Funded by The Research Council of Norway

RCN Project number: 333917

NTNU

zeabuz

MARITIME ROBOTICS

TORGHATTEN

Dynamic Scene Representation for Docking in Urban Waters Using a Stereo Camera

Authors:

Trym Anthonsen Nygård, Edmund Førland Brekke, Rudolf Mester, Annette Stahl

Contact information:

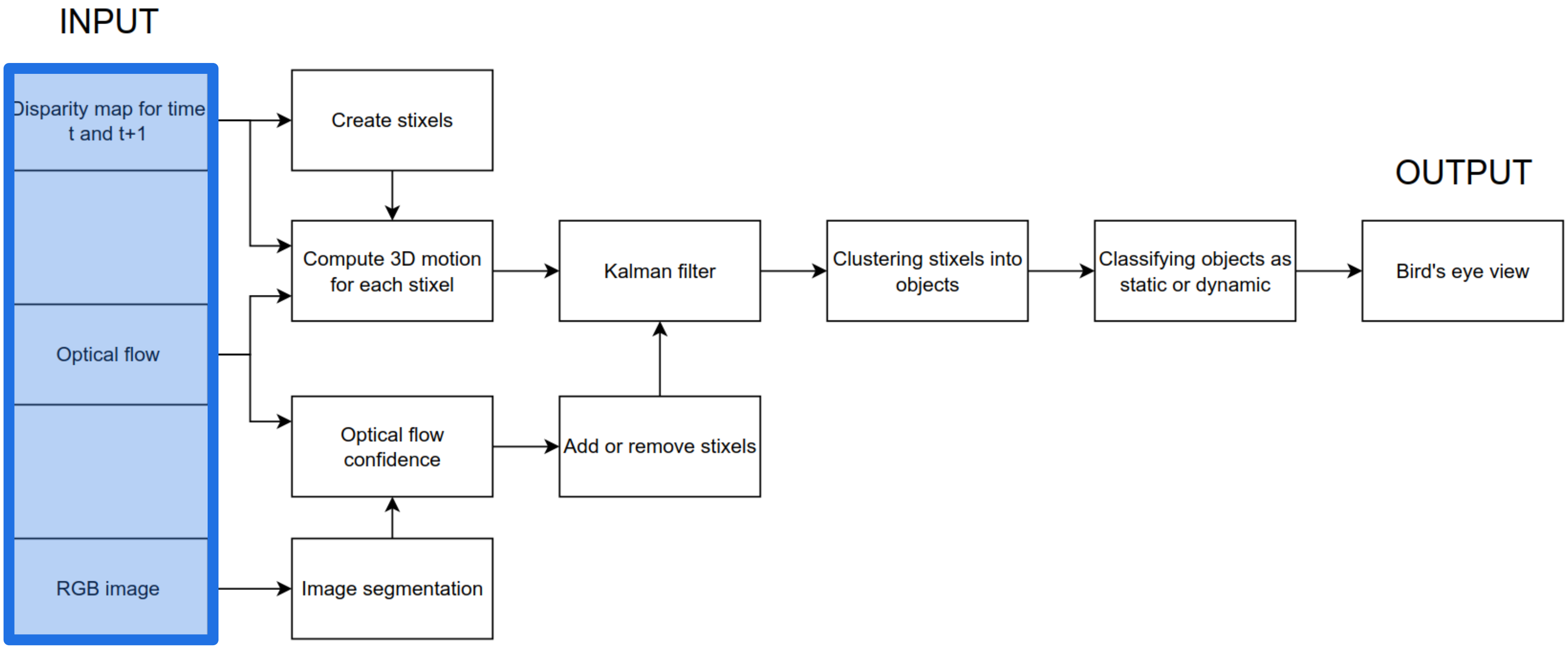
trym.a.nygard@ntnu.no

Norwegian University of Science and Technology

Motivation

- **Why:**
  - Identifying navigable free space and potential obstacles in the maritime domain.
  - Providing crucial information for path planning and collision avoidance systems.
  - Stereo camera provide dense scene information.
- **How:**
  - **Creating stixels in the image:** Reducing information from millions of pixels to a handful of segments.
  - **Tracking stixels in 3D:** To get smooth position and velocity information.
  - **Clustering stixels into handful of line segments in 3D:** Provide more consistent estimates, less sensitive to outliers.
  - **Classification:** Identify which line segments that are static, and which are dynamic.

Pipeline

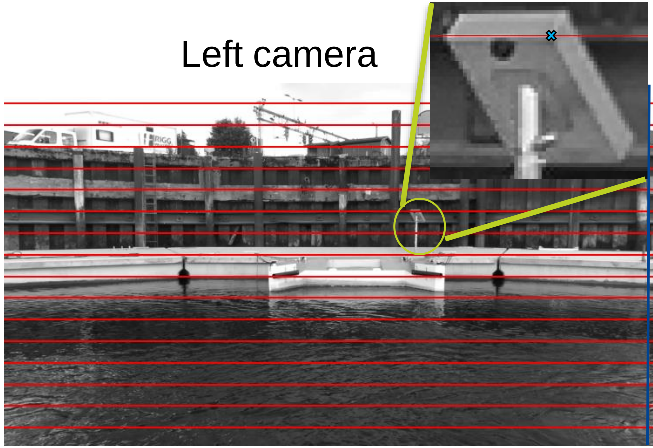


INPUT
Disparity map for time t and t+1
Optical flow
RGB image
Create stixels
Compute 3D motion for each stixel
Optical flow confidence
Image segmentation
Kalman filter
Add or remove stixels
Clustering stixels into objects
Classifying objects as static or dynamic
OUTPUT
Bird's eye view

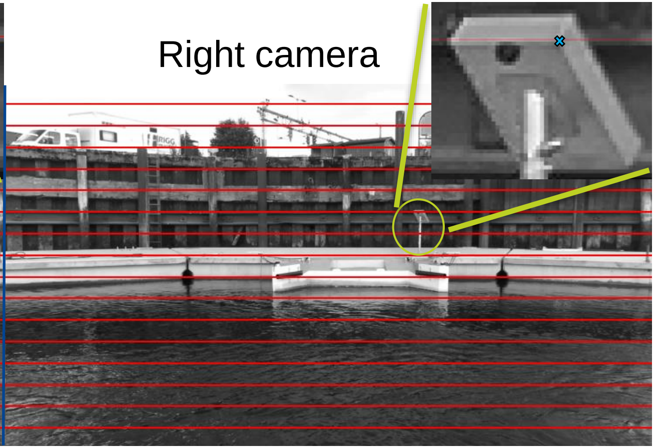
Disparity image

- **Rectified images:** warp and transform the images such that corresponding points are only displaced horizontally.
- **Disparity:** Horizontal pixel displacement between corresponding pairs of pixels from two images.

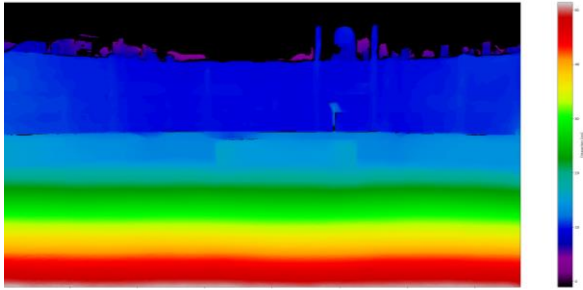
Left camera



Right camera



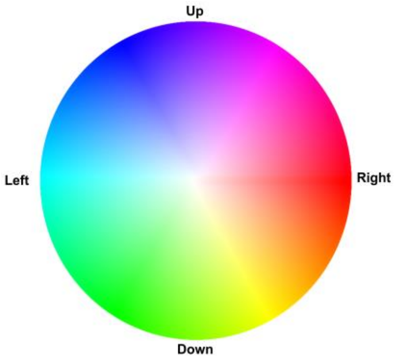
Dense disparity image



Optical flow

- **Optical flow:** Apparent pixel motion in the image.
- Assuming constant brightness.
- Estimated by tracking corresponding pixels over time.

Direction of movement

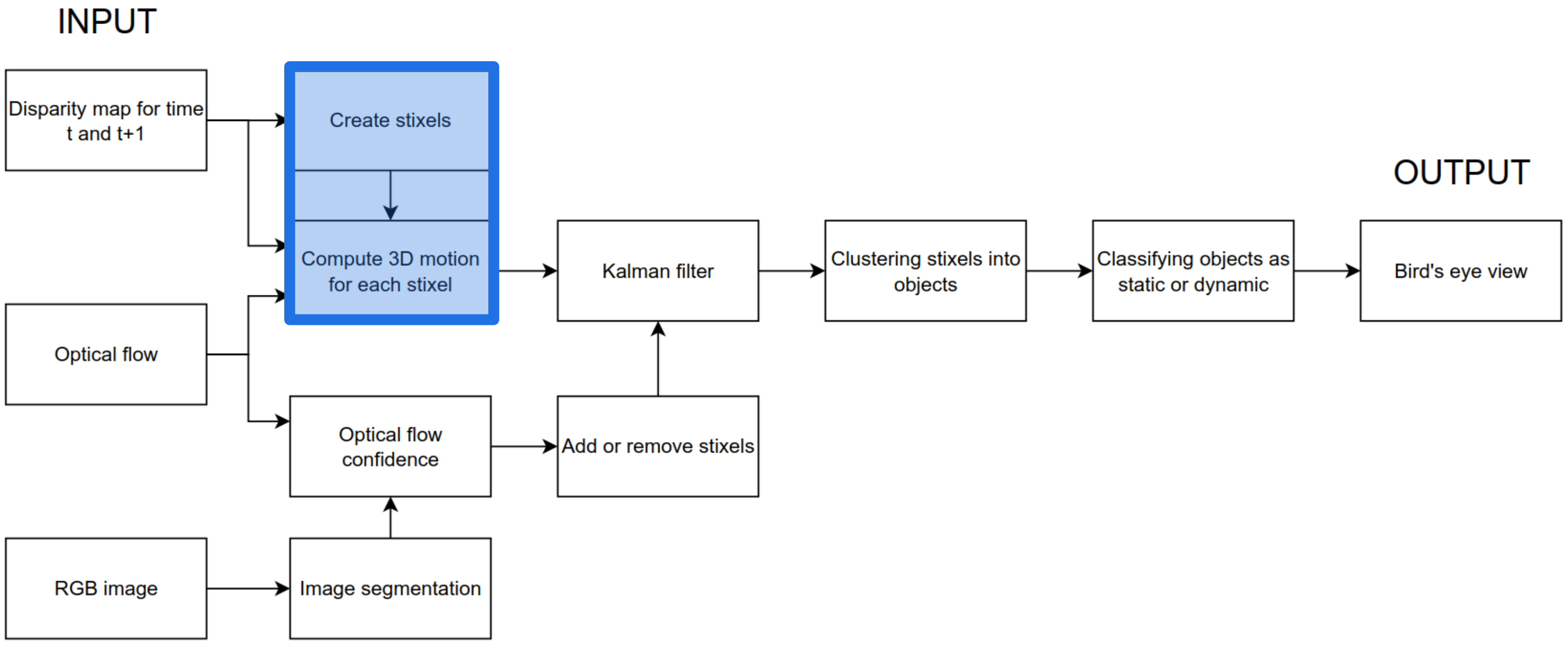


t+1

t

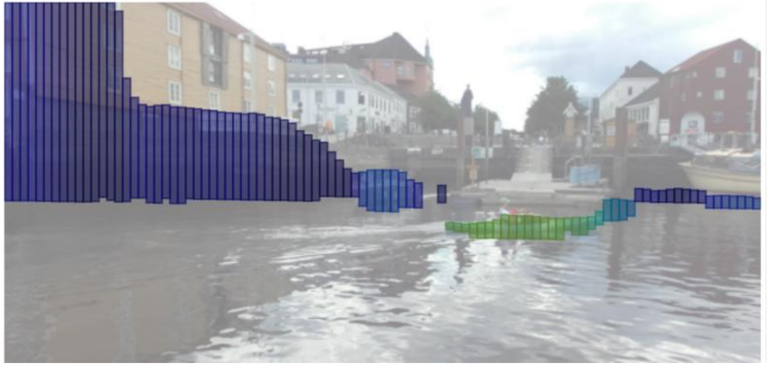
Dense optical flow field

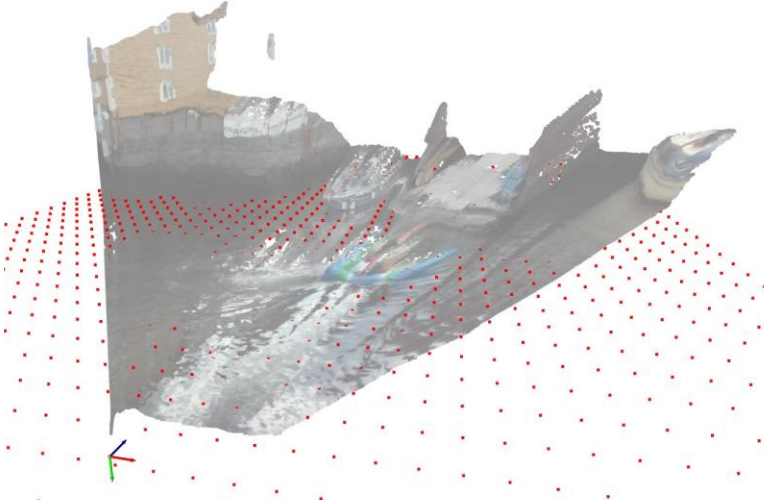
Pipeline


INPUT
Disparity map for time t and t+1
Create stixels
Compute 3D motion for each stixel
Optical flow
Optical flow confidence
RGB image
Image segmentation
Kalman filter
Add or remove stixels
Clustering stixels into objects
Classifying objects as static or dynamic
OUTPUT
Bird's eye view

Create stixels

- **Assumption:** Obstacles in the maritime domain can be represented as a set of planes.
- **Stixels:** Vertically oriented rectangular segments representing the planes.
- Identifying water surface
  - Plane fitting with RANdom SAmple Consencus (RANSAC).

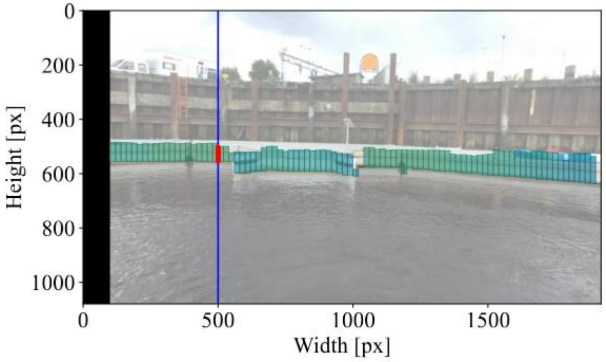


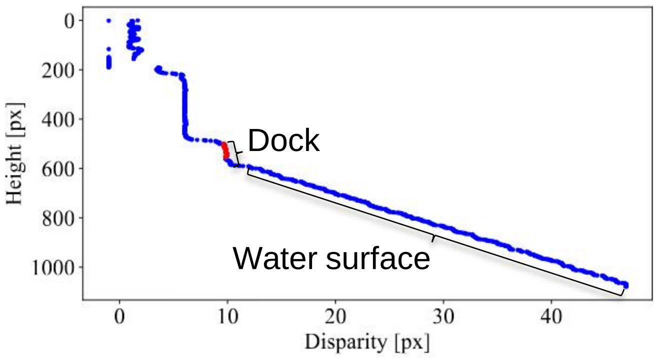


T. A. Nygård, N. Dalhaug, R. Mester, E. Brekke, and A. Stahl, “Stereo Camera-based Free Space Estimation for Docking in Urban Waters,” vol. 45, no. 2, pp. 51–63, Jan. 2024. Available: https://www.mic-journal.no/ABS/MIC-2024-2024-2.asp

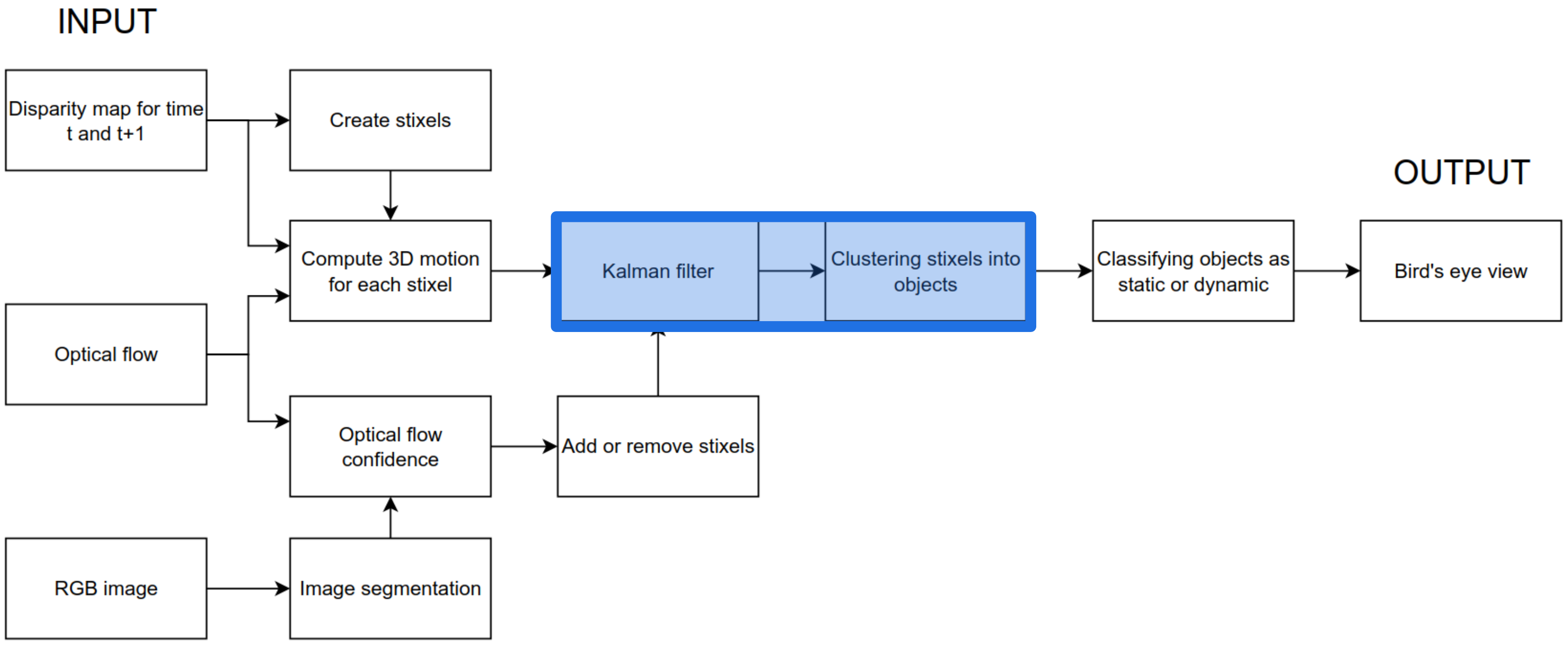
Create stixels

- Horizontal surfaces – sloped lines in disparity space.
- Vertical surfaces – straight vertical lines in the disparity space.
- Identify the vertical segments.
- 3D motion (Scene flow) computed from two consecutive disparity images and dense optical flow.





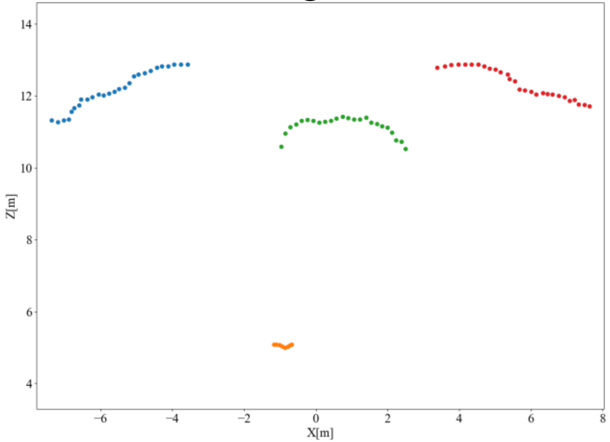
Pipeline


INPUT
Disparity map for time t and t+1
Create stixels
Compute 3D motion for each stixel
Optical flow
Optical flow confidence
RGB image
Image segmentation
Add or remove stixels
Kalman filter
Clustering stixels into objects
Classifying objects as static or dynamic
OUTPUT
Bird's eye view

Tracking and clustering of stixels

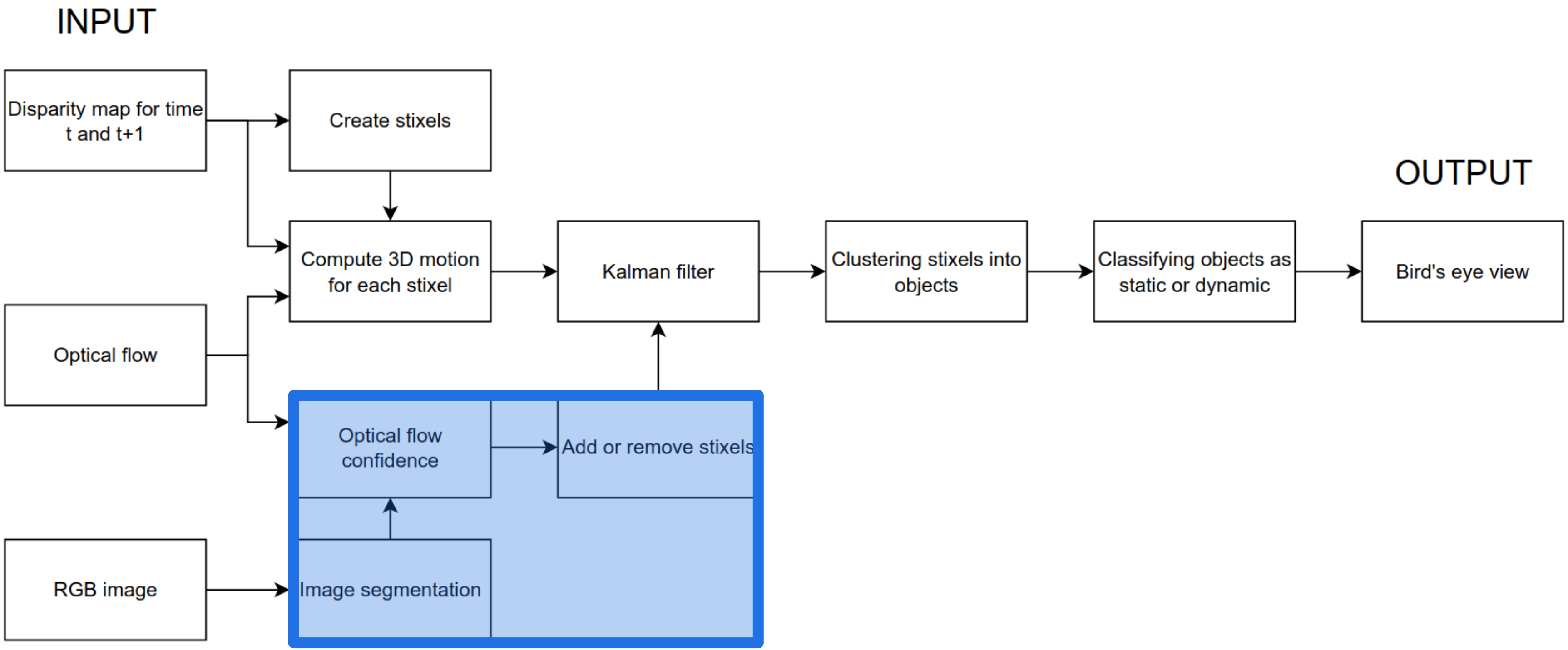
- Kalman filter for smooth and consistent position and velocity estimates.
- Initial clustering in XZ plane with Density Based Spatial Clustering of Applications with Noise (DBSCAN).
- From stixels to line segments in 3D
  - Identify the major axis representing the orientation of the point cluster using Singular Value Decomposition.
- Merging criteria:
  - Consistent motion.
  - Euclidean distance.

Initial clustering with DBSCAN



NTNU

Pipeline



INPUT
Disparity map for time t and t+1
Create stixels
Compute 3D motion for each stixel
Optical flow
RGB image
Optical flow confidence
Add or remove stixels
Image segmentation
Kalman filter
Clustering stixels into objects
Classifying objects as static or dynamic
OUTPUT
Bird's eye view

Update stixels

- Remove stixels that are difficult to track.
- Create a confidence map. Two step approach:
  - Water segmentation using Fast Segment Anything Method (FastSAM).
  - Identify motion boundaries by computing the local variance in optical flow.

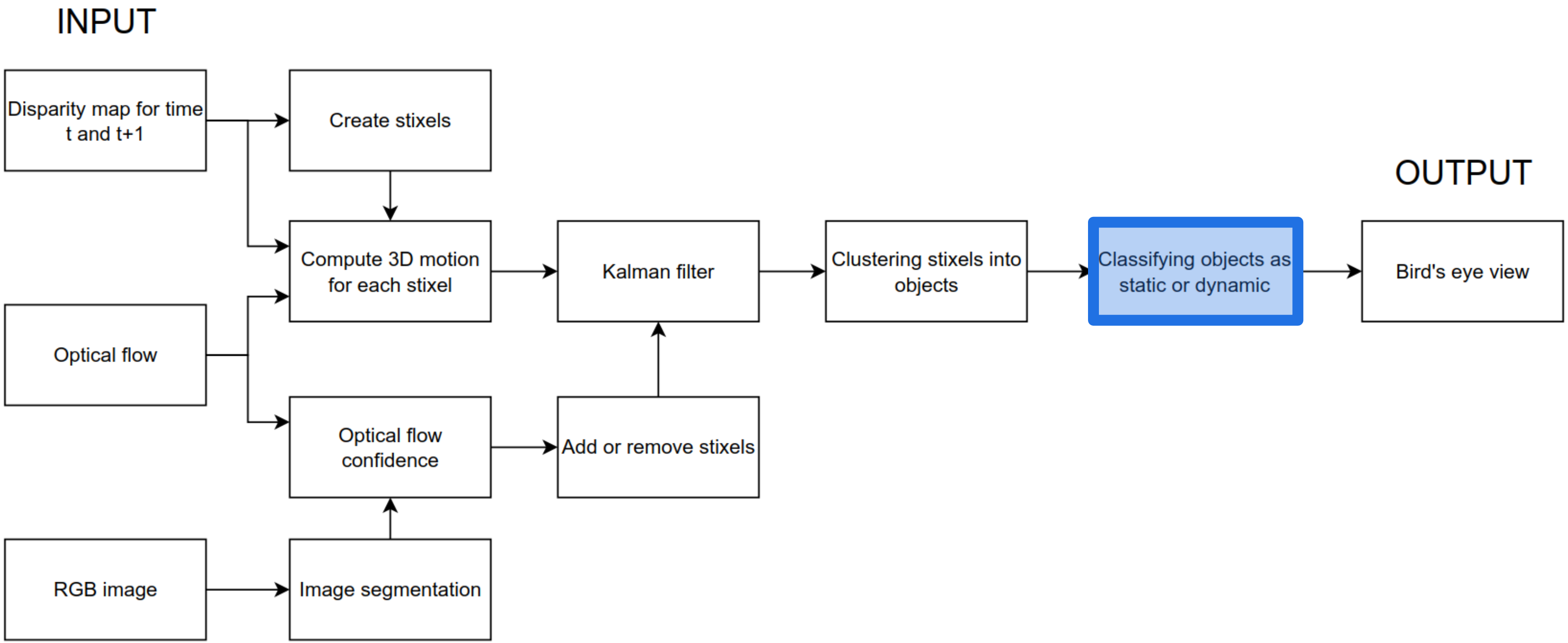
Segmentation masks from FastSAM

T. A. Nygård, N. Dalhaug, R. Mester, E. Brekke, and A. Stahl, “Stereo Camera-based Free Space Estimation for Docking in Urban Waters,” vol. 45, no. 2, pp. 51–63, Jan. 2024. Available: https://www.mic-journal.no/ABS/MIC-2024-2024-2.asp

Confidence map

NTNU

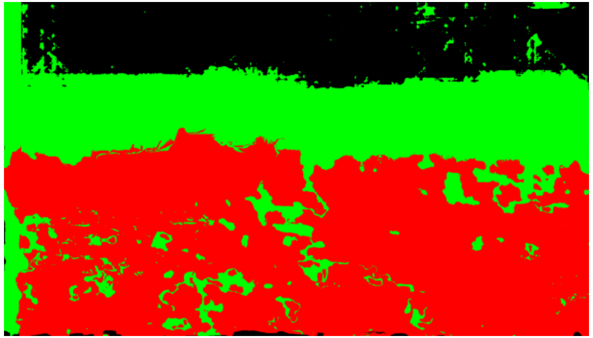
Pipeline


INPUT
Disparity map for time t and t+1
Create stixels
Compute 3D motion for each stixel
Optical flow
Optical flow confidence
RGB image
Image segmentation
Kalman filter
Add or remove stixels
Clustering stixels into objects
Classifying objects as static or dynamic
OUTPUT
Bird's eye view

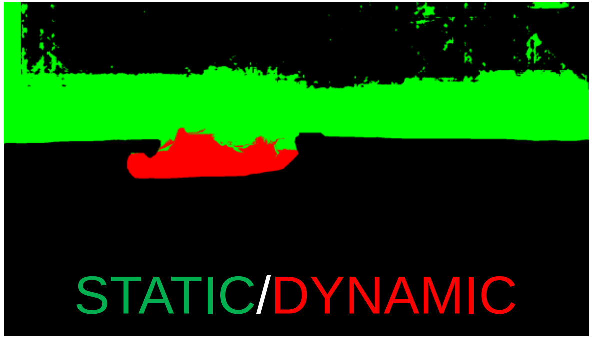
Classification: Static or Dynamic?

- Residual motion likelihood.
- Error propagation and computation of Mahalanobis distance.
- Line segments can be classified by thresholding the Mahalanobis distance.

Motion likelihood (after thresholding)

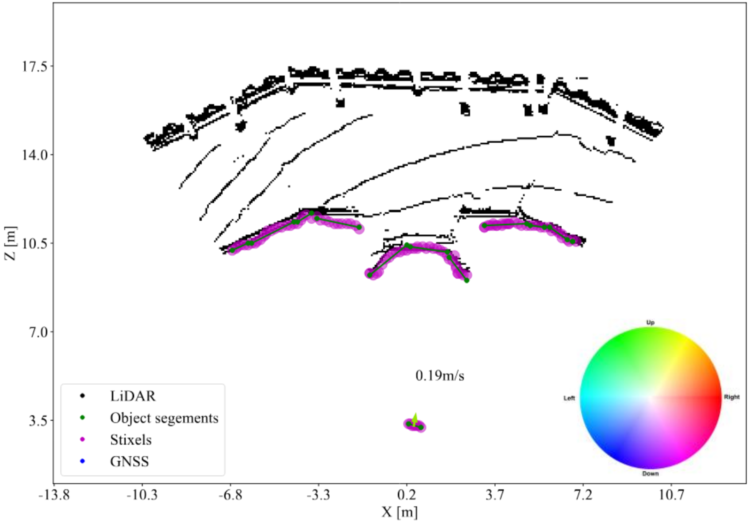


Water surface removed



Results – scenario 1

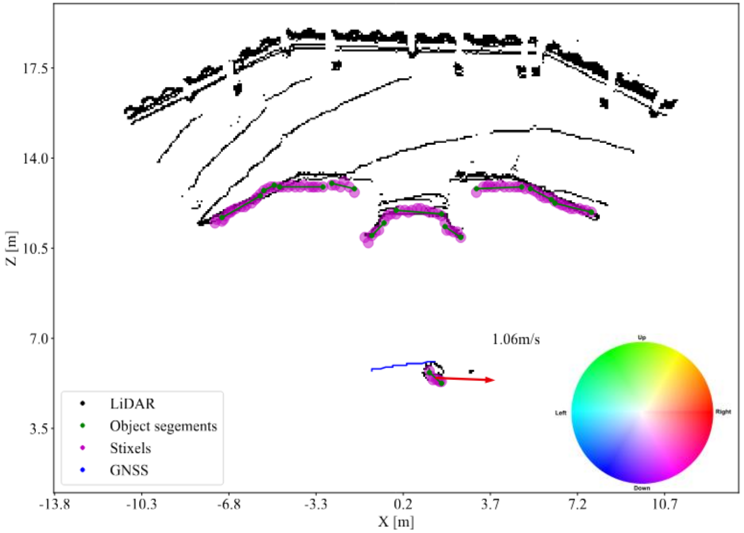
- Detection of sudden changes in the environment.
- Life jacket thrown into the water.
- Outside the LiDAR's field of view.



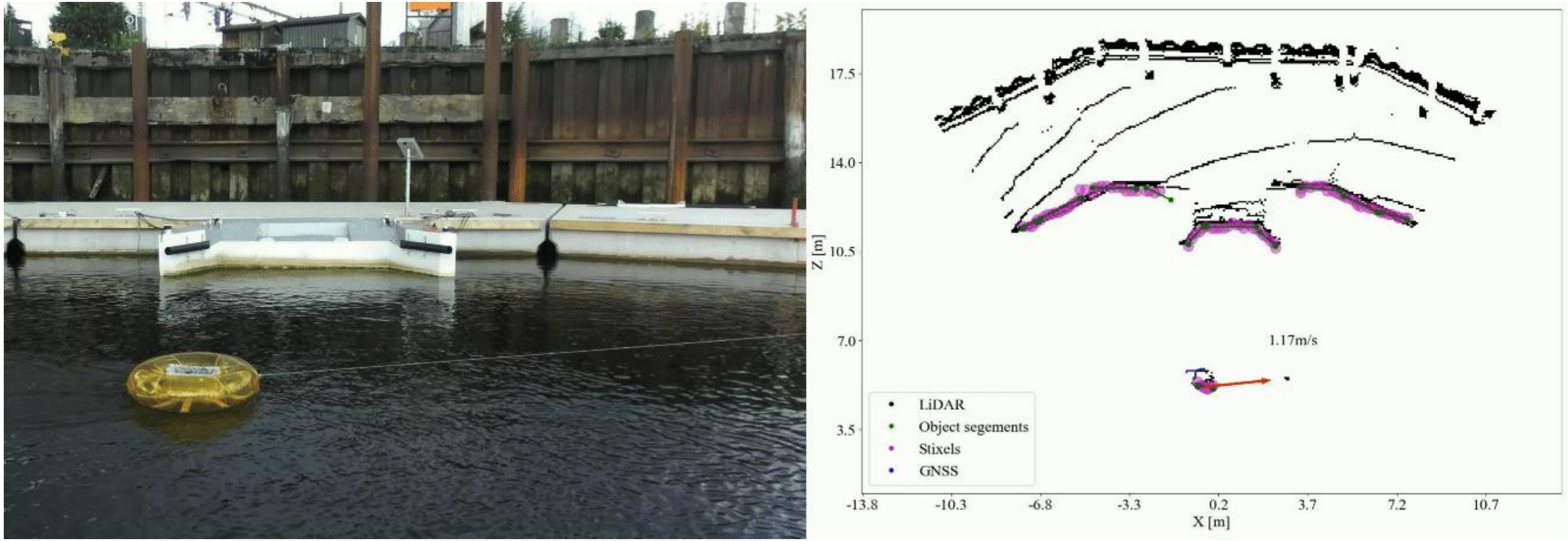
NTNU

Results – scenario 2

- Inflatable ring pulled from left to right.
- Inflatable ring located relatively close to the camera.
- GNSS receiver placed in the center of the ring.

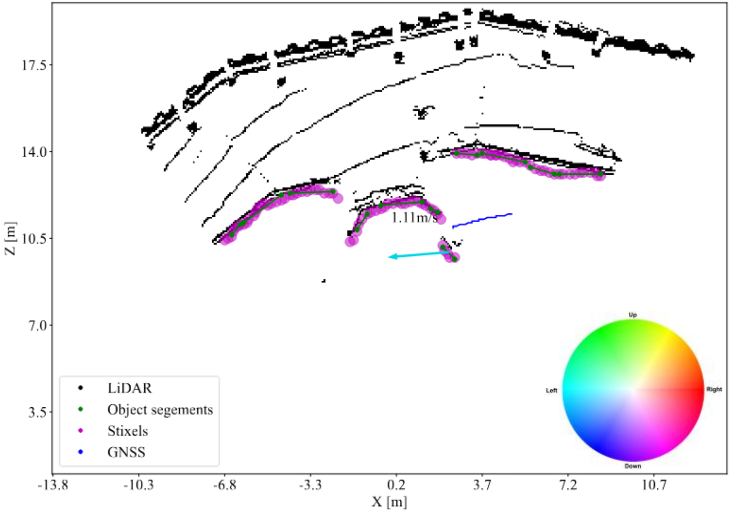


Results - video



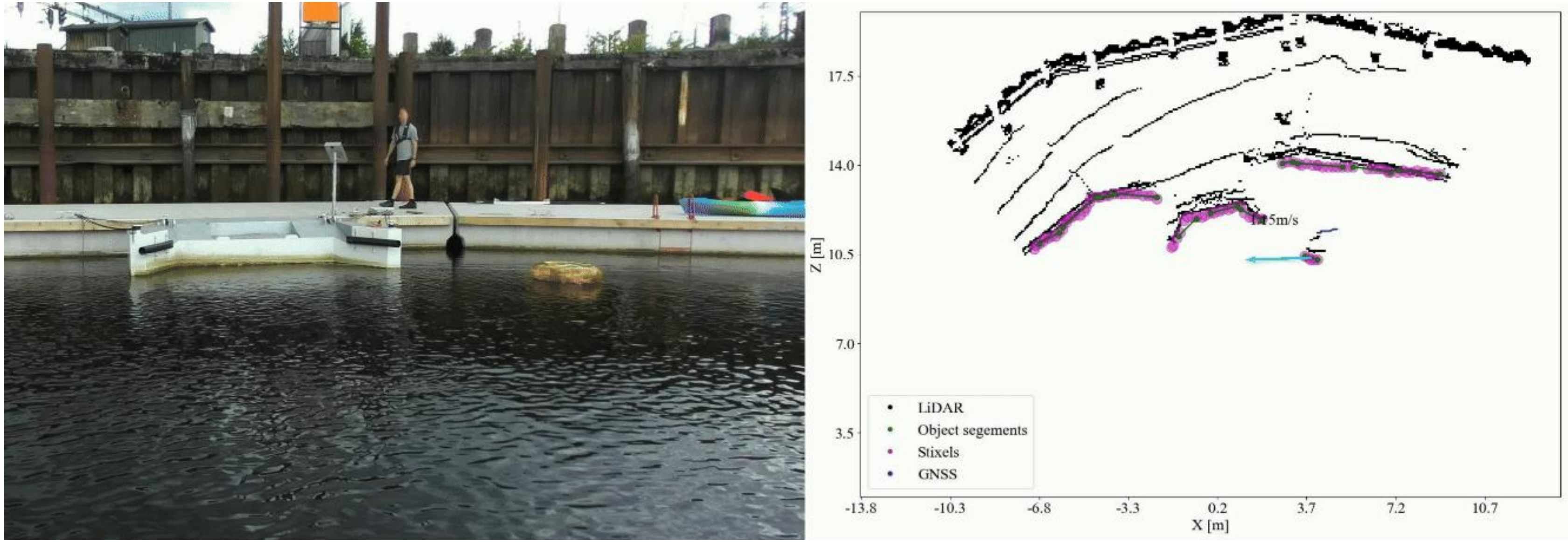
Results – scenario 3

- Inflatable ring pulled from right to left.
- Inflatable ring located further away from the camera.



NTNU

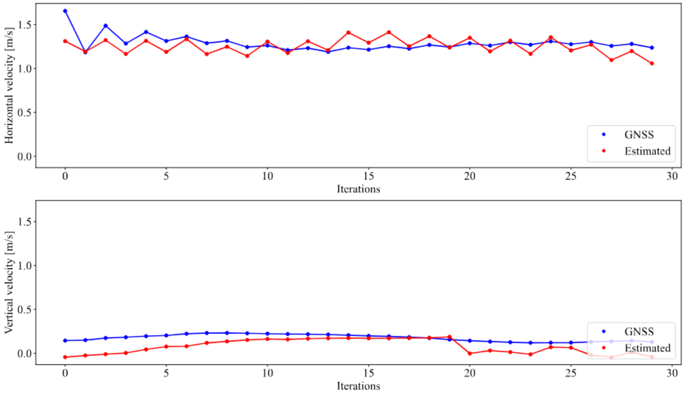
Results - video



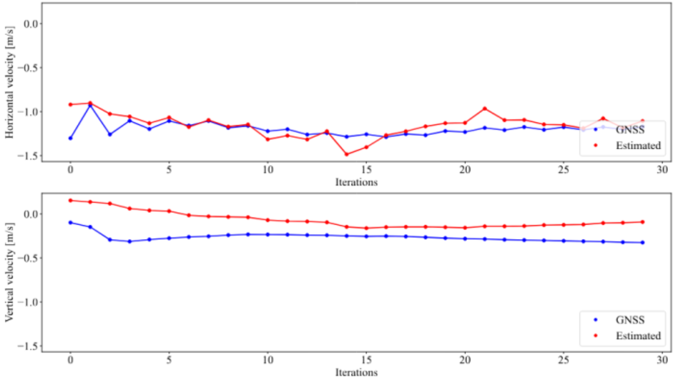
Results – linear velocity

- GNSS velocity ground truth estimated with a Kalman filter.

Scenario 2



Scenario 3



Conclusion

- Presented a dynamic scene representation for docking that only rely on information from a stereo camera.
- Demonstrated the proposed approach using a real dataset recorded with a stereo camera rig mounted on an autonomous vessel.