POLITECNICO
MILANO 1863

Road point cloud estimation from static monocular view

Image analysis and computer vision final project

Andrea Bertogalli
Niccolò Balestrieri
Nicolò Tombini

Accademic year 2023/2024

Prof. V. Caglioti

The task

The task for the project is to reconstruct the 3D point cloud (possibly with surface interpolation) of the road surface exploiting the trajectory of the contact point of moving vehicles' wheels and the road. For what concerns the camera setting a single static camera is considered. The task is very challenging in fact we cannot rely on stereo vision or SLAM.

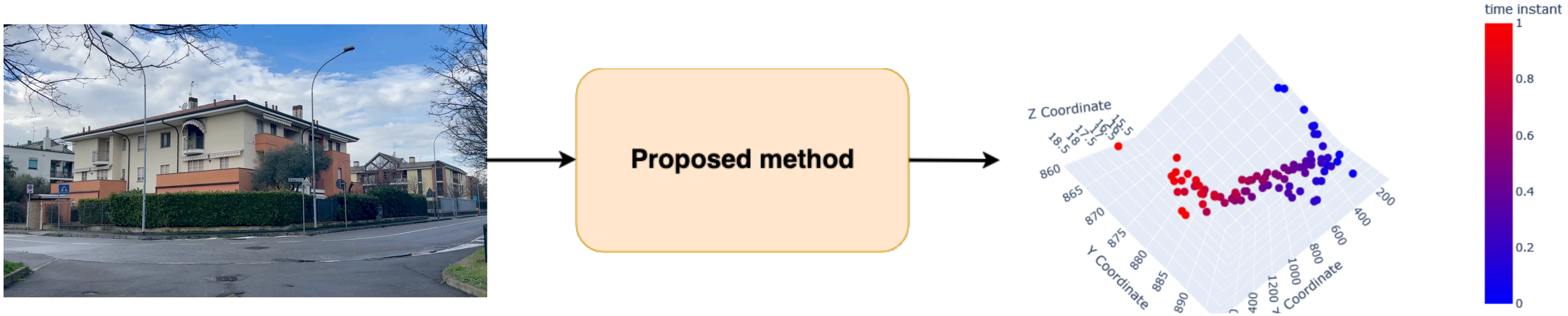


Fig.1: The task requested by the project (short video for presentation purposes).

The proposed pipeline

To archive the result we rely on two different pipelines, one based exclusively on deep learning and one which combines deep learning with classic geometric computer vision.

Both pipelines are composed by **three** steps:

1. Wheels ROI proposal (Deep Learning)
2. Wheels detection
3. 3D point cloud estimation (Geometric / Deep)

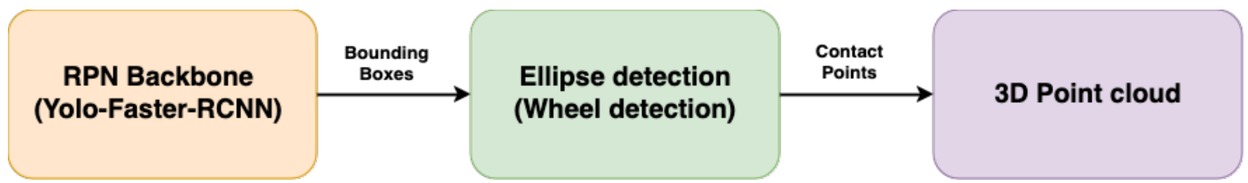


Fig.2: The proposed pipeline.

Both pipelines shares the same RPN backbone to propose a wheel ROI and the wheel detection (Ellipses detection) while they differ on the point cloud estimation.

REFERENCES:

A definition of point cloud

Least Squares based ellipse detection, A. W. Fitzgibbon et al. 1999

Camera intrinsic calibration

As preliminar but fundamental step we performed an intrinsic calibration of our camera (iPhone 14 Pro Max) by using the well known Zhang method.

$$K = \begin{bmatrix} 1.69948770e+03 & 0.00000000e+00 & 5.16343085e+02 \\ 0.00000000e+00 & 1.72316579e+03 & 7.99990899e+02 \\ 0.00000000e+00 & 0.00000000e+00 & 1.00000000e+00 \end{bmatrix}$$ is the calibration matrix

$$d = \begin{bmatrix} 0.30698233 \\ -0.85344961 \\ -0.05681402 \\ -0.00679352 \\ 0.46338888 \end{bmatrix}$$ is the distortion coefficient

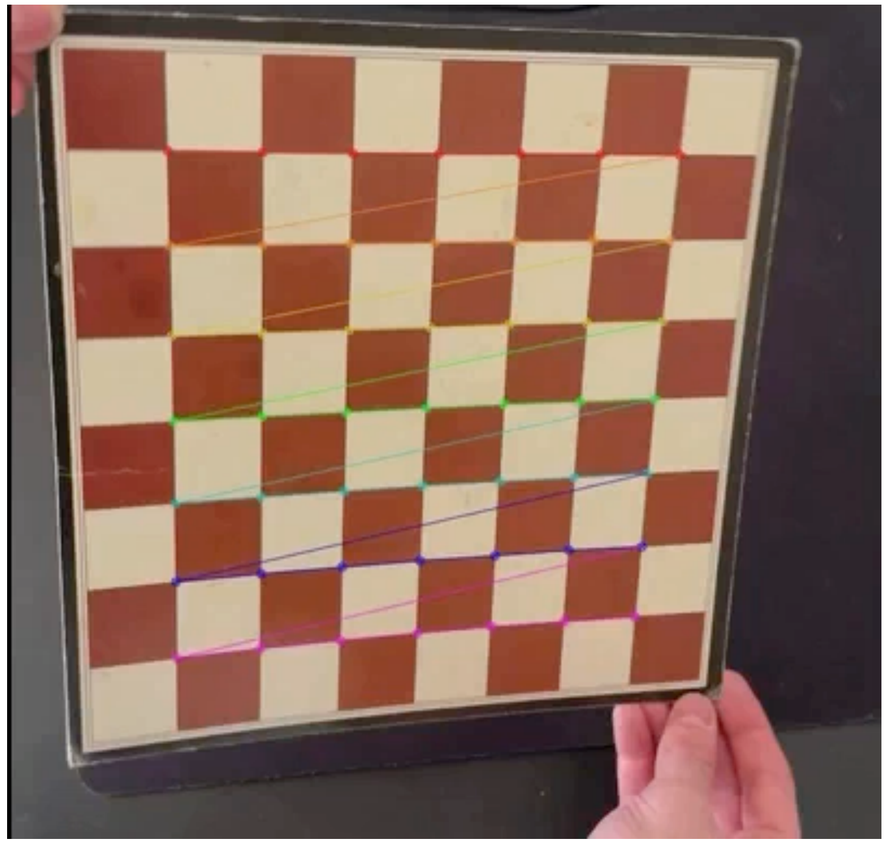


Fig.3: The Zhang method in practice.

REFERENCES:

A flexible new technique for camera calibration, Z. Zhang, 2000

Multiple View Geometry in Computer Vision, R. harley and A. Zisserman, 2003

Wheels RPN network

As first step of our pipeline we employ an object detection model (fine tuned on wheels) to extract patches (ROI) of the wheels. This allow us to isolate the wheels region and leads to better ellipses detection performances.

We tested two different models, in particular:

1. SSD-ResNet50 V1
2. Yolo V5

Both have been fine-tuned to detect wheels, from our tests Yolo V5 archives better performances

Fig.4: Wheels detection task.

REFERENCES:

Deep Residual Learning for Image Recognition, K. He et al. 2016
You Only Look Once: Unified, Real-Time Object Detection, J. Redmon et al., 2015

Contact point extraction

After the ROI extraction we need to perform ellipse detection in each patch. For this we employed the following approach:

1. Binarize the patch.
2. Compute edges with Canny algorithm.
3. Find contours from edges.
4. Fit ellipses using Least Squares.
5. Filter out bad ellipses (multiple filters).
6. Contact point extraction.

Fig.5: The Ellipses and contact point detection result on a curved road.

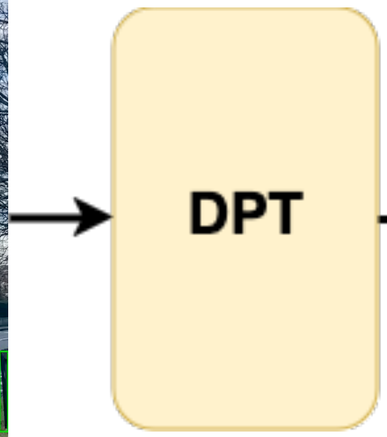
REFERENCES:

Least Squares based ellipse detection, A. W. Fitzgibbon et al. 1999

Deep learning point cloud

For the 3D point estimation the first method proposed is the use of a monocular depth estimation model, in particular we employed Dense Prediction Transformer (a.k.a DPT) an improved version of ViT (Vision Transformer). This model produce an accurate depth map given a single image. From the depth map is trivial to extract the z-coordinate of the contact point.

Fig.6: The input to DPT.



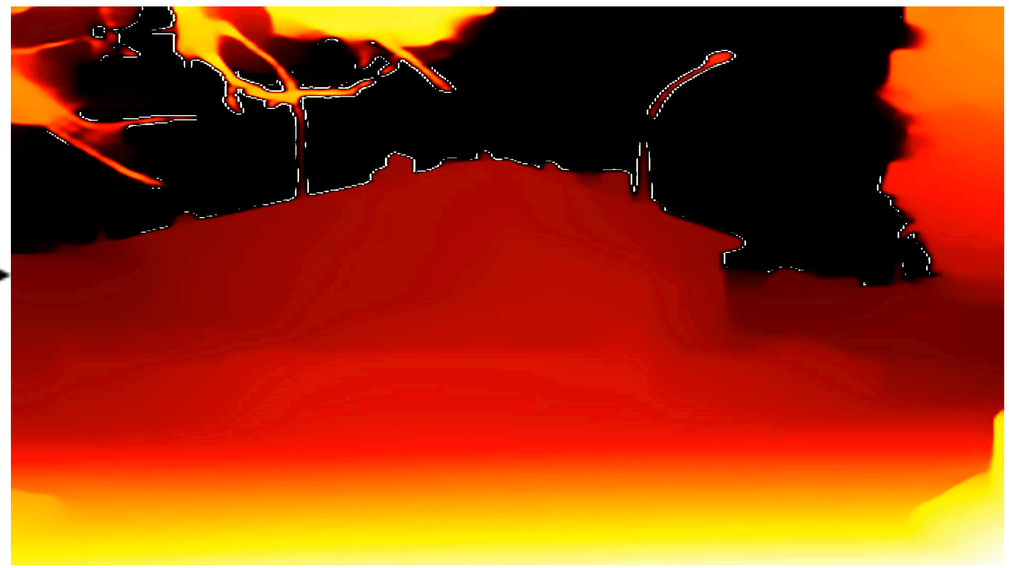


Fig.7: The depth map extracted with DPT.

REFERENCES:

Vision Transformers for Dense Prediction, René Ranftl et al. 2021
An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, Alexey Dosovitskiy et al., 2021

Geometric point cloud

As an alternative to the depth estimation approach we designed a classical pipeline by exploiting geometric properties of the scene. By assuming the radius of a wheel we can in fact infer a 3D model of a generic wheel, which allows us to employ the well known PnP method (Prespective-n-Points).

For each frame:

1. Compute 4 feature points on the wheel.
2. Use PnP (Levenberg-Marquardt algorithm)
3. Use *R* and *t* to estimate word points.
4. Ensemble the estimates in a unique point cloud.

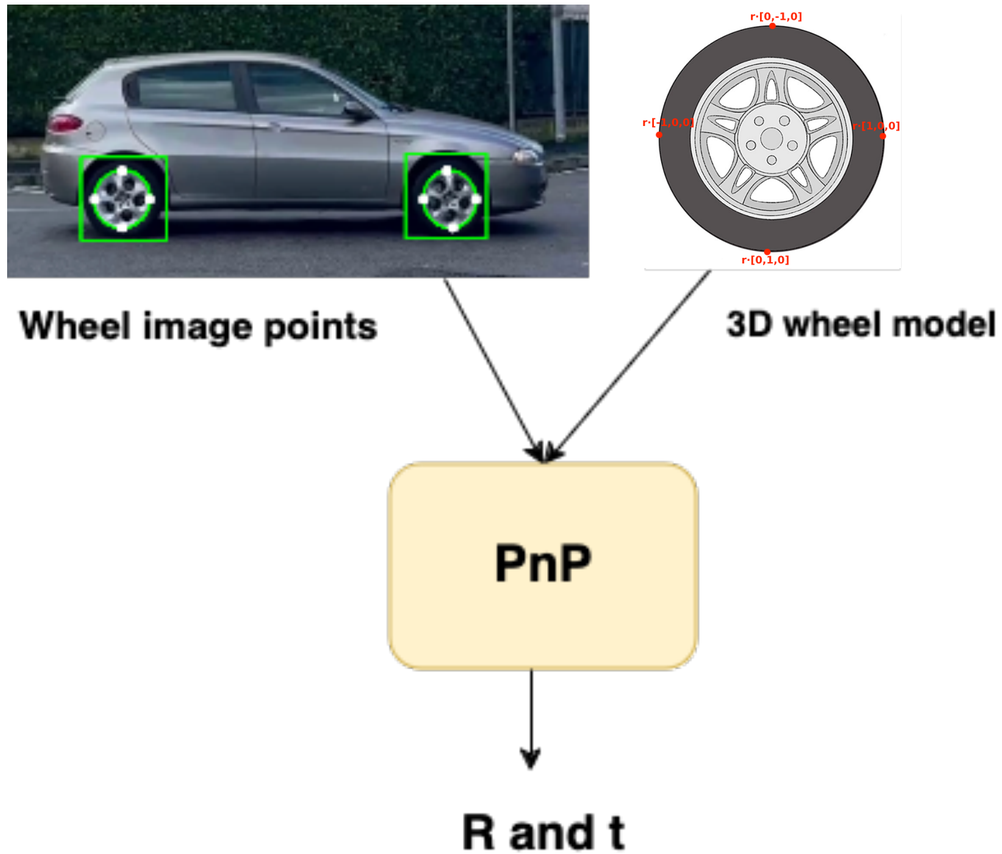


Fig.8: PnP method scheme.

REFERENCES:

Revisiting the PnP Problem: A Fast, General and Optimal Solution, Yinqiang Zheng et al., 2013

EPnP: Efficient Perspective-n-Point Camera Pose Estimation, V. Lepetit et al., 2009

Results

By analysing the results qualitatively we notice a performance improvement when using the Yolo V5 network, in fact it detects much better bouncing boxes also in difficult conditions. For what concerns the 3D estimation we archived better results with the deep learning approach, but the geometric approach is much faster.

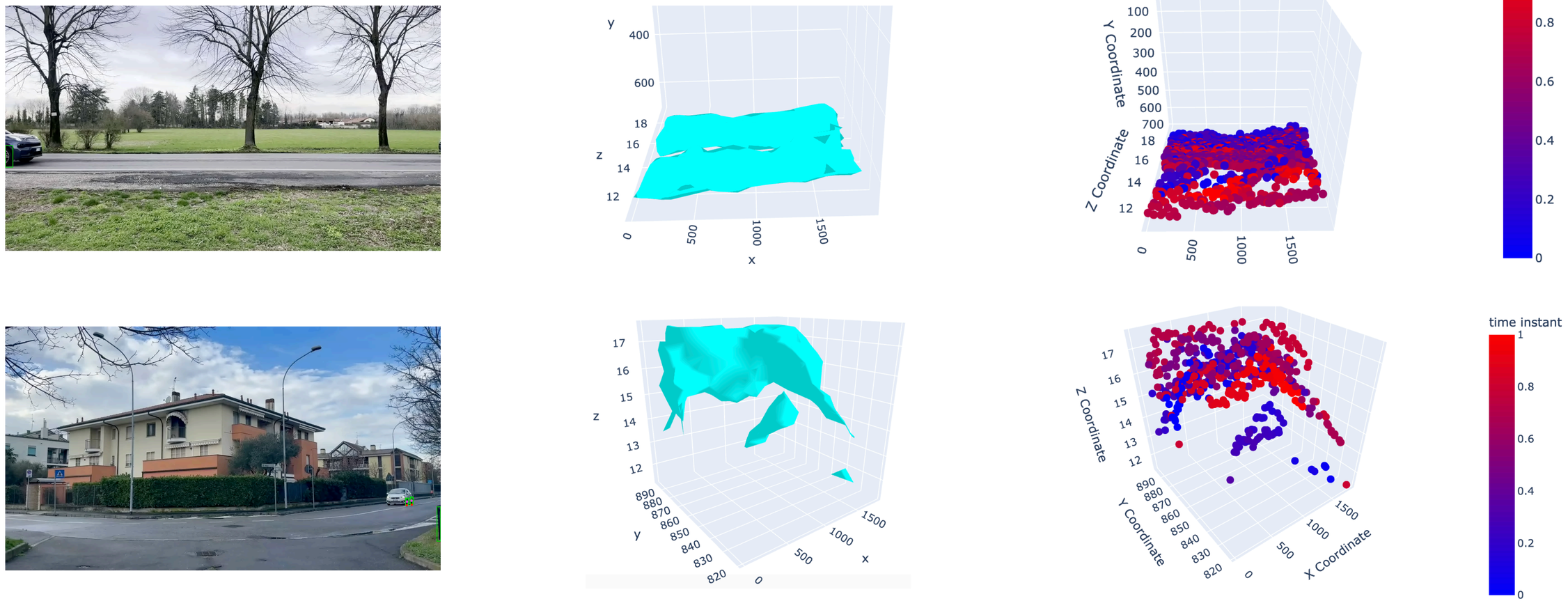


Fig.9: Results obtained.

POLITECNICO
MILANO 1863

THANKS

Andrea Bertogalli
Niccolò Balestrieri
Nicolò Tombini

Accademic year 2023/2024