Institute of Parallel and Distributed Systems
Applied Computer Science - Image Understanding
University of Stuttgart
Universitaetsstraße 38
D–70569 Stuttgart

Fachstudie Nr. 169

Tracking of persons with camera-fusion technology

C. O. Anika Holtermueller Joerg Ploedereder

Course of Study:	Computer Science
Examiner:	Oliver Zweigle
Supervisor:	Kai Haeussermann
Commenced:	Juli 01, 2012
Completed:	September 30, 2012
CR-Classification:	I.4.8, I.2.10, I.2.9

Abstract

The idea of a robot tracking and following a person is not new. Different combinations of laser range finders and camera pairings have been used for research on this subject. In the last years stereoscopic systems have been developed to compensate shortcomings of laser range finders or ultra sonic sensor arrays in means of 3D recognition. When Microsoft began the distribution of the Microsoft Kinect in the year 2010 they released a comparatively cheap system, that combines depth measurement and a color view in one device. Though the system was intended as a new remote controlling system for games, tackling the market launch of motion sensing wireless controllers such as the Nintendo Wiimote and Sony Playstation 3 move some developers saw more in this technology.

And so it did not take long until the first hacks for the Microsoft Kinect were published after the initial release. More and more people started to create own software, ranging from shadowpuppets [TW10] to remote controlling home cinema systems [Nar11]. Microsoft and PrimeSense soon recognized the potential and released free drivers and sdks for the use of the camera device with PCs.

With Prime Sense publishing the drivers as open source a lot of possible uses came up for the Microsoft Kinect. Some companies used this event to enter the market of camera-fusion technology. The most comparable system to the Microsoft Kinect is the Xtion Pro Live by Asus.

These devices merging the depth measurement and color view with computation on a device-internal system reveal new possibilities concerning the tracking of persons and enabling them to even give the robot commands using gestures. This paper shall inquire to what extend the Microsoft Kinect or the Asus Xtion Pro Live can be used as substitute for stereoscopic cameras/laser range finder systems in context of a tracking and control device for human robot interaction scenarios with person following applications for service robots.

Contents

List of Figures

List of Tables

1 Introduction

The reason for the research of this paper is the idea of having a service robot following a human guide.

The service robot shall perform a certain task at designated points and so help easing up the work of its guide for instance by carrying heavy loads or placing an object at a precise location in an unknown environment. The service robot is to be guided to the designated points by the human guide and shall begin the task automatically on reaching the intended destination. For that purpose a robot-chassis with omni-directional drive and a tracking system for the service robot is needed. Furthermore some image recognition system or special signal is required for the robot to know at what point it should stop following the guide and begin the task. Also some way of reinitializing the following mode after completing a task or loosing the guide is needed. A minimization of direct physical interaction between the human guide and the robot is desired. Therefore it is intended that the robot controls the initiation of the task to be performed after reaching its designated position.
The approach is to utilize either the Microsoft Kinect or the Asus Xtion Pro Live as camera devices for tracking guides. Both devices shall further be referenced as devices or camera systems unless further specification is mandatory. Using the systems ability to create a RGB picture and a depth map via laser grid mapping as well as the skeleton detection implemented in the drivers of the camera systems it is intended to build the tracking and control system with either camera system, depending on their usability for the tasks at hand.
To minimize the need of direct physical interaction between human guide and robot the system is required to be able to track the guide and detect his gestures, if he makes any. These requirements are met by Microsoft Kinect and Asus Xtion Pro Live using their native software. With the objective of following a guide, there are a few challenges and problems that must be considered.

Organisation

This document is organized into the following chapters:

Chapter 2 – Challenges and Problems: First of all the problems and challenges for the complete system are introduced and discussed according to their importance for the task.

Chapter 3 – Technical data of cameras: Here we describe the technical design and data of the two camera systems.

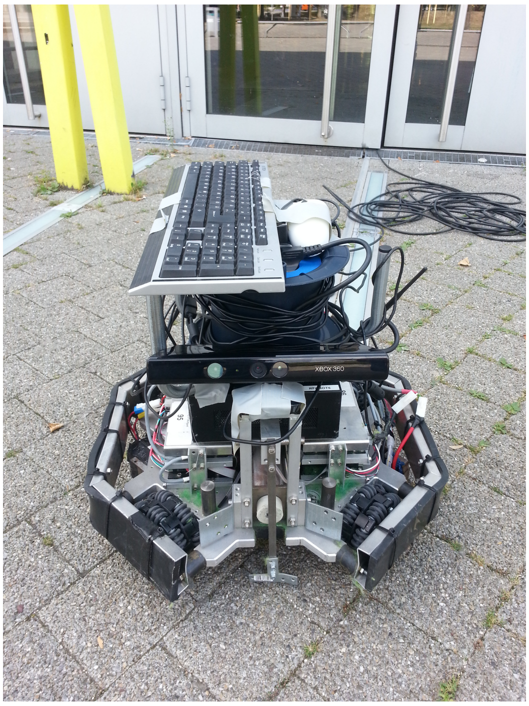

Figure 1.1: Assemble of the service robot prototype with Microsoft Kinect and needed power cord (without 3D environment scanner

Chapter 4 – Comparison of the camera systems in use: The data collected from testing the different camera systems in changing environments and situations is evaluated, presented and compared within the third chapter.

Chapter 5 – Conclusion Finally we present our verdict on which camera suites the task at hand the best and also show further possible work needed on the subject.

2 Challenges and Problems

There are different challenges and problems for the whole system on behalf of the tracking and interaction functionality. Obstacles, stairs, detecting and tracking the correct person and the correct detection of gestures to name the most important problems. Those and further problems shall now be analyzed and given a possible solution.

2.1 Initializing a guide

Beginning with the most prominent of challenges, the detection of the guide, we must distinguish the guide from the rest of the scene. Different ways to achieving this objective can be followed. First and easiest would be to let the robot idle until a user steps up into its field of view.The next possibility would be, that the robot turns around its center axis to find a guide. Note that there is a certain threshold to the rotation speed of the robot, which is given by the camera systems ability to recognize a possible guide correctly against a constant changing motion-blurred fore- and background [TSY$^+$08]. A recommendable technique is a mixture of idling and rotating. To minimize the problem of receiving motion blurred data, causing the robot to misinterpret the environment and missing its guide, the robot rather only scans its field of view for a guide. If no guide is found the robot turns a predefined angle to scan from anew. The angle is derived from a defined overlay which two adjoining field of views should have. Either way we recommend requiring either an initialization pose from potential guides or setting a predefined marker that the guide must wear, so that bypassers do not "hijack" the robot accidentally whenever no guide is determined.
To determine the guide from the surroundings the functionality of the camera systems with their RGB image and depth map is used. These functions enable the system to differ objects from each other. To then decide which object forms the guide, the skeletal overlay has to be made over all visible objects. Sometimes the skeletal overlay falsefully identifies an object as person. To avoid having the robot focusing on this object rather then the real guide the initializing pose or the marker would also be rather helpful.

2.2 Tracking

Having found the guide the tracking begins and with it come the main problems of keeping the guide in sight and not letting people who cross the path of the guide and the robot "hijack"

the robot.
Due to certain architectural or natural features it can happen that the guide is lost from the field of view. This happens mostly out of the reason, that the line of sight to the guide is obstructed. Should the guide be lost, the robot is recommended to move to the last known position of the guide, rather then stopping at its current position. On reaching this position the robot must rescan the environment to regain its guide. If the guide cannot be found within the frontal field of view, the robot must begin with a rotating search for the guide. As with the initializing rotation this rotation can only be executed anglewise [TSY+08].
Relying on the results of resaearch on robots using stereoscopic cameras as visual guiding devices the problem of distinguishing a person from surrounding objects during rotation has been classified as hard to solve. Therefore this feature must be observed with imminent care regarding the camera systems intended to be used.
To increase performance of reacquiring the guide, the robot should be able to distinguish if its guide was more to the right or the left of the last image. Depending on the guides alignment in the image the robot begins its rotation routine in the according direction.
To ensure finding the correct guide and guarantee user comfort, the robot is rather to rely on a color matching functionality or marker searching then on an initialization pose. The RGB matching for the guide minimizes the chance of choosing a wrong guide in combination with the skeletal overlay even further. Iocchi of the University “La Sapienza” of Roma shows that this the most promising operating procedure [CIL07]. The RGB matching enables to either predefine certain textile colors the guide must be wearing or has the possibility to save the color combination worn by the guide at the moment of his initializing pose. Either way, it ensures a better tracking, since the system compares the colors of all visible objects under skeletal overlay and thereby distinguishes the correct guide from interfering entities possibly recognized as guides.
The idling concept used for initialization is not of use in this case, because it forces the guide to constantly return to the front of the robot to reinitialize the tracking. Idling therefore is if at all only useful for the first initialization or for reinitializing the robot after it has preformed a 3D scan of the area. Under all other circumstances the robot must try find its guide on its own if he is lost from sight.
After considering all above situations one problem can never be eliminated. The problem of usage in a envirnoment with dress code. This dress code can result in conflict with the color matching algorithm in a way that no guide can be assured to an absolute.

2.3 Obstacles

Obstacles appear in the way of the robot when following the guide. The depth map ensures the detection of obstacles in the field of view of the camera system. To ensure that the robot does not collide with objects outside its field of view, either ultra sonic or laser range sensors have to be installed at the side and back of the robot. For further safety against missed detections of stairs or similar hazardous objects in the path of the robot another sensor is installed in front. This sensor is preferable an ultra sonic sensor, since laser range finders do not work reliably when glasswalls are in the vicinity. Laser range sensors penetrate glass and therefore

do not provide adequate data for a needed safety stop. Ultra sonic sensors however reflect on the glass, therefore giving a reliable signal for needed safety stops even with the problem of interference through sonic turbulence from the environment.
However being not part of the question if Microsoft Kinect-like systems are capable as tracking device for robotics, we shall only recommend most safety sensors and their setup yet not be testing them. The only obstacle recognition we shall be testing is achieved by our camera systems.

2.3.1 Stairs

Stairs need to be taken into special account with the robot system. While the camera system is able to detect stairs in its field of view, the situation can arise that the robot must turn while tracking the guide and have stairs in front of itself yet outside the field of view. To distinguish if there are stairs, the front sensor applied to distinguish frontal obstructions is not aligned parallel to the floor. It is rather arranged in a angle towards the floor below the field of view of the camera system. A sudden change in measured distance with a greater value then a defined epsilon-value indicates a stair in front of the robot. The defined epsilon-value helps the robot in differentiating between a ramp or stairs, so that the robot does not falsely stop in front of ramps. Futhermore the definition of an epsilon-value allows us to hinder the robot on taking ramps that might bring the robot to topple due to a high weight point provided by the 3D scanning device on top of the robot.

3 Technical data of cameras

Since the intention of this paper is to find out which camera system is better suited for the task at hand, we have to determine the similarities and differences of the two camera systems. The compared systems are the Microsoft Kinect for PC and the Asus Xtion Pro Live.

Both systems are comprised of two cameras, a projector and a certain count of microphones. One of the cameras supplies RGB valued images, whilst the other is an infrared camera intended to capture the reflective points of the infrared laser grid emitted by the projector.

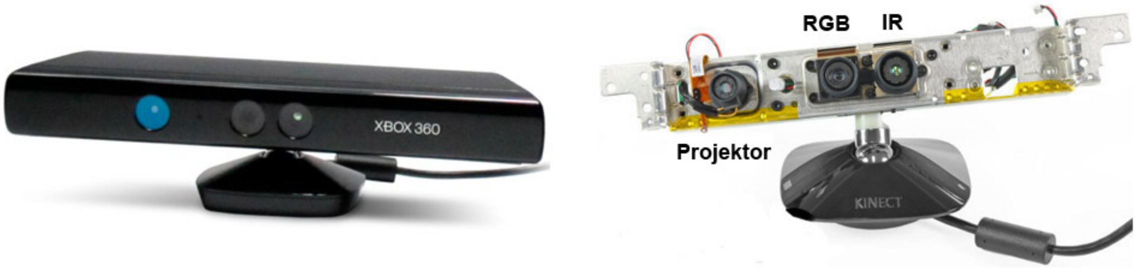

Figure 3.1: Assemble of the Microsoft Kinect[Kof11]

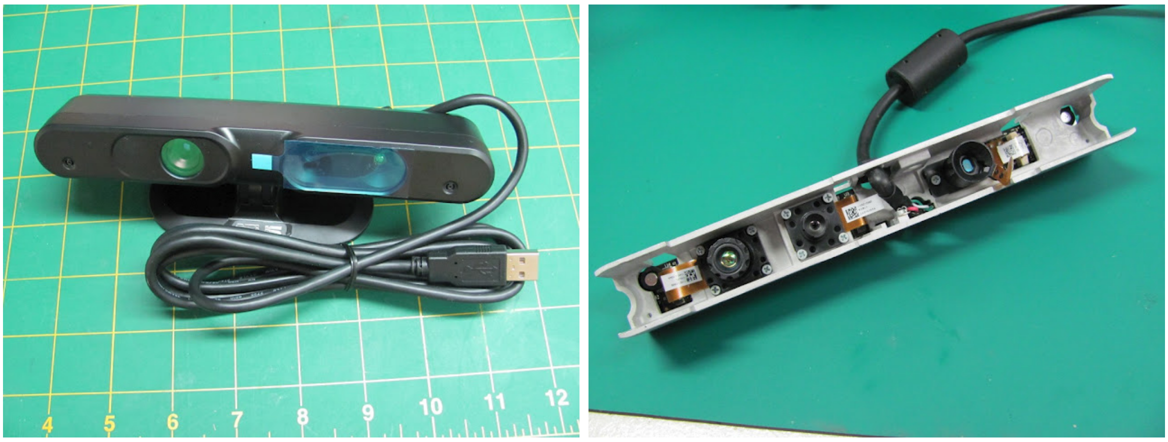

Figure 3.2: Assemble of the Asus Xtion Pro Live [http://www.iheartrobotics.com/2011/09/asus-xtion-pro-live-unboxing.html]

The laser grid is emitted with 830nm [cad11, ope] of wavelength. And the RGB image capable cameras have individual resolutions.

The following sections shall show the individual cameras technical data in detail, before comparing the two systems to each other.

3.1 Microsoft Kinect

3.1.1 Design and features

The build of the Microsoft Kinect is rather simple. It combines a VGA camera for normal RGB image streams, a depth sensor to measure distances of objects to the Microsoft Kinect and a microphone array. All is packed into a device having a complete measurement of 30 by 6 by 8 centimeters, based on a stand that holds a motor enabling the sensor array to tilt automatically. Due to the motor the Microsoft Kinect is in need of more electricity then a standard USB-port is able to supply. Depending on the model of the Microsoft Kinect it is either needed to use a special USB bridge cable combining the USB-cable to the robot and an external power supply or having the USB-cable and the power supply on different ports, as the Microsoft Kinect Xbox 360S enables with an AUX-port for the power supply.
Sadly Microsoft has not build the Microsoft Kinect to function as motion tracking device without the external power supply. Without the power supply only the normal VGA camera has function.

3.1.2 Field of view

The Microsoft Kinect has a field of view of 57° in the horizontal axis and 43° in the vertical axis. The motor in the stand enables the Microsoft Kinect to tilt up to 54° in total, tilting either up or down for 27°. Defined by these values a certain distance must be held, so that the Microsoft Kinect can properly register a user and interpret his gestures. Microsoft's datasheet recommends 1.2 meters (3.9 feet) as minimum distance and 3.5 meters (11 feet) as maximum distance.
It is however possible to deactivate the auto-calibration of the sensing range, to force the Microsoft Kinect to work between the distance of 0.7 and 6 meters (2.3-20 feet). Depending on the environment lighting and the predefined needed skeletal parts to interact with the program the tracking can be rather poor in the regions outside the predefined 1.2 to 3.5 meters (3.9-11 feet).
If the full body is needed the Microsoft Kinect running on its SDK will not allow any closer range that takes the head or the legs above the knee out of its field of view. Therefore the different heights of people result in different minimum distances. For our tests our smallest proband was 1.65 meters (5.4 feet) resulting in a minimum full body tracking distance of 2.3 meters (7.5 feet) and a maximum distance of 3.2 meters (10.5 feet) under full artificial lighting. Our largest proband was 1.95 meters (6.4 feet) resulting in a minimum full body tracking distance of 3.1 meters (10.2 feet) and a maximum distance of 3.2 meters (10.5 feet) under full artificial lighting.

3.1.3 Resultion

The resolution of both depth and VGA sensor are to be at 640x480 streaming with a 30Hz output. The VGA camera sends an 8bit stream of Bayers color encoded images. Whereas the depth sensor sends a monochrome 11bit stream with 2.048 levels of sensitivity.
While Microsoft claims that the Microsoft Kinect can uphold the 640x480 resolution on all sensors at all times, tests on the PC with non-Microsoft products have fallen short of coming up to the 640x480 resolution on the depth sensor with 30Hz. This resolution tends to jump around bringing up problems for some programs reading out the streams, since they have to recalculate the individual images to match the same resolution. This recalculations seldom occur within realtime so reoccurring lags bring up problems that can result in up to the loss of the user. Stable results can be achieved when the depth sensor is restricted to a 320x240 resolution. The stream keeps running with 30Hz.

3.1.4 Audio

The Microsoft Kinect has 4 microphones distributed over the sensorhull so that it can discern the direction of any incoming audible input. The microphones record in a 16bit audio stream with a 16kHz sampling rate.
It is possible to use the microphone array as further means to give orders to the robot, however for our purpose this is not intended.

3.1.5 Further features for PC

Compared the the Microsoft Xbox PCs are intended to being able to use up to four Microsoft Kinects simultaneously. This would enable the possibility of a 228° horizontal field of view. Thereby minimizing the loss of the guide or the need of either ultrasonic sensors or laser range finders all around the robot. However the need of 4 extra power supplies would seriously drain the power resources of the robot or even bind it into an environment where it can have constant power supply not depending on batteries [Geo11]. And the camera laser grid projections can interfere with each other degrading the image quality [4.5].

3.2 Asus Xtion Pro Live

3.2.1 Design and features

The build of the Asus Xtion Pro Live is quite similar to the Microsoft Kinect but some important differences apply. It combines a VGA camera for normal RGB image streams, a depth sensor to measure distances of objects to the Asus Xtion Pro Live and a pair of microphone. All is packed into a device having a complete measurement of 18 by 3.5 by 5 centimeters, based on a stand enabled for manual tilt. Power supply is achieved over the

USB-cable from a simple USB-port.

3.2.2 Field of view

The Asus Xtion Pro Livehas a field of view of 58° in the horizontal axis and 45° in the vertical axis. The manual tilt can be up to 90°.Defined by these values a certain distance must be held, so that the Asus Xtion Pro Live can properly register a user and interpret his gestures. The Asus Xtion Pro Live datasheet recommends 0.8 meters (2.6 feet) as minimum distance and 3.5 meters (11 feet) as maximum distance.
Depending on the environment lighting and the predefined needed skeletal parts to interface with the program the tracking can be rather poor in the regions outside the predefined 0.8 to 3.5 meters (3.9-11 feet). These measurement tests were achieved using the PrimeSense SDK. If the full body is needed the Asus Xtion Pro Live must be initialized by the predefined pose with the body standing inside the field of view. After the initialization of the skeleton it is possible to move freely inside the distance of 0.8 and 3.5 meters (3.9-11 feet) without losing track. However some gestures might not be recognized due to decreasing size of the field of view when closing in on the sensor.

3.2.3 Resultion

The resolution of the VGA sensor is under SXGA definition at 1280x1024 streaming with a 30Hz output. The depth image size can be varied between a VGA resolution of 640x480 with 30Hz and a QVGA resolution of 320x240 with 60Hz.

3.2.4 Audio

The Asus Xtion Pro Live has 2 microphones distributed over the sensorhull. The microphones record in a 16bit audio stream with a 16kHz sampling rate.
It is possible to use the microphone array as further means to give orders to the robot, however it is not intended within the robots function.

3.2.5 Further features for PC

Having been developed purely for the use on the PC itself, no further features then the developed bundle are given on the PC.

3.3 Compairing technical data of the camera systems

A simple technical comparison is based on these values.

	Microsoft Kinect	Asus Xtion Pro Live
Field of view		
Horizontal:	57°	58°
Vertical:	43°	45°
Depth range:	1.2 - 3.5m	0.8 - 3.5m
Data Stream		
Depth	640x480 16-bit @ 30 frames/sec (XBox Specification) 320x240 16-bit @ 30 frames/sec (PC based)	640x480 16-bit @ 30 frames/sec 320x240 16-bit @ 60 frames/sec
Color	640x480 32-bit @ 30 frames/sec	1280x1024 @ 30 frames/sec
Power Supply	12V DC + 5V USB connection	5V USB connection
Audio	4 Microphones	2 Microphones

Table 3.1: The table shows the comparable technical data of the camerasystems

Alone by these values we can tell, that the Asus Xtion Pro Live has a rather grand technical advantage against the Microsoft Kinect. It's field of view is greater. The Microsoft Kinect has a field of view size of 1.3x0.95 meters (4.3x3.12 feet) at minimum distance and 3.8x2.76 meters (12.5x9 feet) at maximum. The Asus Xtion Pro Lives field of view is 1.33x0.99 meters (4.4x3.3 feet) at the Microsoft Kinects minimum distance and 3.9x2.9 meters (12.8x9.5 feet) at maximum. This is a clear point for the Asus Xtion Pro Live.
The only advantage the Microsoft Kinect has against the Asus Xtion Pro Live on the field of view is its motor allowing it to tilt up to 54° in the vertical direction. Thereby increasing its possible field of view on the vertical axis to a full 97°. This field however can only be used to search for a guide or if the camera were to check safety sensor reading on obstacles in front of the robot. Latter is rather inconvenient for it has a very high risk of losing the guide and therefore having to restart the initialization. It does however come in hand when initializing the guide, since the tilt can enable the Microsoft Kinect to detect guides of tall build in close proximity.
The Asus Xtion Pro Live allows a tilt, but only manually. So the user must either define the best angle beforehand on a spare monitor or rely on his adjustment to bring him best results.

The depth range of both devices also show, that the Asus Xtion Pro Live has a greater focus span. Again as mentioned beforehand it is possible to force the Microsoft Kinect to keep its depth focus up in a distance between 0.8 and 1.2m making it at least as capable as the

Asus Xtion Pro Live in range. However to achieve this the autofocus of the Microsoft Kinect must be disabled, decreasing the Microsoft Kinects overall quality in depth range perception. Again the Asus Xtion Pro Live makes the race under these aspects.

The most grave difference between the two camera systems is the resolution. Whilst Microsoft claims that the Microsoft Kinect is able to achieve a resolution of 640x480 on the RGB and depth range channel, only the resolution of the RGB channel can be confirmed on a PC environment. The depth channel resolution of the Microsoft Kinect is only able to achieve an unstable frame rate. If the Microsoft Kinect is forced to keep up a predefined frame rate of 30 frames per second it constantly switches resolution size depending on the amount of movement perceived. This constant switching results in a computational challenge when comparing the RGB and the depth measure input in different resolution to differ all entities in the field of view.
To ensure a quality of resolution and the quantity of 30 frames per second the PC driver forces the Microsoft Kinects resolution to 320x240. The Asus Xtion Pro Live however has more than double the Microsoft Kinects RGB channel resolution, with 1280x1024. This enables amongst others a much smoother and more determined color matching. Its depth channel provides either guaranteed a 640x320 with 30 frames per second or if wanted a higher frame rate of 60 at the resolution of 320x240.
This concludes that using comparable resolutions, the Asus Xtion Pro Live has double the frame rate, enabling a far better computation of possible obstacles whilst traveling. Also movement does not affect the Asus Xtion Pro Live as much as the Microsoft Kinect due to its higher frame rate or higher resolution as different test with variation of frame rate and resolution had shown. Therefore the Asus Xtion Pro Live is enabled to better compute changes in the environment.

Allowing us to come to the subject of power supply: The Microsoft Kinect is as mentioned unable to run on only the power provided by the USB-port due to the motor in its socket. It needs an extra 12 Volts direct current to work as a full system. The Asus Xtion Pro Live on the other hand is content with the power provided by a simple USB-port. Again the Asus Xtion Pro Live is preferable.

The only point that the Microsoft Kinect can completely call for itself is the fact, that it has 4 microphones enabling it to pinpoint the source of a sound quite well. The Asus Xtion Pro Live only has 2 microphones, also enabling it to receive voice commands, but not being able to map the origin of the sound on a certain point. Thereby commands uttered by non-guiding persons may easily interfere with the Asus Xtion Pro Live guided robot. Yet since voice commands were not intended, this feature does not serve as a high priority argument.

3.3.1 Conclusion based on technical data

Based on the comparison of the technical data and the intended field of use the Asus Xtion Pro Live has a clear advantage over the the Microsoft Kinect. Its higher resolution and grander field of view with broader focus region gives the system a more precise amount of data, on

which it may react. Futhermore the Asus Xtion Pro Live is smaller in size, allowing it to be mounted more easily onto the robot. Its lack in motor within its base allows to detach the stand without damaging the device itself.
Therefore based only on the facts of technical data, the momentarily recommendation would lie on the Asus Xtion Pro Live.
We shall however first proceed with the evaluation and comparison of there achievements in the field of work before we come to a final verdict.

4 Comparison of the camera systems in use

The testsystem was setup using an omnidirectional driving chassis. The computing system is a Intel Core2Duo 2.2 GHz with 1 GB RAM and an Ubuntu 12.04 OS. The driver was implemented for the ROS Framework making use of the tf-tracker and the ROS OpenNI Support.
The camera systems where placed in a height of 35cm (1,15ft) with an upward tilt of 15°. Sadly the Asus Xtion Pro Live is not able to run native on the OpenNI node from ROS, due to the stream being send in Bayer 32 encoding. Its depth stream is fully available, which leaves us with most of our test possible. With exception of the automated following mode due to need of colormatching all tests where however possible in a manual way.

4.1 Creation of depth images:

As mentioned in the technical details the Microsoft Kinect and Asus Xtion Pro Live devices are fitted with a Class1 Laser operating at 830nm wavelength.
Instead of travel-time measurements as often suspected the camera systems use a variation of structured light. The laser beam itself is diffused into a dot pattern and projected into the room. The reflection of the grid points are read by the IR-camera or more precisely by a monochrome infrared CMOS Sensor (Aptina MT9M001). PrimeSense uses a patented variation of structured light, this technique is called Coded Light [pri11] [cod]. A reference plane at a fixed range is used to determine the distance of objects in its field of view. The re-emission of the coded light pattern is then evaluated using triangulation between the IR-laser and IR-camera and delivers rather accurate depth information with a maximum error of 4cm [Kho].

M.Kofler [Kof11] who disassembled a Microsoft Kinect to research its' components found, that the light pattern is reproduced nine times in a 3x3 square with an overall depth resolution of 640 x 480 @ 30fps

4.2 Moving the camera systems across the room:

Camera systems like the Microsoft Kinect were not intended to be moved in the first place. The intent of the creators was to have a immobile sensor focused on an immobile surrounding containing moving users. Our intent to clarify the possible use of such a camera system as a guiding system for a moving robot is clearly against the intent of the use of such camera

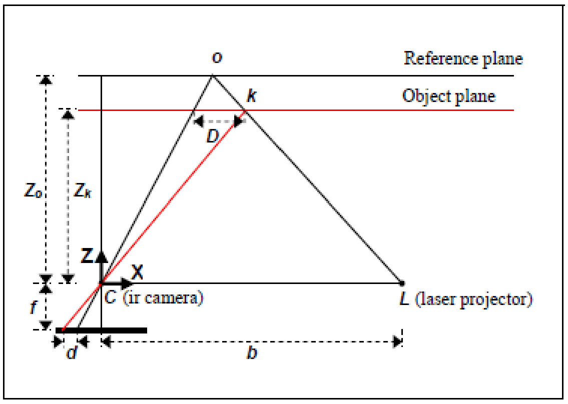

Figure 4.1: Schematic representation of depth-disparity relation [Kho].

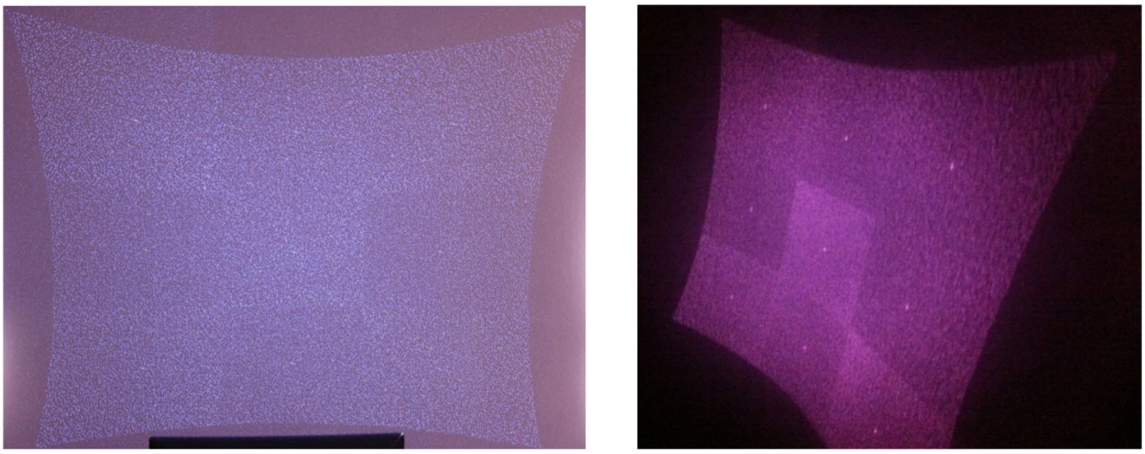

Figure 4.2: Projection of laser grid from a Microsoft Kinect(light coded) [Kof11]. Projections from the Asus Xtion Pro Live look similar

system. Thus more or less rapid movement of the camera system results in loss of users.
The Microsoft Kinect with its' 30 frames per second has the most problems. Its' tracking is not continuous and users ‘’beam” to their new position, since the distance of the surveilled person regarding the prior and actual frame is larger than the fault tolerance threshold of the tracker algorithm. This problem especially occurs when the robot turns or accelerates rapidly, but also on uneven surfaces as the sensor is shaken by the impulse. And even though the Microsoft Kinect is fitted with a motor the regulate the tilt,this motor is not able to compensate the shakes due to its slow reaction time.
The Asus Xtion Pro Live gives us better results on uneven terrain. As long as the guide stayed within the boundaries of the image set by top, left and right edge the Asus Xtion Pro Live beheld its guide. This was tested up to a point where the speed of the rotation forces the images to suffer under motion blurring, where the guide was lost.
In most cases this speed is however never reached in a guided mode, since the speed for motion blurring lies over the required 3 meters per second (10 feet per second) achieved by the guide

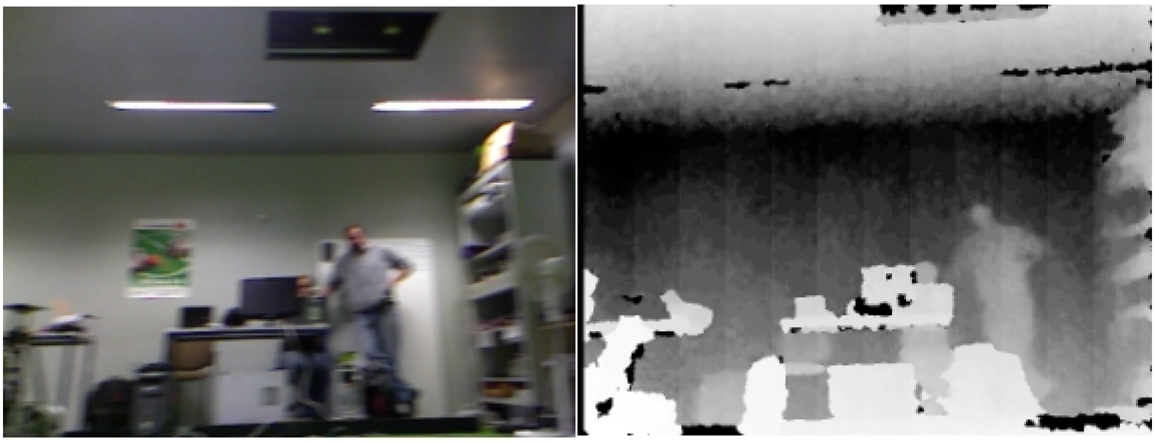

Figure 4.3: Images are result of rotation of the robot with Microsoft Kinect. Even this minor motion blurring suffices to loss guide recognition

moving within the focus boundaries [3.2.2]. The cases in which the guide was lost where either when the guide was faster then the required 3 meters per second or with a 1:1 chance, if the guide was closer then 1 meter (3,3 feet) to the camera (prerequiring that the guide was within the TLR boundaries).
This test could however only be achieved manually, since the Asus Xtion Pro Live‘s RGB stream was unusable for automatic guide following.

Remaining to be said is that during all our tests we have come to the conclusion that an initialization pose seldom has an advantage. It only saves against “hijacking” in low populated regions or must be so complicated to ensure that it can not be achieved by chance, be it alone or in combination of obstacles misinterpreted due to near proximity to the “hijacker”. Tests with a color marker defining a guide have proven to be much more “hijack”-resistant and faster in initializing and finding the guide again after losing him.
During testing we did however discover a discrepancy arising if the robot had a object with a similar marker on it in its field of view. As long as no movement occurred close to that object it was deemed as obstacle and circumvented. If however a person passed the object in close proximity the object could become a guide and the robot would drive up to it. To regain the control the guide had to reinitialize himself as true guide.

4.3 Use in direct and indirect sunlight:

Since the service robot should be able to work in different environments varying ambient light is a natural occurrence. Thus it is expected from the service robot to focus its guide in night scenery as well as in bright sunlight and shall not be distracted by flashing lights.A wide variety of environments and outdoor use must be taken into consideration.

Due to the conception of the Microsoft Kinect being toys and intended for indoor use only exposal of the IR-camera to direct sunlight was not considered originally. Since the Asus Xtion Pro Live is based on the concept of the Microsoft Kinect, the Asus Xtion Pro Live has the same intended field of use with not reconsidering a possible use outdoors.
This results in the fact, that even though the laser grid is emitted with a wave length of 830 nm, the IR camera is only fitted with a spectrum filter, allowing all wave lengths over 780 nm to pass. Which completely suffices indoors where the IR radiation is rather weak. However outdoors light provided by the sun has a far intense force, even in a completely shadowed region. This difference in intensity is the crucial point. The weak intensity indoors does not interfere with the measurable grid, since the laser has a higher intensity. The laser intensity compared with the intensity of the suns radiation is however insignificant [4.1]. Testruns outdoors during daytime result in interference from the intense sunlight overlaying the laser. This interference prevents the IR-camera from interpreting the light pattern correctly, resulting in no depth image at all.

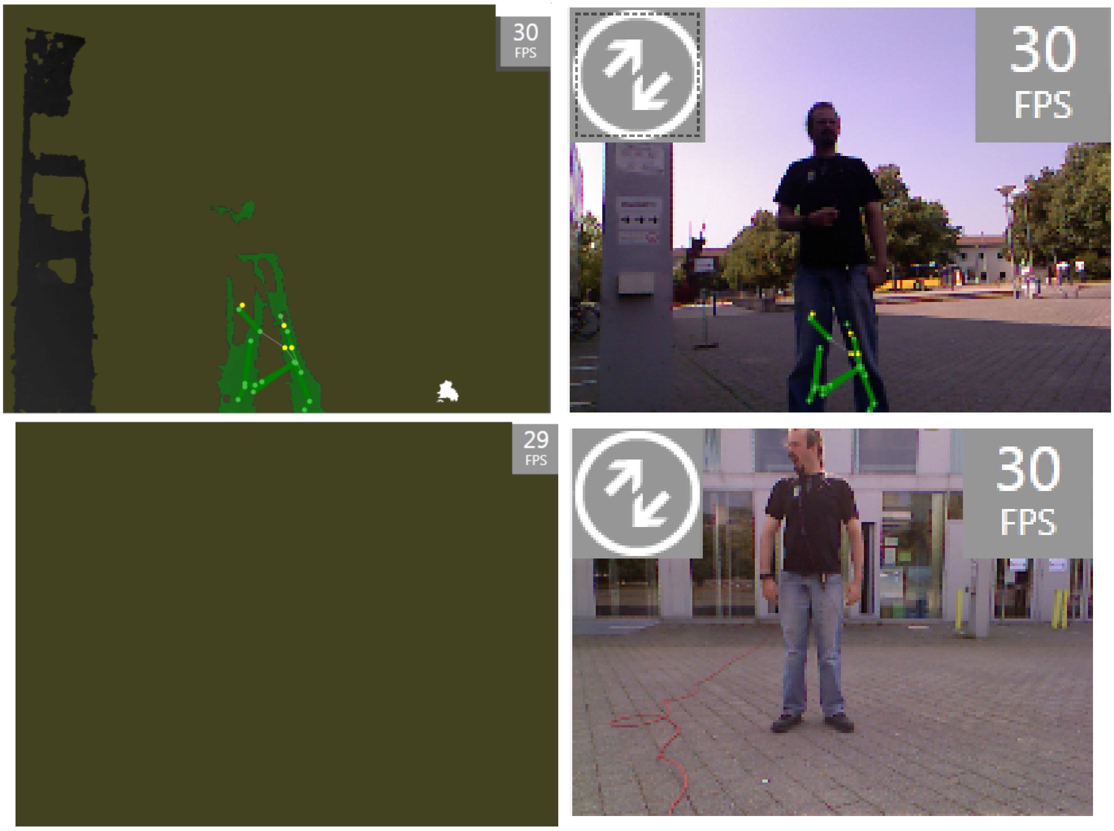

Figure 4.4: Effects of unfiltered sunlight on the perception of the camera systems

With this disadvantage clearly at hand the camera systems in their original set up disqualify for the given task.
However except for sunlight the camera systems performed quite well under varying light conditions hardly relying on it's RGB-camera for user detection. In fact the depth sensor is

not distracted by any source of neither artificial nor indirect sunlight. Only direct sunlight makes problems, even within buildings.

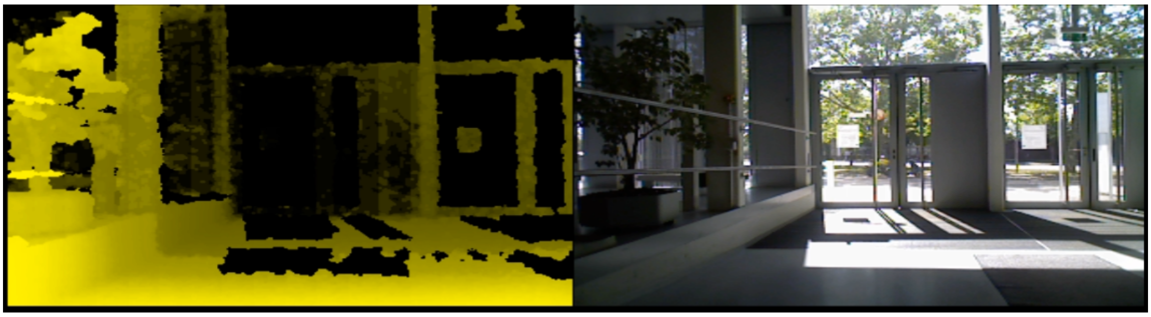

Figure 4.5: Direct sunlight can even effect the camera systems through double layered glass windows. Visible through the black surfaces in the depth image where the sun light hits the floor on the RGB image.

Other effect triggered by differentiating light concern the RGB images. If the light is highly differentiated in intensity it can produce mach bands depending on the angle of incidence. These mach bands falsify either the color values thereby creating false negatives and positives due to its induced alpha value. Or the depth values in darker regions, throwing of the triangulation value by more then the known 4cm possible offset.

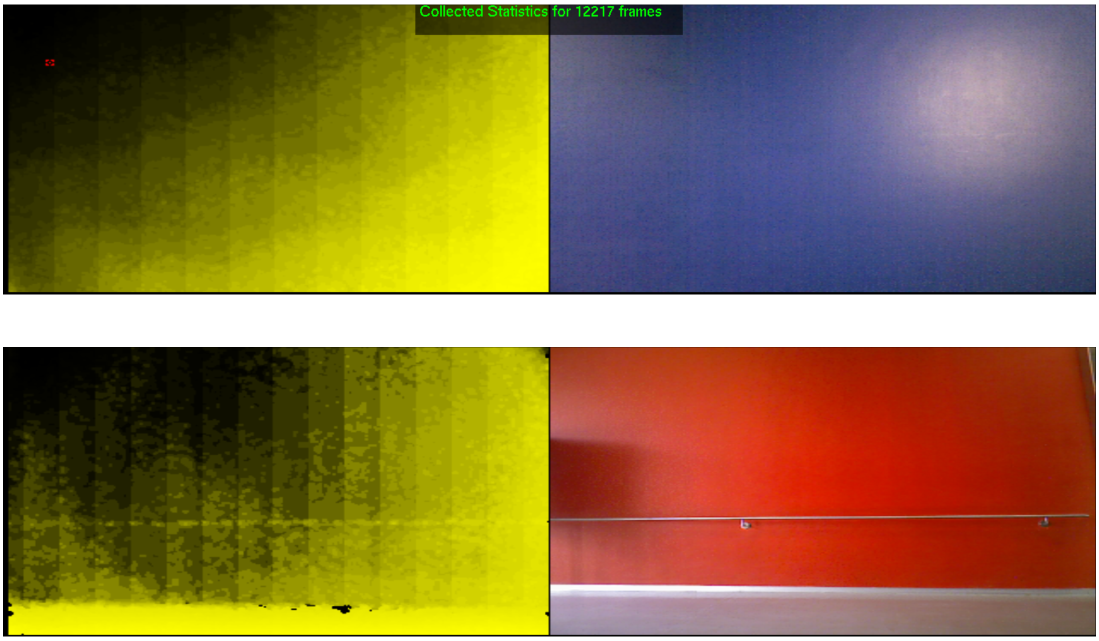

Figure 4.6: Mach band effecting depth and RGB perception of a #0000FF blue and a #FF0000 red. Due to image inversion the light source on one side of the camera appears on the other side in the image.

This is of concern since the driver used for this paper combined the user detection with colour recognition to follow a guide. Being conditioned to follow red items in combination with a user skeleton the different ambient light proved to be a problem, as also pink, orange and certain shades of grey occured in a redish colour when the backlight was to intense. Luckily the angle of incidence between the two different lighting intensities must lie at 180° to each other with an 90° angle to the cameras orientation. If no other lightsource is around then the mach bands take effect.

Tests during the night or in lowlight environments showed that the system can operate inside and outside without problems concerning the depth measurement. To achieve the automated guiding the red marker must be either of a neon like colortone that compensates for minimal lighting or highlighted by a light, so that the RGB camera can pick up the marker.

4.4 Problems with obstacles

The camera systems are clearly able to differ between the guide and other persons or obstacles in its field of view. Therefore it is able to circumvent these obstacles unless they are made of glass or in certain cases have mirroring properties.
Glass is not recognized by the depth measurement, since it laser penetrates the glass rather than being refracted and thereby throwing of the depth measurements. Hence complete glass walls or windows with a low frame are not detected and might be subject to collision, if the guide just steps around them, relying on the robot to follow on the fastest path.
Unexpected side effects occurred during the scenery tests when either camera system was facing a mirror or other reflecting surface. It was expected from the specification of the drivers, that a certain depth difference was of need to detect a guide. But not only was the guide detected but also the mirror image of the guide was acknowledged, resulting in false distance measurements and possible collision of the robot with the reflecting surface trying to approach the guide's mirror image, if the guide was not in the robot's direct field of view.

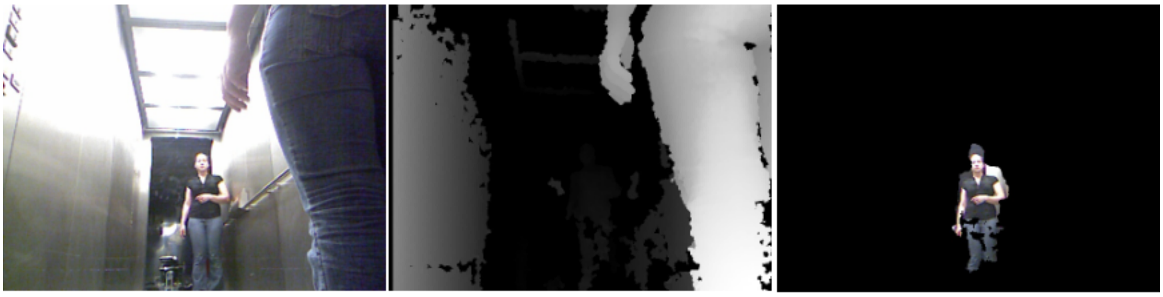

Figure 4.7: Mirror taking effect on guide recognition and "hijacking" the robot.

As we can see by this, it is obviously not needed to provide depth differences to receive a guide acknowledgment, posing the problem of unwanted "hijacking" in a room equipped with mirrors. The only possible solution to this would be to have the guidemarker on only one side

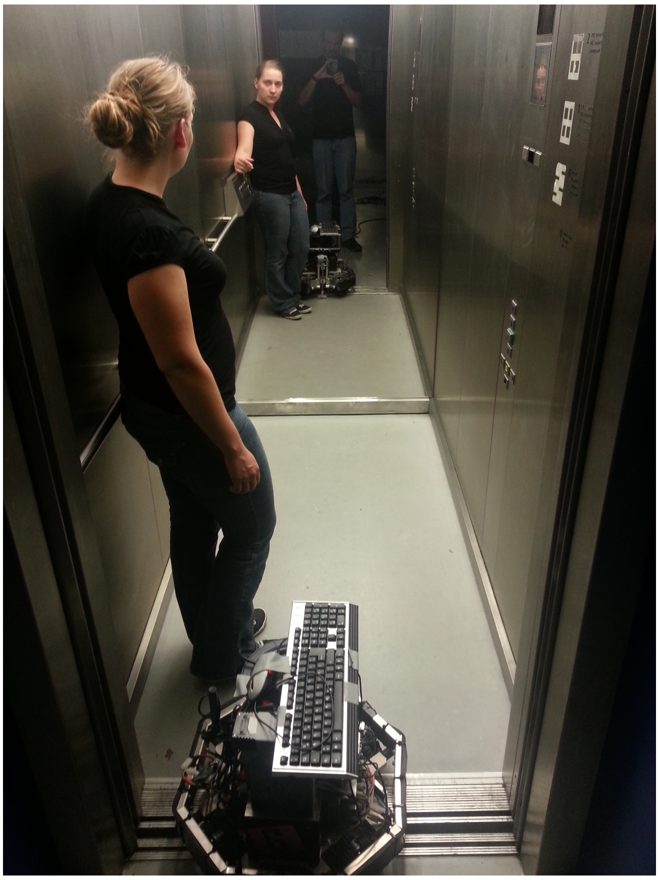

Figure 4.8: Setup of complete situation with a mirror.

of the guide, making it nearly improbable to reflect from a mirror in the field of view of the robot.

4.5 Problems with the setup

As put in [3.1.5] it is possible to use multiple Microsoft Kinects at once. If it is possible with the Asus Xtion Pro Live could not be tested, since only one device was issued for this project. However multiple camera devices aligned to achieve the largest field of view without gaps in the picture result in a reflective feedback in the edge regions of the depth image. This feedback diminishes the quality of the depth image in those regions being a result of the IR laser projection of one system reflecting into the others where they interfere with their own grid reflection. When the camera systems align the depth with the RGB image this interference can result in not recognizing a guide standing in an edge region of any single camera field of view [SSR+11].

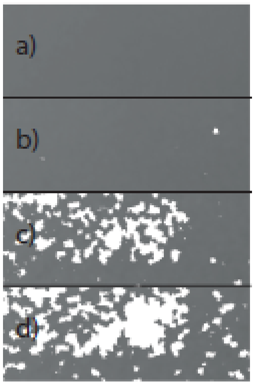

Figure 4.9: The image quality degradation with different numbers of Kinects facing the same surface.
a) 1 Kinect, b) 2 Kinects, c) 3 Kinects, d) 4 Kinects. Each stripe is from the same area of the image of one of the Kinects [SSR+11].

5 Conclusion

Having been confronted with the problematic of possible usage of a Microsoft Kinect-like technology as sensor for tracking and following of a guide in robotics different problem have clearly arisen.
Some have been solved.
Such as the guide detection based on color matching combined with a predefined colored marker and the following of this marked guide. This has proven to be the most stable and fastest way to ensure correct guide detection and following.
Or the obstacle detection and avoidance which work with only two flaws that the setup with camera systems alone could not manage to solve.
"Hijacking" through mirroring is preventable by wearing the marker only on one bodyside disabling a mirror infront to reflect the marker and becoming a guide.

Some have been solved only theoretically.
Like the problem arising from direct sunlight rendering the depth measurement camera obsolete. The solution in theory is to equip the IR camera with a spectrum filter narrowing down the wave band to around 830 nm to filter out the suns interference.

Some were not able to be solved
Be it the extra needed power supply for the Microsoft Kinect. The obstacles that posed a problem where either made of glass, for which ultrasonic sensor detection is needed to add the camera systems as guiding system. Also obstacles having the defined marker on them and a person passing by in close proximity. In this case the only solution to that is to reinitialize the real guide. We recommend choosing the marker after determining the environmental input so that this "hijacking" can be minimized.

Our data clearly stats that there is the possibility of using a camera system like the Microsoft Kinect or the Asus Xtion Pro Live as a visual guiding system in robotics. However so far we can only see the systems use indoors or at night.

In this case our recommendation lies with the Asus Xtion Pro Live. The technical data clearly shows that the Asus Xtion Pro Live makes the run and has certain advantages over the Microsoft Kinect.
Better resolution with higher framerates make it the better choice to be in an constantly changing environment. And its lack of need for an extra powersource than an USB-Port makes it much more usable on a system running solely on batteries.
All the possible testes, due to the unadapted openni node, show clearly that the Asus Xtion

Pro Live has the advantage over the Microsoft Kinect to meet up with the requirements.

Future Work

Based on our research the camera system can be chosen with the needed background knowledge.

To build the requested system for the 3D scanning system a few things might be researched to see if an all environment system can be build using the Asus Xtion Pro Live technology. These requests also regard systems considering the use of the Microsoft Kinect.

First off the theory behind the possible use of a spectrum filter for the IR camera should be checked.
Also noticing that sudden increase or decrease in speed of the robot coincide with a tilt of the camera system and result in possible loss of the guide it might be recommendable to set the camera on a steadicam system. Through which the camera rests in a balanced position ignorant to the changing speed.
Furthermore the sensor setup adding in ultrasonic sensors would need final conception and alignment testing to secure against running against glass obstacles or down stairs or ridges.
The complete system should also be build considering the avoidance of obstacles following the concept introduced in the article "Person Following Robot with Vision-based and Sensor Fusion Tracking Algorithm" [TSY$^+$08]. Thereby hazards like driving into a glass object are minimized even without ultrasonic scanners as extra sensor.

Bibliography

[cad11] Cadet Microsoft Kinect. 2011. URL http://www.cadet.at/wp-content/uploads/2011/02/kinect_tech.pdf. (Cited on page 13)

[CIL07] D. Calisi, L. Iocchi, R. Leone. Person Following through Appearance Models and Stereo Vision using a Mobile Robot. 2007. (Cited on page 10)

[cod] (Cited on page 21)

Kinect pattern uncovered. URL http://azttm.wordpress.com/2011/04/03/kinect-pattern-uncovered/.

[Geo11] J. Georg. Kinect als Eingabegerät. 2011. URL http://homepages.thm.de/~hg6458/semsts/Kinect%20als%20Eingabegeraet%20-%20Ausarbeitung_Georg.pdf. (Cited on page 15)

[Kho] K. Khoshelham. Accuracy analysis of kinect depth data. URL http://www.isprs.org/proceedings/XXXVIII/5-W12/Papers/ls2011_submission_40.pdf. (Cited on pages 6, 21 and 22)

[Kof11] M. Kofler. Intebriebnahme und Untersuchung des Kinect Sensors. 2011. URL http://rrt.fh-wels.at/publications/technical_papers/MP1_Kinect_Kofler_2011.pdf. (Cited on pages 6, 13, 21 and 22)

[Nar11] H. Narayanan. Controlling the TV system with kinect. 2011. URL http://code42tiger.blogspot.de/2011/02/controlling-tv-and-set-top-box-with.html. (Cited on page 3)

[ope] URL http://openkinect.org/wiki/Hardware_info. (Cited on page 13)

[pri11] PrimeSense technology. 2011. URL http://www.primesense.com/en/technology. (Cited on page 21)

[SSR+11] Y. Schröder, A. Scholz, K. Ruhl, D. rer. nat. Stefan Guthe, P. D.-I. M. Magnor. Multiple Kinect Studies. 2011. URL http://www.cg.cs.tu-bs.de/media/publications/multikinects_1.pdf. (Cited on pages 6, 27 and 28)

[TSY+08] Takafumi, Sonoura, T. Yoshimi, M. Nishiyama, H. Nakamoto, S. Tokura, N. Matsuhira. *Person Following Robot with Vision-based and Sensor Fusion Tracking Algorithm*. 2008. (Cited on pages 9, 10 and 30)

[TW10] E. G. Theo Watson. World's largest shadowpuppet. 2010. URL http://vimeo.com/16985224. (Cited on page 3)

All links were last followed on October 30, 2012.

Declaration

All the work contained within this thesis, except where otherwise acknowledged, was solely the effort of the author. At no stage was any collaboration entered into with any other party.

(C. O. Anika Holtermueller Joerg Ploedereder)