Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Combining PMD and Stereo camera for Motion Estimation of a Mobile Robot

C. Netramai * **O. Melnychuk** ** **C. Joochim** ** **H. Roth** **

* *Center for Sensor Systems (ZESS). Paul-Bonatz-Strasse 9-11 57076.*
(netramai@ipp.zess.uni-siegen.de)
** *Institute of Automatic Control Engineering*
Department of Electrical Engineering and Computer Science.
Hoelderlinstr. 3 57068.
({oleksandr.melnychuk,chanin.joochim,hubert.roth}@uni-siegen.de)
University of Siegen,Germany

Abstract: This paper introduces an improvement to the 6D motion estimation solution by combining the Photonic Mixer Device (PMD) and Stereo Camera. The main feature of the stereo camera is the higher resolution of the 2D image compared to the one from the PMD camera. Whereas the depth information derived from a PMD is usually far superior to the result from a stereo camera. This work hence proposes a combination scheme for the PMD and stereo camera in order to improve the results of the motion estimation. The combined setup was placed on a mobile robot and carried on the 6D motion estimation task using a provided artificial landmark. The robot was moved around while the combined setup simultaneously captures 3D images of the landmark from each new position. The motion was estimated based on the matching of the captured 3D images between two successive positions. The classical singular value decomposition (SVD) algorithm was used to solve the matching problem. The referent points for the SVD algorithm were extracted from the landmark using a robust corner detection algorithm. The experiments were fulfilled using 1) stereo 2) PMD and 3) combination of stereo and PMD camera. The results of these three arrangements are compared and the outcome of the comparing are presented.

Keywords: Stereo vision; Motion estimation; Mobile robots; Singular value decomposition;Range images; PMD camera.

1. INTRODUCTION

Estimation of 6D motion using visual sensor is a challenging problem. One approach to estimate the motion of the visual sensor through the environment is to refer its movement to the set of reference points found within the environment. By referring to these reference points the complete motion of the sensor along its trajectory can then be estimated. Since the real world is 3D, the visual sensor should therefore be able to perceive the environment in the complete 3D Euclidean notation in order to provide the fullest advantage for motion estimation task.

The stereo camera is one of the most well known depth-enabled camera which exists. It has been used widely in robotics and automation applications since decades, particularly the short baseline stereo camera which is small in size and can be easily integrated to many applications. While the problem of motion estimation using stereo camera is perfectly feasible, the depth measurement precision of the stereo camera is usually limited due to its stereoscopic design and the software computational complexity. High precision depth measurement can be achieved using large baseline stereo system with high resolution imagers and complex algorithms. But this would hinder the use of such systems on many applications where space and computational power are limited. Therefore one always has to find a good compromise in order to bring the most benefits from the stereo camera system.

Recently, a new type of 3D camera, called Photonic Mixer Device (PMD), is becoming more and more attractive due to its state of the art depth measurement using Time of Flight (TOF) principle. The PMD camera has its own modulated light source and thus can work independently from the distracting lights in the environment. The PMD camera is also better than the stereo camera in the way that it can measures depth of the surface where stereo camera might fail to work, e.g. a big surface patch with uniform intensity and color. But since the PMD camera is still in its maturing state, the main drawback of the current PMD camera is the sensor resolution which is still low compared to another type of cameras available. However, this technological drawback will be eliminated once the device has become popular and produced in a larger scale.

Since the PMD and stereo camera do share one important purpose in common, namely their purpose to acquire the depth of the 3D scene, therefore there have been some works which compare and combine the output of both camera systems in several ways to gain advantages over the use of a single camera system. Ghobadi et al. [2006] made a comparison for both camera systems for the classification of moving objects task and describes the characteristics

978-1-1234-7890-2/08/$20.00 © 2008 IFAC 10.3182/20080706-5-KR-1001.3529

of both cameras in detail. Hahne and Alexa [2007] used the depth measurement from PMD camera to aid the depth calculation of a stereo camera system and found improved benefit of such configuration. Kuhnert [2006] attempted to fuse the depth data between both camera systems to yield a better surface reconstruction. However, an intensive study about the use of such combination systems for motion estimation task is still missing.

This work therefore seeks the way to combine the output from PMD and stereo camera for the 6D motion estimation task. A data combination method which compensates the strong and the weak features from each camera is presented. The suggested method is implemented on the real combined camera system. Carefully designed experiments are presented in order to demonstrate and evaluate the improvement of the result over the conventional single camera system.

2. MOBILE ROBOT TOM3D

The implementation and evaluation of the proposed combined PMD and stereo camera system was done on an actual mobile robot. In this case, the mobile robot TOM3D (Tele Operated Machine with 3D PMD-Camera) with differential drive was used. The robot uses data from wheel encoders for the rough pose estimation. The velocity control is realized using a PI-algorithm in a closed loop controls independently for each wheel. These functions are carried by the micro controller C167, which gives the control PWM signals to the motor controllers. The robot is also equipped with an embedded PC for the entire image processing functions.

The prototype of mobile robot Tom3D is illustrated in figure 1. The hardware configuration of robot is divided into three sections. The top section consists of the PMD camera located at the front of the robot, the middle section consists of stereo camera, mini-computer and wireless communication module while the bottom section comprises of micro controller, the power electronics and DC motors.

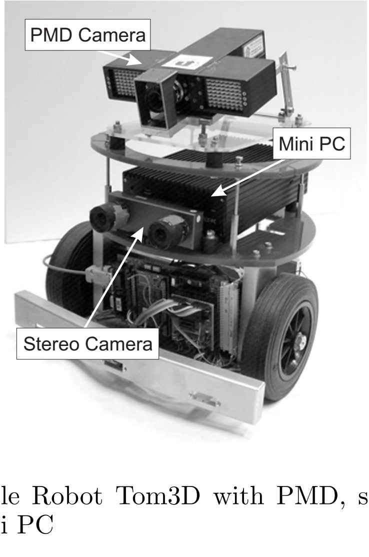

Fig. 1. Mobile Robot Tom3D with PMD, stereo camera and mini PC

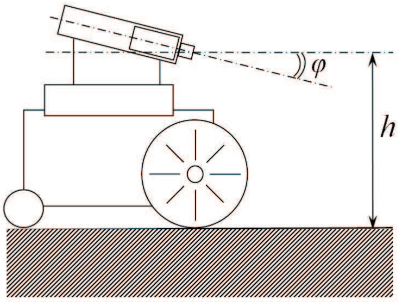

Fig. 2. Schematic of the PMD camera on the mobile robot

3. CAMERA SYSTEMS

3.1 PMD camera

The recently invented PMD camera makes it possible to obtain the depth image of the environment. In general the PMD is the semiconductor structure, based on CCD- or CMOS-technology, Xu et al. [1998]. The basic PMD-System consists of a source of modulated light and PMD pixels, which can be seen as a correlation elements. The evaluation of distances is based on the time of flight principle. Two arrays of the LEDs illuminate the scene with incoherent, modulated infrared light and the reflected optical signal is measured and compared with the reference signal. The depth data results from the phase shift of the outcoming and the incoming signals.

The equation for the autocorrelation is:

$$c(\tau) = \int_0^T x(t)x(t-\tau)dt \qquad (1)$$

Where T is time of integration. To demodulate the signal und find the phase shift, four samples of $c(\tau)$ with time interval of $\pi/2$ are used. $C_i = c(\tau_i)$, $\tau_i = \frac{\pi}{2} \cdot i$, $i = 0, ..3$:

$$\phi = arctan\left(\frac{C_3 - C_1}{C_0 - C2}\right) \qquad (2)$$

the distance can be easily calculated to:

$$d = \frac{c_0 \cdot \phi}{4\pi \cdot f_{mod}} \qquad (3)$$

Where c_0 is the speed of light and f_{mod} the modulation frequency. A common value of the modulation frequency is 20 MHz. Since the maximum phase shift could be 2π, distance up to 7.5 m can be measured. This maximal distance defines an unambiguous depth range for the PMD.

Due to the limitation of the unambiguous range, the PMD camera must be mounted on the robot with an inclination angle (figure 2), so that the maximal way of the modulated optical signal is not longer than 7.5 m.

In order to calculate the inclination angle(φ) and the height of assembly (h) of the PMD camera, a depth image of a ground without any obstacle is taken. The values of the inclination angle and the height of the camera are used for rotation and translation of the depth image, to shape a data in the form, which is suitable to combine this data with one from the stereo camera

3.2 Stereo camera

The stereo camera system used in this work has parallel optical axes and a short baseline of 90 mm. It consists of two CMOS imagers which are able to produce maximum image size at 1280×960 pixels, although during the experiment the images are captured at 320×240 pixels. Both imagers are equipped with 8 mm fixed-focal length C-mount lenses. The imagers are hardware synchronized and are connected to the host computer via IEEE1394 connection. The stereo disparity is computed within a software library installed on a host computer. The disparity computation uses correlation algorithm and post-filtering using a texture filter. This stereo computation is designed to compromised a real time performance and it runs at 30 Hz on a standard personal computer (Konolige [2007]). Once the disparity value is available, the depth value at each pixel is derived from the following equation

$$Z = \frac{fb}{D} \tag{4}$$

where f is the focal length, b is the baseline distance and D is the disparity value.

The calibration of the stereo camera is performed using a planar calibration objects and the rectified gray scale images were used as input for the stereo computation and feature detection during the entire experiment.

3.3 Calibration for the combined PMD and stereo system

The PMD and stereo camera are both fixed on a rigid platform on a mobile robot (figure 1). Both cameras are positioned to look at the same direction, i.e. toward the front direction of the robot. The stereo camera is mounted with its optical axis parallel to the ground while some inclination angle is introduced to the PMD camera. In order to combine the data from PMD and stereo camera, the relative position between both cameras must be found. This relative position can be summarized by the 3D transformation matrix R_{rel} and t_{rel} and they can be used to transform the data from one camera's reference frame to another as follow:

$$P_{stereo} = R_{rel} P_{PMD} + t_{rel} \tag{5}$$

Where P_{stereo} and P_{PMD} are point sets from stereo and PMD camera respectively.

A standard 9×7 (243×189 mm) chessboard pattern was used as a calibration target. To do the calibration, the calibration target was placed in front of both cameras. Images of the calibration target were taken and the reference points within the calibration target were marked manually. Since the dimension and geometry of the calibration target are known, the referent points can therefore be used to determine the transformation matrix R_{rel} and t_{rel} using least square optimization technique.

4. MOTION ESTIMATION

In this work, the motion estimation relies solely on the information received from the camera system. Since the purpose of this work is to demonstrate and evaluate the combined PMD and stereo camera, three different configurations were arranged to feed the motion estimation algorithm with different information. The first and the second configuration make use of single PMD and single stereo camera system respectively while in the third configuration the combined information from both camera systems is given.

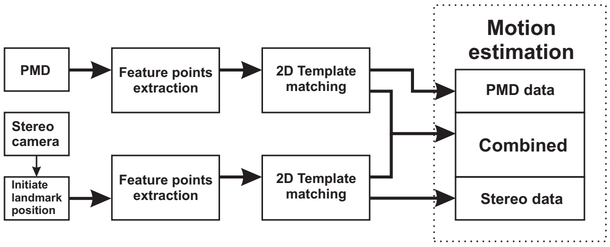

Fig. 3. Motion estimation process

The actual motion estimation process can be divided into several steps: firstly, the present of the landmark is initiated using the 2D image from the stereo camera. This initiation is done in order to roughly guide the heading of the robot toward the landmark. Once the present of the landmark is confirmed ahead of the robot, the corner detection routine is then used to extract strong visual feature points from the landmark. Template matching is then applied on the feature points in order to validate the position and orientation of the landmark in the 2D image. Once the landmark is correctly identified on the 2D images, the 3D information of the corresponding feature points is used for the full 6D motion estimation. These steps were done separately on both camera systems as illustrated in figure 3.

Please note that an artificial landmark was used during the entire experiment in order to control the measurement parameters which are important for the evaluation of the different setups.

The rest of this section provides description of each components in the motion estimation process in more detail.

4.1 Initialization of the landmark position

The artificial landmark is used as a reference for the mobile robot and it is located in the center point of a half circle trajectory of the movement. The high resolution 2D image from the stereo camera is used to roughly initiate the landmark position. Recently, rapid object detection using a Boosted cascade of Simple Features is introduced by Viola [2001] and improved by Lienhart [2002]. They developed a reliable method to detect frontal upright faces, this method can be calculated extremely rapidly and achieving high detection rates in real time. Viola [2001] approach is separated in three steps contributions. The first is image representation called an integral image that allows a very fast feature evaluation. In order to compute these features very rapidly at many scales, the integral image can be computed from an image using a few operations per pixel. The second is a learning algorithm, based on AdaBoost. This algorithm provides an effective learning algorithm and strong bounds on generalization performance by selecting a small number of important features. The third is a method for combing successively more complex classifiers in a cascade structure which increases the speed of the detector by focusing attention on promising regions of the image. Inspired by this approach, the artificial landmark detection is adopted to be a reference point of mobile robot trajectory curve movement.

The landmark is recognized by PMD camera using the amplitude data from the scene. The white regions on the landmark provide a maximal amplitude and the black regions have a very low amplitude value. Due to the low resolution and high noise level of PMD camera it was only possible to detect the corner points from the landmark using both corner and edge detection algorithms. Due to the fact that measurements with low amplitude are improper, the detected black corners of the landmark must be fitted to the plain of the landmark, estimated with high amplitude points.

4.2 2D feature points extraction

The artificial landmark consists of several black and white rectangles. The black rectangles are laid on the rim and the white one is on the center. It gives the evident 8 corner points, thereby corner points are defined as a showning combining experiment results. The Harris corner detection is used to extract certain all 8 corner points from the 2D image because it is robust, upon strong invariance to rotation, scale, illumination variation and image noise,Harris [1988] and Derpanis [2004].

4.3 Template matching

The correct geometrical data of the landmark is used as a template in order to match with the 2D feature points obtained from the corner detection process. The successful matching confirms the present of the required landmark as well as its position and orientation in the 2D image. It also gives us the true alignment of the corresponding 2D points that are needed for the motion estimation process. Once the correct match is found, the complete 3D information of the feature points is included and ready to be used as input to the motion estimation algorithm.

4.4 Motion estimation using SVD

In this work, motion refers to the change of orientation and translation of the robot between two positions. Two different sets of 3D feature points that are successfully extracted from the landmark from two successive robot positions are used as input of the estimation algorithm. Singular value decomposition (SVD) allows the robust and easy implementation of the methods for matching the point sets from the two locations of the robot according to the knowledge of the landmark in order to estimate the motion between the locations. The application of SVD to motion estimation was first published by Arun et al. [1987],Alter et al. [2000] .The two sets of points: Model $Q : q_i$ and Data $P : p_i$ are the point sets from the two locations of the robot were matched according to the knowledge of the landmark. In order to estimate the motion between the locations, we minimize

$$\sum_{i=1}^{n} |p_i - (Rq_i + t)|^2 \tag{6}$$

where p and q are the two corresponding point sets and (R, t) is the optimum transformation. The first step of the computation is to decouple the calculation of the rotation from the translation using the centroid of the points belonging to the matching.

$$c_q = \frac{1}{N} \sum_{i=1}^{N} q_i \tag{7}$$

$$c_p = \frac{1}{N} \sum_{i=1}^{N} p_i \tag{8}$$

$$Q' : q'_i = q_i - c_q \tag{9}$$

$$P' : p'_i = p_i - c_p \tag{10}$$

The registration calculates the optimal rotation by

$$R = VU^T \tag{11}$$

Hereby, the matrices V and U are derived by the singular value decomposition of a correlation matrix

$$H = U\Lambda V^T \tag{12}$$

This matrix is given by

$$H = \sum_{i=1}^{N} p'_i q'^T_i = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \tag{13}$$

where

$$S_{xx} = \sum_{i=1}^{N} q'_{ix} p'_{ix} \tag{14}$$

$$S_{xy} = \sum_{i=1}^{N} q'_{ix} p'_{iy} \tag{15}$$

With known rotation we can easily calculate a translation:

$$t = c_q - Rc_p \tag{16}$$

4.5 Single camera and combined system

In the single camera arrangement only the data from one camera is used for the SVD algorithm:

$$(R_{Stereo}, t_{Stereo}) = SVD\,(P_{Stereo}, Q_{Stereo}) \tag{17}$$

$$(R_{PMD}, t_{PMD}) = SVD\,(P_{PMD}, Q_{PMD}) \tag{18}$$

For the combined PMD and stereo camera system the following assumptions were made:

- The 2D spatial resolution of the stereo camera is higher than the PMD
- The depth measurement of the PMD is better than the stereo camera

In order to follow these assumptions, it is wise to combine the precise depth measurement from the PMD camera and the accurate detection of the landmark from a stereo camera. The data is added together in the following way: the 3D data set from the stereo camera is extended with distance values of the PMD camera and finally the SVD matches two extended sets of 3D points.

$$(R_{Combined}, t_{Combined}) = SVD\left(P_{Combined}, Q_{combined}\right) \tag{19}$$

where

$$P_{Combined} = \{P_{x_{Stereo}}; P_{y_{Stereo}}; P_{z_{Stereo}} \cup P_{z_{PMD}}\} \tag{20}$$
$$Q_{Combined} = \{Q_{x_{Stereo}}; Q_{y_{Stereo}}; Q_{z_{Stereo}} \cup Q_{z_{PMD}}\} \tag{21}$$

5. EXPERIMENT DESCRIPTION

During the experiment, the robot is moved to different positions within the allowed space while both camera systems are allowed to acquire images of the landmark at each new positions. The experiment simulates a movement of a robot along a half circle around the landmark. The relationship between each rotation angle on translation axes of the robot along this movement is illustrated by figure 5. This movement provides a completely experiment result of 3D rotation and translation. The starting point of the experiment is on point A, thence mobile robot moves to point B, C up to point F respectively. Therefore the relative movement between A and B, A and C up to F has five different positions. The motion between two successive positions is calculated by using the landmark as a reference. The hand measurement data that tells the true translation and rotation of the mobile robot was used as a reference value for the evaluate the motion estimation result.

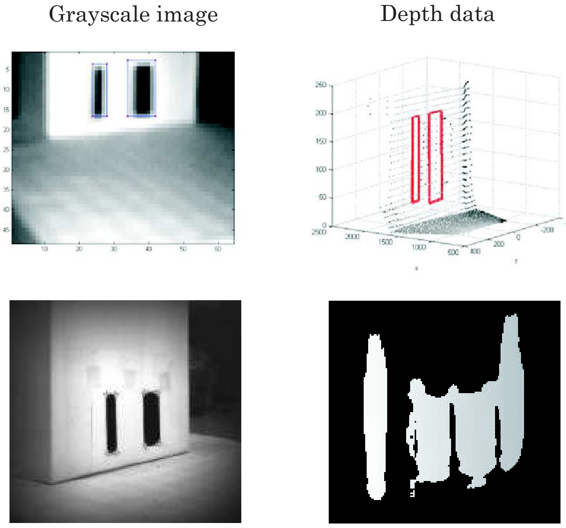

Fig. 4. Gray scale image and depth data from PMD(top), Stereo(bottom) camera

6. RESULTS

The actual and the measured of the robot's movement from stereo, PMD camera and combination are illustrated in figure 5.

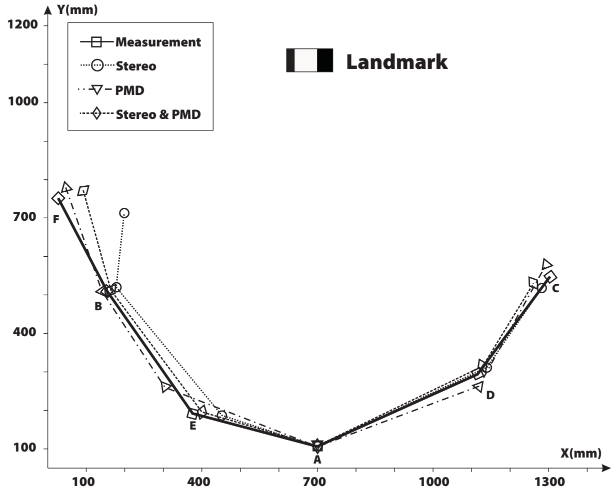

Fig. 5. Result of trajectory curve movement

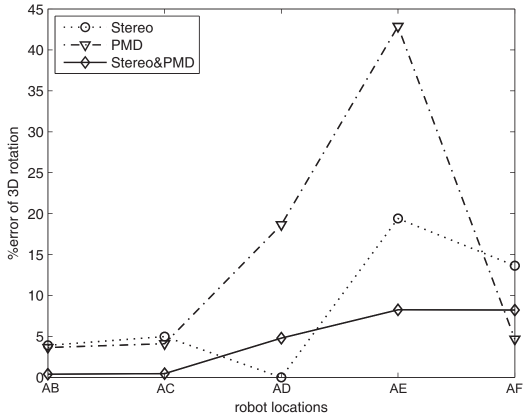

Fig. 6. Percentage error of 3D rotation

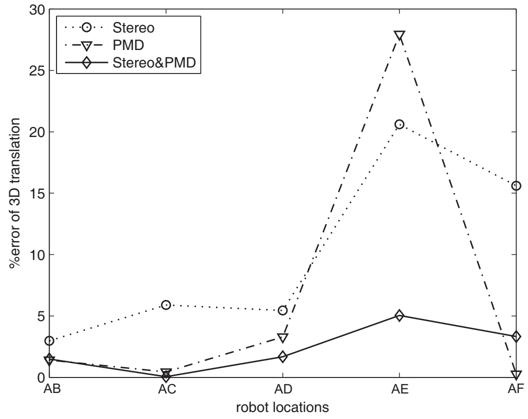

Fig. 7. Percentage error of 3D translation

6.1 Analysis for stereo data

It can be seen from the experiment result (figure 6) that the stereo camera seems to provide a slightly better performance due to its higher optical resolution which helps locating the exact position of the feature points on the landmark. However, (figure 7) shows how the stereo camera has failed to defeat the PMD in term of depth measurement due to its limitation of the stereoscopic design which delivers only modest result compared to the excellent depth measurement using time of flight principle.

6.2 Analysis for PMD data

The low resolution of the PMD camera does not allow the correct detection of the corners of the artificial landmark. It has a consequence that the rotation of the robot was estimated with higher error comparing with results from the stereo camera. However the depth measurement performed a better result for the translation.

6.3 Analysis for combined data

Due to the fact that the combined data has inherited the strong sides from both cameras, mainly the better depth measurement from the PMD over stereo depth data, the resulting error from computing 3D rotation and 3D translation using the combined data sets is less than one from using stereo or PMD camera along. This can be seen from figures 6 and 7.

7. CONCLUSION

This paper compares the results of the motion estimation by using PMD and stereo camera systems. A test environment with an artificial landmark was constructed for the experiments. The gray scale images and the depth data from both cameras were used as input. Three motion estimation results were obtained using only the PMD camera, stereo camera and the combination of both cameras. The results form stereo, PMD and combination stereo&PMD can be investigated that the stereo and PMD camera provide the results with almost comparable accuracy. The PMD camera can though measure 3D points with better depth accuracy than a stereo camera but the low resolution of gray scale image from PMD camera (64×48 pixels) is not excellence enough to precisely locate the corners within the artificial landmark. The best results of motion estimation were obtained from using a combination of both PMD and stereo, where the advantages of both cameras were combined, that is, the precise corner detection from 2D high resolution stereo camera and the accurate depth data from the PMD camera.

ACKNOWLEDGEMENTS

This work was partly supported by the German Research Foundation (DFG), KO-2044/3-1. Furthermore, we would grateful like to thank German Academic Exchange Service (DAAD) for providing financial support in the form of scholarship grants.

REFERENCES

S. E. Ghobadi, K. Hartmann, W. Weihs, C. Netramai, O. Loffeld and H. Roth. Detection and Classification of Moving Objects-Stereo or Time-of-Flight Images. In *2006 International Conference on Computational Intelligence and Security (CIS'2006)*, Guangzhou, China, November 3-6, 2006.

U. Hahne and M. Alexa. Combining Time-of-Flight Depth and Stereo Images without Accurate Extrinsic Calibration. *Dynamic 3D Imaging (Workshop in conjugation with DAGM'07)*, Heidelberg, Germany, 2007.

K. Kuhnert and M. Stommel. Fusion of Stereo-Camera and PMD-Camera Data for Real-Time Suited Precise 3D Environment Reconstruction. *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference*, pages 4780–4785, 2006.

Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum and T. Ringbeck. Smart pixels -Photonic Mixer Device(PMD). In *Mechatronics and Machine Vision in Practice*, pages 259–264, 1998.

K. Konolige. Stereo Vision System. www.videredesign.com

P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *Conference on Computer Vision and Pattern Recognition*, 2001.

R. Lienhart and J. Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. *IEEE ICIP 2002, Vol.1*, pages 900–903, September 2002.

C. Harris and M.J. Stephens. A Combined Corner and Edge Detector. In *Alvey Vision Conference*, pages 147–152, 1988.

K.G. Derpanis. The Harris Corner Detector. October 27, 2004.

K. S. Arun, T. S. Huang and S. D. Blostein. Least Square Fitting of two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(5)*, pages 698–700, 1987.

O. Alter, P.O. Brown, D. Botstein. Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling. *Proceedings of the National Academy of Sciences USA*, pages 10101–10106, 2000.