Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Posture Detection Method for Augmented Reality Aided Assembly Based on YOLO-6D

Liping Zhang
Feng XU (xufeng@nuaa.edu.cn)
Nanjing University of Aeronautics and Astronautics https://orcid.org/0000-0001-6372-635X
Yu Liu
Dan Zhang
Liang Gui
Dunwen Zuo

Research Article

Keywords: Pose estimation, Deep learning, 3D Registration, Augmented reality, Assembly guidance

Posted Date: June 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1740576/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Manuscript

A Posture Detection Method for Augmented Reality Aided Assembly Based on YOLO-6D

Liping Zhang[1], Feng Xu[1*], Yu Liu[1], Dan Zhang[1*], Liang Gui[1] and Dunwen Zuo[1]

[1]College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China.

*Corresponding author(s). E-mail(s): xufeng@nuaa.edu.cn; zhd@nuaa.edu.cn;
Contributing authors: zhanglp@nuaa.edu.cn; liuyusx1905012@163.com; xianlianggui@163.com;

Abstract

The assembly of small batch electrical and mechanical products still relies on manual operation, with the problems of high rechecking time and low assembly efficiency and quality. The augmented reality technology can be used to assist assembly to improve efficiency. In view of the high professionalism, labor and time cost for the traditional posture detection method, a posture detection method is proposed to match the assembly posture by pre-calibrating the assembly posture template. YOLO-6D deep learning network is brought in to increase the accuracy of virtual information tracking registration. And the detection from both translation and rotation perspectives is designed to enhance the adaptability to different assembly tasks. For the generation of training datasets, a weighted sampling method based on part features is proposed to optimize the accuracy of position estimation with the limited training samples. Taking the assembly process of typical electronic products as an example, the developed augmented reality aided assembly system is used to boost the assembly efficiency of assembly operators. It is illustrated that the proposed positional detection method is effective in registering virtual assembly guidance information in real scenarios and can be applied to augmented reality assembly guidance of products.

Keywords: Pose estimation, Deep learning, 3D Registration, Augmented reality, Assembly guidance

1 Introduction

As the concept of Industry 4.0 continues to advance, the integration, packaging and complexity of electromechanical products are increasing [1, 2]. At present, the overall assembly automation level of small batch electromechanical products is low, and its assembly process basically belongs to the fixed assembly of connectors and the connection of related cables. At the same time, the staff are also faced with the problems of numerous assembly steps and the need for technical experience support. Therefore, assembly is still a high labor-intensity industry with long training cycles and high costs [3, 4]. In this context, Augmented Reality (AR) technology has emerged as a solution to this engineering problem. Among them, vision-based augmented reality, which is carried by removable or wearable display devices [5], has been regarded as a powerful technological tool for improving efficiency in the field of manual assembly, because of its ability to help workers establish a connection between the physical world and the digital information environment [6–9].

The posture estimation of assembled parts is the basis for the registration of virtual information and the detection of assembly posture. Lai et al. [10] developed an intelligent augmented reality assisted assembly system towards assembly workers, and used the richly textured marker images as the basis for tracking registration. Chen et al. [11] applied a trained classifier to measure the depth image and perform pixel classification to obtain the corresponding RGB values , which were compared them with the RGB values of the marker image to identify each part in the assembly process and it can be inferred the assembly position. Hinterstoisser [12, 13] proposed the LINE2D algorithm which is only based on RGB image color gradient information and the LINEMOD algorithm which combines RGB image color gradient information and depth image normal vector information. By establishing templates at different viewpoints and scales to match the templates formed by real-time images, the estimation of the object's posture can be achieved. Konishi et al. [14] proposed a perspectively cumulated orientation feature to estimate the posture through the use of offline templates that match the target object contour in the image, which can reduce the number of templates while ensuring the accuracy of assembly registration. In the case of the above-mentioned research results, the traditional posture detection method is mostly based on the template matching algorithm from the perspective of image processing. This method requires technical personnel with specialized knowledge to design the corresponding feature operator for a specific assembly scene, which is highly specialized and a large number of template images need to be collected as a template base, thus the labor and time costs required are large.

In recent years, deep learning-based pose estimation techniques have been developed rapidly and have higher accuracy and better scalability than other traditional vision methods [15–18]. However, the huge amount of training samples and labeled data for matching usage scenarios are still an important constraint for deep learning in industrial applications [19, 20]. Many scholars have proposed the use of random domains, domain adaptation, and physical

engine synthesis [21–23] to expand the training samples, which can effectively save labor cost, improve efficiency, as well as have some practical significance. The technology of posture estimation based on deep learning is still in the rapid development stage as a whole, while the mature application cases on the aided assembly system are relatively few.

To address the limitations of current assembly posture detection methods, this paper proposes a posture template-based, non-contact assembly posture detection method that requires only a simple 2D calibration of an assembled image to determine the state of the assembly through the posture relationship. In summary, a monocular camera is used and a YOLO-6D [24] deep learning networks are brought in to regress the 2D image coordinates of the labeled 3D control points by learning the features in the sample images. Based on the positional information, a designed positional matching strategy for detection from both translation and rotation perspectives is used to achieve 6D positional estimation. For the limited assembly datasets, a weighted sampling method based on the part features is proposed to improve the correct rate of pose estimation by filtering the sampled frames with low similarity based on the features of the parts. Finally, the effectiveness of the proposed method is verified by using the developed augmented reality aided assembly system with a typical electronic product assembly example.

2 Related works

In augmented reality, 3D tracking and registration technology is the basis for virtual information guidance and assembly posture detection, which is the most direct part of users' perception. This technology acquires the pose of the target object in the real field of view in real time by means of a camera or other type of sensor and reestablishes the spatial coordinate system according to the user's viewpoint, superimposing the virtual information accurately on the real object. The performance of its algorithm determines the effectiveness of the application of augmented reality technology [25]. Among them, the vision-based tracking and registration technology has been the focus of researchers' efforts due to its characteristics of cost-effectiveness, easy portability and broad application prospects. Salonen, Reiners, Boulanger et al. [26–28] applied marker-based tracking registration methods to AR assembly systems, while Yuan, Andersen, Alvarez et al. [29–31] performed tracking registration based on different features of the scene. As in Table 1, the advantages and disadvantages of each method of vision-based 3D tracking registration technology are provided. At present, the marker-based method is still the most widely used tracking registration method in AR assembly by virtue of its robustness and good real-time performance, but its problems such as marker obscuration and secondary contamination of parts also restrict its further development.

The 3D tracking registration method based on deep learning predicts the positional information of the target in the video frame by learning the relationship between the features in the image and the 6D poses of the target object

Table 1 Comparison of 3D tracking registration technology methods

Methods		Benefits	Limitations
Marker-based 3D tracking registration technology [32]		High robustness, less affected by illumination, great real-time performance.	Markers are easily obscured and can easily cause secondary contamination of parts.
Marker-less based 3D tracking registration technology	Feature-based [33]	Higher accuracy and faster speed.	Affected by the workpiece's own texture characteristics and light greatly.
	Template-based [34]	Suitable for weakly textured scenes.	Reliance on hardware devices and poor real-time performance.

[35]. The deep learning networks built with such methods generally consist of two parts, first for target detection and then for regression to the 6D poses of the target object. In the process of regression 6D poses, it can be divided into direct regression 6D poses and indirect regression 6D poses. As shown in Fig. 1, the flow chart of these two types of regression methods is presented.

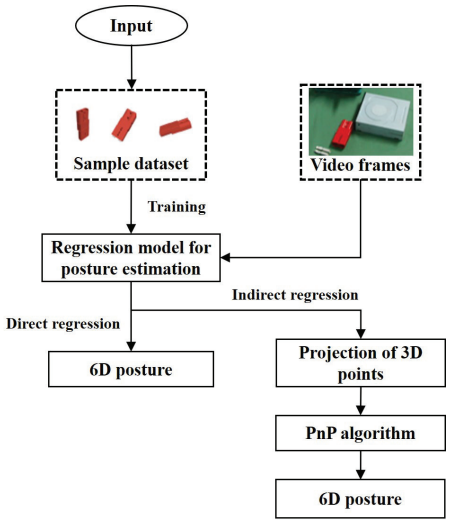


Fig. 1 Flow chart of 6D pose estimation based on deep learning

The technical framework of the proposed augmented reality aided assembly posture detection consists of two sections, as shown in Fig. 2. For the offline part, the weighted sampling method based on the part features is sampled when the weakly textured joint plug-in 3D model is known, and a virtual synthetic data set is generated by the Unreal engine to reduce the collection and labeling of real samples. For the online part, the registration of the virtual

model of the assembled part is achieved based on the aruco [36, 37] module first, and the 3D poses of the virtual model are manually pre-adjusted and calibrated to achieve real-time guidance in the assembly process. In addition, according to a proposed assembly posture detection method that matches by pre-calibrating the assembly posture template, it detects the assembly position from two perspectives: translations and rotations. Finally, the feedback information is used to reduce the re-inspection time of employees and improve the assembly efficiency.

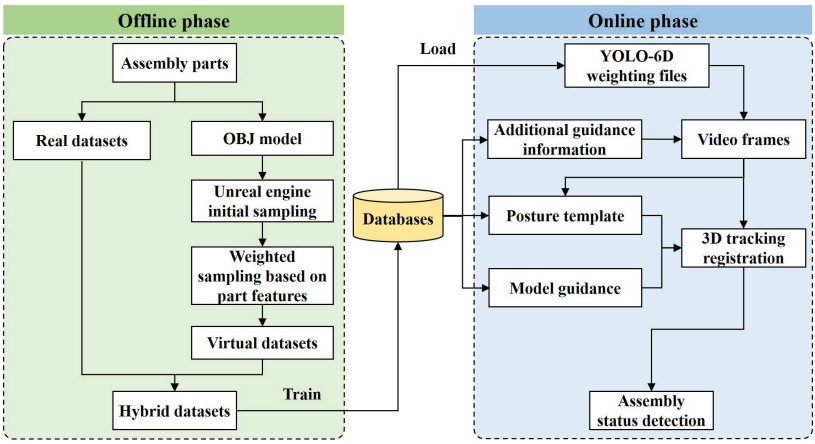


Fig. 2 Framework for posture detection in augmented reality aided assembly

3 The proposed approach

3.1 Offline processing

In order to associate the sampling viewpoints with the characteristics of the sampled objects, and to obtain more samples with large positional variability for model training with a limited number of samples, this paper chooses to improve on the typical sampling method [12] to obtain an adaptive sampling viewpoint distribution of the objects.

3.1.1 SSC equation

In order to strengthen the connection between the sampling viewpoint and the estimated object's own features, as well as to obtain higher quality learning samples, the features of the parts should be selected for analysis first. The research objects in this paper are connectors of electromechanical products, which generally have a regular, weakly textured shape. From the perspective of image processing, the surface texture and contour features of the estimated objects are extracted mainly by SIFT algorithm. Among them, the matching degree ρ_{sift} for surface texture features is defined as the ratio of the number

of matching points based on the Euclidean distance to the total number of matching pairs. The match IoU_{cp} for contour features is defined as the intersection over union of the areas of the part objects after center alignment in the sampled frame.

According to the similarity evaluation metrics ρ_{sift} and IoU_{cp} determined above, the CRITIC approach in the objective weighting evaluation method is used to measure the objective weights of the metrics. The flow chart of the CRITIC algorithm [38] is shown in Fig. 3. In order to quantify the indicators and to combine the intrinsic connection between the indicators at the macro level, the number of operators extracted by the SIFT algorithm in each sampling frame and the contour area of the target object are used as the values of the evaluation samples in the data matrix, respectively. Then for each item the objective weight can be determined by the Eq. 1.

$$\begin{cases} C_j = S_j \sum_{i=1}^{p} (1 - r_{ij}) = S_j \times R_j \\ W_j = \frac{C_j}{\sum_{j=1}^{p} C_j} \end{cases} \tag{1}$$

Where, r_{ij} is the correlation coefficient between evaluation objective i and j , R_j is the correlation matrix, S_j is the standard deviation of the jth objectives, and W_j is the weight of the jth objectives.

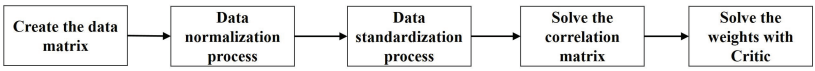


Fig. 3 Flow chart of CRITIC algorithm

In combination with the weights of each indicator, the similarity of surface features and contour features (SSC) can be expressed as:

$$SSC = \alpha \rho_{sift} + \beta IoU_{cp} \tag{2}$$

Where α and β represent the weights of surface features similarity ρ_{sift} and contour features similarity IoU_{cp} respectively, and the sum of the both is one.

3.1.2 Filtering path

For the typical sampling method based on latitude and longitude grid has the advantages that the distribution of viewpoints is more regular and dense , besides the position of the target object at the same latitude with adjacent viewpoints has only one rotational variation at one latitude. This paper is based on the latitude and longitude grid and defines the paths of comparison and filtering of adjacent sampling frames. That is, the comparison and compression sequence of a certain sampled viewpoint on the same latitude and longitude sampling sphere with the neighboring viewpoints. The path diagram of the sampled frame compression is shown in Fig. 4.

In this study the spatial coordinate point sets $\{R_1, R_2, \cdots, R_n\}$ are constructed for sampled viewpoints at the same latitude. In the point set R_n ,

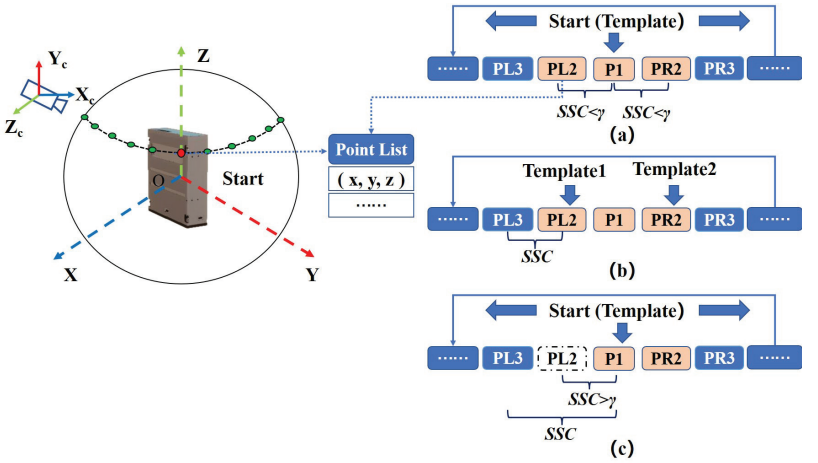


Fig. 4 Schematic diagram of the path to sample frame compression

a spatial point P_1 is randomly selected as the starting point of the path and stored in the chain of spatial points corresponding to the keyframe created. And the point on the left side of P_1 is recorded as PL_n and the point on the right side is recorded as PR_n . Next, the sampling frame corresponding to the P_1 point is used as the template image, compared to the left and right sides of the diffusion, and the SSC is calculated. If the SSC is less than the set threshold γ , this point set is added to the spatial point set link list, as shown in Fig. 4 (a). At the same time, the template of the side that meets the condition of $SSC < \gamma$ is switched to the sampling frame corresponding to this spatial point, as in Fig. 4 (b). If the SSC is not less than the set threshold γ , the corresponding spatial point is skipped, and it continues to compare the sampling frames of the spatial point to the left or right of this point, as in Fig. 4 (c). Finally, the same latitude point set is looped along this keyframe filtering path until the compared point sets overlap.

3.1.3 Sampling process

The crucial of the proposed weighted sampling method based on part features is to merge highly similar sampling frames in a limited number of learning samples. As the Fig. 5, the flow chart of the weighted sampling based on part features is shown. And the detailed steps are described as follows:

Step 1: With a known OBJ model, import the model of the sampled object and the coordinates of the spatial points sampled based on the latitude and longitude grid in Unreal Engine, with the number of samples being N.

Step 2: Set the threshold γ for the SSC and the compression ratio ω for the sampled frames, and compress them along the keyframe compression path. Generally, the threshold γ is initially set to 0.7 and the compression ratio ω is initially set to 3/4.

Step 3: Compare the size of the number of compressed keyframes n with the number of samples N. If the condition of $|\mathrm{N}-\mathrm{n}|<\mathrm{N}\cdot\omega\cdot 5\%$ is satisfied, go to step 4; otherwise, adjust the value of SSC threshold γ and continue with step 2.

Step 4: Uniformly distribute the Fibonacci sampling viewpoints of the same radius sampling sphere among different latitudes, with the number of $|\mathrm{N}-\mathrm{n}|$.

Step 5: Reimport the spatial points in the final generated sampling link list into Unreal Engine to generate the sampling image.

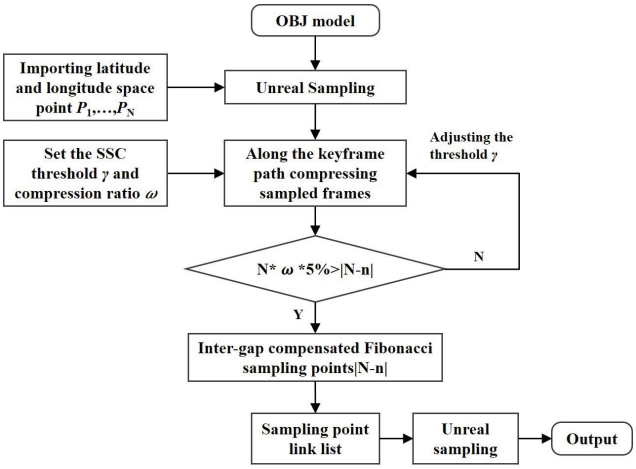


Fig. 5 Flow chart of weighted sampling based on part features

3.2 Online detection

The assembly posture detection method proposed in this paper contains two stages, which mainly decompose the correct assembly position of parts into rotational and translational movements for matching degree calculation, and the technical route of this method is shown in Fig. 6. Where the matching errors for the rotational motion are defined as the absolute orientation errors based on the rotation matrices, and the matching for the translational motion is defined as the intersection rate of the minimum bounding box of the virtual model used for guidance and the minimum bounding box of the 3D bounding box estimated by the YOLO-6D network.

3.2.1 Calibration of posture templates

In this paper, the YOLO-6D-based pose estimation is performed on the part to be assembled, and the result of the pose estimation is the RT transformation matrix from the world coordinate system to the camera coordinate system where the part is located. Then the local coordinate system of the part can be regarded as coincident with the world coordinate system. Thus, the rotational

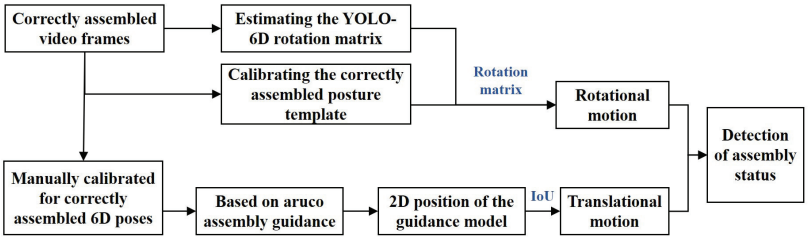


Fig. 6 The technical route of assembly state detection method based on posture template

motion posture of each component in the correct assembly position needs to be manually calibrated to extract the rotational transformation matrix in the correct assembly position.

Extracting the displacement matrix of the part to be assembled from the image of the correct assembly is a typical 3D-to-2D motion solution problem, which is solved in this paper with the Perspective-n-Point (PnP) algorithm [39]. The calibration process of the rotation matrix is shown in Fig. 7. Firstly, the world coordinates of the 3D feature points of the model are selected by the online tool 3d-on-2d. Secondly, the 2D points corresponding to the 3D point set on the video frame are calibrated with the self-developed 2D calibration tool based on Qt development platform. Eventually, the RT transformation matrix from the world coordinate system to the camera coordinate system is calculated by the PnP algorithm, where the rotation matrix is saved as a template.

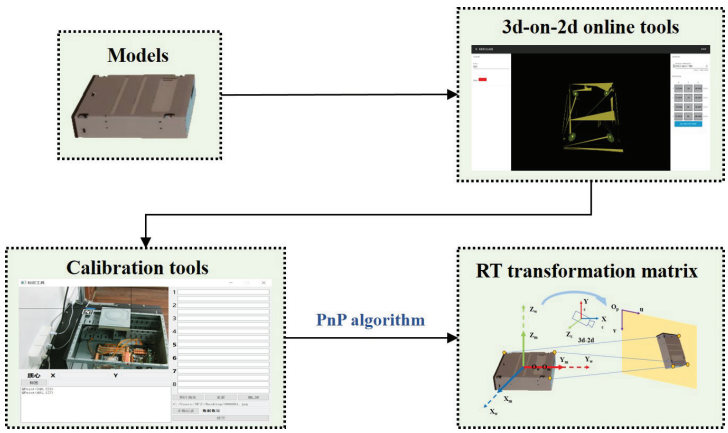


Fig. 7 Flow chart of calibrating the rotation matrix

3.2.2 A strategy for posture matching

The arbitrary posture of the part to be assembled can be decomposed into two motion steps: translation and rotation. The result of the pose matching can be

obtained from the comparison of the translational and rotational movements of the part in the template for each frame. It is shown that the flow chart of the pose matching strategy for assembly posture detection in Fig. 8.

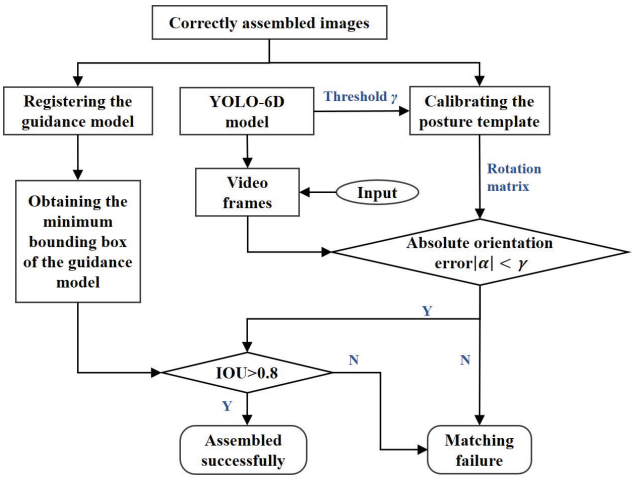


Fig. 8 Flow chart of the pose-matching strategy for assembly state detection

For translational motion, in order to better quantify the error of the parts to be assembled in different directions, the correctness of the translational motion during assembly is determined by calculating the intersection rate of the minimum bounding box of the 3D bounding box obtained from the guidance model and the YOLO-6D estimation.

If the intersection rate is greater than 0.8, then the translation poses of the assembly are correct as default. As shown in Fig. 9, the pink area is the estimated positional bounding box and the yellow area is the overlap area. Where the virtual model used for guidance is calibrated based on the markers for the position and rendered by the aruco module.

For rotational motion, the error can be compared between the rotation matrix obtained by estimating the part pose in YOLO-6D and the rotation matrix in the pose template. The quantitative analysis was performed with Eq. 3 proposed by Sattler T et al [40]. The average error of the test sets calculated during the training of the YOLO-6D model was set as the threshold value. If the absolute orientation error is less than the threshold value, the rotation position of the assembly is correct by default.

$$2\cos(|\alpha|) = \text{trace}\left(\mathrm{R_{gt}}^{-1}\mathrm{R_{est}}\right) - 1 \tag{3}$$

Where, $|\alpha|$ is the absolute directional error in degrees. R_{est} and R_{gt} are the estimated and true rotation matrices.

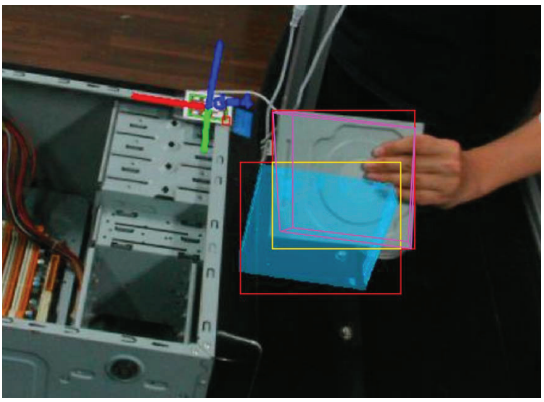


Fig. 9 Schematic diagram of translational motion posture matching

4 Implementation and experiments

In order to verify the effectiveness of the method proposed above, this section designs and validates relevant experiments based on the YOLO-6D network model with the plug-in electronics as the experimental object.

4.1 Verifying the effectiveness of sampling method

4.1.1 Assessment datasets and training schemes

The measurement dataset is composed of part of the public dataset in LINEMOD [11] and a self-made plug-in dataset. The model for testing and validation is shown in Fig. 10.

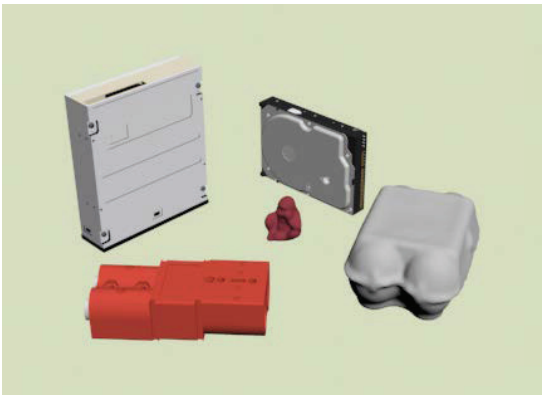


Fig. 10 Rendering of the tested and validated model

The comparison experiments of typical sampling methods and weighted sampling methods based on part features were conducted by using GTX 1660S

graphics card with 6 GB of video memory in Windows environment, respectively. Among the typical sampling methods include the latitude and longitude grid-based and the Fibonacci grid-based sampling methods [41, 42]. According to the size of the model used in the datasets, sampling was performed on a sampling sphere with a radius of 60 cm to 100 cm. Each of these classes of objects is sampled with 800 samples based on each of the three sampling methods. From these samples, 70% are randomly selected for training, and the remaining 30% samples of each method are mixed and then 30% are randomly selected as the overall test sample.

4.1.2 Experimental results

In Table 2, the comparison results of 800 samples training are shown.

It is observed from Table 2 that for the 6D Pose [40] and 5 cm 5° [43] metrics, the mean values improved by 8.76%, 8.27% and 9.17%, 9.86%, respectively, compared to the two classical sampling methods in the case of a small number of samples. For the judging metric of 5px 2D Protection [44], by using the weighted sampling method based on part features proposed in this paper to estimate the percentage of correct poses consistently approaches or even exceeds the higher percentage of the other two methods. Thus it demonstrates the better robustness of the method against different sampled objects.

4.2 Verifying the effectiveness of assembly posture detection method

4.2.1 Experiment design

Taking the assembly position detection of optical disk driver in a computer case as an example, firstly, the synthetic samples are obtained by the sampling method above and mixed with a certain number of real calibration samples to generate the training hybrid datasets. And the specific production process is shown in Fig. 11. After training the YOLO-6D network, the average absolute orientational error of the test sets was obtained as $|\alpha| = 7.92$, and the threshold value of the rotation matrix comparison was $\gamma = 7.92$.

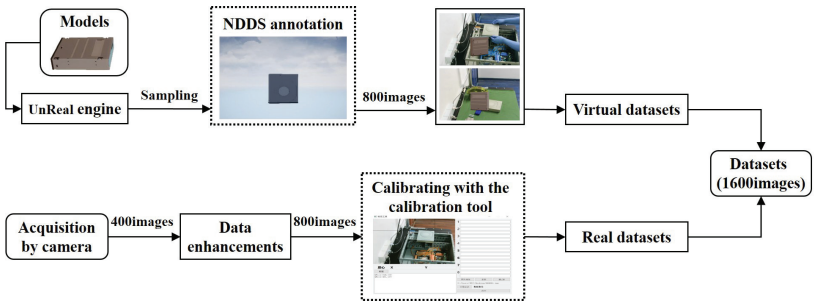
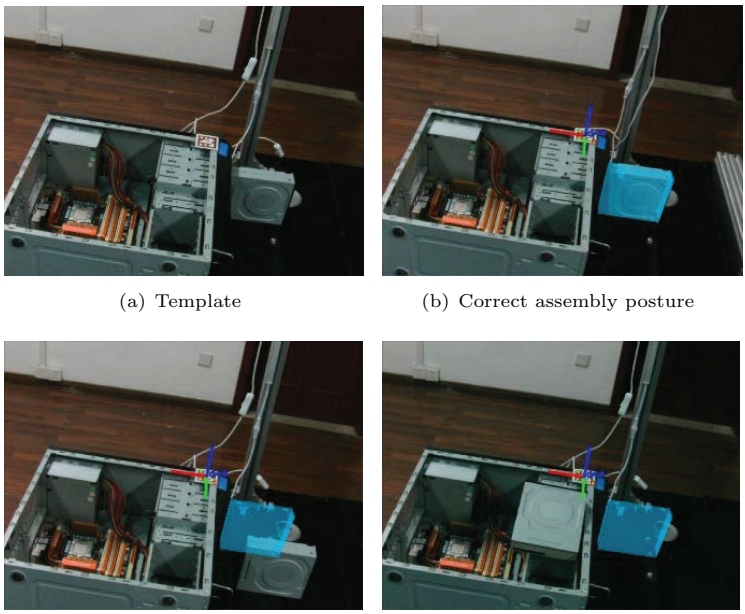


Fig. 11 Flow chart for the generation of datasets

Table 2 Comparison of the three sampling methods

Objects	Based on latitude and longitude grid			Based on the Fibonacci grid			Based on weighted sampling of part features		
	5px 2D Protection (%)	6D Pose (%)	5 cm 5° (%)	5px 2D Protection (%)	6D Pose (%)	5 cm 5° (%)	5px 2D Protection (%)	6D Pose (%)	5 cm 5° (%)
ape	99.71	26.13	58.33	100	46.36	62.86	100	49.79	79.25
eggbox	99.07	84.85	90.3	98.85	63.81	82.86	97.2	79.33	84.94
Optical Disk Driver	61.46	72.43	77.64	79.86	78.33	70.32	79.67	82.3	81.74
Hard Disk Drive	69.62	74.68	79.58	82.92	79.58	72.57	80.91	82.16	82.57
Connector Plugs	84.51	55.46	57.4	85.62	43.43	66.7	86.25	63.75	76.11
average	82.87	62.71	72.65	89.45	62.3	71.06	88.81	71.47	80.92

In order to have better comparison results in the posture detection experiment, a correct assembly posture and two wrong assembly postures are designed respectively. Taking the assembly of optical disk driver as an example, a sample of the assembly posture detection is shown in Fig. 12. Where Fig. 12(a) shows the template image of correct assembly; Fig. 12(b) shows the correct assembly pose; the wrong assembly pose includes the following two kinds: correct rotational motion and wrong translational motion, as in Fig. 12(c); and wrong rotational motion, as in Fig. 12(d).



(a) Template (b) Correct assembly posture

(c) Wrong assembly posture1 (d) Wrong assembly posture2

Fig. 12 Templates and samples for assembly posture detection. a Template. b Correct assembly posture. c Wrong assembly posture1. d Wrong assembly posture2

4.2.2 Experimental results

The above samples are used to calculate the matching results based on the positional matching strategy in Section 3.2.2. As shown in Table 3, for the assembly state detection results of the samples.

The above assembly posture detection results for the samples show that for both correct and incorrect assembly posture images of the assembly process, the rotation and translation information can be detected exactly using the posture template-based detection method, and correct matching results are obtained.

Table 3 The results of the posture detection of the samples

Samples	Rotational motion (°)	Translational motion	Results
Correct assembly posture	7.3	0.89	Position correctly
Wrong assembly posture1	5.9	0.17	Position error
Wrong assembly posture2	169.5	-	Position error

5 Application Cases

A Qt-based design and development of an augmented reality aided assembly system for mechanical and electrical products with small batches, large bodies, and multiple connectors is presented in this paper. Fig. 13 shows the sketch of the system integration operation and application. This section takes the assembly of a computer case as an example to specifically demonstrate the actual whole assembly effect of the case. The goal of the assembly is to use the corresponding tools to assemble the six components provided to the case, namely the optical drive, hard disk 1, hard disk 2, power supply, fan and CPU.

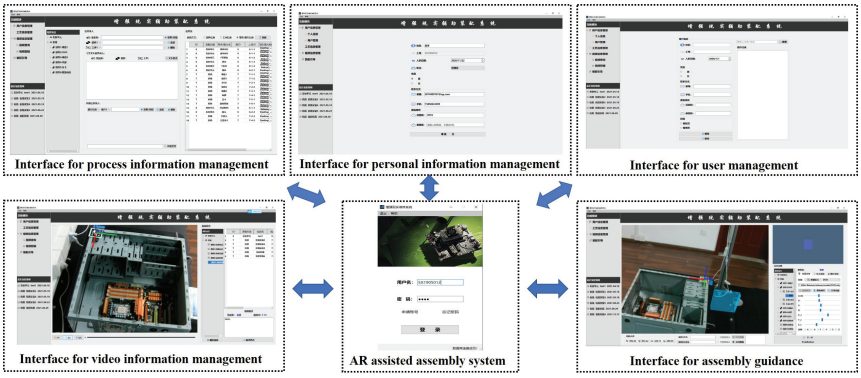
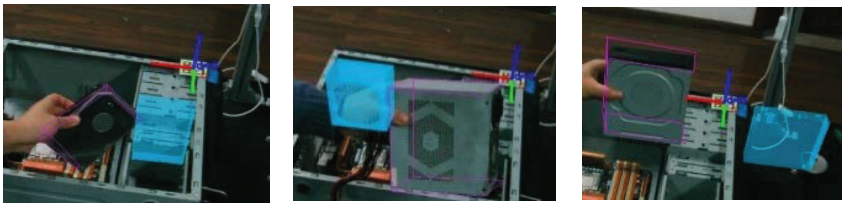


Fig. 13 Sketch of the system integration operations and applications

5.1 Application

Eight people with no prior knowledge of the assembly were selected and divided into two groups. Four assemblers from each group stood at the workbench to do the experiments on the whole assembly of the case. One group was required to assemble by referring to a paper manual, and the other group was required to assemble using an integrated YOLO-6D-based augmented reality aided assembly system. For the assemblers used the system, after specifying the assembly target of the case, they enter the information management module of the system to register the assembly process information and store it in the databases. Then, the poses of the assembled parts are adjusted and calibrated

through the assembly guidance module. Subsequently, the posture template is calibrated in the assembly posture detection module and all this information is recorded in the databases. Finally, the system decided whether to automatically call the guidance information in the databases of the next step by the feedback from the posture detection module. Throughout the whole assembly process, the assembly guidance module assisted in the whole assembly of the case by interacting with the 3D tracking registration module and the assembly posture detection module in real time. The effect of position estimation of the case assembly is shown in Fig. 14.



(a) (b) (c)

Fig. 14 The effect of the estimated position of the chassis assembly

5.2 Results and analysis

From the perspective of productivity, the time to completion of assembly and the number of errors are the most critical performance metrics in assembly operations. Table 4 lists the three types of assembly errors that are considered to be the most common errors in manual mechanical assembly processes [45]. Therefore, during the chassis assembly experiment, it is required to record the completion time of assembly by the assemblers as well as the number and type of assembly errors when they occurred. Also, statistics and comparison of the experimental results with using paper manual and using augmented reality assisted assembly system are shown in Tables 5 and 6.

Table 4 Three types of assembly errors

Number	Types of errors	Description
1	Selection of component error	e.g. when assembling hard disk 1 took into hard disk 2, screw selection error, etc.
2	Sequence of assembly error	e.g. first to put all the parts in the respective positions and then to fix the whole, etc.
3	Fixation/installation error	e.g. incorrect fixing of the data cable to the slot, missing fixing screws, etc.

Table 5 The assessment results of the group with the paper manuals

Number of assemblers	1	2	3	4	Average value	Total
Totality of errors	5	3	4	5	-	17
Quantity of error 1	1	1	2	2	-	6
Quantity of error 2	2	1	0	1	-	4
Quantity of error 3	2	1	2	2	-	7
Time of assembly (s)	435	396	408	457	424	-

Table 6 The assessment results of the group used the system

Number of assemblers	1	2	3	4	Average value	Total
Totality of errors	1	2	3	2	-	8
Quantity of error 1	0	1	1	1	-	3
Quantity of error 2	0	0	0	0	-	0
Quantity of error 3	1	1	2	1	-	5
Time of assembly (s)	324	369	377	348	355	-

Compared to using paper manuals, assembly completion time and the number of assembly errors made during assembly can be reduced by 16.3% and 52.9%, respectively, with the augmented reality instructions of the augmented reality-assisted assembly system. Among them, the error rates of two types of errors, part selection errors and assembly fixing errors, were reduced by 50% and 28.6%, respectively. And the errors in assembly sequence were reduced by 100%. It is due to the fact that the system can identify the components to be assembled at each step, which is a better guidance for the assembly sequence. As for the system, it is not yet equipped with recognition and detection functions for tiny parts and flexible cables. And there are not enough samples of mixed datasets for each component, which makes the occurrence of incorrect estimation of the position posture and forms a disturbance to the assemblers.

In summary, the developed augmented reality aided assembly system based on YOLO-6D in this paper has practicality in real assembly environment. It also combines the recognition and state detection functions of deep learning, which reduced the assembly error rate and the rechecking time, thus it has a certain potential for development.

6 Conclusion

In this paper, considering the problems of current assembly posture detection methods, a posture detection method based on the 6D posture estimated by YOLO-6D is proposed, and a posture matching strategy is introduced to detect the posture from both translation and rotation perspectives, which effectively reduces the rechecking time of the assembly and improves the assembly efficiency. A weighted sampling method based on part features is designed to address the limitations of typical sampling methods, and it is verified through

comparative experiments that the method can enhance the sampling quality of parts with limited training samples, reduce distortion caused by too much synthetic data, and optimize the effectiveness of the training model. To verify the practicality of the proposed approach, an augmented reality aided assembly system for small batch mechanical and electrical products is designed and developed. The system is applied to an assembly case of a typical mechanical and electrical product. Through user testing, it can be observed that with the assistance of the augmented reality aided assembly system compared to the use of paper manuals, the assembly completion time and the number of assembly errors made during assembly can be reduced by 16.3% and 52.9% respectively, where errors of the assembly sequence type can be reduced by 100%, thus effectively strengthening the efficiency and quality of mechanical and electrical product assembly.

Acknowledgments. Thanks to the Joint Funds of the National Natural Science Foundation of China (Grant No. U20A20293) and the National Natural Science Foundation of China (No. 51705246) for supporting this paper.

Declarations

- Funding
  This work was supported by the Joint Funds of the National Science Foundation of China (Grant No. U20A20293) and the National Natural Science Foundation of China (No. 51705246).
- Conflict of interest/Competing interests
  The authors declare no competing interests.
- Ethics approval
  Not applicable.
- Consent to participate
  Not applicable.
- Consent for publication
  The authors give their consent for publication.
- Availability of data and materials
  All data generated or analysed during this study are included in this published article.
- Code availability
  The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.
- Authors' contributions
  All authors participated, read and approved the final manuscript.

References

[1] Bottani, E., Vignali, G.: Augmented reality technology in the manufacturing industry: A review of the last decade. IISE Transactions **51**(3), 284–310 (2019). https://doi.org/10.1080/24725854.2018.1493244

[2] Lee, J., Bagheri, B., Kao, H.-A.: A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing letters **3**, 18–23 (2015). https://doi.org/10.1016/j.mfglet.2014.12.001

[3] Li, W., Wang, J., Liu, M., Zhao, S.: Real-time occlusion handling for augmented reality assistance assembly systems with monocular images. Journal of Manufacturing Systems **62**, 561–574 (2022). https://doi.org/10.1016/j.jmsy.2022.01.012

[4] Chiew, J.H., Sung, A.N.: Augmented reality application for laptop assembly with assembly complexity study. The International Journal of Advanced Manufacturing Technology **120**(1), 1149–1167 (2022). https://doi.org/10.1007/s00170-022-08751-x

[5] Wang, X., Ong, S., Nee, A.Y.-C.: Multi-modal augmented-reality assembly guidance based on bare-hand interface. Advanced Engineering Informatics **30**(3), 406–421 (2016)

[6] Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., Barbaray, R.: The industrial management of smes in the era of industry 4.0. International Journal of Production Research **56**(3), 1118–1136 (2018). https://doi.org/10.1080/00207543.2017.1372647

[7] Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D.: Deep learning for smart manufacturing: Methods and applications. Journal of manufacturing systems **48**, 144–156 (2018). https://doi.org/10.1016/j.jmsy.2018.01.003

[8] Radkowski, R., Herrema, J., Oliver, J.: Augmented reality-based manual assembly support with visual features for different degrees of difficulty. International Journal of Human-Computer Interaction **31**(5), 337–349 (2015). https://doi.org/10.1080/10447318.2014.994194

[9] Wang, X., Ong, S.K., Nee, A.Y.: A comprehensive survey of augmented reality assembly research. Advances in Manufacturing **4**(1), 1–22 (2016). https://doi.org/10.1007/s40436-015-0131-4

[10] Lai, Z.-H., Tao, W., Leu, M.C., Yin, Z.: Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. Journal of Manufacturing Systems **55**, 69–81 (2020). https://doi.org/10.1016/j.jmsy.2020.02.010

[11] Chen, C., Tian, Z., Li, D., Pang, L., Wang, T., Hong, J.: Projection-based augmented reality system for assembly guidance and monitoring. Assembly Automation (2020). https://doi.org/10.1108/AA-02-2019-0028

[12] Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N.: 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 (2012). https://doi.org/10.1007/978-3-642-37331-2_42. Springer

[13] Hinterstoisser, S., Holzer, S., Cagniart, C., Ilic, S., Konolige, K., Navab, N., Lepetit, V.: Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In: 2011 International Conference on Computer Vision, pp. 858–865 (2011). https://doi.org/10.1109/ICCV.2011.6126326. IEEE

[14] Konishi, Y., Hanzawa, Y., Kawade, M., Hashimoto, M.: Fast 6d pose estimation from a monocular image using hierarchical pose trees. In: European Conference on Computer Vision, pp. 398–413 (2016). https://doi.org/10.1007/978-3-319-46448-0_24. Springer

[15] Le, T.-T., Le, T.-S., Chen, Y.-R., Vidal, J., Lin, C.-Y.: 6d pose estimation with combined deep learning and 3d vision techniques for a fast and accurate object grasping. Robotics and Autonomous Systems **141**, 103775 (2021). https://doi.org/10.1016/j.robot.2021.103775

[16] Ming, Y., Meng, X., Fan, C., Yu, H.: Deep learning for monocular depth estimation: A review. Neurocomputing **438**, 14–33 (2021). https://doi.org/10.1016/j.neucom.2020.12.089

[17] Vafadar, S., Skalli, W., Bonnet-Lebrun, A., Khalifé, M., Renaudin, M., Hamza, A., Gajny, L.: A novel dataset and deep learning-based approach for marker-less motion capture during gait. Gait & Posture **86**, 70–76 (2021). https://doi.org/10.1016/j.gaitpost.2021.03.003

[18] Zhang, X., Jiang, Z., Zhang, H.: Real-time 6d pose estimation from a single rgb image. Image and Vision Computing **89**, 1–11 (2019). https://doi.org/10.1016/j.imavis.2019.06.013

[19] Zhang, W., Zhang, X., Lan, L., Luo, Z.: Enhancing unsupervised domain adaptation by discriminative relevance regularization. Knowledge and Information Systems **62**(9), 3641–3664 (2020). https://doi.org/10.1007/s10115-020-01466-z

[20] Yang, B., Yuen, P.C.: Learning adaptive geometry for unsupervised domain adaptation. Pattern Recognition **110**, 107638 (2021). https://doi.org/10.1016/j.patcog.2020.107638

[21] Zhu, X., Li, Y., Sun, J., Chen, H., Zhu, J.: Learning with noisy labels method for unsupervised domain adaptive person re-identification. Neurocomputing **452**, 78–88 (2021). https://doi.org/10.1016/j.neucom.2021.04.120

[22] Zhao, G., Li, G., Xu, R., Lin, L.: Collaborative training between region proposal localization and classification for domain adaptive object detection. In: European Conference on Computer Vision, pp. 86–102 (2020). https://doi.org/10.1007/978-3-030-58523-5_6. Springer

[23] Aziz, E.-S.S., Chang, Y., Esche, S.K., Chassapis, C.: Virtual mechanical assembly training based on a 3d game engine. Computer-Aided Design and Applications **12**(2), 119–134 (2015). https://doi.org/10.1080/16864360.2014.962424

[24] Tekin, B., Sinha, S.N., Fua, P.: Real-time seamless single shot 6d object pose prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 292–301 (2018)

[25] Wang, K., Liu, D., Liu, Z., Duan, G., Hu, L., Tan, J.: A fast object registration method for augmented reality assembly with simultaneous determination of multiple 2d-3d correspondences. Robotics and Computer-Integrated Manufacturing **63**, 101890 (2020). https://doi.org/10.1016/j.rcim.2019.101890

[26] Salonen, T., Sääski, J., Hakkarainen, Kannetis, T., Perakakis, M., Siltanen, S., Potamianos, A., Korkalo, O., Woodward, C.: Demonstration of assembly work using augmented reality. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval, pp. 120–123 (2007). https://doi.org/10.1145/1282280.1282301

[27] Reiners, D., Stricker, D., Klinker, G., Müller, S.: Augmented reality for construction tasks: Doorlock assembly. Proc. Ieee And Acm Iwar **98**(1), 31–46 (1999)

[28] Boulanger, P.: Application of augmented reality to industrial tele-training. In: First Canadian Conference on Computer and Robot Vision, 2004. Proceedings., pp. 320–328 (2004). https://doi.org/10.1109/CCCRV.2004.1301462. IEEE

[29] Yuan, M., Ong, S., Nee, A.: Augmented reality for assembly guidance using a virtual interactive tool. International journal of production research **46**(7), 1745–1767 (2008). https://doi.org/10.1080/00207540600972935

[30] Andersen, M., Andersen, R., Larsen, C., Moeslund, T.B., Madsen, O.: Interactive assembly guide using augmented reality. In: International Symposium on Visual Computing, pp. 999–1008 (2009). https://doi.org/10.1007/978-3-642-10331-5_93. Springer

[31] Alvarez, H., Aguinaga, I., Borro, D.: Providing guidance for maintenance operations using automatic markerless augmented reality system. In: 2011

10th IEEE International Symposium on Mixed and Augmented Reality, pp. 181–190 (2011). IEEE

[32] Mahmood, B., Han, S., Seo, J.: Implementation experiments on convolutional neural network training using synthetic images for 3d pose estimation of an excavator on real images. Automation in Construction **133**, 103996 (2022). https://doi.org/10.1109/ISMAR.2011.6092385

[33] Danielsson, O., Holm, M., Syberfeldt, A.: Augmented reality smart glasses in industrial assembly: Current status and future challenges. Journal of Industrial Information Integration **20**, 100175 (2020). https://doi.org/10.1016/j.jii.2020.100175

[34] Choi, C., Christensen, H.I.: Rgb-d object pose estimation in unstructured environments. Robotics and Autonomous Systems **75**, 595–613 (2016). https://doi.org/10.1016/j.robot.2015.09.020

[35] Zhang, H., Cao, Q.: Texture-less object detection and 6d pose estimation in rgb-d images. Robotics and Autonomous Systems **95**, 64–79 (2017). https://doi.org/10.1016/j.robot.2017.06.003

[36] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition **47**(6), 2280–2292 (2014). https://doi.org/10.1016/j.patcog.2014.01.005

[37] La Delfa, G.C., Catania, V., Monteleone, S., De Paz, J.F., Bajo, J.: Computer vision based indoor navigation: A visual markers evaluation. In: Mohamed, A., Novais, P., Pereira, A., Villarrubia González, G., Fernández-Caballero, A. (eds.) Ambient Intelligence - Software and Applications, pp. 165–173. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19695-4_17

[38] Diakoulaki, D., Mavrotas, G., Papayannakis, L.: Determining objective weights in multiple criteria problems: The critic method. Computers & Operations Research **22**(7), 763–770 (1995). https://doi.org/10.1016/0305-0548(94)00059-H

[39] Lepetit, V., Moreno-Noguer, F., Fua, P.: Epnp: An accurate o (n) solution to the pnp problem. International journal of computer vision **81**(2), 155–166 (2009). https://doi.org/10.1007/s11263-008-0152-6

[40] Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J., *et al.*: Benchmarking 6dof outdoor visual localization in changing conditions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8601–8610 (2018)

[41] González, Á.: Measurement of areas on a sphere using fibonacci and latitude–longitude lattices. Mathematical Geosciences **42**(1), 49–64 (2010). https://doi.org/10.1007/s11004-009-9257-x

[42] Baselga, S.: Fibonacci lattices for the evaluation and optimization of map projections. Computers & Geosciences **117**, 1–8 (2018). https://doi.org/10.1016/j.cageo.2018.04.012

[43] Shotton, J., Glocker, B., Zach, C., Izadi, S., Criminisi, A., Fitzgibbon, A.: Scene coordinate regression forests for camera relocalization in rgb-d images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2930–2937 (2013)

[44] Brachmann, E., Michel, F., Krull, A., Yang, M.Y., Gumhold, S., *et al.*: Uncertainty-driven 6d pose estimation of objects and scenes from a single rgb image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3364–3372 (2016)

[45] Hou, L., Wang, X., Bernold, L., Love, P.E.: Using animated augmented reality to cognitively guide assembly. Journal of Computing in Civil Engineering **27**(5), 439–451 (2013)