Automatic Construction of Real-World Datasets for 3D Object Localization using Two Cameras

Joris Guérin **Olivier Gibaru** Eric Nyiri
Stéphane Thiery Jorge Palos

LISPEN, Arts et Métiers ParisTech, Lille FRANCE

IECON 2018 - October 22, 2018

Outline

Motivations

Proposed approach to dataset construction

Conclusions and future work

Stereo vision for robotic manipulation

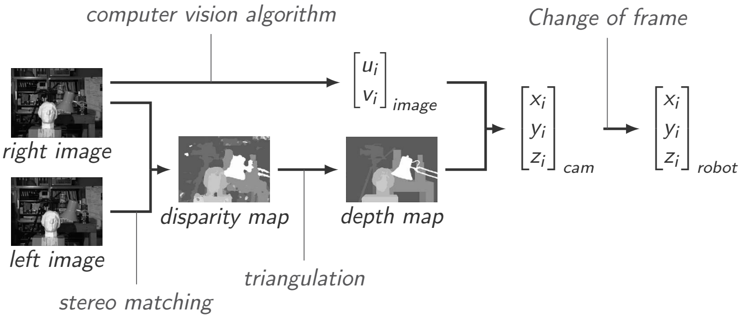
3D object localization using stereo vision has been used for various tasks in robotics

Examples:

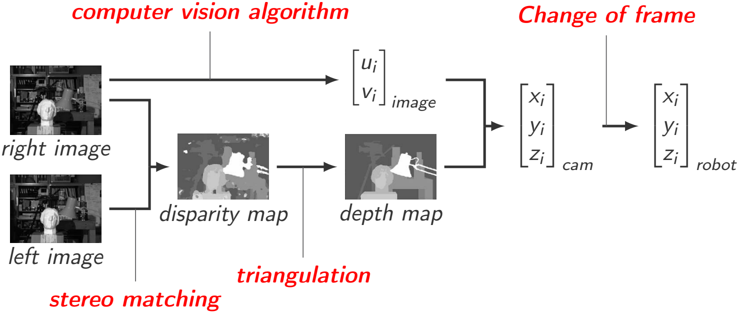
- Grasping ([Azad et al., 2007], [Morales et al., 2006]),
- Contact-reach tasks ([Hudson et al., 2012]),
- Playing soccer ([Käppeler et al., 2010]),
- ...

Classical approach to stereo localization

Stereo localization = Locate a rigid body object in space.



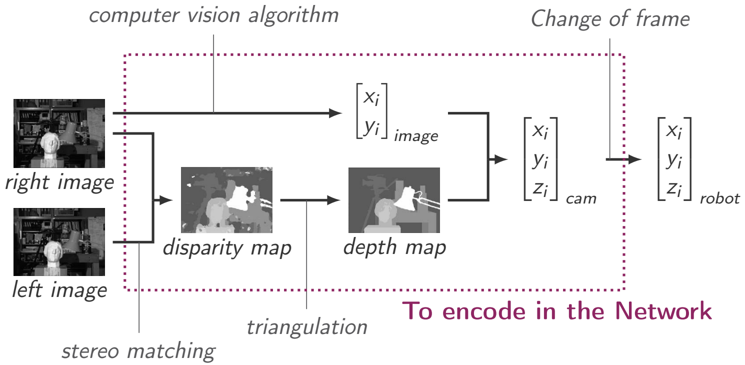
Limitations for robotics manipulation



Many potential sources of errors that can accumulate:

- Cameras calibration / relative calibration,
- Matching errors,
- CV algorithm errors.

Towards an end-to-end approach



Many successes for end-to-end approaches in robotics:

- manipulation ([Levine et al., 2016]),
- self-driving cars ([Bojarski et al., 2016]),
- obstacle avoidance ([Muller et al., 2006]),
- ...

Outline

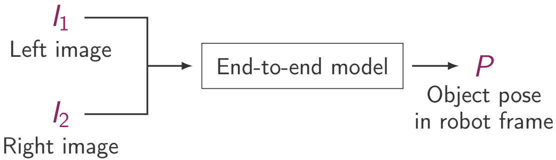
Motivations

Proposed approach to dataset construction

Conclusions and future work

Dataset construction for stereo localization

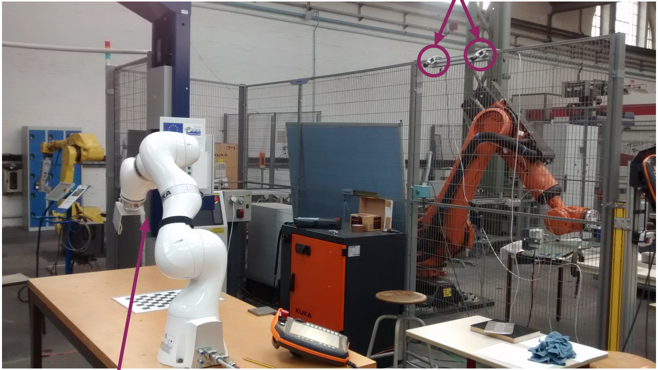
Methodology leveraging a collaborative industrial robot to build a dataset for 3D object stereo localization



$$\text{Dataset} = \text{Tuples } \{I_1, I_2, P\}$$

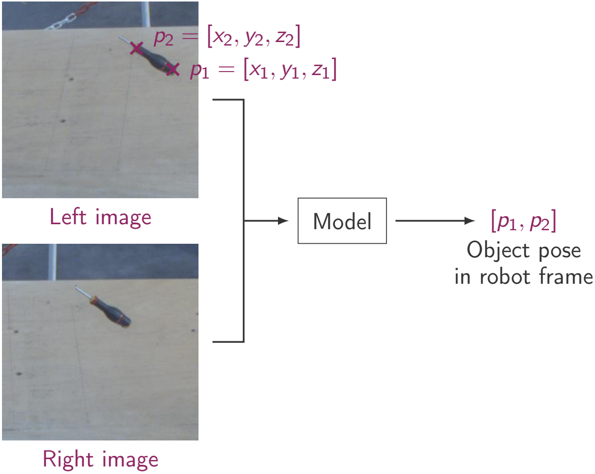
Practical setup

Pair of uncalibrated cameras
looking at the same area



collaborative robot manipulator
handling object to locate

Practical use case: Screw driver localization dataset



$[p_1, p_2]$ in robot frame to be directly usable for manipulation.

Precisely locate the object in the robot frame

Use the precise torque sensors of the iiwa to detect contacts and estimate offsets with the tool holder in every direction.

Using the direct kinematic, the positions of the two points on the screw driver can be computed.

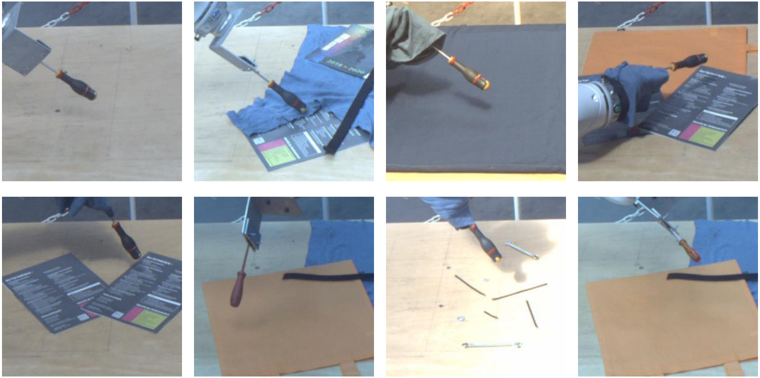
Data collection

Generate random robot movements within the field of view of the two fixed cameras

Synchronized recording of image pairs and tool poses.

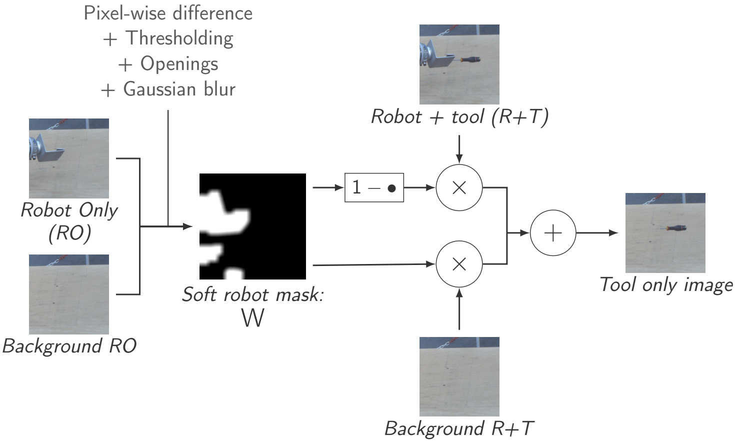
Generate diversity

- Different backgrounds
- Different lighting conditions
- Adding distractors



Avoid strong biases

Remove robot + tool holder from some of the images:
Physically (hide with clothes) and computationally



Outline

Motivations

Proposed approach to dataset construction

Conclusions and future work

Conclusion

Main contributions:

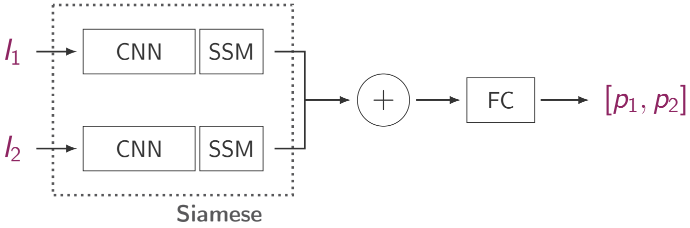
- Methodology for automatic dataset creation with collaborative robot.
- Release of a dataset containing $\approx$ 5000 pairs of screwdriver images with 3D position labels.

Going further:

- Using this dataset to demonstrate efficiency of end-to-end methods

Training a deep CNN: first trials

Architecture:



CNN: conv net / SSM: Spatial softmax / FC: Fully connected net

First results: $\approx$ 2 cm average precision.

Perspective of improvement

- Different representation of the object pose (one point + orientation)
- Data augmentation in simulated environment
- Pretrained CNN weights for initialization

Acknowledgements

This work was supported by the European Union's 2020 research and innovation program under grant agreement No.688807, project ColRobot (collaborative robotics for assembly and kitting in smart manufacturing).

References I

Azad, P., Asfour, T., and Dillmann, R. (2007).
Stereo-based 6d object localization for grasping with humanoid robot systems.
In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 919–924. IEEE.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., et al. (2016).
End to end learning for self-driving cars.
arXiv preprint arXiv:1604.07316.

Hudson, N., Howard, T., Ma, J., Jain, A., Bajracharya, M., Myint, S., Kuo, C., Matthies, L., Backes, P., Hebert, P., et al. (2012).
End-to-end dexterous manipulation with deliberate interactive estimation.
In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 2371–2378. IEEE.

References II

Käppeler, U.-P., Höferlin, M., and Levi, P. (2010).
3d object localization via stereo vision using an omnidirectional and a perspective camera.
In *Proceedings of the 2nd Workshop on Omnidirectional Robot Vision, Anchorage, Alaska*, pages 7–12.

Levine, S., Finn, C., Darrell, T., and Abbeel, P. (2016).
End-to-end training of deep visuomotor policies.
The Journal of Machine Learning Research, 17(1):1334–1373.

Morales, A., Asfour, T., Azad, P., Knoop, S., and Dillmann, R. (2006).
Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands.
In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 5663–5668. IEEE.

Muller, U., Ben, J., Cosatto, E., Flepp, B., and Cun, Y. L. (2006).
Off-road obstacle avoidance through end-to-end learning.
In *Advances in neural information processing systems*, pages 739–746.

Thank you for your attention.