convolutional neural networks for object segmentation and pose estimation. On: IEEE Conference Publication, IEEE Applied Imagery Pattern Recognition Workshop 2017. DOI:10.1109/AIPR.2017.8457942

[5] G. Blais and M. D. Levine: Registering multiview range data to create 3D computer objects. IEEE Trans. Pattern Analysis Machine Intell. *17* (1995), 820–824. DOI:10.1109/34.400574

[6] E. Brachmann, F. Michel, A. Krull, M. J. Yang, S. Gumhold, and C. Rother, Carsten: Uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image. IEEE Xplore (2016), 3364–3372. DOI:10.1109/CVPR.2016.366

[7] M. Brunot, A. Janot, P. C. Young, and F. Carrillo: An instrumental variable method for robot identification based on time variable parameter estimation. Kybernetika *54* (2018), 1. DOI:10.14736/kyb-2018-1-0202

[8] S. Chen, L. Liang, W. Luming, and H. Foroosh: 3D pose tracking with multi-template warping and SIFT correspondences. IEEE Trans. Circuits Systems Video Technol. (2015). DOI:10.1109/tcsvt.2015.2452782

[9] Y. Chen and G. Medioni: Object modelling by registration of multiple range images. Image Vision Comput. *10* (1992), 145–155. DOI:10.1016/0262-8856(92)90066-c

[10] M. Delavari, A. H. Foruzan, and Y.-W. Chen: Accurate point correspondences using a modified coherent point drift algorithm. Biomedical Signal Process. Control *52* (2019), 429–444. DOI:10.1016/j.bspc.2017.02.009

[11] J. Deng, Y. Pan, T. Yao, W. Zhou, H. Li, and T. Mei: Single shot video object detector. IEEE Trans. Multimedia *23* (2021), 846–858. DOI:10.1109/TMM.2020.2990070

[12] D. Ding, R. A. Cooper, P. F. Pasquina, and L. Fici-Pasquina: Sensor technology for smart homes. Maturitas *69* (2011), 131–136. DOI:10.1016/j.maturitas.2011.03.016

[13] W. Gao and R. Tedrake: FilterReg: Robust and Efficient Probabilistic Point-Set Registration Using Gaussian Filter and Twist Parameterization. Massachusetts Institute of Technology, 2019. DOI:10.1109/cvpr.2019.01135

[14] L. A. Giefer, J. D. Castellanos, M. M. Babr, and M. Freitag: Deep learning-based pose estimation of apples for inspection in logistic centers using single-perspective imaging. Processes *7* (2019), 424. DOI:10.3390/pr7070424

[15] Z. He, Z. Jiang, X. Zhao, S. Zhang, and Ch. Wu: Sparse template-based 6-D pose estimation of metal parts using a monocular camera. IEEE Trans. Industrial Electronics *67* (2020), 390–401. DOI:10.1109/tie.2019.2897539

[16] H. K. Hong, H. Yu, and B.-H. Lee: Regeneration of normal distributions transform for target lattice based on fusion of truncated Gaussian components. IEEE Robotics Automation Lett. *4* (2019), 684–691. DOI:10.1109/lra.2019.2891493

[17] G. Hua, and L. Li, and S. Liu: Multipath affinage stacked—hourglass networks for human pose estimation. Frontiers Computer Sci. *14* (2020). DOI:10.1007/s11704-019-8266-2

[18] A. Kendall, M. Grimes, and R. Cipolla: PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In: 2015 IEEE International Conference on Computer Vision (ICCV). DOI:10.1109/iccv.2015.336

[19] L. Kevin, B. Liefeng, and D. Fox: Unsupervised feature learning for 3D scene labeling. In: IEEE International Conference on Robotics and Automation (ICRA), 2014.

[20] N. A. Kilyen, R. F. Lemnariu, I. Muntean, and G. D. Mois: An Indoor Localization System for Automotive Driving Competitions, Int. J. Computers Commun. Control *19* (2024), 1. DOI:10.15837/ijccc.2024.1.6030

[21] D. W. Leng and W. D. Sun: Contour-based iterative pose estimation of 3D rigid object. IET Computer Vision *5* (2011), 291. DOI:10.1049/iet-cvi.2010.0098

[22] M. Li, T. Oncel, V. Ashok, and Ch. Rama: Fast directional chamfer matching. Pennsylvania State University, 2010. DOI:10.1109/cvpr.2010.5539837

[23] D. Liu, S. Arai, J. Miao, J.Kinugawa, Z. Wang, and K. Kosuge: Point pair feature-based pose estimation with multiple edge appearance models (PPF-MEAM) for robotic bin picking. Sensors *18* (2018), 2719. DOI:10.3390/s18082719

[24] P. Liu, M. R. Lyu, I. King, and J. Xu: SelFlow: Self-supervised learning of optical flow. CVPR (2019). DOI:10.1109/cvpr.2019.00470

[25] K. Liu, W. Wang, K. Thiagalingam, and J. Wang: Dynamic vehicle detection with sparse point clouds based on PE-CPD. Institute of Electrical and Electronics Engineers *20* (2019), 1964–1977. DOI:10.1109/tits.2018.2857510

[26] T. Liu, J. Zheng, Z. Wang, Z. Huang, and Y. Chen: Composite clustering normal distribution transform algorithm. Int. J. Advanced Robotic Systems, SAGE Publish. *17* (2020). DOI:10.1177/1729881420912142

[27] A. Myronenko and S. Xubo: Point Set Registration: Coherent Point Drift. IEEE Trans. Pattern Anal. Machine Intell. *32* (2010), 2262–2275. DOI:10.1109/tpami.2010.46

[28] R. Opromolla, G. Fasano, G. Rufino, and M. Grassi: A model-based 3D template matching technique for pose acquisition of an uncooperative space object. Sensors *15* (2015), 6260–6382. DOI:10.3390/s150306360

[29] K. Picos, V. H. Diaz-Ramirez, V. Kober, A. S. Montemayor, and J. J. Pantrigo: Accurate three-dimensional pose recognition from monocular images using template matched filtering. Optical Engrg. *55* (2016). DOI:10.1117/1.oe.55.6.063102

[30] T. G. Phillips and P. R. McAree: An evidence-based approach to object pose estimation from LiDAR measurements in challenging environments. J. Field Robotics *35* (2018), 921–936. DOI:10.1002/rob.21788

[31] D. Presnov, M. Lambers, and A. Kolb: Robust range camera pose estimation for mobile online scene reconstruction. IEEE Sensors J. *18* (2018), 2903–2915. DOI:10.1109/jsen.2018.2801878

[32] S. Quan, J. Ma, Jie, F. Hu, B. Fang, and T. Ma: Local voxelized structure for 3D binary feature representation and robust registration of point clouds from low-cost sensors. Inform. Sci. *444* (2018), 153–171. DOI:10.1016/j.ins.2018.02.070

[33] J. Schlobohm, A. Pösch, E. Reithmeier, and B. Rosenhahn: Improving contour based pose estimation for fast 3D measurement of free form objects. Measurement *92* (2016), 79–82. DOI:10.1016/j.measurement.2016.05.093

[34] A. Singh, J. Sha, K. S. Narayan, S. Karthik, T. Achim, and A. Pieter: BigBIRD: A large-scale 3D database of object instances. In: International Conference on Robotics and Automation, 2014. DOI:10.1109/icra.2014.6906903

[35] K.-T. Song, Ch.-H. Wu, and S.-Y. Jiang: CAD-based Pose estimation design for random bin picking using a RGB-D camera. J. Intell. Robotic Systems *87* (2017), 455–470. DOI:10.1007/s10846-017-0501-1

[36] X. Teng, Q. Yu, Qifeng, J. Luo, G. Wang, and X. Zhang: Aircraft pose estimation based on geometry structure features and line correspondences. Sensors *19* (2019), 2165. DOI:10.3390/s19092165

[37] M. I. Thorbjorn G. B. Anders, N. Krüger, and D. Kraft: Shape dependency of ICP pose uncertainties in the context of pose estimation systems. Lecture Notes Computer Sci. (2015), 303–315. DOI:10.1007/978-3-319-20904-3-28

[38] Ch.-Y. Tsai, K.-J. Hsu, and H. Nisar: Efficient model-based object pose estimation based on multi-template tracking and PnP algorithms. Algorithms *11* (2018), 122. DOI:10.3390/a11080122

[39] B. Wang, F. Zhong, and X. Qin: Robust edge-based 3D object tracking with direction-based pose validation. Multimedia Tools Appl. *78* (2018), 12307–12331. DOI:10.1007/s11042-018-6727-5

[40] T. Yang, Q. Zhao, X. Wang, and Q. Zhou: Sub-pixel chessboard corner localization for camera calibration and pose estimation. Appl. Sci. *8* (2018), 2118. DOI:10.3390/app8112118

[41] A. Zeng, K.-T. Yu, S. Song, D. Suo, E. Walker, A. Rodriguez, and J. Xiao: Multi-view self-supervised deep learning for 6D pose estimation in the Amazon Picking Challenge. IEEE Xplore (2017), 1386–1383. DOI:10.1109/ICRA.2017.7989165

[42] R. Zhao, H. Ali, and P. van der Smagt: Two-stream RNN/CNN for action recognition in 3D videos. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017. DOI:10.1109/iros.2017.8206288

[43] X. Zhang, Z. Jiang, H. Zhang, and Q. Wei: Vision-based pose estimation for textureless space objects by contour points matching. IEEE Trans. Aerospace Electron. Systems *54* (2018), 2342–2355. DOI:10.1109/taes.2018.2815879

Mehmet Akif Alper, complete postal addresses of author
e-mail: mehmetakifalper@karatekin.edu.tr