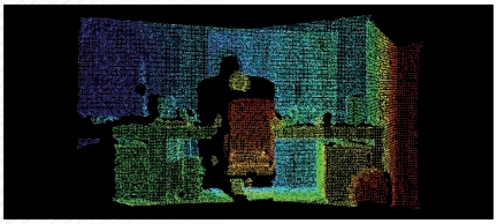
FOCUS ON IMAGE SENSORS

Robots need sensors to operate properly. Using a single image sensor, various aspects of a robot operating in its environment can be measured or monitored. Over the past few years, image sensors have improved a lot: frame rate and resolution have increased, while prices have fallen. As a result, data output has increased and in a number of applications data transfer to a processing unit has become the limiting factor for performance. Local processing in the sensor is one way of reducing data transfer. A report on the Vision in Robotics and Mechatronics project.



JAAP VAN DE LOOSDRECHT, JOHAN VAN ALTHUIS, JOS GUNSING, DANIËL TELGEN, MARK STAPPERS AND PETER KLIJN

Vision in Robotics and Mechatronics was a so-called 'RAAK' project (Regional Attention and Action for Knowledge circulation); last month the concluding symposium was held. It focused on applications of vision techniques in the field of robotics and mechatronics. The project had a number of different goals. As a first step, a vision application was chosen in which data transfer limits performance. This served as a case study of local processing as a means to reduce data. The possibilities of Open CV (open source computer vision) and the Point Cloud Library were investigated for further processing of the reduced data. The items investigated were implemented in a number of demonstrators.

Applications

The Mechatronics department at Fontys is collaborating with TU/e to develop a robot for home care. This robot must be able to move around autonomously in different environments. A 3D sensor can be used so that the robot can avoid colliding into objects in that environment. Light conditions may vary a lot in this application. In terms of autonomy, power consumption is also a cause for concern.

Mechatronics researchers at Avans are focusing on gripper technology. A 3D sensor is considered a useful sensor for grippers. It can recognise an object and determine the distance between object and sensor, and the moment when the object is within grasp. Such a sensor should be small and have an interface that does not require many connections and that secures reliable data transfer.

The HU department of Micro Systems Technology/ Embedded Systems is working on improving Agile Manufacturing methods. Agile Manufacturing uses production equipment with downloadable functions. The functions of production equipment can be changed during production, resulting in a more efficient use of equipment. HU is developing a relatively cheap and modularly built pick & place robot, the HUniplacer (Figure 1). This HUniplacer will be used for research into reconfigurable machines, ROS (Robot Operating System), Intelligent Agents and Computer Vision.

1 *The HUniplacer is a delta robot.*

AUTHOR'S NOTE

Jaap van de Loosdrecht is manager of the Centre of Expertise in Computer Vision at the NHL University of Applied Sciences in Leeuwarden, the Netherlands. Johan van Althuis is a project engineer at the same centre. Jos Gunsing is the Mechatronics research group leader at Avans University of Applied Sciences in Breda, the Netherlands. He also works as a self-employed consultant with MaromeTech. Daniël Telgen is a senior lecturer in Computer Engineering at the University of Applied Sciences Utrecht (HU), based in Utrecht, the Netherlands. Mark Stappers is a researcher at Fontys University of Applied Sciences for Engineering in Eindhoven, the Netherlands. Peter Klijn is a lecturer and researcher at the Academy for Engineering and Informatics of Avans, and a member of the Mechatronics research group.

plaj.klijn@avans.nl
www.nhl.nl
www.avans.nl
www.hu.nl
www.fontys.nl

A configurable vision sensor compatible with ROS is the perfect solution for Agile Manufacturing. Based on the requirements, it was decided to apply a sensor that can be used as a 3D sensor as well as a 2D sensor. The sensor should be able to generate depth maps as well as (processed) 2D images.

The HUniplacer requires the highest frame rate of the three applications described. A frame rate of 50 fps (frames per second) is sufficient for this application. The vision system must be able to work under widely varying light conditions (indoor as well as outdoor, even in direct sunlight). Power consumption must be minimised. For the further processing of 3D images, NHL investigated the Point Cloud Library. TU/e contributed with overall support on the theory of image processing and its implementation.

3D vision techniques

In general, there are four techniques to generate a 3D image using vision sensors.

Structured light

A source projects a light pattern onto a surface. The image of the pattern is recorded with a 2D sensor. This image is compared to a reference pattern. From the differences between the reference pattern and the recorded pattern, a depth map can be calculated. A well-known sensor that uses this technique is the Microsoft Kinect sensor. It uses an infrared (IR) source and sensor to generate the depth map. It also has an RGB image sensor (for colour detection: Red, Green, Blue). Images have a resolution of 640x480 pixels, at a frame rate up to 30 fps. The sensor has a range of 0.6 to 6 m. Accuracy depends on distance to the object. At 1 m, it is in the order of magnitude of a few mm. As they operate with IR light, the sensors do not work in direct sunlight.

Scanning-based

This technique uses the same principle as the structured-light approach. A source of light projects a spot or a sheet of light onto a surface. A 2D image sensor records the spot or sheet of light and compares it to the image on a reference plane. From the differences between these two images, the distance can be calculated (for example, using triangulation). With a spot or a sheet of light, you only get a 1D or 2D image. A scanning movement in one or two directions must be added to get a 3D image. Accuracy, range and reliability depend on the system's set-up.

Time of flight (TOF)

A TOF camera calculates distance by measuring the time of flight of a light signal. The camera transmits a light signal and records the reflected light. The time between transmission and reception of the light is used to calculate the distance of an object. Two approaches can be distinguished: 'range-gated' and Photonic Mixer Devices (PMD). In the first case, short light pulses are emitted and the total amount of light transmitted and received is measured. PMDs generate a sine in the RF range (radio frequency) and measure the phase difference between the transmitted and the received wave. The MESA SwissRanger 4000 is an example of a PMD. It has 176x144 pixels and a maximum frame rate of 50 fps. Accuracy is approximately 10 mm and range at this accuracy is 0.8 to 5 m.

Stereovision

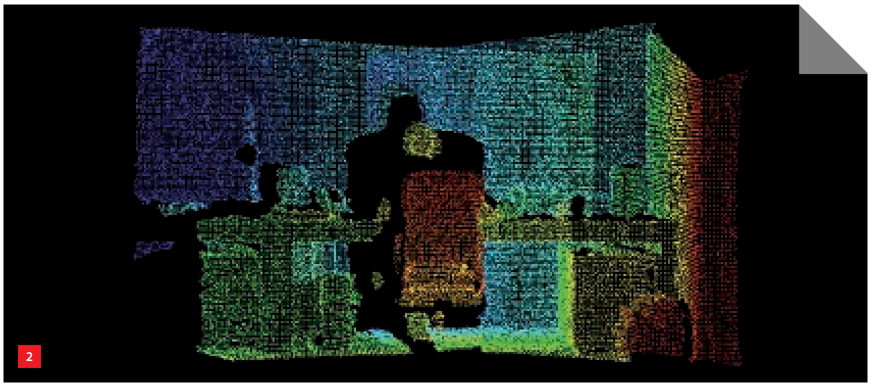
Stereovision uses two 2D image sensors that are placed a distance apart. A point of an object will have a certain location in the image of both sensors. The distance can be calculated from the difference in location of the point in the two images:

$$z = \frac{f \cdot b}{d}$$

Here f is the focal length of the lens, b (baseline) is the distance between the two image sensors, and d the so-called disparity, i.e. the difference in position of the same object point in the two images. The sensors record a large number of object points simultaneously, of course. Therefore, it is possible to generate a 3D image for the field of view that the two sensors share. Accuracy and reliability of such a system depend on the image sensor, lens quality and baseline as well as the object and its environment. The frame rate depends on the camera/image sensor and processing unit used, as well as the data transfer rate. Stereovision algorithms are available for CPUs. The frame rate, however, is restricted by the data transfer between the camera and the processing unit. This could be resolved by implementing the stereovision algorithm in an FPGA (Field Programmable Gate Array, programmable electronics that can have parallel communication with a camera).

Participants in the project

- NHL Centre of Expertise in Computer Vision.
- Fontys University of Applied Sciences for Engineering, Mechatronics research department.
- University of Applied Sciences Utrecht (HU), department of Micro Systems Technology/Embedded Systems.
- Eindhoven University of Technology (TU/e).
- Academy for Engineering and Informatics at Avans University of Applied Sciences, Mechatronics research department.



2 *Example of a point cloud*

Choice of sensors

It was decided to develop a vision system that uses stereovision and an FPGA to locally process the data. This decision was made because of the required frame rate, the desire to minimise power consumption, the varying light conditions, and the desire to have 2D and 3D images. To investigate the possibilities of the Point Cloud Library (PCL, see below), a Kinect sensor was used. By using this off-the-shelf sensor, it was possible to do the PCL work in parallel with the actual development of the vision system. This is why commercially available camera systems are also used in the demonstrators.

Division of work

The project was divided into different chunks of work, assigned to the various parties.

Stereovision system

• Hardware selection and design	Avans and Fontys
• Hardware implementation (VHDL in FPGA)	Avans
• Software development/test set-up	Avans

3D data-processing software

• Point Cloud Library	NHL

Demonstrators

• Jenga robot	Avans
• HUniplacer	HU
• Localising objects using fiducials & QR codes	HU
• Object recognition	HU
• 3D images	HU

Stereovision

A stereovision system is being developed. Signal processing will be done using an FPGA and PC. RGB cameras will be used that have VGA resolution (Video Graphics Array). These cameras are equipped with a global shutter. The global shutter will be used to synchronise the different cameras. It will be possible to connect four of these cameras to the system. It will also be possible to use cameras with a rolling shutter. The system will also allow for four of those cameras to be connected.

The research system set-up is based on an existing FPGA development board. It needs two types of camera boards, i.e. a camera with a rolling shutter and a camera with a global shutter – a global shutter refreshes all the image lines simultaneously (snapshots), while a rolling shutter refreshes the image lines one by one (scanning). There is a parallel interface between the cameras and the development board. A break-out board is needed to connect the camera boards to the development board. The FPGA takes care of the interface with the cameras and the process steps of the stereovision process up to the point that a disparity map has been generated.

The results from each process step are written to RAM memory on the development board. The interface between the development board and the PC is achieved via ethernet and UDP (User Datagram Protocol). The PC provides a user interface, visualisation of data and a way to calibrate the vision system. The user can also use the PC to write settings to the camera.

Point Cloud Library

A point cloud is a collection of n-dimensional points (Figure 2). In the case of 3D vision, this is usually a collection of 3D points. To acquire a point cloud, a 3D camera system is used. This could be a TOF camera, a stereovision system or a structured-light camera. PCL is a cross-platform open source C++ library for processing point clouds. It is also available for ROS.

The NHL Centre of Expertise in Computer Vision is integrating the PCL into an existing 2D vision program. The result will be a test facility that enables experiments with 3D as well as 2D algorithms.

Demonstrators

Jenga robot

Jenga is a game that consists of a number of wooden blocks that are used to build a tower. Players take turns to remove a block from the tower. The removed block must be placed at the top of the tower. When the tower collapses, the game is over. The tower then has to be rebuilt. This game is being used as a demonstrator to highlight the possibilities of 3D vision.

The system uses an ABB robot and a TOF camera (MESA SR4000). The players order the robot to remove a block from the tower. The robot places this block on top of the tower in a position given by the players. Once the tower collapses, the robot rebuilds it. The main tasks of the vision system concern the recognition of blocks, their position and orientation, deciding which block has to be taken first and telling the robot controller this.

Agile manufacturing

The HUniplacer is a delta robot that was developed by the HU. It is being used as a platform for a demonstrator. In the demo, four crates are being used. Two of those crates are empty, while one of the crates is full of blue marbles and the final crate is full of red marbles. In the demo, the delta robot distributes the marbles over the four crates according to several predetermined patterns. The system needs to identify and locate the crates. The system must monitor whether a crate is moved or removed from the field of view. It uses a camera that provides a top view and a camera that provides a bottom view. The vision sensors used are USB cameras. HU has written a Linux driver for the cameras in order to communicate with OpenCV. Each crate has a QR code that is used to identify and locate the crate.

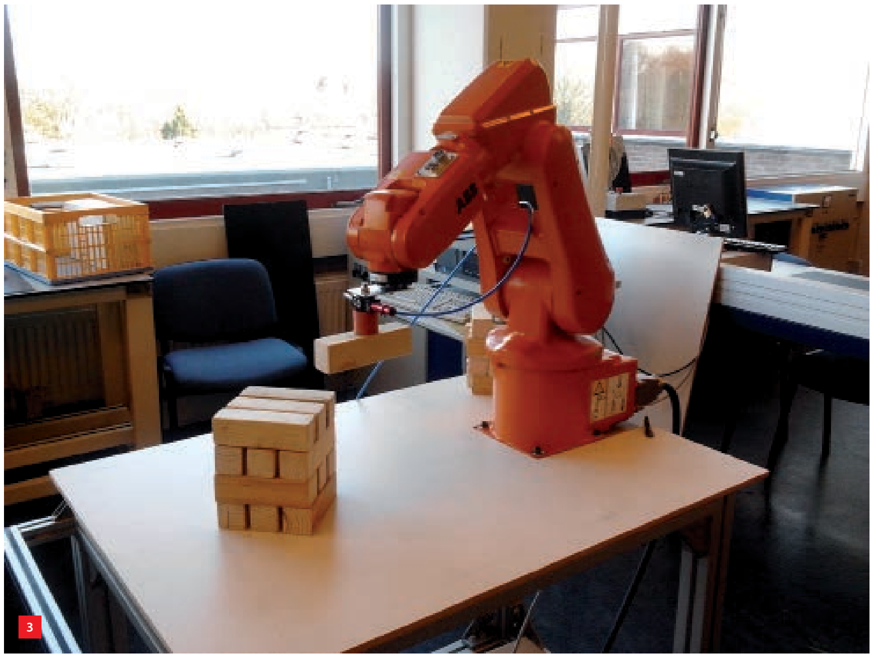
Status

Stereovision

The camera boards and the break-out board have been designed and are being manufactured. The firmware for a system with cameras with a global shutter has been written and is ready to be tested. The test software is not completely functional yet. The system calibration still has to be implemented in the software. After successful tests, the vision system will be used in follow-up projects (i.e. Medical Robotics, Agile Manufacturing, Jenga Robot, Adaptive Robotics) in a production environment. A schematic that combines the camera boards and a development board has been developed.

Jenga robot

A first version of the Jenga robot has been tested (Figure 3). The vision algorithms work well. The camera had some distortion because of reflections. Therefore, several camera shots were averaged, resulting in a lower frame rate. This had no effect on overall system performance. The resolution of the camera in the X-Y-plane is not enough for the required accuracy of the system. This problem will be solved mechanically. A camera system with higher resolution would be preferable though. In the near future the TOF camera will be replaced by the stereovision system.



3 *The Jenga robot in action.*

Point Cloud Library

A substantial part of the PCL is available in the test facility. Functionality has been checked by experiments in several demonstrators. This functionality will be used extensively in further experiments, including those with stereovision.

Agile Manufacturing

A fiducial and QR recognition module has been developed for OpenCV. The demonstrator works as defined. In the future, the demonstrator may be fitted with the camera system that is being developed. ■

REFERENCES

- agilemanufacturing.nl/products/vision-for-mechatronics-robotics/
- agilemanufacturing.nl/media/
- www.willowgarage.com
- www.mesa-imaging.ch/swissranger4000.php
- www.ros.org/wiki/openni_kinect/kinect_accuracy
- www.ncbi.nlm.nih.gov/pmc/articles/PMC3304120