Data flow in object detection and tracking – the R-ZWEI KICKERS approach 2022

Thomas Jäger, Mouctar Sanoussy Camara, Desmond Krämer, Ket Yee Lee, Adrian Müller – UAS Kaiserslautern

Contact: adrian.mueller@hs-kl.de

Abstract

● Our approach is based on a cascaded set of individual modules, arranged in a sequential pipeline, with no feedback loops. The output of several modules is compared with ground truth data, stemming from the game controller..

●For Object Detection we use a pretrained YOLO Network (COCO Dataset) and fine-tuned (transfer learning) it on the RoboCup Dataset

●For detecting the field lines we use a DeepLabV3+ Network. It is pretrained on a large soccer dataset which is publicly available on https://www.soccer-net.org/download.

●We made our on small dataset with annotated RoboCup field lines and use this dataset for fine-tuning wrt. field size and the fraction of visible lines.

●With the detected field lines, we estimate a homography matrix with OpenCV.

●We decompose the homography matrix to estimate the camera parameter

●For tracking the robots and the ball, we use a slightly modified SORT algorithm.

●For identifying the number of the robots, we use the GC data, enriched with data augmentation in a pre-processing step (cropping, rotation), and a multi-layer CNN.

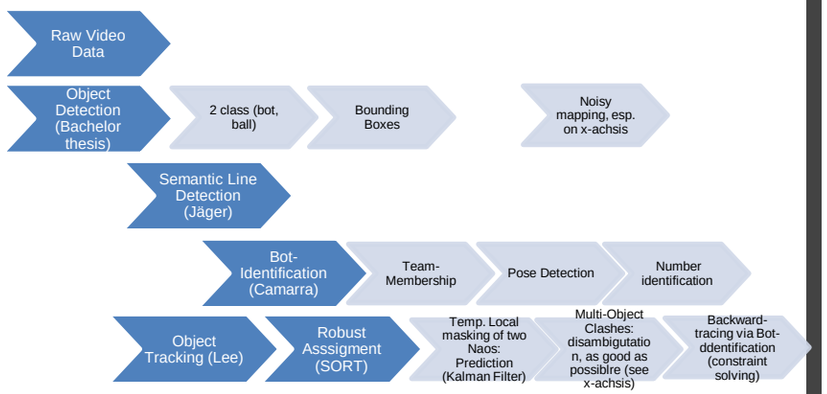


Fig. 1: Data flow of the R-ZWEI KICKERs approach

Object Detection

The video challenge is about recognizing the trajectory of soccer players, in this case the Naos. The course of the movement paths of the individual players is to be recognized and visually represented.

Before entering the challenge, we already set up a two class detection network. In a bachelor thesis, we trained a YOLOv5 Network, which performed reasonable well, even on the Naos CPU.

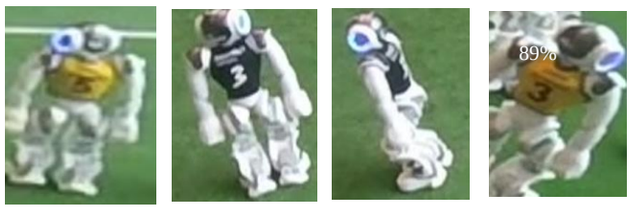
The original dataset contains 5000 images extracted from video clips in the R-Two KICKERS database, the resolution of each image is 1920 * 1080. The training data for Numbers and Colors Classification comes from the object recognition of the YOLO network, providing the bounding box coordinates of each recognized robot.

Yolo5 network for object detection

Fig. 2: Output of the Yolo5. network. The labelling for the subsequent training phasse was done semi-automatically by means of tools and manual correction.

For jersey number and color recognition, a new dataset is created for the learning procedure, with a single bot on each image. The top of each robot B-box is converted to grayscale, then cropped and resized to 40 * 40.After that, the images are labeled in the database to determine whether the jersey number and color are visible or invisible. The best classified images are then used to manually label the jersey colors and numbers. The classification " invisible" is given to images where the recognition problems exist.

Jersey number and color recognition



Due to the relatively small training data set, we applied data mining techniques:
The model was trained with 60% training data and 40% test data.
The mean value of the confusion matrix with 1107 images is 91%.
The learning curves (Accuracy and Loss) through 100 epochs trained with 55137 images (Train:4430, Val:1107).

Tracking Algorithm

Frame t-1 \ Frame t	1	2	3	4	5
1	0.5	0.9	0.6	0.9	0.1
2	0.5	0.8	0.6	0.1	0.7
3	0.7	0.6	0.3	0.6	0.8
4	0.7	0.1	0.8	0.8	0.9
5	0.1	0.6	0.7	0.6	0.5

Semantic line detection, camera calibration

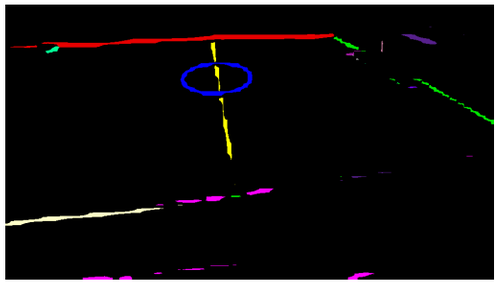


Fig. 3. Line detection as mult-class labelling task

This is the output of the DeepLabv3 network, a mask with multi-class labeled pixels. It distinguished between 13 different variants of line. The aggregated lines are used in combination with four reference points to compute the homography matrix of the camera.

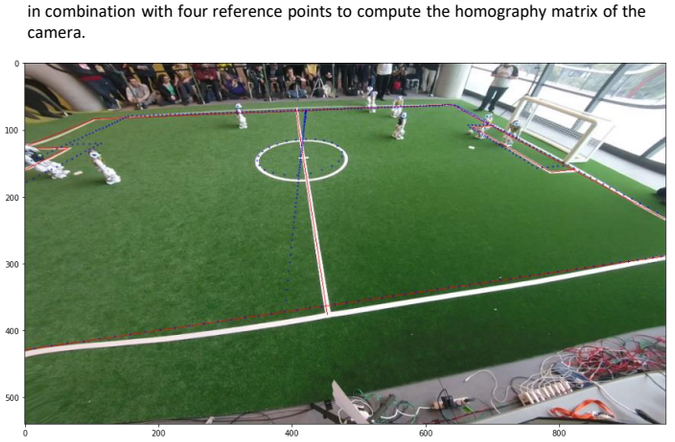
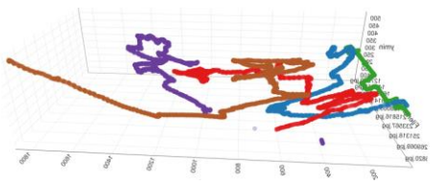


Fig. 4: Overlay of the corrected line segments (in red)



Next Steps

- Future work will feed the raw analysis data into a data mining analysis (“scouting”), so we can synthetically replay a certain team at its strategy in B-Human’s SimRobot, just from the video recording. The 3D plots show the movements of one team over 10min time, playing right to left.
- We made first experiments in applying our approach to human soccer matches, with very promising results.
- Collecting training data is a cumbersome work. We are recently implementing a form of sub-sampling a large pre-trained net: we use it to augment soccer images, and feed the labelled data to our test collection.