

# Stereo and Robot Navigation

## Author

Mikkel D. Olsen, Stud. M.Sc. MMC  
DTU Informatics, [s072260@student.dtu.dk](mailto:s072260@student.dtu.dk)

## Supervisor

Henrik Aanæs, Associate Professor  
DTU Informatics, [haa@imm.dtu.dk](mailto:haa@imm.dtu.dk)

## Supervisor

Anders Dahl, Assistant Professor  
DTU Informatics, [abd@imm.dtu.dk](mailto:abd@imm.dtu.dk)

## Supervisor

Morten Rufus Blas, Team Leader  
CLAAS, [rufus.blas@claas.com](mailto:rufus.blas@claas.com)

## Project description and Structure from Motion

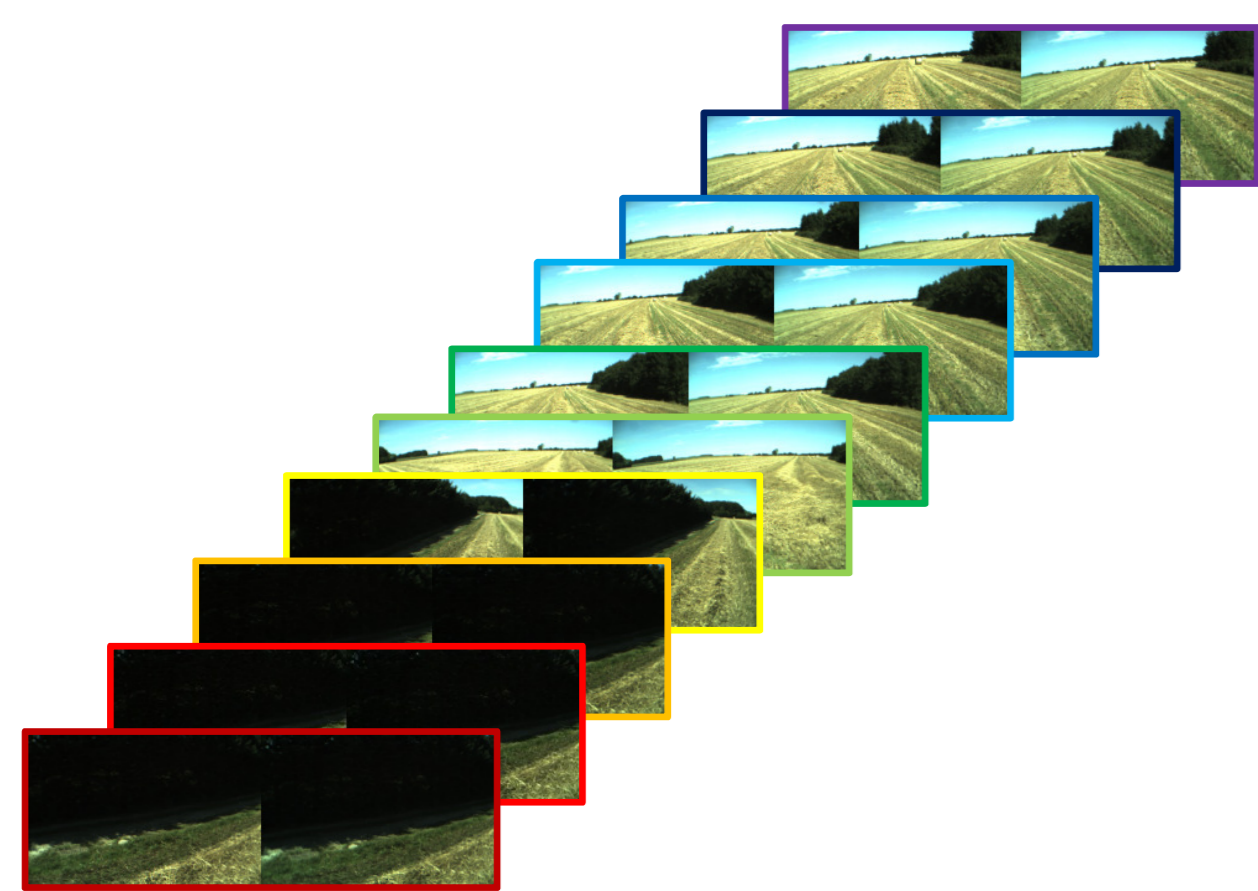


Figure 1. A sequence of stereo images taken from a camera mounted on the front of a harvester/tractor.

CLAAS Agrosystems is currently using stereo cameras, mounted on the front of the agricultural machines, to reconstruct the observed structure seen in the camera, as well as the motion of the machine. This is done using a sequence of stereo images, as the one seen in Figure 1. (Only some of the images from the full image sequence are illustrated)

By finding and matching features in each pair of stereo images, a number of 3D points can be reconstructed, thus resulting in a point cloud for each pair of stereo images. These point clouds can be combined, resulting in a sparse structure of the world, observed by the stereo camera.

In Figure 2, a point cloud has been reconstructed, using the image sequence, where the color of the points, indicate the time when the 3D point was seen by the camera. (red is in the start of the sequence).



Figure 2. Using the sequence of stereo images, a point cloud can be reconstructed for each pair of stereo images.

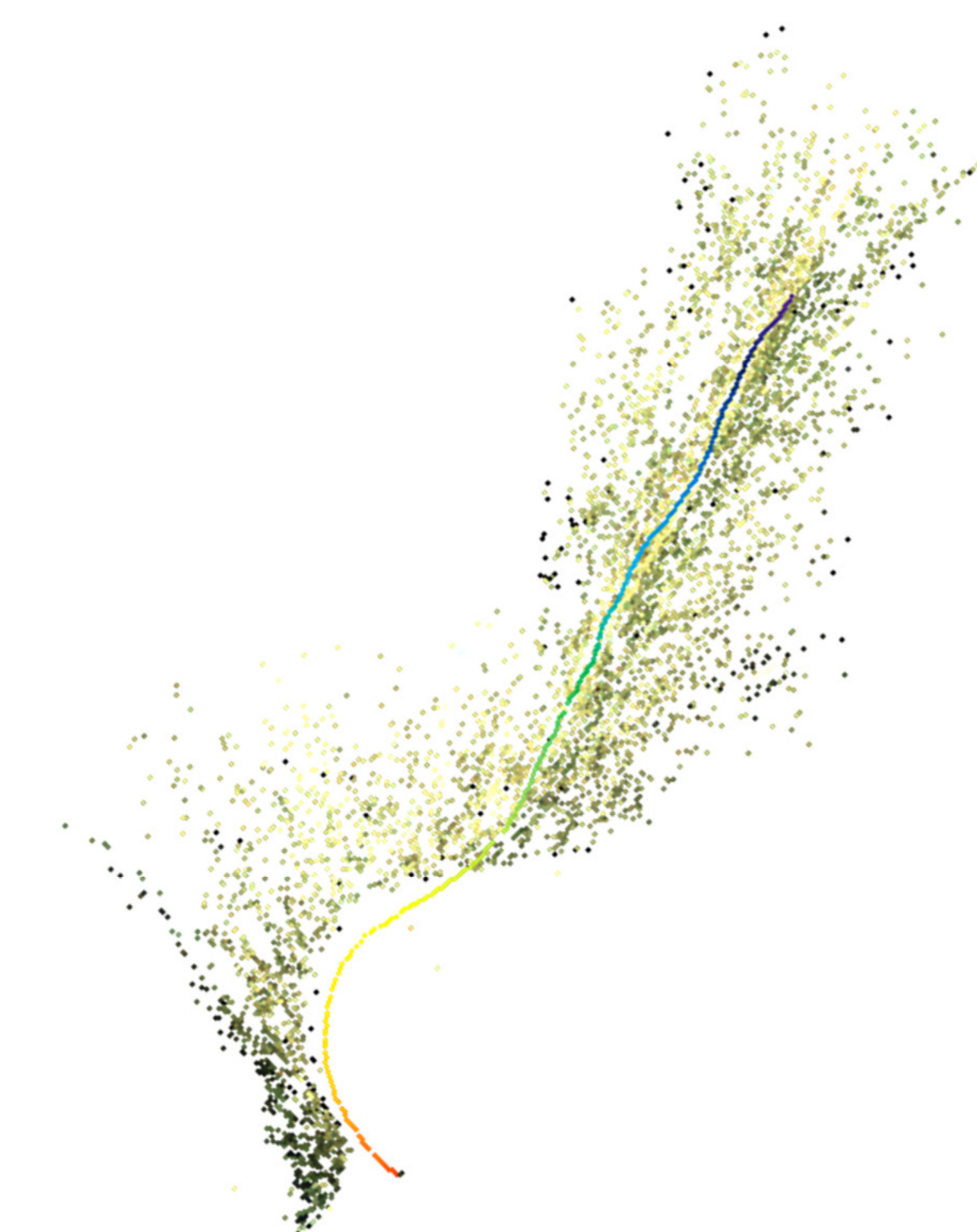


Figure 3. The reconstructed path of the camera, within the virtually reconstructed world. The color of the path describes the time when the position was reconstructed.

Using the color-information of the images, as well as the relative orientation between successive stereo images, the path of the camera can be reconstructed within a virtual world (c.f. Figure 3). An even sparser structure, can be estimated, using disparity estimates of the stereo images, i.e. estimating the disparities the left and the right images from the stereo camera. In this project, the concept of combining images in the time domain is considered, by use of the fact that two successive pairs of the stereo images will contain almost the same features.

## Image rectification



Figure 4. In the case, where the camera/machine has moved in a forward motion, the epipoles are located within the images.

In order to obtain a dense matching between the images, a rectification is usually applied to the images, such that the corresponding pixels share the same vertical coordinate. The most commonly method, is to linearly transform the images using a homography, in order to align the scan lines in the two images. However, as the camera moves in a forward direction, the epipoles lies within the images and the commonly used rectification methods can not be applied. Instead, a polar-based rectification method is used, where the rectified images are constructed by sampling along the epipolar lines. The two rectified images can be seen in Figure 5, where the first and second axes are the polar coordinates.

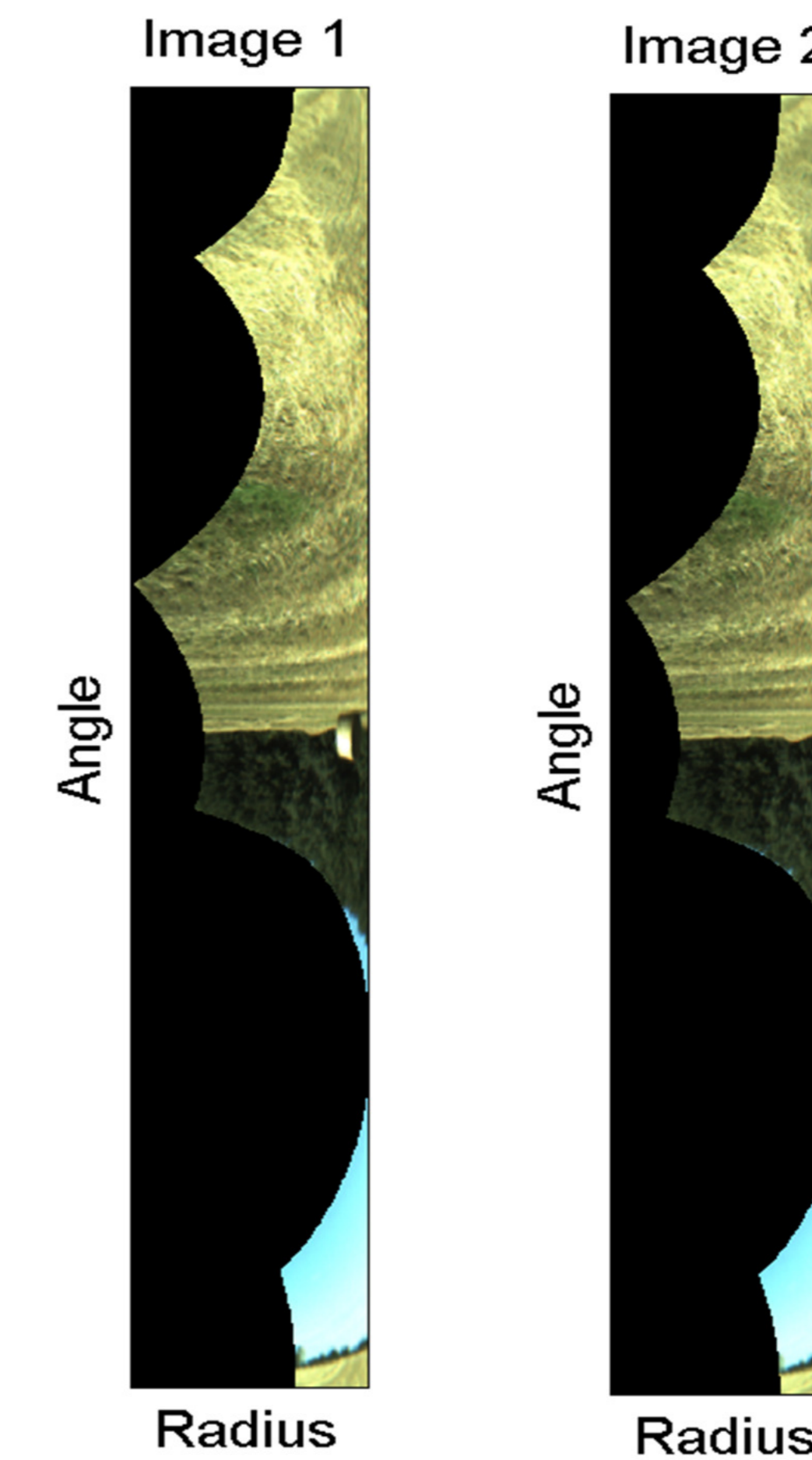


Figure 5. The two rectified images, by use of the polar-based rectification.

The next step, is to estimate the disparities between the two rectified images, i.e. finding the (horizontal) translation of one pixel in the first image, in order to get to the corresponding pixel in the second image. This has been done in Figure 6. The result is a matching between the two rectified images and can thus be used to match epipolar lines in the original images. This limits the matching to not cover the total domain of the original images, but only lines originating from the epipoles. However, the disparity matching still gives a denser matching, compared to the feature based matching, used for the structure from motion part.

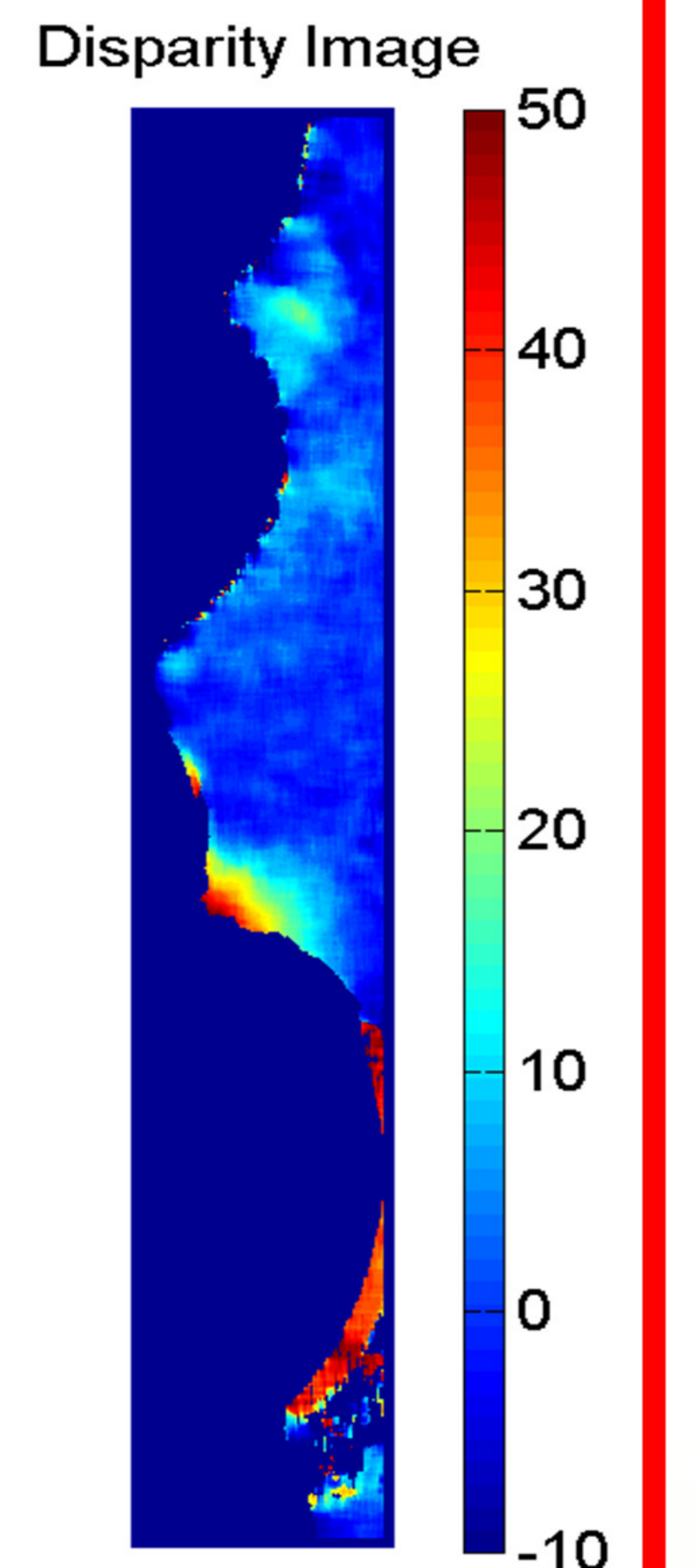


Figure 6. The estimated disparities between the rectified images.

## Camera matching

In order to combine the cameras in a good way, a measure of distance between the cameras, must be defined. In the plots below, a number of different measurements are calculated between all the cameras. This include the distance between camera positions, differences in viewing direction and number of matched image features. Combining these measurements into one single cost-value, can be used to solve the problem using graph-theory. The goal is thus to find the shortest path from the first camera to the last camera, while minimizing the cost of connecting the cameras.

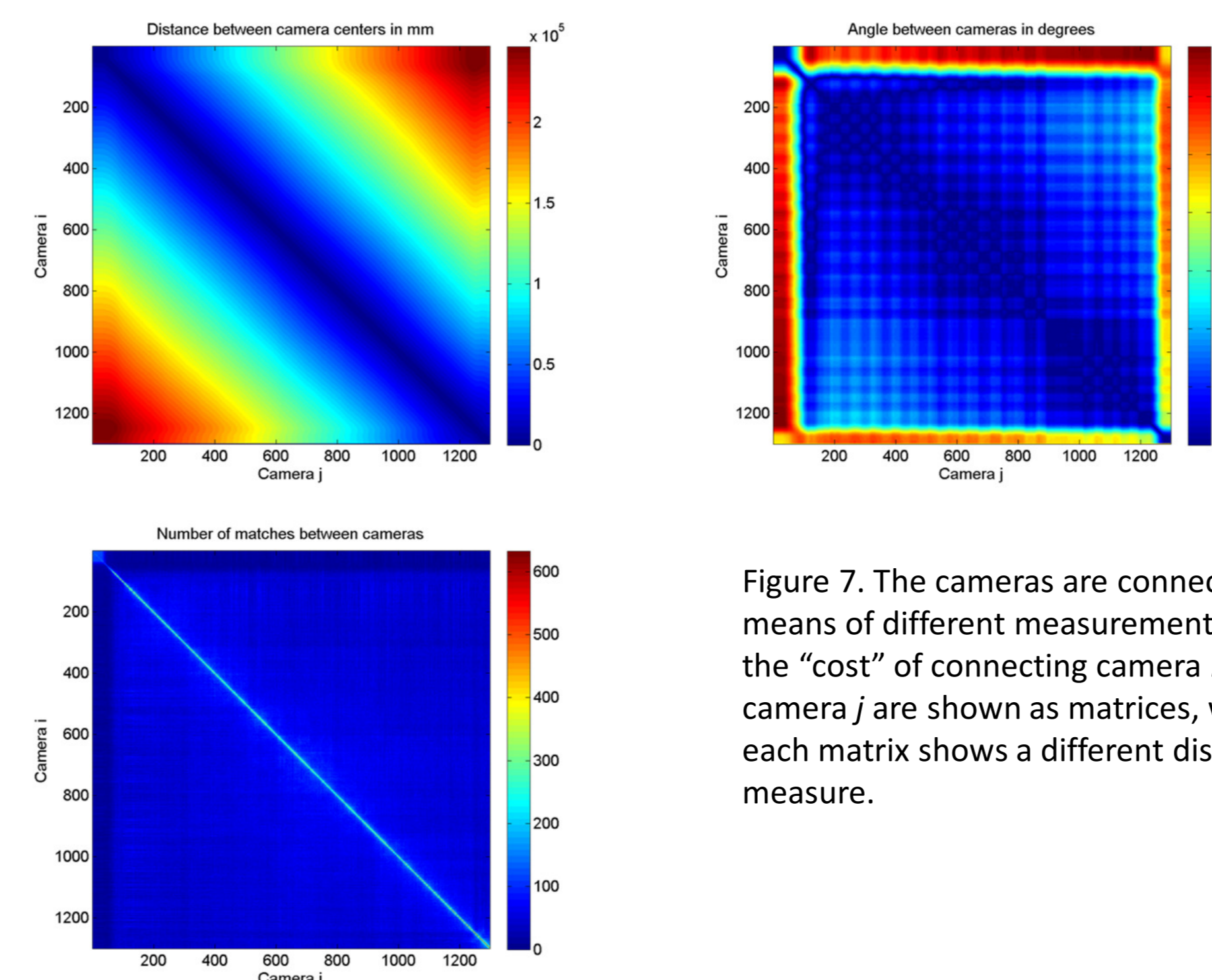


Figure 7. The cameras are connected by means of different measurements. Here the "cost" of connecting camera  $i$  with camera  $j$  are shown as matrices, where each matrix shows a different distance measure.

## Dense Reconstruction

The final step is to use the dense matching and the good combinations of cameras, to reconstruct a dense, but accurate, 3D point cloud of the observed structure. In a near future, the reconstructed point cloud can be used for guiding the agriculture machines to do the farmers tasks more or less automatically. It is thus possible for one farmer to operate and monitor multiple machines, by combining computer vision and robotic technology.

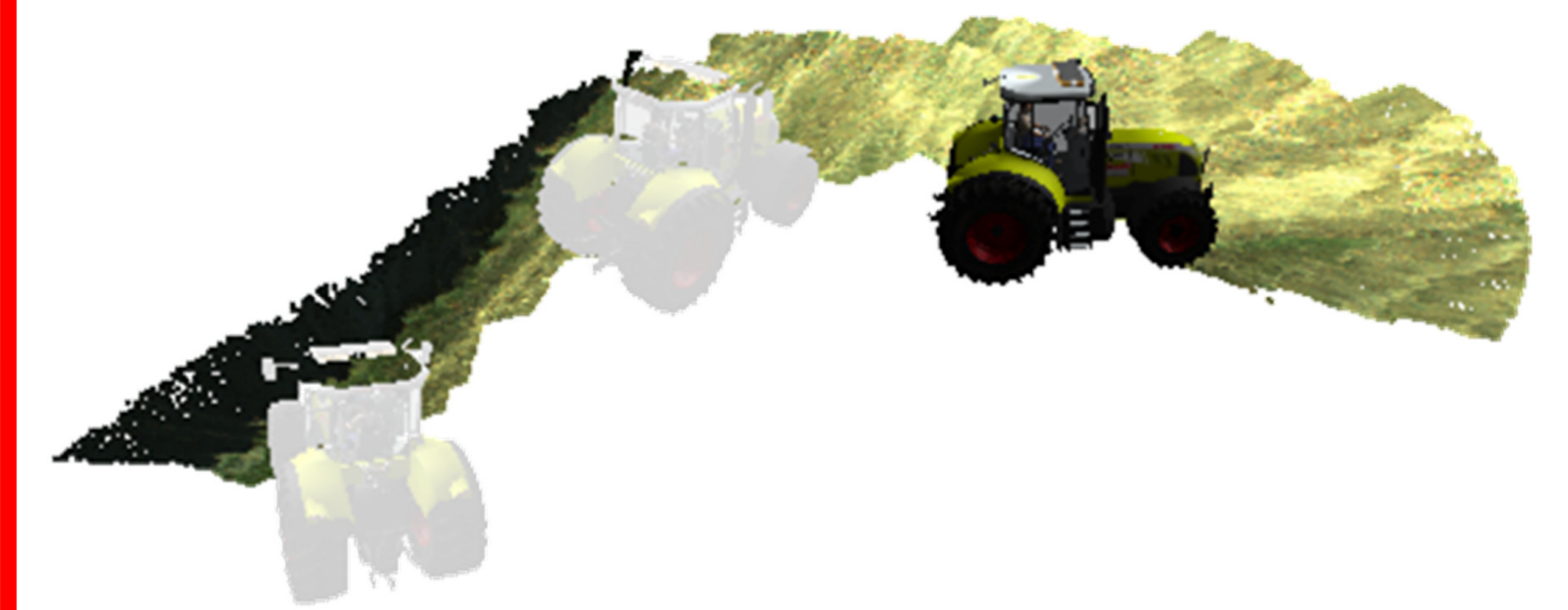


Figure 8. As a final result, a very dense point cloud can be reconstructed, by combining the results from the dense matching. This point cloud can be used for automatic guidance of the agriculture machines, in order to ease the tasks of the farmers.