

# TRansPose: Large-scale multispectral dataset for transparent object

The International Journal of  
Robotics Research  
2023, Vol. 0(0) 1–8  
© The Author(s) 2023  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/02783649231213117  
[journals.sagepub.com/home/ijr](https://journals.sagepub.com/home/ijr)



Jeongyun Kim<sup>1,\*</sup> , Myung-Hwan Jeon<sup>2,\*</sup> , Sangwoo Jung<sup>1</sup> ,  
Woosong Yang<sup>1</sup> , Minwoo Jung<sup>1</sup> , Jaeho Shin<sup>1</sup>  and Ayoung Kim<sup>1</sup> 

## Abstract

Transparent objects are encountered frequently in our daily lives, yet recognizing them poses challenges for conventional vision sensors due to their unique material properties, not being well perceived from RGB or depth cameras. Overcoming this limitation, thermal infrared cameras have emerged as a solution, offering improved visibility and shape information for transparent objects. In this paper, we present TRansPose, the first large-scale multispectral dataset that combines stereo RGB-D, thermal infrared (TIR) images, and object poses to promote transparent object research. The dataset includes 99 transparent objects, encompassing 43 household items, 27 recyclable trashes, 29 chemical laboratory equivalents, and 12 non-transparent objects. It comprises a vast collection of 333,819 images and 4,000,056 annotations, providing instance-level segmentation masks, ground-truth poses, and completed depth information. The data was acquired using an FLIR A65 thermal infrared camera, two Intel RealSense L515 RGB-D cameras, and a Franka Emika Panda robot manipulator. Spanning 87 sequences, TRansPose covers various challenging real-life scenarios, including objects filled with water, diverse lighting conditions, heavy clutter, non-transparent or translucent containers, objects in plastic bags, and multi-stacked objects. Supplementary material can be accessed from the following link: <https://sites.google.com/view/transpose-dataset>.

## Keywords

Dataset, multispectral cameras, object recognition, transparent object, object pose estimation

## 1. Introduction

Over the past decade, object recognition has garnered significant attention, with datasets (Kasper et al. 2012; Calli et al. 2017; Novkovic et al. 2019) contributing greatly to advancements in the robotics research (Saxena et al. 2008; Sinapov et al. 2011; Li and Kleeman, 2011; Zeng et al. 2022). These datasets, however, mainly provide RGB images and capture only a limited representation of transparent objects. Transparent objects are common in daily life, recycling centers, as well as chemical and household settings, raising the requirement to perceive these objects for robots and automation. When recognizing transparent objects with vision sensors, they show insurmountable challenges caused by their inherent material. The transparent objects are absent of distinct features and textures caused by high dependency on the background, making inconsistent visual appearances. In addition, the usage of depth measurement obtained from the distance sensors is arduous since the transparent objects break the Lambertian reflectance.

Many existing methods have mainly targeted RGB-based approaches, proposing datasets for boosting the research of transparent object recognition, such as segmentation, detection,

and pose estimation (Chen et al. 2018, 2022; Liu et al. 2020, 2021; Xie et al. 2020; Sajjan et al. 2020; Fang et al. 2022). Still, these RGB-based datasets suffer from the inconsistent texture of transparent objects, which are susceptible to background interference. To tackle these challenges, alternative sensors have been examined, including light-field (Xu et al., 2015; Zhou et al., 2020) and polarized cameras (Mei et al. 2022). Despite their potential, these light-dependent sensors may lead to longer processing times or higher noise levels due to optical effects.

Recently, another alternative sensor, thermal infrared (TIR) camera, has been introduced in RGB-T dataset (Huo et al. 2023) for their robust sensing capability on transparent objects (specifically glasses). TIR cameras function by measuring the temperature emitted from the object surface and operate within the wavelength range of 8–12  $\mu\text{m}$ , which does not penetrate the typical material of transparent

<sup>1</sup>Department of Mechanical Engineering, SNU, Seoul, Korea

<sup>2</sup>Institute of Advanced Machines and Design, SNU, Seoul, Korea

\*The authors contributed equally to this paper

## Corresponding author:

Ayoung Kim, Department of Mechanical Engineering, SNU, Bldg 301 Rm 1515, 1 Gwanak-ro Gwanak-gu, Seoul 08826, Korea.  
Email: [ayoungk@snu.ac.kr](mailto:ayoungk@snu.ac.kr)

objects. Valuing a similar potential of TIR but not being limited to glasses, our dataset aims to expand the potential of the TIR imaging for a broad range of transparent objects. As illustrated in Figure 1, the distinctive features of TIR cameras facilitate a straightforward perception of the overall shape of transparent objects, thereby simplifying tasks such as segmentation and pose estimation. Moreover, TIR cameras prove useful in real-life scenarios as they can

penetrate vinyl, enabling observation of objects enclosed in plastic bags (Figure 1). Our dataset encompasses such instances captured to support further research in using TIR for transparent objects.

In our work, we propose a large-scale multispectral dataset for transparent object recognition, TRansPose (TIR-**R**GB dataset for **T**ransparent object **P**ose). We exploit two RGB-D cameras and one TIR camera attached to the end-

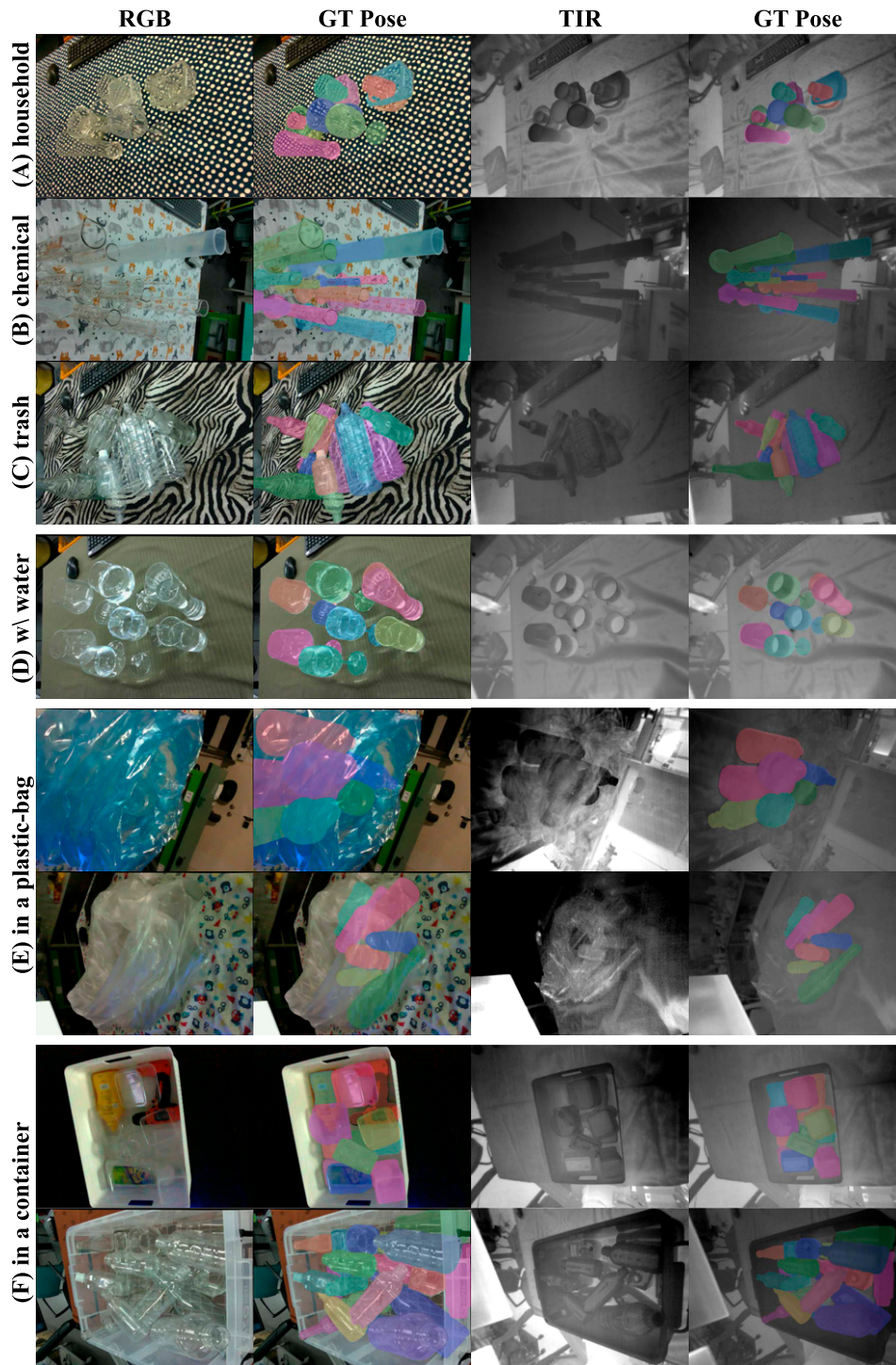


Figure 1. Sample RGB images and TIR images with mask images rendered using the provided annotations.

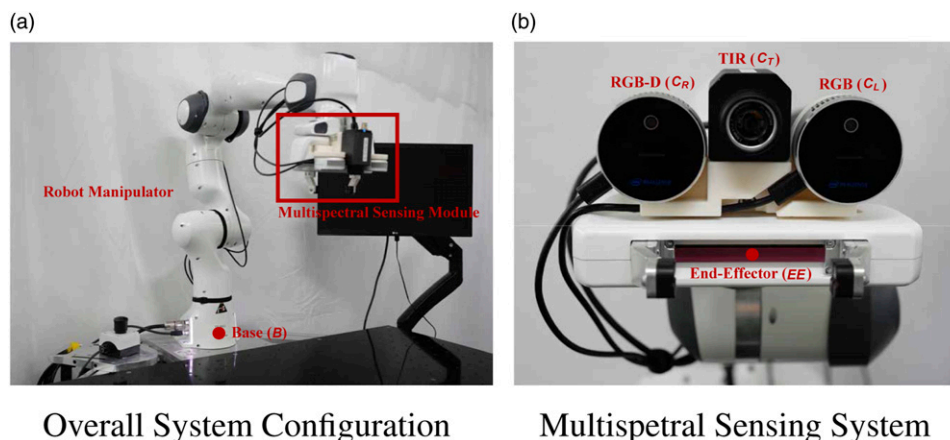
effector of the robot manipulator (Figure 2). Using this multispectral sensing system, we collected data sequences with accurate camera poses relative to the base of the robot manipulator. The proposed dataset includes 99 transparent objects of three types: 43 household objects, 27 recyclable trashes, and 29 chemical laboratory equipment. We provide 87 diverse data sequences with required annotations, including object class, segmentation, and 6D pose. We also offer rendering code to generate result plots as Figure 1. Furthermore, our dataset encompasses typical yet highly demanding scenarios involving transparent objects enclosed within rigid or deformable containers (e.g., objects in a container, objects in a plastic bag, and multi-stacked objects). The proposed dataset aims to push the boundaries of transparent object recognition research and facilitate advancements in the field of robotics, opening up new possibilities for the perception and manipulation of transparent objects. Compared with existing datasets, the proposed dataset possesses the following key contributions:

- (1) We introduce a large-scale multispectral transparent object dataset, TRansPose, incorporating two RGB-D cameras and one TIR camera. This camera configuration makes it straightforward to exploit the object shape information.
- (2) We provide 99 transparent objects of three types (43 household objects, 27 recyclable trashes, and 29 chemical laboratory equipment) and 12 non-transparent objects.
- (3) We provide 87 data sequences on the table-top setup for practical robot manipulation tasks. These sequences include various challenging environments: diverse lighting conditions, heavy clutters, objects in the container, objects in a plastic-bag, object filled with water, and multi-stacked objects.
- (4) For each data sequence, we provide multispectral images, instance-level segmentation, ground-truth pose, completed depth, as well as 3D object models.

## 2. Related works

In recent years, datasets for transparent object recognition have proliferated. Table 1 summarizes recent datasets for transparent object recognition. As mainstream, most researchers were focused on the RGB-based datasets (Bashkirova et al., 2022; Chen et al., 2018, 2022; Dai et al., 2022; Fang et al., 2022; Jiang et al., 2022; Jiang and Shan Li, 2022b; Lin et al., 2021; Liu et al., 2020, 2021; Mei et al., 2020; Proença and Simoes, 2020; Sajjan et al., 2020; Xie et al., 2020; Xu et al., 2022). These datasets have necessitated background clutters using additional artifacts to overcome the attribute of the transparent objects heavily dependent on the background. Notably, ClearPose (Chen et al. 2022) made significant strides by providing a comprehensive collection of 62 transparent objects, including those commonly encountered daily and used in controlled environments. This dataset not only offers precise pose annotations but also includes instance segmentation. What sets ClearPose apart is its inclusion of diverse objects, encompassing variations in size, shape, and category. However, these RGB-based datasets do not adequately address the challenges arising from the unpredictable surface characteristics of transparent objects and their susceptibility to disruptions caused by the background.

Some studies have also incorporated alternative sensor modalities to tackle challenges provoked by light reflection and refraction in transparent objects. Notably, light-field cameras capable of capturing both the intensity and direction of light have proven valuable in capturing the position and shape of transparent objects (Xu et al., 2015; Zhou et al., 2020). Similarly, polarized cameras have emerged as a suitable choice for overcoming issues related to light reflection and glare while simultaneously enhancing contrast (Kalra et al. 2020; Mei et al. 2022; Wang et al. 2022). However, the nature of all visible light, such as reflections and refractions within transparent objects and susceptibility to background interference, invokes challenges when applying datasets to transparent object perception tasks, resulting in noise-contaminated and incomplete outcomes.



**Figure 2.** System configuration for the TRansPose dataset. A coordinate system is defined based on the base of the robot manipulator. (a) Overall System Configuration and (b) Multispectral Sensing System.



**Table 1.** Comparison With Existing Transparent Datasets on Sensor Modality, the Number of Transparent Objects (#transparent obj) and Total Object (#total obj), the Number of Frames in the Real-World (#frame), the Total Number of Pose Annotations (#pose annotation), the Total Number of Sequences (#seq), and object classes. The Mark  $\times$  Indicates That the Dataset Consists of Unspecified Number of Objects Like Glass Walls or Window.

Dataset	Modality	#transparent obj	#total obj	#frame	#pose annotation	#seq	Object classes
Tom-Net (Chen et al. 2018)	RGB	14	14	876	seg only	—	Household
GDD (Mei et al. 2020)	RGB	$\times$	$\times$	3,916	seg only	—	In-the-wild glass
Trans10K (Xie et al. 2020)	RGB	$\times$	$\times$	~10K	seg only	—	Household In-the-wild glass
GSD (Lin et al. 2021)	RGB	$\times$	$\times$	4,012	seg only	—	In-the-wild glass
TACO (Proença and Simoes, 2020)	RGB	$\times$	$\times$	1,499	seg only	—	Trash
TransTouCh (Jiang et al. 2022c)	RGB	9	9	180	seg only	—	Household
ZeroWaste (Bashkirova et al. 2022)	RGB	$\times$	$\times$	~11K	seg only	—	Trash
TransCG (Fang et al. 2022)	RGB-D	51	51	~58K	~0.2M	130	Household
ClearGrasp (Sajjan et al. 2020)	RGB-D	10	10	286	736	—	Household, Trash
ClearPose (Chen et al. 2022)	RGB-D	63	72	~350K	~5.0M	51	Household Chemical Equipment
TODD (Xu et al. 2022)	RGB-D	6	6	~15K	~0.1M	49	Chemical Equipment
Dex-NeRF (Ichnowski et al. 2022)	RGB-D	6	6	516	$\times$	8	Household Chemical Equipment
TRANS-AFF (Jiang et al. 2022a)	RGB-D	8	8	1,346	seg only	—	Household
STD (Dai et al. 2022)	RGB-D	22	50	~27K	~139K	30	Household
StereoObj1M (Liu et al. 2021)	Stereo RGB	7	18	~393K	~1.5M	182	Chemical Equipment
TOD (Liu et al. 2020)	Stereo RGB-D	15	15	~48K	~0.1M	40	Household
Polarised (Kalra et al. 2020)	RGB Polarisation	6	6	1,600	seg only	—	Household
RGB-P (Mei et al. 2022)	RGB Polarisation	$\times$	$\times$	4,511	seg only	—	In-the-wild glass
PhoCaL (Wang et al. 2022)	RGB-D Polarisation	8	60	3,951	~91K	20	Household
TransCut (Xu et al. 2015)	Light-field	7	7	49	seg only	—	Household
ProLIT (Zhou et al., 2020)	Light-field	6	6	300	421	—	Household
RGB-T (Huo et al. 2023)	RGB TIR	$\times$	$\times$	5,551	seg only	—	In-the-wild glass
TRansPose (ours)	Stereo RGB-D TIR	<b>99</b>	<b>111</b>	<b>~100K</b>	<b>~3.9M</b>	<b>87</b>	Household, Trash Chemical Equipment

To address these limitations, Huo et al. (2023) proposed RGB-T dataset which exploits RGB and TIR camera pair. Utilizing the inherent characteristic of TIR cameras with their zero transmittance for transparent objects, this dataset is harnessed to detect transparent entities, but only for window glasses.

Most of the existing datasets are restricted to specific domains due to the deficient number of objects. These also suffer from a lack of realism, artificially arranged and labeled objects and scenes deviating from real-world conditions. Since these datasets fail to capture the full spectrum of transparent objects found in everyday scenarios, it is arduous to reflect the real-world scenarios commonly encountered daily, such as occlusions, diverse lighting conditions, and complex object

arrangements. In our work, we provide 99 transparent objects considering more categories than other existing datasets and challenging scenes, including objects filled with water, diverse lighting conditions, heavy clutters, objects in non-transparent or translucent containers, objects in plastic-bags, and multi-stacked objects. Additionally, our work provides about 3.9M pose annotations for target objects presented in all sequences.

### 3. System description

#### 3.1. System setup

Figure 2 shows an overall system configuration. We incorporate two Intel RealSense L515 RGB cameras and an

FLIR A65 TIR camera as a multispectral sensing module. To avoid laser interference between these two sensors, only one camera  $C_R$  obtains depth maps. The robot manipulator is installed on the workspace table, which is  $1250 \times 850$  mm. The multispectral sensing module is mounted on the end-effector of the robot manipulator. The sensor specifications used in the proposed dataset are summarized in Table 2.

### 3.2. System calibration

To utilize the image set obtained from the multispectral sensing module, we need to estimate camera parameters and the extrinsic relationship between all sensors.

**3.2.1. Intrinsic calibration.** For camera parameters of RGB-D cameras, camera calibration is performed using a checkerboard (Zhang, 1999). For the TIR camera, we encounter a challenge as it cannot accurately detect the corners of a conventional checkerboard. To surmount this limitation, we utilize a manually designed checkerboard composed of different materials (Saponaro et al. 2015), each capable of undergoing selective heating. The average re-projection calculated with the estimated camera parameters is approximately 0.2 pixels.

**3.2.2. Extrinsic calibration.** We execute the hand-eye calibration (Tsai and Lenz, 1989) for the extrinsic calibration between all sensors, obtaining three transformations ( ${}^{EE}T_{C_R}$ ,  ${}^{EE}T_{C_T}$  and  ${}^{EE}T_{C_L}$ ). We fix a calibration pattern on the workspace table and manually capture the images by moving the robot manipulator. For the RGB-D cameras, we use a ChArUco board. For the TIR camera, we use the same checkerboard with the process of intrinsic calibration. The average rotation and translation error are approximately  $0.3^\circ$  and 2 mm, respectively.

### 3.3. Data collection

3D CAD models of the objects are generated using a 3D scanner. To scan the transparent objects, we cover their surface with a suitable material to ensure complete mesh generation. Additionally, for objects like wine glasses or cylinders where scanning the interior is difficult, we fill them out with additional materials to create a solid mesh representation.

For sequential data acquisition, synchronization between sensors is crucial for the proper usage of image set. However, we confront two problems that prevent the synchronization of all sensors. First, the framerate of each sensor is different, so synchronization is required through external signals. Unfortunately, the multispectral sensing module composed in this work does not inherently support this. Second, the TIR camera is susceptible to noise caused by internal heat, resulting in increased image noise over time. To address this issue, this sensor incorporates Non-Uniformity Correction (NUC) function to compensate for the inherent noise and improve image quality. This process causes a delay of approximately 1s in the data acquisition, making it challenging to synchronize with other sensors.

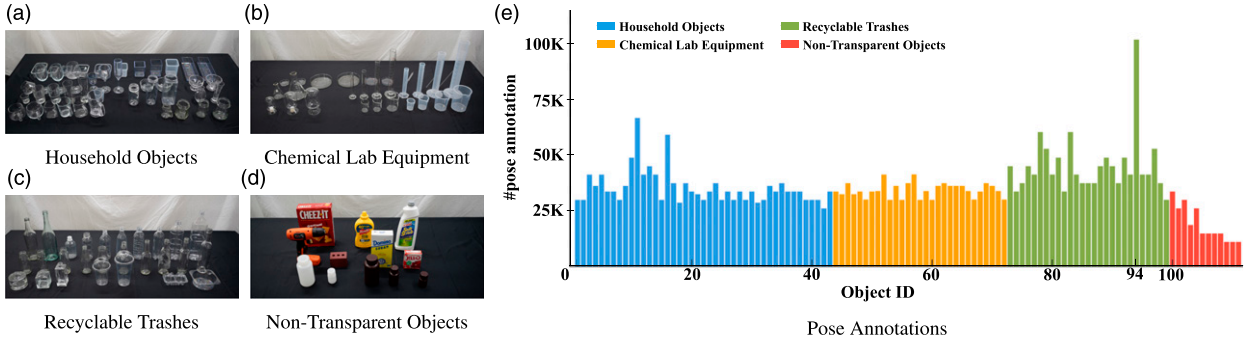
To overcome the abovementioned problems, our work involves discrete data capturing. In advance, we pre-define 14 waypoints as the trajectory of the robot manipulator. We set up 90 steps between each waypoint, capturing data at every step. As a result, joint angles of the robot manipulator, two RGB images, a depth image, and a TIR image can be obtained simultaneously at 30fps without motion blur.

## 4. Data annotation

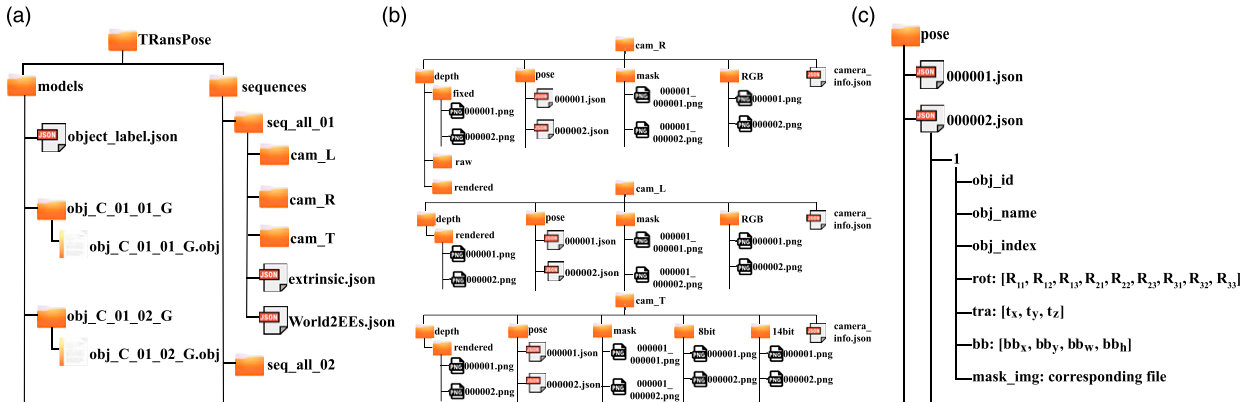
Most target objects have transparent material types, so we cannot obtain proper depth maps, making it challenging to align the 6D object poses. To tackle this, we employ ProgressLabeller to annotate the 6D object poses ( ${}^{C_R}T_{obj}$ ,  ${}^{C_T}T_{obj}$  and  ${}^{C_L}T_{obj}$ ), instance-segmentation masks, and ground truth depth maps. ProgressLabeller is a manual labeling tool for 6D object pose plugged into Blender, proposing a sophisticated multi-view silhouette matching technique to align objects in a 3D interactive workspace. Since ProgressLabeller utilizes the monocular RGB-D sequence only for object pose labeling, we enhance the ProgressLabeller to utilize stereo RGB and TIR images and the end-effector pose of the robot manipulator simultaneously. We import the following: stereo RGB and TIR images, 3D CAD models, camera poses, and intrinsic and extrinsic parameters. Since ProgressLabeller utilizes visual simultaneous localization and mapping (SLAM) to estimate the camera poses, it cannot always ensure to estimate accurate camera poses. In our work, we utilize the end-effector pose of the robot manipulator and extrinsic transformations ( ${}^{EE}T_{C_R}$ ,  ${}^{EE}T_{C_T}$  and  ${}^{EE}T_{C_L}$ ) between the end-effector of the robot manipulator

**Table 2.** Sensors Used in the Multispectral Sensing Module.

Sensors	Type	Manufacturer	Model	FPS	Resolution	Data used	
Camera	$C_R$	RGB-D	Intel RealSense	L515	30 Hz	$640 \times 480$	RGB, depth
	$C_T$	TIR	FLIR	A65	30 Hz	$640 \times 512$	14-bit TIR
	$C_L$	RGB-D	Intel RealSense	L515	30 Hz	$640 \times 480$	RGB
Robot manipulator	Positional encoder	Franka Emika	Panda	1000 hz	14bit	Joint angle	



**Figure 3.** (a)–(d) The objects included in the TRAnsPose dataset. (e) The number of pose annotations per object. In some sequences, a multiple-instance of the object 94 appears, its pose annotations are numerous.



**Figure 4.** Structure of provided dataset. (a) TRAnsPose includes models and sequences folder. Models folder includes 3D CAD models and object\_label.json which has a dictionary paired with the ID and name of objects. In sequences folder, all data are provided according to the sequence index. (b) All data are separated depending on the sensor. For all sensors, images, masks, depth maps, poses, and camera information are provided. Additionally, for the right RGB camera, raw and completed depth maps are provided. In the case of TIR camera, we provide 14-bit raw images and 8-bit images applied simple min–max normalization. (c) Pose annotations are provided as JSON extension.

and three cameras. As a result, we can produce more accurate and consistent annotations.

Manually aligning the projected object model across all multi-view images is time-consuming and requires a lot of delicacy to make accurate annotations. To alleviate these issues, we implement object-wise mask-based optimization. For this process, we manually annotate object mask for approximately five to eight images captured from various viewpoints, encompassing both TIR and stereo RGB. This process is executed in only each single image so that it is more easier than the aligning process across all multi-view images. Then, we optimize the 6D object pose, minimizing the loss  $\mathcal{L}$  between the images obtained by projecting the CAD model of the target object and the labeled mask.

We define a rendering operator as  $S = \text{Render}(K, T, P)$ , which renders object points  $P$  given camera intrinsic  $K$  and camera pose  $T$  into an object mask  $S$ . Assuming that  ${}^{EE}T$  and  ${}^{World}T$  are already known and  ${}^X_{Obj}T_i$  is obtained by manual labeling process,  ${}^{World}T$  is estimated by minimizing below loss.

$$\mathcal{L} = \sum_{i,X} \left\| \text{Render} \left( K_X, {}^{EE}T^{-1} {}^{World}T^{-1} {}^{World}T, P_{Obj} \right) - \text{Render} \left( K_X, {}^X_{Obj}T_i, P_{Obj} \right) \right\|^2 \quad (1)$$

$P_{Obj}$  and  $K_X$  represent a set of object points and camera intrinsic parameters, respectively.  $X$  represents the type of camera, with  $X \in \{C_L, C_T, C_R\}$ .

Object-wise mask-based optimization proves to be more effective in challenging scenes such as multi-stack or container scenes in terms of accuracy and time consumption, while manual labeling is only effective in simpler scenes.

Based on the labeled 6D object pose, we produce all annotations for object-wise instance segmentation masks and completed depth map. The completed depth maps are generated by replacing the object parts of raw depth images with rendered depth images using object CAD models.

## 5. Dataset description

As shown in Figure 3, TRansPose dataset consists of 99 transparent objects and 12 non-transparent objects. All objects are further categorized as follows:

- 43 Household objects (Figure 3(a)): 2 vases, 2 dishes, 2 bowls, 3 water bottles, 4 jars, 6 cups with handles, 6 cups without handles, 7 wine glasses, and 11 containers.
- 27 Recyclable Trashes (Figure 3(c)): 2 disposable cups, 2 pet bottles, 3 perfume containers, and 20 beverage bottles.
- 29 Chemical Laboratory Equipment (Figure 3(b)): 1 pipette, 2 Erlenmeyer flasks, 2 glass stirring rods, 3 seeds bottles, 6 Petri dishes, 7 beakers, and 8 cylinders.
- 12 Non-Transparent Objects (Figure 3(d)): 5 bottles and 7 general objects used in YCB-Object (Calli et al. 2017).

We provide 87 sequences (333,819 images), which are categorized into four groups: 11 scenes with household objects, 11 scenes with chemical laboratory equipment, 15 scenes with recyclable objects, and 50 scenes with all categories of objects. Additionally, we provide 20 different backgrounds, each characterized by various colors, patterns, and materials such as silk and denim. These sequences encompass several challenging scenes: objects filled with water, diverse lighting conditions, heavy clutters, objects in non-transparent or translucent containers, objects in plastic-bags, and multi-stacked objects. A total of 87 sequences are divided into 61 sequences as trainset and 26 sequences as testset.

Figure 4 depicts the structure of the provided data. All data are provided in a compressed file with the tar.gz extension. The 3D CAD models and model ID of all provided objects are located in `models` folder. All sensor data, related annotations, and calibration values for each sensor are stored in `sequences` folder. The provided annotations include 6D object poses, instance segmentation masks, bounding-box, and completed depth maps. The distribution of the 6D object pose annotations is shown in Figure 3(e). Each transparent object in the provided dataset has at least 26K object pose annotations.

## 6. Conclusion

In this paper, we have presented TRansPose, the first large-scale multispectral transparent object dataset, encompassing TIR images, stereo RGB-D images, annotated pose, mask, and associated labels. It surpasses existing datasets by offering a wider range of categories and objects, enabling its applicability in various environments, including chemical laboratories, households, and recycling centers. Additionally, TRansPose incorporates challenging scenes featuring diverse lighting conditions, heavy clutter, objects inside translucent or non-transparent containers, objects inside

plastic bags, objects filled with water, and multi-stacked objects. This comprehensive dataset aims to facilitate research and advancements in transparent object recognition across a wide array of real-world scenarios.






### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) No.2022-0-00480, Development of Training and Inference Methods for Goal-Oriented Artificial Intelligence Agents. Also, this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (2020R1C1C1006620).

### ORCID iDs

Jeongyun Kim  <https://orcid.org/0000-0002-5019-2894>  
 Myung-Hwan Jeon  <https://orcid.org/0000-0003-3611-0298>  
 Sangwoo Jung  <https://orcid.org/0000-0003-0187-006X>  
 Wooseong Yang  <https://orcid.org/0000-0001-5105-8833>  
 Minwoo Jung  <https://orcid.org/0000-0002-5623-6288>  
 Jaeho Shin  <https://orcid.org/0000-0003-2121-9311>  
 Ayoung Kim  <https://orcid.org/0000-0001-9829-2408>

### References

- Bashkirova D, Abdelfattah M, Zhu Z, et al. (2022) Zerowaste dataset: towards deformable object segmentation in cluttered scenes. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, Silver Spring, MD, USA, 21–23 June 2022.
- Calli B, Singh A, Bruce J, et al. (2017) Yale-CMU-Berkeley dataset for robotic manipulation research. *The International Journal of Robotics Research* 36(3): 261–268.
- Chen G, Han K, Kwan-Yee Kenneth W, et al (2018) Tom-net: learning transparent object matting from a single image. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, Salt Lake City, UT, USA, 18–23 June 2018.
- Chen X, Zhang H, Yu Z, et al. (2022) Clearpose: large-scale transparent object dataset and benchmark. Proc. European Conf. On Comput. Vision, Tel Aviv, Israel, 23–27 October 2022.
- Dai Q, Zhang J, Li Q, et al. (2022) Domain randomization-enhanced depth simulation and restoration for perceiving and grasping specular and transparent objects. Proc. European Conf. On Comput. Vision, Tel Aviv, Israel, 23–27 October 2022.
- Fang H, Fang HS, Xu S, et al. (2022) Transcg: a large-scale real-world dataset for transparent object depth completion and a grasping baseline. *IEEE Robotics and Automation Letters* 7(3): 7383–7390.

- Huo D, Wang J, Qian Y, et al. (2023) Glass segmentation with RGB-thermal image pairs. *IEEE Transactions on Image Processing* 32: 1911–1926.
- Ichnowski J, Avigal Y, Kerr J, et al. (2022) Dex-nerf: using a neural radiance field to grasp transparent objects. Proc. Conf. On Robot Learning, Auckland, New Zealand, 14–18 December 2022
- Jiang J, Cao G, Do TT, et al. (2022a) A4t: hierarchical affordance detection for transparent objects depth reconstruction and manipulation. *IEEE Robotics and Automation Letters* 7(4): 9826–9833.
- Jiang J, Cao G, Butterworth A, et al. (2022c) Where shall I touch? Vision-guided tactile poking for transparent object grasping. *IEEE* 28(1): 233–244.
- Jiang J and Shan LI (2022b) Robotic perception of object properties using tactile sensing. *Tactile Sensing, Skill Learning, and Robotic Dexterous Manipulation*. Amsterdam: Elsevier.
- Kalra A, Taamazyan V, Rao SK, et al. (2020) Deep polarization cues for transparent object segmentation. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, Seattle, WA, USA, 13–19 June 2020.
- Kasper A, Xue Z and Dillmann R (2012) The KIT object models database: an object model database for object recognition, localization and manipulation in service robotics. *The International Journal of Robotics Research* 31(8): 927–934.
- Li WH and Kleeman L (2011) Segmentation and modeling of visually symmetric objects by robot actions. *The International Journal of Robotics Research* 30(9): 1124–1142.
- Lin J, He Z, Lau RWH, et al. (2021) Rich context aggregation with reflection prior for glass surface detection. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, Nashville, TN, USA, 20–25 June 2021.
- Liu X, Jonschkowski R, Angelova A, et al. (2020) Keypose: multi-view 3d labeling and keypoint estimation for transparent objects. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, Seattle, WA, USA, 13–19 June 2020.
- Liu X, Iwase S, Kitani KM, et al. (2021) Stereobj-1m: large-scale stereo image dataset for 6D object pose estimation. *Proc. IEEE Intl. Conf. On Comput. Vision, Montreal, Canada*, 11–17 October 2021.
- Mei H, Yang X, Wang Y, et al. (2020) Don't hit me! glass detection in real-world scenes. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, Seattle, WA, USA, 13–19 June 2020.
- Mei H, Dong B, Dong W, et al. (2022) Glass segmentation using intensity and spectral polarization cues. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, New Orleans, LA, USA, 18–24 June 2022.
- Novkovic T, Furrer F, Panjek M, et al. (2019) CLUBS: an RGB-D dataset with cluttered box scenes containing household objects. *The International Journal of Robotics Research* 38(14): 1538–1548.
- Proença PF and Simoes P (2020) Taco: Trash Annotations in Context for Litter Detection. arXiv preprint arXiv: 2003.06975.
- Sajjan S, Moore M, Pan M, et al. (2020) Clear grasp: 3d shape estimation of transparent objects for manipulation. *Proc. IEEE Intl. Conf. On Robot. and Automat*, Philadelphia, PA, USA, 23–27 May 2022.
- Saponaro P, Sorensen S, Rhein S, et al. (2015) Improving calibration of thermal stereo cameras using heated calibration board Proc. Intl. Conf. On Image Processing, Quebec City, Quebec, Canada, 27–30 September 2015.
- Saxena A, Driemeyer J and Ng AY (2008) Robotic grasping of novel objects using vision. *The International Journal of Robotics Research* 27(2): 157–173.
- Sinapov J, Bergquist T, Schenck C, et al. (2011) Interactive object recognition using proprioceptive and auditory feedback. *The International Journal of Robotics Research* 30(10): 1250–1262.
- Tsai RY and Lenz R (1989) A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation* 5(3): 345–358.
- Wang P, Jung HJ, Li Y, et al. (2022) Phocal: a multi-modal dataset for category-level object pose estimation with photometrically challenging objects. Proc. IEEE Conf. On Comput. Vision and Pattern Recog, New Orleans, LA, USA, 18–24 June 2022.
- Xie E, Wang W, Wang W, et al. (2020) Segmenting transparent objects in the wild. Proc. European Conf. On Comput. Vision, Glasgow, UK, 23–28 August 2020.
- Xu Y, Nagahara H, Shimada A, et al. (2015) Transcut: transparent object segmentation from a light-field image. *Proc. IEEE Intl. Conf. On. Santiago, Chile: Vision, Comput.*, 7–13. (accessed December 2015).
- Xu H, Wang YR, Eppel S, et al. (2022) Seeing glass: joint point-cloud and depth completion for transparent objects. *Proc. Conf. On Robot Learning*, London, UK, 8–11 November 2022.
- Zeng A, Song S, Yu KT, et al. (2022) Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. *The International Journal of Robotics Research* 41(7): 690–705.
- Zhang Z (1999) Flexible camera calibration by viewing a plane from unknown orientations. *Proceedings of IEEE International Conference on Computer Vision* 1: 666–673.
- Zhou Z, Chen X and Jenkins OC (2020) Lit: light-field inference of transparency for refractive object localization. *IEEE Robot. and Automat. Lett* 5(3): 4548–4555.