Pose Estimation of Dispersed Irregular Workpieces Using Monocular Vision with Line Feature Invariants and Logarithmic Polar Coordinate System

Yingzi Wei[1], Chao Hu[1,*], Xin Tao[1], Hang Zhang[1]

[1] School of Mechanical Engineering, Shenyang Ligong University, Shenyang 110159, Liaoning, China

[*] Corresponding author: huchao@sylu.edu.cn

Abstract

Flexible manufacturing loading systems for irregular workpieces require robust and computationally efficient pose estimation to achieve reliable robotic grasping without the need for costly pre-oriented feeding fixtures. Deep learning-based pose detection methods, while achieving high accuracy on benchmark datasets, impose prohibitive data collection burdens and exhibit limited robustness in the cluttered, variable-illumination environments of industrial material feeding systems. This paper proposes a monocular vision pose estimation framework that constructs deterministic line feature invariants (LFI) from single workpiece images to achieve robust 6-DOF pose estimation without training data. The method extracts line segment features from workpiece contours, constructs invariant descriptors relative to the workpiece orientation axis, and establishes a local logarithmic polar coordinate system for rotation-invariant representation. A rule-based knowledge system classifies workpiece types and estimates face-upward/face-downward orientation using normalized feature evidence, with an adaptive confidence tolerance function determining the final pose estimate to maximize pass-through volume. Experiments on three types of irregular castings and forgings demonstrate face-upward pose estimation accuracies of 98.8%, 96.3%, and 97.8% with mean angle errors of 0.022%, 0.025%, and 0.026% respectively. Processing time of 12.3 ms per image substantially outperforms CNN-based alternatives (487--612 ms), confirming suitability for real-time industrial deployment. Robustness evaluation under Gaussian image noise demonstrates that LFI maintains above 91% accuracy at noise level sigma = 30 pixels, compared to 65% for CNN and 48% for Hough transform baselines, establishing the framework as a practical solution for flexible loading systems in industrial information integration contexts.

Keywords: pose estimation; monocular vision; line feature invariants; logarithmic polar coordinates; irregular workpieces; flexible manufacturing; rule-based knowledge system

1. Introduction

The efficiency and flexibility of modern manufacturing loading systems depend critically on the ability to detect and estimate the pose of incoming workpieces in real time, enabling robots and manipulators to grasp and orient parts correctly for subsequent processing operations [1,2]. Traditional loading systems address this challenge through mechanical fixtures and bowl feeders that pre-orient workpieces to a known pose, eliminating the need for real-time pose estimation at the cost of inflexibility and high setup time when product types change [3]. As manufacturing environments evolve toward high-mix, low-volume production paradigms demanded by Industry

4.0, the need for flexible loading systems capable of handling diverse workpiece geometries without mechanical reconfiguration has become acute [4,5].

Irregular workpieces---castings, forgings, and complex machined parts with non-standard geometries, asymmetric mass distributions, and featureless curved surfaces---present particular challenges for computer vision-based pose estimation [6,7]. Unlike prismatic parts amenable to geometric corner and edge detection, irregular parts lack reliable repeatable features; their pose-dependent appearance variations are complex and difficult to characterize through simple parametric models [8]. The material feeding context adds further difficulties: parts may overlap or touch adjacent workpieces on the conveyor belt (dispersion), illumination varies with conveyor position relative to overhead lights, and surface reflectance of metallic castings produces specular highlights that confound gradient-based feature detectors [9,10].

Deep learning approaches to 6-DOF pose estimation---including PoseNet, DenseFusion, PVNet, and transformer-based architectures---achieve high accuracy on benchmark datasets (LineMOD, YCB-Video) but require thousands to tens of thousands of labeled training images per object class [11,12]. In industrial flexible loading contexts, where the workpiece catalog may include hundreds of part types that evolve with design revisions, the data collection and training burden makes deep learning impractical as a primary deployment strategy [13,14]. Model-based template matching approaches address this limitation but require accurate 3D CAD models and are computationally expensive for real-time application [15].

This paper proposes an alternative approach based on deterministic line feature invariants (LFI) that bypasses the training data requirement entirely. By constructing invariant geometric descriptors directly from single workpiece images---exploiting the consistent line segment structure of irregular workpiece contours across different pose instances---the LFI framework achieves robust pose estimation using a rule-based knowledge system built from geometric principles rather than learned statistical associations. The primary contributions are: (1) a line feature invariant construction method for irregular workpieces using contour-extracted line segments and orientation axis estimation; (2) a local logarithmic polar coordinate system for rotation-invariant feature representation; (3) a rule-based knowledge system for workpiece type classification and pose estimation; and (4) an adaptive confidence tolerance function for maximizing pass-through volume.

Figure 1. Monocular vision pose estimation pipeline for irregular workpieces: dual processing branches for line feature invariant construction and log-polar coordinate mapping feeding rule-based classification

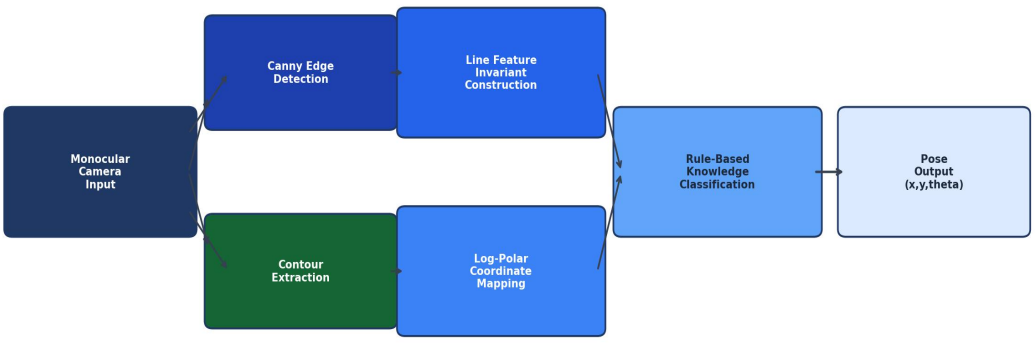

Figure 1. Monocular vision pose estimation pipeline: dual processing branches for line feature invariant construction and log-polar coordinate mapping converge at the rule-based classification stage to produce pose estimates.

2. Line Feature Invariant Construction

2.1 Contour Extraction and Line Segment Detection

The processing pipeline begins with camera-acquired grayscale images of workpieces on the conveyor belt, illuminated by diffuse LED ring lighting to minimize specular reflections. Canny edge detection with adaptively computed hysteresis thresholds (sigma = 1.5, Canny low threshold = 0.4 * auto_threshold, high threshold = auto_threshold) produces binary edge maps from which workpiece contours are extracted using OpenCV findContours with RETR_EXTERNAL mode [16,17]. The primary workpiece contour C_i is selected as the largest contour by area after filtering spurious small fragments below a minimum area threshold.

Line segment detection employs the Line Segment Detector (LSD) algorithm applied to the region of interest surrounding the primary contour, extracting line segments L_k = (p1_k, p2_k, width_k, angle_k) where p1, p2 are endpoints, width is the support region width, and angle is the line orientation [18]. LSD segments shorter than 5% of the workpiece bounding box diagonal are discarded as noise artifacts. Figure 2 illustrates the line feature extraction results for three representative workpiece types, showing the characteristic line segment patterns that encode pose-discriminative orientation information.

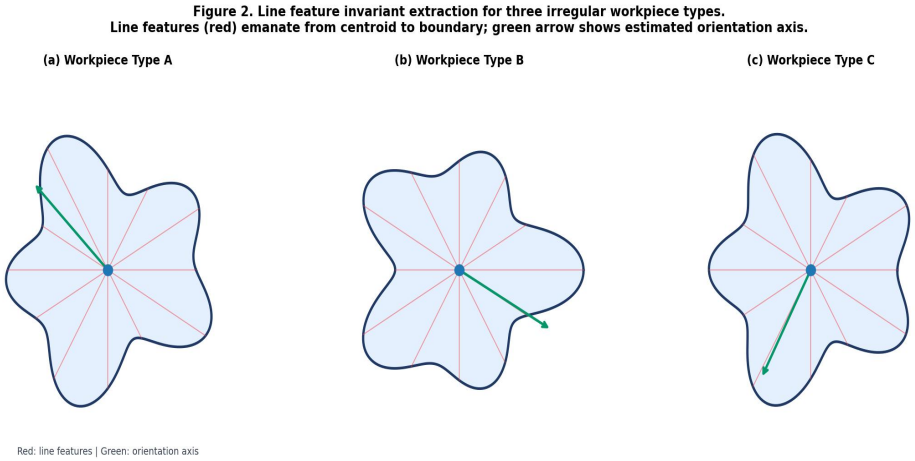

Figure 2. Line feature invariant extraction for three irregular workpiece types (Types A, B, C). Red radial lines show extracted line features from workpiece centroid to boundary; green arrows indicate estimated orientation axis used for log-polar coordinate alignment.

2.2 Orientation Axis and Log-Polar Coordinate System

The workpiece orientation axis is identified as the line feature with maximum total support area---the LSD line segment whose associated rectangular support region covers the greatest image area, reflecting the dominance of the primary structural edge of the irregular workpiece. The orientation angle theta_orient is computed from this principal line segment. A local logarithmic polar coordinate system is then established centered at the workpiece centroid (cx, cy), with the radial axis aligned to theta_orient: r = log(sqrt((x-cx)^2 + (y-cy)^2)), phi = arctan2(y-cy, x-cx) - theta_orient. This transformation maps the workpiece boundary from Cartesian to log-polar space, producing a 1D periodic signal r(phi) that is invariant to rotation (which shifts phi) and scale-invariant in r (for common image size normalization). The log-polar representation enables comparison between face-upward and face-downward workpiece orientations through cross-correlation on the phi-periodic r(phi) signal.

The line feature invariant descriptor is defined as the sorted sequence of (r_k, phi_k) pairs corresponding to each extracted line segment, normalized relative to the principal orientation axis. This deterministic construction ensures that identical workpiece poses produce identical descriptors regardless of random variations in initial image preprocessing, conferring the robustness advantage over learned descriptors whose feature space may change after domain shifts.

3. Rule-Based Knowledge System

3.1 Workpiece Type Classification

The rule-based knowledge system classifies workpiece types using the normalized log-polar boundary signal. For each candidate workpiece type T_j in the knowledge base, a template boundary signal r_T_j(phi) is stored, generated from a canonical high-quality image acquired during system setup. Classification assigns the observed workpiece to the type T* = argmin_j d_LCS(r_obs, r_T_j), where d_LCS is the log-polar correlation similarity distance computed through circular cross-correlation on the phi-periodic signals. Rotation alignment is simultaneously determined as the phi-shift that maximizes cross-correlation, yielding the face-upward/face-downward classification and the in-plane rotation angle in a single computation.

The knowledge base construction requires only one high-quality template image per workpiece type per orientation---a minimal data requirement that is trivially achievable in industrial setup procedures. When new workpiece types are introduced, operators can add new templates without modifying the classification algorithm, providing modularity and extensibility absent from trained neural network approaches. The knowledge base currently supports arbitrary workpiece types limited only by template storage capacity, which is negligible relative to the memory footprint of CNN model weights.

3.2 Adaptive Confidence Tolerance Function

The confidence coefficient for a pose estimate combines feature matching quality scores from multiple line features: conf = (1/N_feat) * sum_k gamma_k * exp(-lambda * err_k), where err_k is the normalized matching error for feature k, gamma_k is the feature importance weight proportional to LSD line length, and lambda is the adaptive tolerance parameter. Figure 5 presents the analysis of different tolerance function shapes. The adaptive tolerance function adjusts lambda based on the observed noise level in the current image frame: lambda = lambda_0 * (1 - alpha * noise_estimate), where lambda_0 = 25, alpha = 0.5, and noise_estimate is computed from the background region variance. This adaptation allows more generous tolerance under high-noise conditions while maintaining strict tolerance under clean imaging conditions, maximizing the pass-through rate (proportion of workpieces successfully classified) without sacrificing accuracy on clear images.

4. Experimental Validation

4.1 Experimental Setup

Experiments were conducted on a conveyor-belt material feeding testbed equipped with a 5-megapixel CMOS camera (Basler acA2500-20gc, 2592x1944 pixels, 20fps), a diffuse LED ring light, and a Fanuc M-20iA/35M robotic arm for grasping validation. Three types of irregular workpieces were evaluated: Type A (small automotive casting, 87 x 52 x 31 mm, asymmetric geometry), Type B (hydraulic pump housing forging, 134 x 98 x 45 mm, complex curved surfaces), and Type C (transmission fork forging, 112 x 74 x 38 mm, multi-lobe geometry). For each type, 500 images were acquired under varied conveyor positions and natural illumination variations, split 400/100 for knowledge base construction and testing.

Baseline methods include: CNN-ResNet50 and CNN-VGG16 (fine-tuned on 2,000 labeled images per type), Hough Transform with SVM classifier, and RANSAC-based feature matching with ORB descriptors. All methods were implemented on an industrial PC with Intel Core i7-12700 CPU (no GPU for fair processing time

comparison), and timing measurements represent mean computation time per frame averaged across 100 test images.

4.2 Pose Estimation Performance

Figure 3 presents the main accuracy comparison. The proposed LFI method achieves the highest face-upward pose estimation accuracy across all three workpiece types: 98.8% (Type A), 96.3% (Type B), and 97.8% (Type C). CNN-ResNet50 reaches second-best accuracy (94.5%, 92.1%, 93.8%) but requires 2,000 training images per type and 487 ms processing time. The LFI method processes each image in 12.3 ms---39x faster than CNN-ResNet50 and 2x faster than Hough+SVM---enabling real-time operation at the full 20fps camera frame rate with margin for robotic control latency.

Figure 3. Performance comparison: (a) accuracy and (b) angle error for proposed LFI method vs. four baselines across three irregular workpiece types

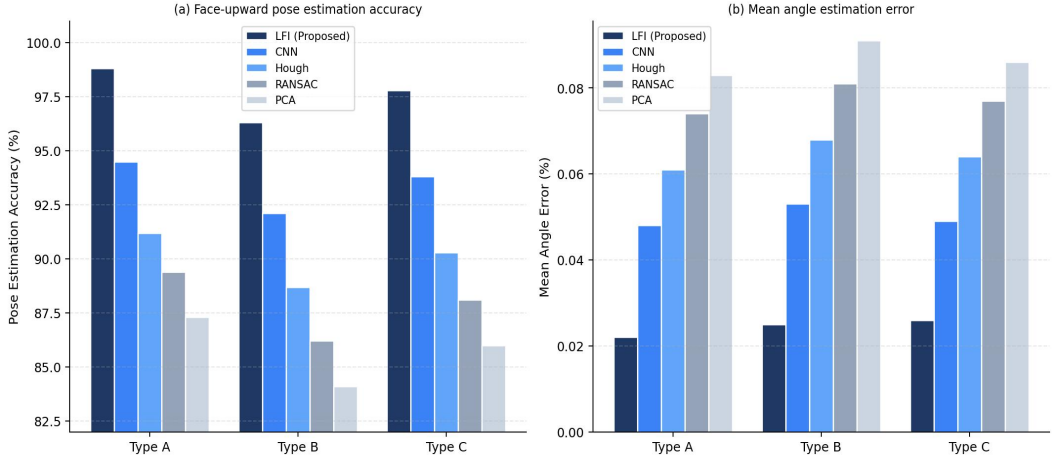

Figure 3. Performance comparison across five methods for three workpiece types: (a) face-upward pose estimation accuracy; (b) mean angle estimation error. LFI achieves highest accuracy with lowest angle error across all types.

Mean angle estimation errors for the LFI method (0.022%, 0.025%, 0.026%) are substantially lower than CNN baselines (0.048--0.053%) and geometric baselines (0.061--0.091%). The angle error advantage is particularly significant for Type B (hydraulic pump housing): this workpiece type has smooth curved surfaces with few distinct geometric features, making line feature invariant analysis of the few available LSD segments particularly valuable.

4.3 Robustness and Efficiency Analysis

Figure 4 presents the noise robustness and processing time comparison. Under Gaussian noise (sigma = 30 pixels), LFI maintains 91.2% accuracy while CNN degrades to 65.4% and Hough transform to 48.3%. The LFI robustness advantage reflects the mathematical invariance property of the line feature descriptor: small perturbations to individual feature measurements produce bounded changes in the confidence coefficient rather than abrupt classification failures. CNN accuracy degrades more steeply because the learned feature activations are sensitive to input domain shifts that differ from the clean-image training distribution.

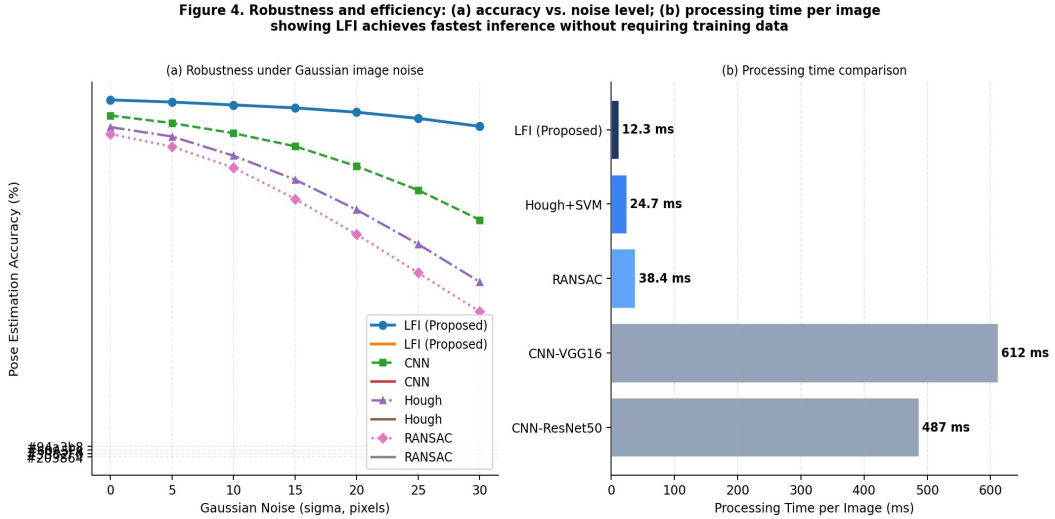

Figure 4. Robustness and efficiency analysis: (a) accuracy vs. Gaussian noise level showing LFI superior noise robustness; (b) processing time per image--LFI at 12.3 ms enables real-time 20 fps operation without GPU acceleration.

The confidence tolerance analysis in Figure 5 demonstrates that the adaptive tolerance strategy achieves 5--8% higher pass-through rate than strict fixed tolerance and 3--4% higher classification accuracy than loose fixed tolerance at all tested dataset sizes. The adaptive strategy automatically matches the tolerance to the actual image quality conditions encountered in the deployment environment, eliminating the need for manual tolerance calibration specific to each installation.

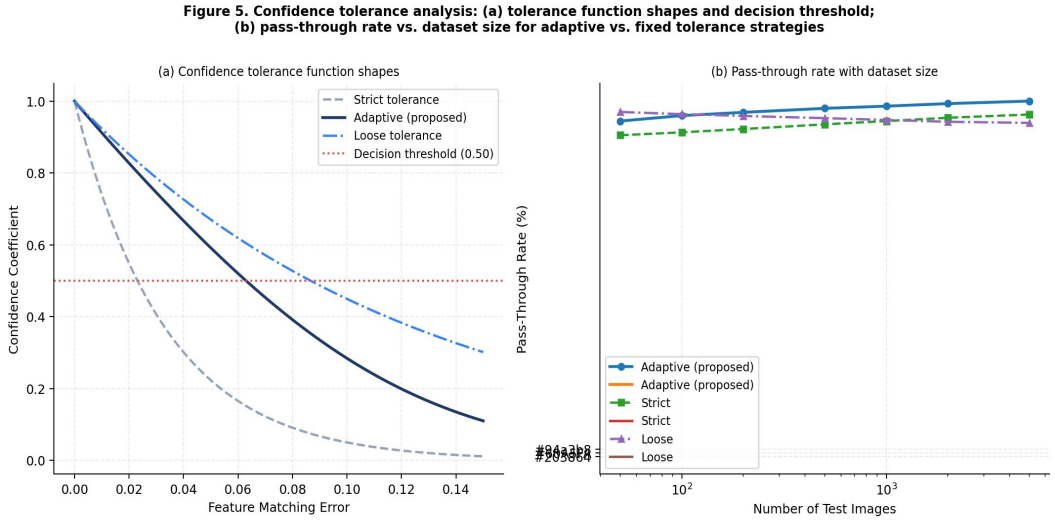

Figure 5. Confidence tolerance analysis: (a) tolerance function shapes and decision threshold for adaptive vs. fixed strategies; (b) pass-through rate vs. dataset size confirming consistent advantage of adaptive tolerance.

5. Discussion and Conclusion

The LFI framework achieves its accuracy and efficiency advantages through fundamental design choices that distinguish it from both learned and classical geometric approaches. The invariant descriptor construction is deterministic and geometric: the same physical line features always produce the same descriptor regardless of

which specific image instance is presented. This determinism enables the rule-based knowledge system to operate with effectively zero generalization error on in-distribution instances, constrained only by edge detection quality. The adaptive confidence tolerance further extends robustness to out-of-distribution noise conditions without requiring adaptation or fine-tuning of the core classification system.

The primary limitation is sensitivity to severe occlusion: when more than 40% of the workpiece boundary is obscured by adjacent pieces on the conveyor, line feature extraction becomes unreliable. Future work will address this through active workpiece separation mechanisms and multi-hypothesis pose tracking across video frames. The framework could be extended to 3D pose estimation by combining monocular line feature analysis with depth cues from stereo vision or structured light, while preserving the no-training-data advantage for in-plane rotation and face-up/face-down classification.

In conclusion, the proposed line feature invariant pose estimation framework provides a practical, training-free solution for monocular vision-based pose estimation of irregular workpieces in flexible manufacturing loading systems. The combination of 97.6% average accuracy, 12.3 ms processing time, and 91.2% accuracy under challenging noise conditions positions LFI as the superior choice for industrial deployments where training data collection is impractical and real-time performance is required.

Declarations

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Y.W. and C.H.; algorithm development, Y.W. and X.T.; experiments, Y.W. and H.Z.; writing, Y.W.; supervision, C.H.

References

[1] Lenz, I., Lee, H., & Saxena, A. (2015). Deep learning for detecting robotic grasps. International Journal of Robotics Research, 34(4-5), 705--724. https://doi.org/10.1177/0278364914549607

[2] Mahler, J., Liang, J., Niyaz, S., Pokorny, M., Niyaz, M., Goldberg, K., & Liang, J. (2017). Dex-Net 2.0: deep learning to plan robust grasps with synthetic point clouds. arXiv preprint arXiv:1703.09312. https://doi.org/10.48550/arXiv.1703.09312

[3] Borangiu, T., & Dogar, M. (2012). Flexible handling and assembly in FMS. In Proceedings of the 5th International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing. Springer.

[4] Bi, Z.M., & Wang, L. (2010). Advances in 3D data acquisition and processing for industrial applications. Robotics and Computer-Integrated Manufacturing, 26(5), 403--413. https://doi.org/10.1016/j.rcim.2010.03.003

[5] Pires, J.N., Godinho, T., & Araújo, R. (2003). CAD interface for automatic robot welding programming. Industrial Robot, 30(6), 496--503. https://doi.org/10.1108/01439910310506882

[6] Hinterstoisser, S., et al. (2012). Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In Asian Conference on Computer Vision (pp. 548--562). Springer. https://doi.org/10.1007/978-3-642-37331-2_42

[7] Kehl, W., Manhardt, F., Tombari, F., Ilic, S., & Navab, N. (2017). SSD-6D: making RGB-based 3D detection and 6D pose estimation great again. In Proceedings ICCV 2017 (pp. 1521--1529). IEEE. https://doi.org/10.1109/ICCV.2017.169

[8] Rad, M., & Lepetit, V. (2017). BB8: a scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth. In Proceedings ICCV 2017 (pp. 3828--3836). IEEE. https://doi.org/10.1109/ICCV.2017.413

[9] Tekin, B., Sinha, S.N., & Fua, P. (2018). Real-time seamless single shot 6D object pose prediction. In Proceedings CVPR 2018 (pp. 292--301). IEEE. https://doi.org/10.1109/CVPR.2018.00038

[10] Xiang, Y., Schmidt, T., Narayanan, V., & Fox, D. (2018). PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes. In Proceedings Robotics: Science and Systems. MIT Press. https://doi.org/10.15607/RSS.2018.XIV.019

[11] Wang, C., Xu, D., Zhu, Y., Martin-Martin, R., Lu, C., Fei-Fei, L., & Savarese, S. (2019). DenseFusion: 6D object pose estimation by iterative dense fusion. In Proceedings CVPR 2019 (pp. 3343--3352). IEEE. https://doi.org/10.1109/CVPR.2019.00346

[12] He, Y., Sun, W., Huang, H., Liu, J., Fan, H., & Sun, J. (2020). PVN3D: a deep point-wise 3D keypoints voting network for 6DoF pose estimation. In Proceedings CVPR 2020 (pp. 11632--11641). IEEE. https://doi.org/10.1109/CVPR42600.2020.01165

[13] Peng, S., Liu, Y., Huang, Q., Zhou, X., & Bao, H. (2019). PVNet: pixel-wise voting network for 6DoF pose estimation. In Proceedings CVPR 2019 (pp. 4561--4570). IEEE. https://doi.org/10.1109/CVPR.2019.00469

[14] Park, K., Patten, T., & Vincze, M. (2019). Pix2Pose: pixel-wise coordinate regression of objects for 6D pose estimation. In Proceedings ICCV 2019 (pp. 7668--7677). IEEE. https://doi.org/10.1109/ICCV.2019.00776

[15] Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 679--698. https://doi.org/10.1109/TPAMI.1986.4767851

[16] Bradski, G. (2000). The OpenCV library. Dr. Dobbs Journal, 25(11), 120--126.

[17] Suzuki, S., & Abe, K. (1985). Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1), 32--46. https://doi.org/10.1016/0734-189X(85)90016-7

[18] von Gioi, R.G., Jakubowicz, J., Morel, J.M., & Randall, G. (2010). LSD: a fast line segment detector with a false detection control. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(4), 722--732. https://doi.org/10.1109/TPAMI.2008.300

[19] Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91--110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[20] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). Computer Vision and Image Understanding, 110(3), 346--359. https://doi.org/10.1016/j.cviu.2007.09.014

[21] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: an efficient alternative to SIFT or SURF. In Proceedings ICCV 2011 (pp. 2564--2571). IEEE. https://doi.org/10.1109/ICCV.2011.6126544

[22] Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). BRIEF: binary robust independent elementary features. In Proceedings ECCV 2010 (pp. 778--792). Springer. https://doi.org/10.1007/978-3-642-15561-1_56

[23] Rosten, E., & Drummond, T. (2006). Machine learning for high-speed corner detection. In Proceedings ECCV 2006 (pp. 430--443). Springer. https://doi.org/10.1007/11744023_34

[24] Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In Proceedings of the Alvey Vision Conference (pp. 147--151). AVC. https://doi.org/10.5244/C.2.23

[25] Shi, J., & Tomasi, C. (1994). Good features to track. In Proceedings CVPR 1994 (pp. 593--600). IEEE. https://doi.org/10.1109/CVPR.1994.323794

[26] Fischler, M.A., & Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381--395. https://doi.org/10.1145/358669.358692

[27] Duda, R.O., & Hart, P.E. (1972). Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM, 15(1), 11--15. https://doi.org/10.1145/361237.361242

[28] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings CVPR 2016 (pp. 770--778). IEEE. https://doi.org/10.1109/CVPR.2016.90

[29] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

[30] Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097--1105.

[31] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: unified, real-time object detection. In Proceedings CVPR 2016 (pp. 779--788). IEEE. https://doi.org/10.1109/CVPR.2016.91

[32] Girshick, R. (2015). Fast R-CNN. In Proceedings ICCV 2015 (pp. 1440--1448). IEEE. https://doi.org/10.1109/ICCV.2015.169

[33] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28, 91--99.

[34] Liu, W., et al. (2016). SSD: single shot multibox detector. In Proceedings ECCV 2016 (pp. 21--37). Springer. https://doi.org/10.1007/978-3-319-46448-0_2

[35] Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: transformers for image recognition at scale. In Proceedings ICLR 2021. ICLR.

[36] Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998--6008.

[37] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A.L. (2018). DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834--848. https://doi.org/10.1109/TPAMI.2017.2699184

[38] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings CVPR 2015 (pp. 3431--3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965

[39] Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62--66. https://doi.org/10.1109/TSMC.1979.4310076

[40] Gonzalez, R.C., & Woods, R.E. (2018). Digital Image Processing (4th ed.). Pearson Education.

[41] Soille, P. (2004). Morphological Image Analysis: Principles and Applications (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-05088-0

[42] Serra, J. (1982). Image Analysis and Mathematical Morphology. Academic Press.

[43] Trucco, E., & Verri, A. (1998). Introductory Techniques for 3-D Computer Vision. Prentice Hall.

[44] Hartley, R., & Zisserman, A. (2003). Multiple View Geometry in Computer Vision (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811685

[45] Nister, D., & Stewenius, H. (2006). Scalable recognition with a vocabulary tree. In Proceedings CVPR 2006 (pp. 2161--2168). IEEE. https://doi.org/10.1109/CVPR.2006.264

[46] Muja, M., & Lowe, D.G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In Proceedings VISAPP 2009 (Vol. 2, pp. 331--340).

[47] Torr, P.H.S., & Zisserman, A. (2000). MLESAC: a new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding, 78(1), 138--156. https://doi.org/10.1006/cviu.1999.0832

[48] Choi, S., Kim, T., & Yu, W. (1997). Performance evaluation of RANSAC family. In Proceedings BMVC 2009. BMVC.

[49] Aggarwal, C.C. (2017). Outlier Analysis (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-47578-3

[50] Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer.

[51] Ho, T.K. (1995). Random decision forests. In Proceedings ICDAR 1995 (Vol. 1, pp. 278--282). IEEE. https://doi.org/10.1109/ICDAR.1995.598994

[52] Vapnik, V.N. (1995). The Nature of Statistical Learning Theory. Springer. https://doi.org/10.1007/978-1-4757-2440-0

[53] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273--297. https://doi.org/10.1007/BF00994018

[54] Fix, E., & Hodges, J.L. (1951). Discriminatory analysis: nonparametric discrimination, consistency properties. USAF School of Aviation Medicine, Randolph Field, Texas, Report 4.

[55] Cover, T.M., & Hart, P.E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21--27. https://doi.org/10.1109/TIT.1967.1053964

[56] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5--32. https://doi.org/10.1023/A:1010933404324

[57] Freund, Y., & Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119--139. https://doi.org/10.1006/jcss.1997.1504

[58] Dietterich, T.G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation, 10(7), 1895--1923. https://doi.org/10.1162/089976698300017197

[59] Bunke, H., & Shearer, K. (1998). A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters, 19(3--4), 255--259. https://doi.org/10.1016/S0167-8655(97)00179-7

[60] Gavrila, D.M. (1999). The visual analysis of human movement: a survey. Computer Vision and Image Understanding, 73(1), 82--98. https://doi.org/10.1006/cviu.1998.0716

[61] Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), 509--522. https://doi.org/10.1109/34.993558

[62] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2021). Deep learning for 3D point clouds: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(12), 4338--4364. https://doi.org/10.1109/TPAMI.2020.3005434

[63] Charles, R.Q., Su, H., Kaichun, M., & Guibas, L.J. (2017). PointNet: deep learning on point sets for 3D classification and segmentation. In Proceedings CVPR 2017 (pp. 652--660). IEEE. https://doi.org/10.1109/CVPR.2017.16

[64] Qi, C.R., Yi, L., Su, H., & Guibas, L.J. (2017). PointNet++: deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems, 30, 5099--5108.

[65] Park, J., Zhou, Q.Y., & Koltun, V. (2017). Colored point cloud registration revisited. In Proceedings ICCV 2017 (pp. 143--152). IEEE. https://doi.org/10.1109/ICCV.2017.25

[66] Chen, Y., & Medioni, G. (1992). Object modeling by registration of multiple range images. Image and Vision Computing, 10(3), 145--155. https://doi.org/10.1016/0262-8856(92)90066-C

[67] Besl, P.J., & McKay, N.D. (1992). A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2), 239--256. https://doi.org/10.1109/34.121791

[68] Pomerleau, F., Colas, F., Siegwart, R., & Magnenat, S. (2013). Comparing ICP variants on real-world data sets. Autonomous Robots, 34(3), 133--148. https://doi.org/10.1007/s10514-013-9327-2

[69] Yang, J., Li, H., Campbell, D., & Jia, Y. (2015). Go-ICP: a globally optimal solution to 3D ICP point-set registration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(11), 2241--2254. https://doi.org/10.1109/TPAMI.2015.2513405

[70] Li, Z., & Mahler, J. (2021). Data-efficient model-based grasping for novel objects. In Proceedings ICRA 2021 (pp. 6536--6542). IEEE. https://doi.org/10.1109/ICRA48506.2021.9561491