

Jornadas de Automática

Calibración automática mano-ojo mediante marcadores en un robot social

García-Gómez, Miguel^{a,*}, Duque-Domingo, Jaime^a, Gómez-García-Bermejo, Jaime^{a,b}, Zalama, Eduardo^{a,b}

^aDepartamento de Ingeniería de Sistemas y Automática, ITAP, Universidad de Valladolid, C/ Dr Mergelina s/n, 47011 Valladolid, España

^bCARTIF Centro Tecnológico 47151, Valladolid, España

To cite this article: García-Gómez, Miguel, Duque-Domingo, Jaime, Gómez-García-Bermejo, Jaime, Zalama, Eduardo. 2025. Automatic hand-eye calibration using markers on a social robot. *Jornadas de Automática*, 46. <https://doi.org/10.17979/ja-cea.2025.46.12123>

Resumen

El presente trabajo expone un enfoque automatizado para la calibración mano-ojo entre una cámara externa y la base de un robot, utilizando un marcador ArUco como referencia visual común. Se emplea una cámara montada en la muñeca del robot para capturar una imagen del marcador, cuya posición relativa se conoce gracias al URDF del propio robot. Paralelamente, una cámara externa, posicionada libremente en el entorno, toma otra imagen del mismo marcador. A partir de estas dos observaciones y conociendo la transformación entre el marcador y cada cámara, se estima la transformación entre la cámara externa y la base del robot. Este proceso permite alinear sistemas de visión externos con el marco de referencia del robot de forma precisa y sin intervención manual, facilitando el cálculo de la posición de una cámara externa respecto a la base del robot.

Palabras clave: Percepción y detección, Fusión de información y sensores, Robótica integrada, Robots móviles, Tecnología robótica.

Automatic hand-eye calibration using markers on a social robot.

Abstract

This paper presents an automated approach for hand-eye calibration between an external camera and a robot base, using an ArUco marker as a shared visual reference. A camera mounted on the robot's wrist captures an image of the marker, whose relative position is known from the robot's URDF. Simultaneously, an external camera, arbitrarily placed in the environment, captures another image of the same marker. From these two observations and knowing the transformation between the marker and each camera, the transformation between the external camera and the robot base is estimated. This approach enables accurate alignment of external vision systems with the robot's reference frame, eliminating the need for manual calibration, facilitating the calculation of the position of an external camera with respect to the robot base.

Keywords: Perception and sensing, Information and sensor fusion, Embedded robotics, Mobile robots, Robotics technology.

1. Introducción

En la actualidad, muchos sistemas robóticos utilizan cámaras para comprender su entorno e interactuar con él. Estas cámaras pueden estar montadas en el propio robot o colocadas externamente para proporcionar una visión más amplia del espacio de trabajo (Shahria et al., 2022). Para que el robot pueda utilizar la información de estas cámaras, debe saber exactamente dónde se encuentra cada cámara en relación con

su propia base. En esto consiste la calibración ojo-mano: en averiguar la transformación entre el marco de coordenadas del robot y el marco de la cámara (Jiang et al., 2022).

Durante las fases de desarrollo de sistemas robóticos, las cámaras utilizadas para la percepción visual no están físicamente fijadas a la estructura del robot. En lugar de eso, pueden estar montadas sobre soportes o trípodes. Esto significa que su posición puede variar levemente o de forma intencionada, lo que rompe la relación conocida con el sistema de coordenadas

* Autor para correspondencia: miguel.garcia.gomez@uva.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

del robot. Cuando esto ocurre, es necesario volver a calibrar la cámara externa con respecto a la base del robot para que los datos visuales puedan integrarse correctamente en el sistema. El problema es que estas calibraciones suelen requerir intervención manual, y no siempre es evidente cuándo la cámara se ha movido lo suficiente como para necesitar una nueva calibración. Esto lleva a un escenario donde, por seguridad, se recalibra más veces de las necesarias, o peor aún, se trabaja con una calibración obsoleta sin saberlo. Ambas situaciones afectan negativamente al rendimiento del sistema.

En este trabajo se propone un método automático para calibrar la posición de una cámara externa respecto a la base del robot sin necesidad de intervención manual. La clave del enfoque está en utilizar un marcador ArUco como referencia común visible para dos cámaras: una que va montada en la muñeca del robot y una cámara externa. Gracias al URDF (*Unified Robot Description Format*) del robot, se conoce con precisión la posición de la cámara de muñeca en el sistema de coordenadas del robot. Al capturar una imagen del mismo marcador con ambas cámaras, es posible calcular sus posiciones relativas respecto al marcador y, a partir de ahí, obtener la transformación entre la cámara externa y la base del robot.

A diferencia de otros métodos que requieren que la cámara externa observe directamente al brazo del robot para estimar su pose (Lee et al., 2020; Tremblay et al., 2020), este sistema solo necesita que ambas cámaras se orienten hacia el mismo marcador. Esto simplifica mucho la operación y da más libertad sobre dónde colocar la cámara externa. Además, permite repetir la calibración de forma rápida y frecuente, lo que es especialmente útil en entornos de desarrollo donde la configuración física del robot aún no es definitiva.

Con este método, se facilita el uso de visión externa en sistemas robóticos de forma práctica, permitiendo integrar información visual externa sin depender de calibraciones largas o configuraciones fijas. También sienta una base para extender la calibración a sistemas multivisión en futuras etapas del desarrollo.

2. Trabajos relacionados

El problema de saber “dónde están los ojos del robot” no es nuevo. Tras décadas de investigación en robótica, la calibración mano-ojo se ha consolidado como una de esas tareas que, aunque conceptualmente bien entendidas, siguen siendo tediosas de aplicar en entornos reales. La teoría está resuelta desde hace tiempo, pero la práctica sigue llena de matices.

Con el avance de la visión por computador, comenzaron a surgir técnicas basadas en la detección de patrones visuales conocidos, como tableros de ajedrez o, más recientemente, marcadores tipo ArUco o AprilTags (Kalaitzakis et al., 2021). Estos marcadores permiten extraer de forma automatizada correspondencias entre puntos 3D y 2D, lo cual facilita enormemente el cálculo de la pose relativa entre el robot y su entorno. En particular, la familia ArUco ha demostrado ser especialmente útil en entornos de laboratorio debido a su rápida detección, resistencia a occlusiones parciales y precisión (Garrido-Jurado et al., 2014). Un ejemplo de su uso se encuentra en el trabajo de (Yin et al., 2024), quienes propusieron un marco de calibración para robots industriales en espacios de trabajo de

gran tamaño utilizando mapas de marcadores ArUco combinados con visión monocular, permitiendo su reconfiguración sin intervención humana directa.

Muchos métodos de calibración basados en visión emplean el propio brazo robótico para posicionar la cámara externa. Estos enfoques suelen aprovechar características geométricas o puntos de referencia definidos del robot, como las articulaciones, para establecer correspondencias espaciales entre el marco de coordenadas del robot y el de la cámara. Esto es aceptable cuando puedes fijar la cámara en un punto del espacio que permita observar al robot, pero se vuelve restrictivo cuando se quiere que una cámara funcione como los “ojos” del robot, es decir, situada en una posición más natural a los humanos, como una cabeza o un punto elevado en el entorno. Métodos como los de (Lu et al., 2023; Tang et al., 2024; Li et al., 2024), que emplean sensores RGB-D y detección de extremidades del robot, ofrecen soluciones interesantes pero siguen asumiendo esa visibilidad directa. Además, estas técnicas se suelen basar en nubes de puntos e incluso redes neuronales. Aunque estos métodos son atractivos en teoría, en la práctica suelen requerir hardware específico, como sensores RGB-D, o grandes volúmenes de datos para entrenamiento, lo que limita su aplicación generalizada.

Cuando la cámara no se tiene fija al cuerpo del robot, se puede mover en cualquier momento, cambiando las condiciones del entorno. En ese escenario, los métodos tradicionales, por precisos que sean, resultan poco útiles si dependen de una intervención manual cada vez que se cambia algo.

Frente a este panorama, nuestro enfoque adopta la robustez y simplicidad de los marcadores visuales, utilizándolos de forma estratégica para eliminar la necesidad de la manipulación del entorno por parte del usuario. Al hacerlo, abre la puerta a una calibración automática, repetible y compatible con configuraciones del sistema cambiantes.

En resumen, mientras que los métodos existentes han resuelto bien el “qué” del problema, sigue habiendo espacio para mejorar el “cómo”: cómo hacer que la calibración sea tan transparente como cualquier otro servicio dentro del robot. Esa es la contribución que se busca hacer con el presente trabajo.

3. Configuración del sistema

El sistema propuesto, representado en la Figura 1, está formado por dos brazos robóticos Kinova Gen3 montados sobre una columna con desplazamiento vertical, diseñada para integrarse en una plataforma móvil que le proporcionará capacidades de navegación autónoma. Como parte de su percepción visual, se incorporará una cámara que actuará como los “ojos” del robot. Actualmente, durante la fase de desarrollo del robot, esta cámara se encuentra provisionalmente instalada en un trípode, aunque está previsto que en versiones futuras quede fijada de forma permanente a la base del sistema. Por ello, disponer de un sistema de calibración automática resulta fundamental para evitar tener que repetir manualmente el proceso cada vez que la cámara cambia de posición. Esto permite que el sistema se adapte de forma rápida y precisa sin necesidad de intervención humana.

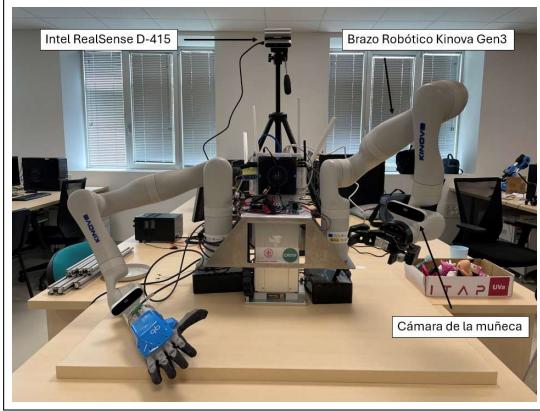


Figura 1: Configuración del sistema real.

Los brazos robóticos tienen una cámara en sus respectivas muñecas. La transformación entre la base del robot y la cámara de muñeca es conocida a través del URDF del robot y se denota como:

$$T_{\text{base} \rightarrow \text{cam.muñeca}}$$

4. Metodología

El objetivo de esta investigación es calcular la transformación entre una cámara externa, que actúa como sistema de visión del robot, y la base del robot, sin necesidad de realizar una calibración manual ni de observar directamente el robot con la cámara externa. Para ello, se utiliza como referencia un marcador ArUco visible por las dos cámaras: la cámara de la muñeca del robot y la cámara externa, como se puede observar en la Figura 2. En la Figura 3 se muestra la vista del marcador desde ambas perspectivas.

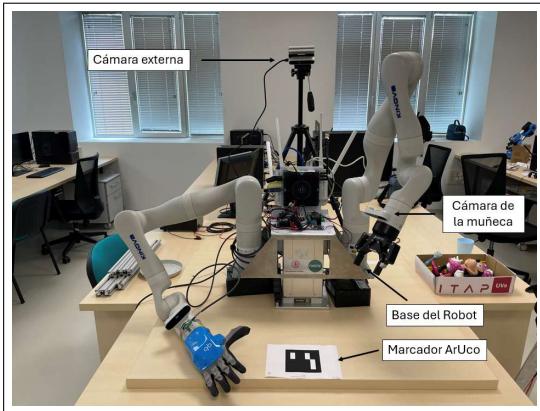


Figura 2: Ambas cámaras visualizan el marcador ArUco.

4.1. Captura de datos en múltiples posiciones

Para mejorar la robustez de la calibración, se capturan múltiples observaciones del marcador moviendo el brazo robótico a distintas configuraciones. En cada posición se obtienen las siguientes transformaciones:

- De la cámara de muñeca respecto al marcador:

$$T_{\text{cam.muñeca} \rightarrow \text{aruco}}$$

- De la cámara externa respecto al marcador:

$$T_{\text{cam.externa} \rightarrow \text{aruco}}$$

- De la base a la cámara de muñeca:

$$T_{\text{base} \rightarrow \text{cam.muñeca}}$$

Para obtener las poses relativas entre el marcador y cada cámara, se resuelve el problema de proyección perspectiva a partir de un conjunto de correspondencias entre puntos 3D del marcador y sus proyecciones 2D en la imagen. Sea $\mathbf{X}_i \in \mathbb{R}^3$ el punto i del marcador en su sistema de coordenadas (en este caso, las esquinas del marcador ArUco), y $\mathbf{x}_i = (u_i, v_i)$ su proyección en la imagen. La relación entre ambos está dada por el modelo de cámara pinhole en coordenadas homogéneas, mostrado en la Ecuación 1.

$$s \cdot \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{K} \cdot (\mathbf{R} \cdot \mathbf{X}_i + \mathbf{t}) \quad (1)$$

donde:

- $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ es la matriz de parámetros intrínsecos de la cámara, que incluye f_x, f_y, c_x, c_y ,
- $\mathbf{R} \in SO(3)$ es la matriz de rotación que alinea el sistema del marcador con el sistema de la cámara,
- $\mathbf{t} \in \mathbb{R}^3$ es el vector de traslación,
- $s \in \mathbb{R}^+$ es un factor de escala que representa la profundidad del punto proyectado en el espacio de imagen homogéneo.

La estimación de la pose consiste en encontrar los parámetros \mathbf{R} y \mathbf{t} que mejor reproducen estas proyecciones, minimizando el error de reproyección sobre todos los puntos observados (ver Ecuación 2), donde $\hat{\mathbf{x}}_i$ es la proyección estimada del punto 3D \mathbf{X}_i usando el modelo anterior.

$$\min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^N \|\mathbf{x}_i - \hat{\mathbf{x}}_i(\mathbf{R}, \mathbf{t})\|^2 \quad (2)$$

Para resolver este problema, se ha empleado el algoritmo propuesto en Collins and Bartoli (2014), conocido como *Infinitesimal Plane-Based Pose Estimation* (IPPE), que está especialmente diseñado para estimar la pose de objetos planos a partir de correspondencias de puntos entre el plano y la imagen. Una vez obtenidos \mathbf{R} y \mathbf{t} , se construye una matriz de transformación homogénea:

$$T = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \in SE(3) \quad (3)$$

Esta matriz representa la transformación desde el sistema de coordenadas del marcador al sistema de coordenadas de la cámara, y se calcula para ambas cámaras en cada observación.



Figura 3: Marcador ArUco visto desde la cámara externa (izquierda). Marcador ArUco visto desde la cámara del brazo (derecha).

A partir de estos datos, se calcula una transformación estimada de la base a la cámara externa para cada muestra, como se indica en la Ecuación 4.

$$T_{\text{base} \rightarrow \text{cam_externa}}^{(i)} = T_{\text{base} \rightarrow \text{cam_muñeca}}^{(i)} \cdot T_{\text{cam_muñeca} \rightarrow \text{aruco}}^{(i)} \cdot (T_{\text{cam_externa} \rightarrow \text{aruco}}^{(i)})^{-1} \quad (4)$$

4.2. Ajuste de la transformación global

Dado un conjunto de N observaciones, se calcula una transformación única $T_{\text{base} \rightarrow \text{cam_externa}}$ como una aproximación promedio de las transformaciones individuales obtenidas. Para ello, los componentes de rotación y traslación se tratan por separado. La rotación se estima mediante descomposición en valores singulares (DVS), aplicada a las matrices de rotación derivadas de cada observación. La traslación, por su parte, se obtiene como el centroide promedio de todos los vectores de traslación estimados.

4.3. Implementación del sistema

La lógica completa del proceso de adquisición y cálculo de la transformación está implementada como un nodo ROS desarrollado en Python. Este nodo procesa imágenes capturadas por dos cámaras RGB sincronizadas, detecta los marcadores ArUco presentes en la escena y extrae las coordenadas de las esquinas del marcador en cada imagen. Utilizando los parámetros intrínsecos de cada cámara, se reconstruye la pose relativa entre el marcador y cada cámara mediante el procedimiento antes descrito, a partir de correspondencias 2D–3D y ajuste no lineal.

El resultado final se representa mediante una transformación homogénea que define la posición y orientación de la cámara externa con respecto a la base del robot. Esta transformación se guarda en un archivo para su uso posterior, quedando disponible para cualquier módulo del sistema de percepción que la requiera.

4.4. Aplicación del resultado

Una vez estimada la transformación óptima, esta puede publicarse en ROS o integrarse directamente en los nodos de percepción y planificación del robot. De esta forma, la cámara externa queda correctamente alineada con el marco de referencia del robot y sus datos visuales pueden usarse como parte del sistema sensorial. En la Figura 4 se pueden observar los tres sistemas de coordenadas una vez estimada la posición de la cámara externa.

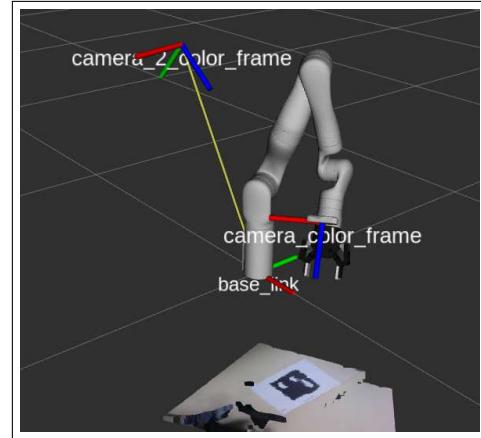


Figura 4: Sistemas de coordenadas de la base del robot y ambas cámaras mostrados en RViz.

Este proceso puede repetirse tantas veces como sea necesario, especialmente si la cámara externa ha sido desplazada o recolocada.

5. Resultados

Una vez obtenida la transformación entre la base del robot y la cámara externa, se ha procedido a validar su precisión mediante un experimento diseñado para evaluar el comportamiento del sistema en una tarea de localización espacial.

Para evaluar la validez de la calibración, se ha utilizado el marcador ArUco colocado en una posición conocida respecto a la base del robot.

El marcador fue observado exclusivamente por la cámara externa. A partir de la imagen obtenida, se estima su posición 3D en el sistema de coordenadas de la cámara como:

$$X = \frac{(x - c_x) \cdot Z}{f_x}, \quad Y = \frac{(y - c_y) \cdot Z}{f_y}, \quad Z = Z \quad (5)$$

Consiguiendo la posición 3D del marcador con respecto a la cámara externa:

$$\mathbf{p}_{\text{cam_externa}} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (6)$$

Posteriormente, se transforma esta posición al marco del robot utilizando la transformación previamente calibrada:

$$\mathbf{p}_{\text{base}} = (\mathbf{T}_{\text{base} \rightarrow \text{cam_externa}})^{-1} \cdot \mathbf{p}_{\text{cam_externa}} \quad (7)$$

La estimación obtenida es comparada con la posición real del marcador, medida manualmente. Se ha calculado el error de traslación como la distancia euclídea entre la posición estimada y la real (ver Ecuación 8).

$$e = \sqrt{(x_{\text{est}} - x_{\text{real}})^2 + (y_{\text{est}} - y_{\text{real}})^2 + (z_{\text{est}} - z_{\text{real}})^2} \quad (8)$$

Este procedimiento se ha repetido en 10 ubicaciones conocidas diferentes dentro del espacio de trabajo.

Los errores observados se resumen a continuación:

- Error medio de posición: 1.5 cm

- Desviación estándar: 0.47 cm
- Error máximo observado: 2.2 cm

Aunque en la mayoría de los casos la posición estimada fue adecuada para tareas de percepción robótica, se observaron desviaciones crecientes al aumentar la distancia entre el marcador y la cámara externa. A mayor distancia, la estimación se desviaba más respecto a la posición real, posiblemente por la pérdida de precisión en la detección del marcador y a un incremento en la incertidumbre de las mediciones de profundidad proporcionadas por sensores RGB-D. Además, la estimación de profundidad puede verse comprometida en superficies oscuras, ya que absorben la luz infrarroja proyectada por la cámara RealSense. Esto impide una correcta reflexión del patrón de luz necesario para calcular la profundidad, generando datos erróneos o incluso pérdida total de información en esas zonas, lo que afecta la precisión en la localización del marcador.

Los resultados del experimento demuestran que la transformación estimada permite ubicar correctamente objetos reales en el entorno del robot utilizando únicamente la cámara externa. Esto confirma que el sistema puede emplearse de manera confiable para su posterior uso en tareas como recogida de objetos, teleoperación mediante realidad virtual o detección de colisiones a través de la cámara externa.

6. Conclusiones

El sistema de calibración automática que se ha desarrollado en esta investigación está diseñado para facilitar la integración de cámaras externas en sistemas robóticos, especialmente en contextos donde la configuración física del robot aún no está completamente definida. Este tipo de solución resulta particularmente valioso durante las etapas tempranas de desarrollo, en las que se requiere una calibración rápida y repetible. En el caso de nuestro proyecto, donde se está diseñando un robot y aún no se ha definido la ubicación final de la cámara, este sistema ha demostrado ser especialmente útil para incorporar de manera fiable la visión externa en el sistema sensorial del robot.

La propuesta presentada, basada en el uso de un marcador ArUco visible desde dos cámaras y la infraestructura de transformaciones de ROS, nos ha permitido calibrar de forma fiable y sin intervención manual la posición relativa de una cámara externa con respecto a la base del robot. El proceso es rápido, puede repetirse fácilmente tras cambios físicos en el sistema, y no impone restricciones sobre la ubicación de la cámara externa, lo que resulta muy práctico cuando aún no se ha decidido su montaje final.

Durante las pruebas, se ha comprobado que el sistema era capaz de estimar correctamente la posición de objetos en el espacio del robot, con errores por debajo de los 2 cm en distancias típicas de trabajo. Este nivel de precisión ha sido satisfactorio para los requisitos del sistema, especialmente en aplicaciones orientadas a percepción aumentada, localización de objetos y visualización desde un punto de vista externo.

No obstante, en aplicaciones que requieran una mayor precisión, este margen de error podría resultar elevado. En estos casos, se propone complementar la visión externa con la información proveniente de la cámara montada en la muñeca del robot. Esta cámara, al tener una vista más cercana de los objetos durante la manipulación, puede proporcionar mediciones más precisas en fases críticas como la aproximación final.

Además, como posible mejora a futuro, se contempla que el propio robot pueda calibrarse de forma automática al encenderse, simplemente observando sus pinzas u otras partes de su estructura. Esto permitiría que el sistema se configure sin necesidad de marcadores, facilitando así una puesta en marcha aún más ágil y adaptable.

Agradecimientos

La investigación que se presenta en este trabajo ha recibido financiación del proyecto ROSOGAR PID2021-123020 OB-I00 financiado por MCI-N/AEI/10.13039/501100011033/FEDER, UE, y del proyecto EIAROB Financiado por la Consejería de Familia de la Junta de Castilla y León - Next Generation EU.

Referencias

- Collins, T., Bartoli, A., 2014. Infinitesimal plane-based pose estimation. *International journal of computer vision* 109 (3), 252–286.
- Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., Marín-Jiménez, M. J., 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47 (6), 2280–2292.
- Jiang, J., Luo, X., Luo, Q., Qiao, L., Li, M., 2022. An overview of hand-eye calibration. *The International Journal of Advanced Manufacturing Technology* 119 (1), 77–97.
- Kalaitzakis, M., Cain, B., Carroll, S., Ambrosi, A., Whitehead, C., Vitzilaios, N., 2021. Fiducial markers for pose estimation: Overview, applications and experimental comparison of the artag, apriltag, aruco and stag markers. *Journal of Intelligent & Robotic Systems* 101, 1–26.
- Lee, T. E., Tremblay, J., To, T., Cheng, J., Mosier, T., Kroemer, O., Fox, D., Birchfield, S., 2020. Camera-to-robot pose estimation from a single image. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 9426–9432.
- Li, L., Yang, X., Wang, R., Zhang, X., 2024. Automatic robot hand-eye calibration enabled by learning-based 3d vision. *Journal of Intelligent & Robotic Systems* 110 (3), 130.
- Lu, J., Richter, F., Yip, M. C., 2023. Markerless camera-to-robot pose estimation via self-supervised sim-to-real transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21296–21306.
- Shahria, M. T., Sunny, M. S. H., Zarif, M. I. I., Ghommam, J., Ahamed, S. I., Rahman, M. H., 2022. A comprehensive review of vision-based robotic applications: current state, components, approaches, barriers, and potential solutions. *Robotics* 11 (6), 139.
- Tang, T., Liu, M., Xu, W., Lu, C., 2024. Kalib: Markerless hand-eye calibration with keypoint tracking. *arXiv preprint arXiv:2408.10562*.
- Tremblay, J., Tyree, S., Mosier, T., Birchfield, S., 2020. Indirect object-to-robot pose estimation from an external monocular rgb camera. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 4227–4234.
- Yin, Y., Gao, D., Deng, K., Lu, Y., 2024. Vision-based autonomous robots calibration for large-size workspace using aruco map and single camera systems. *Precision Engineering* 90, 191–204.