Overview of popular 3D imaging approaches for mobile robots and a pilot study on a low-cost 3D imaging system

Agnieszka Tiszbierek
Faculty of Production Engineering and Logistics, Faculty of Electrical Engineering, Automatic Control and Informatics, Opole University of Technology, Poland
45-036 Opole, ul. Luboszycka 5
a.tiszbierek@po.opole.pl

Michal Podpora
Faculty of Electrical Engineering, Automatic Control and Informatics, Opole University of Technology, Poland
45-272 Opole, ul. Sosnkowskiego 31
m.podpora@po.opole.pl

Abstract— **In this paper authors briefly describe the most popular 3D imaging approaches for mobile robots. The most significant advantages and disadvantages are discussed and a practical example of each conception is given. Authors also describe an idea of a vision-based imaging system. The solution is based on an existing, yet not popular image processing method of using a single video camera for 3D scene reconstruction.**

Keywords—Computer Vision, Machine Vision, Stereovision, Vision Understanding, Image Processing

I. INTRODUCTION

Vision is considered to be the most valuable and significant information source for Human. Image understanding is not limited to 2D still images – the third dimension plays a major role in object recognition and indexing. For human, as well as for robots, it is essential not only for perception of the world, but also for making the interaction with objects more natural and seamless.

Nowadays, generating a reality-based 3D computer model of a scene or object is not only possible, but currently also very popular. There are many various devices for acquisition of 3D properties of objects or environment, offering more or less precise 3D acquisition. Some of them use similar or slightly modified algorithms. Choosing the right method for scanning and digitizing of 3D information has great impact on the quality of the result and on the hardware to be used. This, in turn, implies the financial cost of the acquisition. There are many hardware solutions for 3D acquisition and there are some that give superb 3D imaging quality, but they all have some disadvantages. Usually, for the best possible result, the most precise and valuable hardware should be used. The most problematic disadvantages of high quality 3D imaging systems are: their price and requirement for a laboratory conditions (i.e. special lighting, fixed position or constant rotation, no distraction). Of course, there are some cheaper solutions, but their quality is not always satisfactory. Also, the process of acquisition is often problematic due to the complicated handling of the hardware and/or scanned object or due to some special requirements (e.g. special background) and it disqualifies some of the existing solutions.

While 3D information is the key information in robot navigation systems, many researchers and hobbyists encounter the need to implement 3D scene understanding. Many of them, just like the authors of this paper, decide to build their own system. Before developing a new solution (hardware and algorithms), a brief review of existing solutions should be done. This paper is thought to be a quick guide to the most popular approaches to 3D scene understanding and 3D imaging, as well as an introduction to a low-cost 3D imaging system being developed by the authors.

II. OVERVIEW OF THE MOST POPULAR 3D IMAGING APPROACHES FOR MOBILE ROBOTS

This chapter includes a brief review of the most popular categories of approaches to the acquisition of 3D scenes and/or objects. A short description is given, basic advantages and disadvantages and an example. An important contribution is the authors' subjective evaluation of the quality of a 3D imaging result, if it is easily useful, the accessibility of data and costs of getting and using the hardware device.

A. *Infrared 3D Point Cloud Mapping (Microsoft Kinect)*

One of indubitably the most famous devices for 3D activities and 3D-based interaction is the Kinect Sensor. The Kinect (and its updated version: Kinect2) is produced by Microsoft as an input device for Xbox360 gaming console. Its main task is to track movement of a human player for the purpose of human-computer interaction. The user doesn't need any device to navigate the options or to play games – he becomes the controller himself. He performs some specific (or predefined) movements or gestures to control an application, an interface or a game. The device entered the market on November 2010 and one year later the company gave the possibility to connect it with user-developed Windows applications.

Figure 1. The Kinect Sensor Sensor (A) and the Kinect Sensor 2 (B). [microsoft.com]

The hardware of the Kinect Sensor is rather unique – it includes two cameras (but these are not set up as a stereovision system), infrared transmitter, four directional microphones, accelerometer, and a tilt motor. One of the cameras is a standard, color, RGB camera (resolution 640x480pixels for Kinect and 1920x1080pixels for Kinect2 – both at 30 frames per second) and it is being used for 2D video processing and for Augmented Reality (overlaying computer-generated image over the real-world image recorded by this camera). One of the exemplary uses of this camera is recognizing the face of a user (player). The second camera is only for acquisition of the depth information and it works together with the infrared transmitter. The transmitter projects a structured point cloud [1] of an infrared light onto the player(s) and their environment and the second camera (equipped with an IR filter) records the point cloud [2] to calculate the depth information. The resolution of the IR camera is 300x200 and 512x424 pixels, for Kinect and Kinect2, respectively. The Kinect/Kinect2 Sensor is supposed to operate indoors only, within the range 0.5m – 4.5m and within the field of view 57x43 and 70x60 degrees, for Kinect and Kinect2, respectively. More technical details can be found on the official webpage - [3].

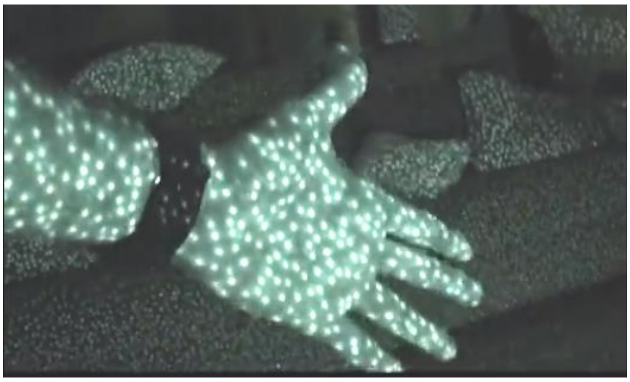

Figure 2. Infrared structured point cloud of Kinect Sensor [2].

The Kinect Sensor is vulnerable to sunlight artifacts, which can impair its function, giving wrong results of depth measurement or wrong skeleton pose estimation.

Another important issue when discussing usability and accessibility of a device is its software. The software that is available with Kinect Sensor is impressively intuitive and it does not have high requirements. It has the possibility to automatically recalibrate to a changed environment, and the software is capable of ignoring non-significant objects (e.g. obstacles, furniture, pets). It also has some skeleton-tracking procedures already implemented (up to 2 players for Kinect Sensor and up to 8 for Kinect2). Users can control an application by a proper movement of their body and/or using voice commands.

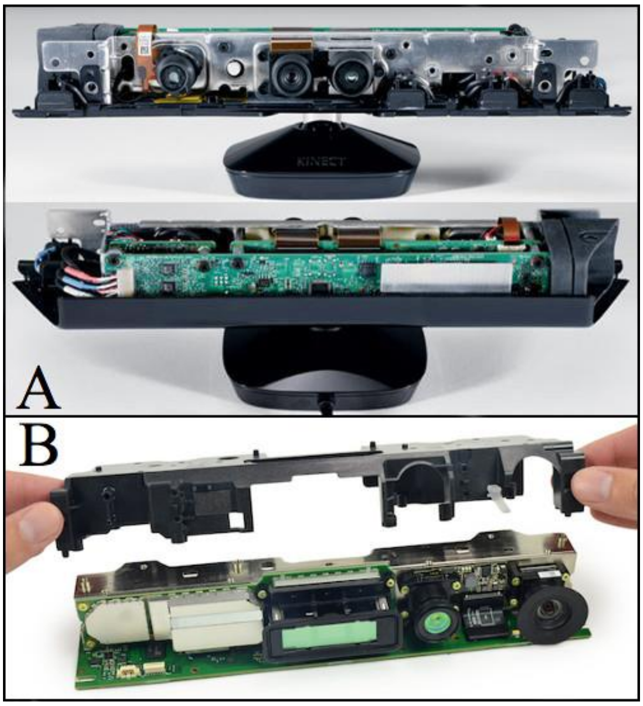

Figure 3. Inside view of the Kinect Sensor (A) [4] and the Kinect Sensor 2 (B) [5].

Of course, the device is currently mainly used in entertainment sector for video games, in which a player controls the game using his/her own body movements. The first application that uses Kinect was the Ricochet game, in which the player interacts with a virtual ball. The device acquires the movement and pose of a player's legs and hands to visualize interaction in virtual environment. Another early Kinect application was a game called Paint Party, in which the player user splashes virtual paint buckets.

Another significant step towards using Kinect as a 3D Scanner was further implementation of recognition algorithms. Peter Molyneux and his team (in Lionhead Studios) developed an AI virtual boy Milo to interact with player. The application uses face recognition for player identification, the voice recognition enables the possibility to use voice-based interaction and to react to player's answers.

Nowadays, beside the commercial uses (i.e. mainly computer games), the Kinect Sensor is being used in computer graphic projects (also in digital art, e.g. [6]) and in robotics (mainly mobile robots). Kinect is often used instead of expensive Laser Range sensors (although the output produced by Kinect is different from Laser Range Finder's output and the processing of data differs) for in-door 3D-enabled projects. The software library is being used not only in Microsoft Windows operating system, but also in other. The library gives the possibility to read vision camera data and depth camera data (as a 3D Point Cloud), an this in turn can be implemented in various object detection and recognition applications, 2D, 2,5D and 3D mapping, navigation and localization systems, inter alia in mobile robots. Another promising use of the sensor is interactive marketing.

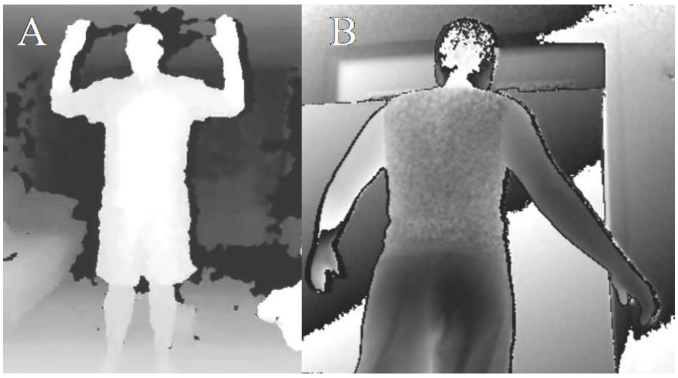

Figure 4. Depth map comparison: Kinect (A) [7] and Kinect 2 (B) [8].

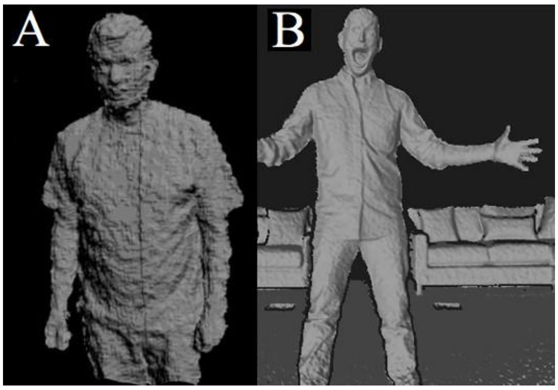

Figure 5. Point Cloud comparison: Kinect (A) [7] and Kinect 2 (B) [9].

Kinect Sensor is easily accessible, its price is affordable, and it seems to be a reasonable choice for a great majority of 3D-enabled applications. It still has its specific disadvantages regarding the embedded infrared light source and IR camera. This implies the in-door use and some requirements regarding sunlight. The algorithm for calculating the depth information requires precise information about distance between points of the Point Cloud in the 2D image acquired from the IR camera. The visualization of real-world objects in the virtual environment can be done more accurately, when the calibration of cameras is checked (and triggered if needed). The Kinect2 has better resolution than Kinect, so the accuracy of the result (precision of the 3D object acquired) should also be much better. The Kinect Sensor is also massively used in various projects involving human posture and movement analysis.

B. Sonar Ring (Adept MobileRobots Pioneer 3DX)

Another popular approach for getting the depth information is using a Sonar Ring. This simple hardware device sends ultrasonic pulses and measures the distance between the device and an object. This solution should not be compared with the Kinect Sensor, because Sonar Ring is designed to be used for completely different purpose – not for object recognition or 3D visual acquisition, but for basic walls/objects detection for a mobile robot. Sonar Ring is often considered to be the fundamental information source for depth data, because it is relatively cheap and easy to read. Sonar Ring is especially useful when using it together with compass and/or GPS (if the localization has enough good coverage for sufficient GPS precision) and/or any other information input, e.g. vision system.

Sonar Ring –based distance calculation is usually implemented as the basic, fail-safe collision avoidance system. This is because all other more sophisticated distance calculation approaches are more CPU power consuming and they are not always fast enough. The device is usually implemented similarly to an emergency stop – it has higher priority (than other movement/control commands) and/or it is checked more frequently (than other sensors) and/or it can execute highest priority avoidance maneuvers.

Sonar Ring can be also used as a coarse depth input device for mobile robot's mapping algorithms. If a mobile robot is supposed to be inexpensive, a Sonar Ring may be sufficient for basic navigation and mapping. Environment mapping (using a Sonar Ring) is usually done by engaging autonomous movement of the robot and acquisition of depth information from the sensors of a Sonar Ring during the movement.

This method can be easily implemented in the Pioneer 3DX [10] mobile robot, which is known to be one of the most popular research and educational robotic platforms.

Figure 6. Pioneer 3DX mobile robot with Sonar Ring [10].

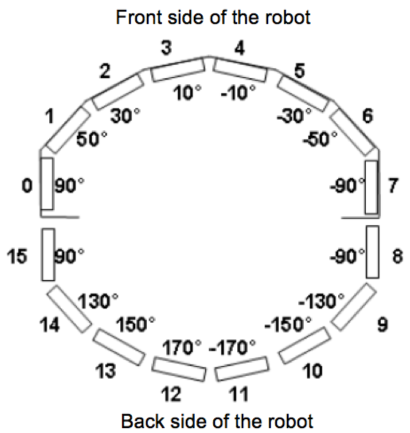

Figure 7. Placement of sonar sensors in a Sonar Ring of a mobile robot.

Sonar Rings are usually produced as a set of sonars [11] around the mobile robot (full Sonar Ring – figure 7) or only at the front half of the robot (half Sonar Ring – figure 6).

The result of 3D mapping procedure, done using a Sonar Ring, usually has a relatively poor quality. The depth information from sonar sensors, coupled with visual information from a vision system, is hardly satisfying even for the lowest expectations. Using a Sonar Ring for more complicated 3D-imaging tasks than navigation is pointless. Sonars are definitely not precise enough to use them in a 3D scanner device for digitization of real 3D objects.

C. Laser Range Finder 2D and 3D Scanners

One of the most advanced techniques for acquisition of depth information is Laser Range Finder (LRF). The measurement is done by sending a beam of light (visible or infrared) emitted by a laser diode and measuring the time interval until it reflects from an object. The term (and acronym) LRF refers to a device used for an optical distance measurement, but it is often used also for Laser Range Finder Scanners.

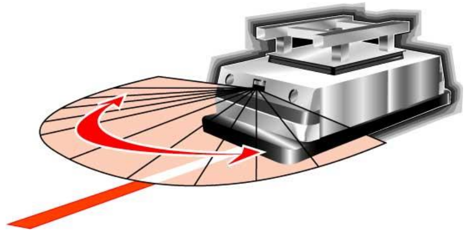

Figure 8. Laser Range Finder 2D Scanner on a mobile robot base [13].

LRF Scanners can manipulate the beam's direction to swipe a 2D surface (as in figure 8) or to scan 3D space. 3D scanning is usually done as a progressive scan (2D – line by line), although there are some implementations of using a (cheaper) 2D LRF Scanner on a motorized tilt head. In that case, the laser swipes horizontally, while the whole LRF Scanner device is moving up and down (bringing the vertical axis aspect to the 2D measurements). Swiping while tilting gives more problematic output, so for better output handling a 3D LRF Scanner should be considered. 3D Scanners are also better for detection/recognition of moving objects and for mobile robots that have to perform 3D mapping of the environment (scene) or objects.

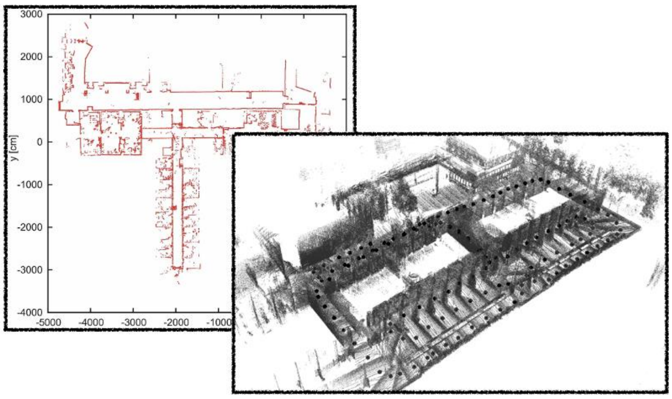

Figure 9. Maps generated using Laser Range Finder Scanners.
Left/back: examplary 2D LRF Scanner map [14],
right/front: examplary 3D LRF Scanner map [google.com].

Figure 10. Examples of the LRF (Laser Range Finder) devices. Left: 2D LRF Scanner SICK TiM3xx. Right: 3D LRF Scanners SICK JEF3xx. [15][16]

In general, Laser Range Finder Scanners are simple, yet very precise, devices. An LRF scanner consists of a laser diode, emitting infrared light and a light detector with a filter for specific wavelength (the same as the laser diode). In addition, LRF Scanner contains at least one rotating mirror (or a system of rotating mirrors) for precise changing of the laser beam's direction. The mirror(s) movement is controlled by a microcontroller. The LRF sends thousands of light impulses per second (up to hundreds of thousands of impulses per second), pointing every light beam into another direction, using the mirror. Every light beam's time is measured and the distance is calculated. The microcontroller is storing the information about the mirror's angle and the distance – for every point in the range (field of view of the mirror/LRF). The distance is calculated using the usual equations for speed of an electromagnetic wave in the air.

LRF Scanners are used not only in stationary Scanners or mobile robots – they are extremely popular, especially in the most demanding applications: Unmanned Aerial Vehicles (UAVs), Unmanned Ground Vehicles (UGVs), also in military applications, and in semi-autonomous and tele-operated robots. The aerial applications usually involve a 3D LRF, whereas ground applications use 2D and 3D [17][18].

If LRF is to be implemented in a 3D Imaging Scanner for digitizing 3D real-world objects to a virtual environment, a 2D LRF Scanner can be used – either as a stationary device with the object rotating on a plate, or as a device moving around a stationary object. 2D LRF Scanner implemented into a 3D Imaging Scanner can be mounted in a vertical position.

In practice, LRF Scanner devices are so precise and universal, that no special background is needed, they can operate in any lighting conditions, they do not require any special position or movement of the object. Using LRF Scanners as 3D Imaging Systems can significantly increase the precision and decrease time needed for every scan. Unfortunately such precise devices are very expensive, which in turn makes this solution not affordable for an average, decent hobbyist.

D. 3D Imaging Scanners

There are many solutions on the consumer market, which are being sold as 3D Imaging Scanners (in short: 3D Scanners). Some of them are stationary devices, and some of them are handheld scanners. Regardless of their construction, they usually have similar features: visible laser light source (green or red), a lens to get a laser line, a camera for acquisition of visual data, a mechanism for rotating the object (or any other kind of generating a relative movement) and a software for generating an object's model, basing on the displacement of the laser light in the images.

For some scanners it is crucial to provide proper lighting conditions, because the best quality (least errors and artifacts) can be achieved in low light.

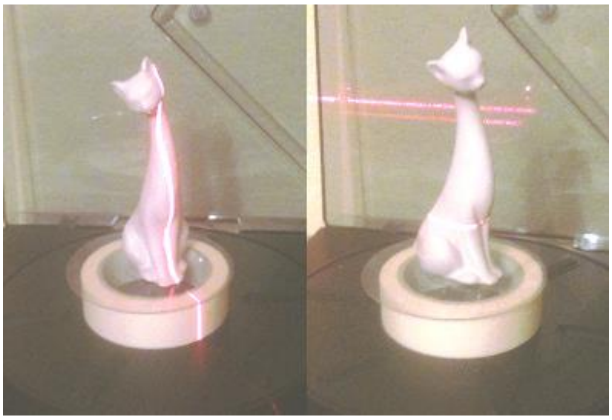

Figure 11. Home-made 3D Imaging Scanner. Left: vertical scanning mode (more useful one), right: horizontal scanning mode [own work].

The appropriate lighting requirement, as well as the appropriate relative position requirement makes it very difficult to implement this technique in a mobile robot application. Besides, a mobile robot should not use a visible laser when interacting with human.

E. Stereovision Camera (Point Grey Bumblebee2)

The last solution, which is discussed in this paper, available on the market, is a stereovision camera. Stereovision (stereoscopic vision) is a method of 3D acquisition using two cameras. The idea of stereovision is biologically-reasoned – people (as well as the majority of animals) have two eyes, which gives the possibility to distinguish objects which are closer. The process of stereoscopic vision is not so intuitive for machines/computers and using a home-made stereovision system may be a huge challenge. Two cameras have to be firmly mounted, aligned, configured, calibrated and then they are ready to be used for acquisition. These tasks (especially: calibration) are complex and they should be done precisely – however, there is no need to perform them if we buy an out-of-the-box solution, just like the most popular BumbleBee2 Stereovision system from Point Grey (formerly: Videre Design). The BumbleBee2 camera uses two Sony 1/3" CCD (Charge Coupled Device) matrices. This gives the opportunity to use progressive scan technology. Beside the Stereovision system, the user gets also developer's software - FlyCapture SDK and Triclops SDK. The FlyCapture software is compatible with various operating systems and its components offer an easy to use and complete library for acquisition, processing, saving and displaying the visual information. The latter library, Triclops SDK, supports additional library tools for image rectification and stereo processing, but the most important functionality is generating precise and real-time-generated depth map [19].

Stereovision is a method of acquisition of the depth information, based on the binocular vision, by finding and showing the binocular disparities. Stereovision gives the possibility to calculate coordinates of points in a 3-dimensional scene, basing on two images acquired from two cameras. Binocular computer vision is based upon the biological binocular vision, and just like the biological version it takes two images from two inputs, calibrates the images, and then calculates the depth information using the disparities.

Figure 12. Stereovision cameras (Point Grey BumbleBee 2).

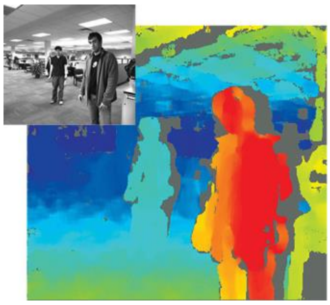

Figure 13. Depth image from a Stereovision camera [20].

In order to obtain one-to-one mapping of the acquired pixels of the real world objects in the 3D space, more information is needed. New pairs of images give further information about the scene, and the mutual relative differences/similarities produce the final knowledge about the depth parameters of the objects in the scene. The exact position of an object in the reality can be then inferred and visualized as a virtual object in the virtual reality, as a 3D model.

The BumbleBee2 camera is one of the most popular and best-ranked stereovision solutions available on the market for 3D visualizing of real-world objects. The results are fairly precise and the depth image is smooth. The camera has superb reviews of its users, despite its high price. The financial cost makes it unaffordable for some hobbyists.

III. Single Camera Low-Cost 3D Imaging System

The authors are currently implementing a software library for vertical 3D Imaging Laser System, and then they plan to modify it so that it could scan the depth information without the laser. This will be achieved by only recording the visual information (frames) and by rotating the plate with a specific angular speed. When the plate (and the object) rotates by a specific angle, the frames will be synchronized and compared with previous frames, so that these two frames (the old frame captured at the old angle and the new frame captured at the new angle) would be treated as a regular stereovision input. They will be synchronized (calibrated) using correlation and then the displacement will be calculated to get 3D information.

The only hardware needed will be a cheap USB HD camera and a rotating plate.

The next step will be to implement this concept in a mobile robot's vision system and to use the robot's movement (instead of a rotating plate) for the camera displacement. This would be similar to the so-called motion parallax. This surely will be more complicated, while the distance between robot and object would be able to change, and because neither the camera would move left/right, nor the object would rotate (with an eligible angular speed).

IV. CONCLUSIONS AND FUTURE WORK

In this paper authors have briefly described the most popular approaches to 3D imaging for mobile robots. The advantages and disadvantages of each solution were discussed. This paper may be helpful to readers willing to building a new 3D Imaging system or a mobile robot.

The horizontal goal of authors' research is to develop a human-centered vision-based Human Computer Interaction System that would be somehow similar to the human vision. Human bodies don't have Sonar Rings or LRF Scanners, yet we perfectly understand the depth information we get. Applying a stereovision system might be a legit way to mimic the biological vision, but human is capable of understanding 3D even using only one eye. Nowadays, when cameras have much better resolution than before, it might be finally possible to implement the proposed approach.

REFERENCES

[1] Fisher M., webpage: "Interpreting Kinect Sensor Values", online, accessed 2014-11: https://graphics.stanford.edu/~mdfisher/Kinect.html

[2] Reza A., video: "Kinect with Nightshot", online, accessed 2014-11: https://www.youtube.com/watch?v=nvvQJxgykcU

[3] Microsoft, Kinect For Windows, 2014, online, accessed 2014-11, http:// www.microsoft.com/en-us/kinectforwindows/meetkinect/about.aspx

[4] Goss P., "Microsoft's Kinect gets dissected", 2010, online, accessed 2014-11: http://www.techradar.com/news/gaming/consoles/microsoft-s-kinect-gets-dissected-708349

[5] Perez D., "Xbox One Kinect Teardown Reveals Its Fairly Easy To Repair", 2013, online, accessed 2014-11, http://www.ubergizmo.com/ 2013/11/xbox-one-kinect-teardown-reveals-its-fairly-easy-to-repair/

[6] Kiefer C., Franke D., "unmanned sculpture" art dance project, documented online, in: Kirn P., "Dancing in Point Clouds: unnamed soundsculpture, Making Of, and Human Element", online, accessed 2014-11, http://createdigitalmotion.com/2012/04/dancing-in-point-clouds-unnamed-soundsculpture-making-of-and-human-element-kinect/

[7] Cogan Y., "First Kinect 2 Depth Image?", 2012, online, accessed 2014-11, http://123kinect.com/first-kinect-2-depth-map-image-2/36985/

[8] Pterneas V., "Kinect for Windows version 2: overview", 2014, online, accessed 2014-11, http://www.codeproject.com/Articles/723665/Kinect-for-Windows-version-overview

[9] brekel.com webpage, kinect2_depth.jpg, online, accessed 2014-11, http://brekel.com

[10] MobileRobots, official MobileRobots webpage, online, accessed 2014-11, www.mobilerobots.com

[11] Teruko Yata, Akihisa Ohya, Shin'ichi Yuta, "Fast and accrate sonar-ring sensor for a mobile robot", Proceedings of IEEE Intl. Conf. on Robotics and Automation, Vol.1, May 1999, accessed: 2014-11, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=773970

[12] Mo3DLS app turns your smartphone into a 3D Scanner, online, accessed 2014-11, http://www.3ders.org/articles/20130225-moedls-app-turns-your-smartphone-into-a-3d-scanner.html

[13] Hokuyo Automatic Co. Ltd., LRF 2D Scanner image, online, accessed 2014-11, https://www.hokuyo-aut.jp/02sensor/07scanner/pbs.html

[14] Biber P., Andreasson H., Duckett T., Schilling A., 3D Modeling of Indoor Environments by a Mobile Robot with a Laser Scanner and Panoramic Camera, online, accessed 2014-12, http://www.gris.uni-tuebingen.de/people/staff/biber/indoor3D/

[15] SICK AG, Sick Sensor Intelligence – 3D laser scanners, online, accessed 2014-11, http://www.sick.com/group/EN/home/products/product_portfolio/laser_measurement_systems/Pages/outdoor_laser_measurement_technology.aspx

[16] SICK AG, Sick Sensor Intelligence – 2D laser scanners, online, accessed 2014-11, http://www.sick.com/group/EN/home/products/product_portfolio/laser_measurement_systems/Pages/indoor_laser_measurement_technology.aspx

[17] LaserTech, "Laser Technology Measurably Superior", online, accessed 2014-11, http://www.lasertech.com/TruPulse-Laser-Rangefinder .aspx

[18] SICK AG, LMS 200 Laser Range Finder user manual, SICK AG Waldkirch.

[19] PointGrey, PointGrey – Bumblebee 2 1394a, online, accessed 2014-11, http://www.ptgrey.com//bumblebee2-firewire-stereo-vision-camera-systems

[20] PointGrey, Triclops SDK, online, accessed 2014-11, http:// www.ptgrey.com/triclops