PMD- Sensoren als Schlüsselkomponenten für die mehrdimensionale Umfelderfassung des Fahrzeuginnenraums

Dr. Thorsten Ringbeck, Paul Frydlewicz, PMDTechnologies GmbH
Oliver Jesorsky, Thomas Sünkel, Elektrobit Automotive Software

Am Eichenhang 50, 57076 Siegen / Germany
www.PMDTec.com / info@PMDTec.com

Frauenweiherstr. 14, 91058 Erlangen / Germany
www.elektrobit.com / info@elektrobit.com

1 Abstract

In recent years, more and more sensor based applications evolve in the field of driver assistance and safety. Besides the activities in environmental sensing, also the passenger compartment gets into the focus of current applications. In this paper the attention is driven to the applications Occupant Classification (OCS), Out-of-Position Sensing (OOP), Driver Drowsiness, Driver Attention Monitoring, and Gesture Control. During the past years more and more attention is spent to an innovative Photonic Mixer Device (PMD) 3D camera technology which has the potential of overcoming many drawbacks of the recent approaches. Basically the entire scene is illuminated with modulated light and the illuminated scene is observed with an intelligent PMD pixel array, where each pixel can individually measure the distance between the sensor and an object point. For those applications where the actually available sensor resolution of 160x120 pixels is not sufficient, the PMD sensor can be combined with a high resolution 2D sensor. The usage of the PMD sensor for the above mentioned applications is addressed and a PMD based gesture recognition system is presented in detail. Complex hand gestures are classified with statistical models and translation invariant features yield very good classification results.

2 Introduction

Currently there is a strong interest in retrieving information about the vehicle's environment. Available sensor based driver assistance applications are for example Adaptive Cruise Control (ACC), Lane Departure Warning (LDW) and Night Vision systems. In a lot of research and pre-series developments, camera, radar and lidar sensors are used to detect and classify e.g. pedestrians and other vehicles.

Besides the activities in environmental sensing, also the passenger compartment gets into the focus of current applications. Sensing the interior enables a system to consider information about the passenger's current state and behaviour. This information allows for example to warn in time in dangerous situations, e.g. if the driver's gaze is not focussed on the road for a longer period of time. In the case of a crash, optimized restraint systems can take properties like the passenger's size and weight into account.

The spectrum of applications is broad. In this paper the attention is driven to the applications Occupant Classification (OCS), Out-of-Position Sensing (OOP), Driver Drowsiness, Driver Attention Monitoring and Gesture Control.

Activities in the field of occupant classification and out of position sensing are mainly driven by United States Regularities. According to the Federal Motor Vehicles Safety Standards and Regulations No. 208 (FMVSS-208) certain injury criteria have to be met for different occupant classes including baby seats. In order to achieve this goal the three airbag deployment strategies ‘suppression in case of presence’, ‘low risk deployment’ and ‘suppression in case of out of position’ are proposed. An out of position situation for example occurs if a passenger leans forward such that the space between the airbag and the passenger becomes less than a predefined threshold (see fig. 1). Applicable strategies depend on the occupancy class, where the classes ‘1 yr old child’, ‘3yr old child’, ‘6yr old child’ and ‘5 percentile female’ are of special interest.

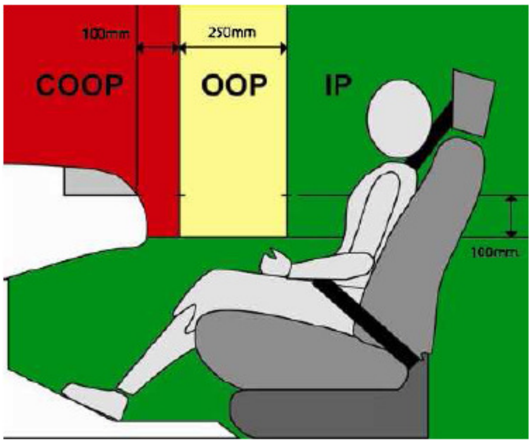

Figure 1: example of out-of-position zones (from [Kro04])

Derived from the stated requirements, an advanced system should be capable of both differentiating between different occupant classes and determining out of position situations. Current systems for example comprise weight sensors and stereo camera systems for the calculation of the desired features.

In contrast to OCS and OOP systems, which are relevant especially in crash situations, driver drowsiness applications are intended to warn the driver in the case of tiredness. According to an analysis of the German Federal Highway Research Institute (BASt), tiredness has a significant causal influence in 6.5% of all accidents. The German Insurance Association (GDV) states that microsleep causes about one quarter of all accidents with fatalities. Current approaches try to derive measures for tiredness from features like head movement and frequency of eye blinking.

But also if the driver doesn’t fall asleep many accidents are caused by distraction of the driver. Driver Attention Monitoring systems try to address this by checking whether the driver’s attention is focused on the driving process. If the driver for example looks at his radio or navigation system for a longer period of time, the system can warn the driver by acoustic or optical signals.

One of the reasons for driver distraction can be the Human Machine Interface (HMI) of the car. The complexity and the number of controllable functions increase with every new car generation. More and more functions like navigation, radio and climate control can be controlled by the driver. Currently haptical input and speech recognition are used as modalities to communicate with the car. In the future gesture recognition might become an additional input channel for modern HMIs. Similar to speech recognition, gesture is a natural form of communication for humans. Section 7 gives a detailed explanation of a realization of a gesture recognition system using a 3D camera system.

3 Sensor technology for in-cabin applications

Pressure sensors are often used for occupant classification and out of position applications. Typically pressure sensing mats are integrated in the seats and the measured pressure distribution is analysed in order to estimate both occupant class and occupant sitting position. Using pressure sensors for occupant classification is quite intuitive as there typically exist a direct correlation between occupant size and weight, especially if you focus on static tests with a fixed set of crash test dummy classes.

Determining the 3D position of the head with respect to the airbag, based on the pressure distribution, is quite more complex. Stereo vision systems seem to be a better solution as their approach is to directly measure 3D information of the occupants. Typically a monocular 2D vision system loses depth information because the 3D object points are projected onto the sensor plane. In a stereo system the depth information can be reconstructed from corresponding feature points in image pairs.

Although many algorithms for Driver Drowsiness and Driver Attention Monitoring rely on 3D information like e.g. gaze, there exist a lot of approaches to solve this task with monocular camera systems. Those systems often make initial assumptions about the observed objects as e.g. head symmetry, head size etc. Those systems work quite well if the initial assumptions are fulfilled but problems can be expected, if not.

During the past years more and more attention is spent to an innovative Photonic Mixer Device 3D camera technology which has the potential of overcoming many drawbacks of the recent approaches. A direct measurement of 3D information based on time of flight calculation can, in addition to the measurement of intensity information, be used in all of the mentioned applications, either as a full replacement of the currently used sensor technology or as an additional sensor channel contributing worth full information.

Their advantage in comparison with monocular camera systems is clear, because the depth data directly delivers an additional information channel. Compared with stereo systems, the PMD approach has some direct benefits because the system directly delivers 3D sensor data without an exhaustive image analysis. In addition to this a PMD measurement is independent from the texture information of the objects in the scene. 3D data can be provided both for homogeneous and fine-textured surfaces.

We give a technical description of the PMD technology in section 4 and present some examples on how this technology can be used for in-cabin applications in section 6. Currently PMD chips are available with a maximum sensor size of approx. 160x120 pixels. If this resolution is not sufficient, sensor fusion with an additional 2D sensor is possible (see section 5 for more details).

Big challenges for all vision based systems, both 2D and 3D systems, occur from the limited space for mounting the cameras within the vehicle. The minimum distance between the passenger and the sensor can be very small, whereas the required field of view can be very large because of the variability of possible seat and sitting positions. In order to cover the complete field of possible head positions including out of position situations, a viewing angle of approx. 120° is needed if the camera is mounted in the car's roof. Additionally a high dynamic range of the sensors is necessary to assure a proper functionality under all illumination conditions.

4 PMD Technology

As the Time-Of-Flight principle has already been described in detail in many technical publications [6,7,8], only the basic principles will be discussed here.

Figure 2 shows the basic Time-Of-Flight principle. In its most simple form, a light pulse is transmitted by a sender unit and the target distance is measured by determining the turn-around time the pulse needs to travel from the sender to the target and back to the receiver. With knowledge of the speed of light the distance can then easily be calculated. However, the receiver needs to measure with picosecond-accuracy the delay between start and stop, if millimetre-precision is required. To realize such system solutions with discrete components, as is done in today's TOF rangefinders, each component in the signal chain must have a very high system bandwidth.

In contrast to Figure 2, Figure 3 shows the basic working principle of a PMD based range imaging camera. Rather than using a single laser beam (which would have to be scanned over the scene to obtain 3D) the entire scene is illuminated with modulated light. The advantage of PMD devices is that we can observe this illuminated scene with an intelligent pixel array, where each pixel can individually measure the turnaround time of the modulated light. Typically this is done by using continuous modulation and measuring the phase delay in each pixel. In addition to this robust method of obtaining 3D without scanning, the realization of the phase measurement in a quasi optical domain offers huge advantages compared to the above mentioned discrete solutions. This is one reason, why we do not require an optical reference channel for most applications.

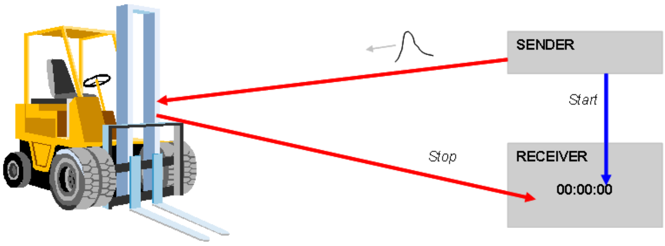

Figure 2: Time-Of-Flight measurement principle with pulsed light

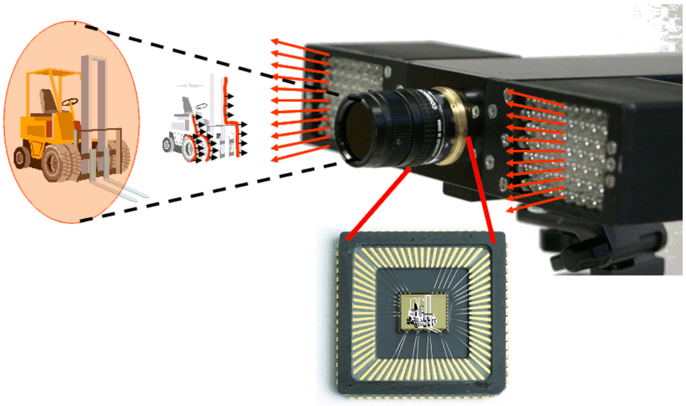

Figure 3: Imaging Time-Of-Flight, measurement based on PMD

Operation principle of PMD

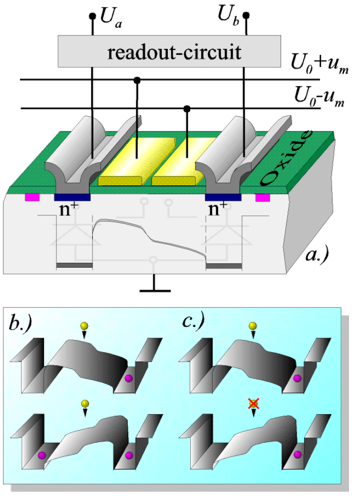

Figure 4:
Operation principle of PMD

This section describes the principle of a simplified PMD sensor realized in CMOS technology. Because the complete mixing process of the electric and optical signal takes place within each pixel we call the PMD elements “smart pixels”. Figure 4 a) shows an illustration of a single pixel PMD sensor element. It is a five-terminal device with an optical input window, i.e. two transparent modulation electrodes in the middle of the illustration. These light sensitive Photogates are isolated from the substrate by a thin oxide layer. The gates are conductive and transparent for the received light. On the left and the right there are readout diodes which are connected to the pixel readout circuitry. In a PMD pixel the movement of generated charge carriers can be controlled by the reference signal applied to the modulation gates. This way one can influence a charge transport to the left or to the right side. The potential distribution in the surface region is influenced by these push-pull voltages leading to a “dynamic seesaw” for the generated charge carriers.

If the incident light is constant and the modulation is a rectangular signal with a duty cycle of 50% the generated charge carriers within a modulation period move to the left and to the right equally. At the end of such a modulation process the output voltages at the readout nodes are the same as those shown in Firgure4 b).

If, however, the incident light is modulated (e.g. also as a rectangular signal), and there is no phase (phase = 0°) delay between modulation of light and detector, then all charge carriers will be moved to one of both readout diodes Figure 4 c). For other phase delays, the modulation of the light and its phase delay compared to the electrical reference signal results in a difference between the two output voltages. This difference of both output nodes is directly dependent on the phase delay between light and pixel modulation. This data can be used to calculate the distance from a light-reflecting object to the sensor [Mö05].

5 Fusion of 3D and 2D data

Currently available PMD-sensors reach a lateral resolution of 160x120 pixels, which is sufficient for many applications because of its third dimension. Nevertheless there are some tasks where a higher lateral resolution is desired due to a high spatial frequency of a scene. For those cases it suggests itself to supplement the distance measurement using PMD-cameras with a high resolution 2D-camera and thus create a 3D-model with a high texture resolution in real time.

Due to the depth measurement of a PMD camera it is possible to solve the correspondence problem between two image planes with only little calculation efforts. It is only necessary to reconstruct the scene based on the information a PMD camera offers, to transform the scene to the coordinate system of the high resolution camera and to project the scene onto its sensor. The resulting coordinates can be used to map the high resolution texture onto the 3D model of the PMD camera. The transformation between the two coordinate systems must be known. The computation of each pixel coordinate is independent from the others and thus can be parallelized very easily.

The amount of information one gets from the fusion of a PMD and a 2D camera is comparable to stereo vision systems, but with a much lower computational effort. Moreover, in contrast to stereo vision both cameras can be mounted close together, it is even possible to use one single lens, which is an advantage for the design of the interior.

6 Usage of PMD for in-cabin applications

In this section we present some sample scenes captured with a PMD[vision] 19k sensor having a resolution of 160x120 pixels. Figure 5 compares the results obtained from a rearward facing baby seat and a person. A clear difference, especially regarding the distance information, can be observed. The 3D structure measured in the distance map is independent from object textures and can therefore be used directly to derivate features for occupant classification.

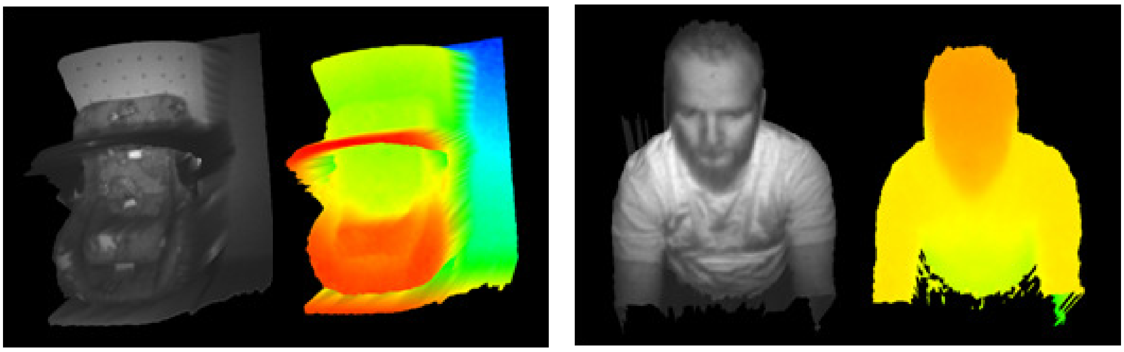

Figure 5: Result amplitude and distance images obtained from a rearward facing baby seat (left image pair) and a person (right image pair). Amplitude information is shown in the left half and distance information is shown in the right half of each image pair.

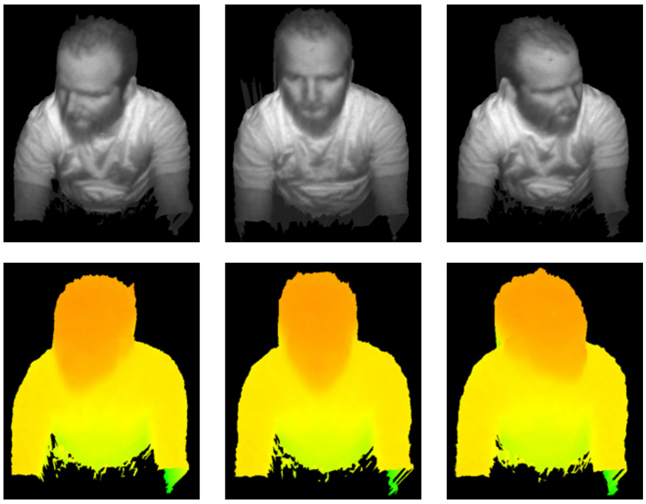

Figure 6: Result amplitude and distance images obtained from a person facing its head in different directions. Amplitude information is shown on the top and distance information is shown on the bottom.

Figure 6 shows amplitude and distance data of a person facing its head in different directions. The images clearly show the resulting differences within the distance data. In addition to the head position, which can be derived from the 3D data, it should also be possible to estimate the head orientation using the PMD data.

For the recordings the sensor has been placed on top of the objects. For out of position, driver drowsiness and driver attention applications it would also be possible to integrate the sensor system within the dashboard.

Section 7 presents a more detailed description of a novel gesture recognition application. More details on the application of time-of-flight technology for occupant classification and out of position sensing can be found e.g. in [Dev 05].

7 Realization of a Gesture Recognition application using a time-of-flight sensor

Gesture in the context of the described system means *hand movements*. Only the *trajectory* of the hand is important, and not its shape. This movements can be quite complex. An example for such a kind of gesture is “writing” the digit “8” into the 3D space (see figure 7). The system is active only, if the current hand position is located inside of a specific volume of interest called *action volume* (see figure 8) .

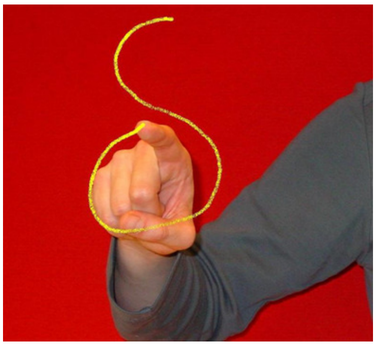

Figure 7: Example gesture sample "8".

The overall system consists of two sub-systems as shown in figure 8. In the *Server System* the PMD camera is employed to track the motion of the actor’s hand in real time. The result is a sequence of 3D positions of the hand representing the performed gesture sample. This sequence is used subsequently as input for the recognition system - the *Client System* – which assigns it to the most probable of a set of trained gesture classes C_i, or rejects it, if no appropriately probable gesture class could be found.

Server System

The *Server System applies* two steps for each 3D frame. The first module preprocesses the raw data of the camera and applies e.g. smoothing algorithms to the data in order to adjust it for further processing.

The *Seqmentation* module determines the 3D barycentre of the hand. This task is solved by segmenting 3D points belonging to the user's hand within a dedicated volume of interest which can be specified freely within the cameras field of view. A *Kalman Filter* [Wel95] is used to predict the hand position for the next frame in order to stabilize and speed-up the hand tracking-performance by reducing the search area. For more details the reader is referred to [Liu06].

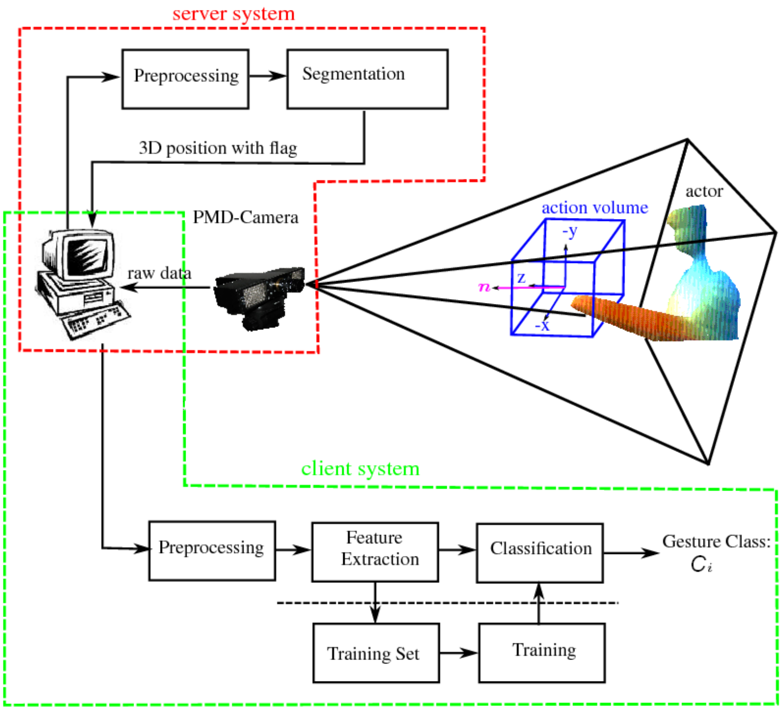

Figure 8: The gesture recognition system consisting of a client and a server system

Client System

The *Client System* recognizes gestures by classifying motion patterns described by 3D trajectories calculated by the Server System with *Hidden Markov Models (HMMs)*. These models are stochastic models used in many pattern recognition systems where time-consecutive features are processed, e.g. in speech recognition systems [Sch95]. *HMMs* have already been used in other gesture recognition systems [Mor98]. Our approach differs from the existing methods in that we calculate our features based on real 3D hand positions which are the results of our *Server System*.

At first a vocabulary of *N* gesture classes $\boldsymbol{C} = C_1, ..., C_N$ is defined. A set of *T* training samples of each gesture class is recorded by different persons. That means *T* trajectories $\boldsymbol{P}_t^{(o)}$ with *t=1, ..., T representing* each gesture class is captured by the *Server System.*

Each of these trajectories is modified in several steps in the *Client's Preprocessing* module, that means transforming the original sequence of 3D positions $\boldsymbol{P}^{(o)} = \boldsymbol{p}_1^{(o)}, \ldots, \boldsymbol{p}_n^{(o)}$ to a feature sequence $\boldsymbol{P}^{(f)} = \boldsymbol{p}_1^{(f)}, \ldots, \boldsymbol{p}_q^{(f)}$.

Different feature types are possible but *Normalized Cartesian Velocities*, which are simply the 3D movement directions of the hand between consecutive frames, deliver the best recognition rates (Section 6.5).

The training samples are used to *train* the *N* models HMM_i, with $i = 1, 2, \ldots, N$. Training means in this case to estimate the parameters of these models. Note that each gesture class C_i is represented by one corresponding HMM_i.

The *N* models are then used to classify previously unknown movement patterns. After running through the same *Preprocessing* and *Feature Extraction* modules as for training these patterns are represented by $\boldsymbol{P}^{(f)}$.

Using our *HMM* approach the probabilities for a set of *N* gesture classes C_I with $i = 1, \ldots, N$ are given by $P_1(\boldsymbol{P}^{(f)} | C_1), \ldots, P_N(\boldsymbol{P}^{(f)} | C_N)$. The values of the probabilities are calculated using the *Viterbi Algorithm* [For73].

These conditional probabilities give an indication of the similarity between the observed movement pattern represented by the feature sequence $\boldsymbol{P}^{(f)}$ and the training sequences of each model HMM_i. The system selects model – i.e. gesture class $C_{classified}$ – which has the biggest similarity to the tested sequence $\boldsymbol{P}^{(f)}$:

$C_{classified} = (argmax_i \{P(\boldsymbol{P}^{(f)} | C_i)\}, i = 1, ..., N.$

The usage of the movement directions for the features leads to the good characteristic that our system is independent of the translation of the performed gestures. The application of the HMM approach makes the system also independent of the speed and the size of the performed gestures. The HMMs can be trained well with much data without losing recognition performance. Using the PMD technology it is also independent of the illumination conditions [Xu98]. In addition to that it does not matter if the left or the right hand is used for the movement. Hence it is quite flexible and robust.

Possible applications in the vehicle are the control of the radio, the CD player or the navigation system.

Experiments and Results

A vocabulary of *N=24* gesture classes was defined, consisting of the ten different digits and 14 other movement patterns, which use all three dimensions of space (the digits actually use only two of them).

A set of *T=75* training samples of each gesture was recorded by 15 different persons (five samples of each gesture by each person) using the a 64x16 time of flight camera (PMD[vision] *1k-S)* running at approximately 20 frames per second.

For testing another 25 samples of each of the 24 classes were captured by another five persons. Using the Normalized Cartesian Velocities features 94 percent of the test sequences were recognized correctly, 3 percent have been rejected and 3 percent have been wrong classified.

8 CONCLUSION

Time-Of-Flight systems based on the PMD-principle give the possibility of fast 3D measurement with customizable resolutions depending on the application. The first sensor systems have already been launched to the mass market - for industrial applications. If the resolution is not sufficient, sensor fusion with an additional 2D sensor is possible. On the basis of sample images major advantages of using 3D camera systems for in-cabin applications like occupant classification and out of position sensing have been shown. As a sample application a prototype system for recognizing complex and isolated hand gestures was described.

References:

[Bau67] L.E. Baum and J.A. Eagon, An Inequality with Applications to Statistical Estimation for Functions of Markov Processes and to a Model of Ecology, Bulletin of American Mathematical Society, volume 73, page(s) 360-363, 1967.

[For73] G. Forney, The Viterbi Algorithm, Proceedings of the IEEE, volume 61, no. 3, page(s) 268-277, March 1973.

[Liu06] B. Liu, T. Sünkel, O. Jesorsky, R. Kompe, J. Hornegger: Gesture Recognition Using Novel Efficient and Robust 3D Image Processing, in 10th AMAA conference, Berlin, April 2006.

[Mor98] P. Morguet and M. Lang, Spotting Dynamic Hand Gestures in Video Image Sequences using Hidden Markov Models, International Conference on Image Processing, page(s) 193-197, Chicago, Illinois, October 1998.

[Sch95] E.G. Schukat-Talamazzini, Automatische Spracherkennung, Vieweg Publisher, ISBN 3-528-05492-1, Braunschweig/ Wiesbaden, 1995.

[Wel95] G. Welch and G. Bishop, An Introduction to the Kalman Filter, Technique Report 95-041, University of North Carolina at Chapel Hill, 1995.

[Xu98] Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, T. Ringbeck, Smart Pixel – Photonic Mixer Device (PMD): New System Concept of a 3D-imaging Camera-on-a-Chip, International Conference on Mechatronics and Machine Vision in Practice, pp. 259-264, Nanjing, China, 1998.

[Mö05] T. Möller, H. Kraft, J. Frey, M. Albrecht, R. Lange; Robust 3D Measurement with PMD Sensors; Range Imaging Research Day, Zürich 2005

[Kra04] H. Kraft, J. Frey, T. Moeller, M. Albrecht, M. Grothof, B. Schink, H. Hess, Universität Siegen; B. Buxbaum, PMDTechnologies GmbH Siegen, "3D-Camera of High 3D-Frame Rate, Depth-Resolution and Background Light Elimination Based on Improved PMD (Photonic Mixer Device)-Technologies", OPTO 2004, AMA Fachverband, Nürnberg 2004.

[Rin03] T. Ringbeck, M. Albrecht, J. Frey, M. Grothof, H. Heß, H. Kraft, T. Möller, J. Mosen, B. Schink, "Time-of-Flight 3D-camera for autonomous navigation and industrial automation" Sensor 2003, Nürnberg

[Bux01] B.Buxbaum, R.Schwarte, T.Ringbeck, M.Grothof, X.Luan, "MSM-PMD as correlation receiver in a new 3D-ranging system", SPIE – Remote Sensing, Laser Radar Techniques: "Ranging and Atmospheric Lidar", Toulouse 2001

[Sch00] R.Schwarte, B.Buxbaum, H.Heinol, Z.Xu, J.Schulte, H.Riedel, R.Steiner, M.Scherer, B.Schneider, T.Ringbeck, New Powerful Sensory Tool in Automotive Safety Systems Based on PMD-Technology", AMAA - Advanced Microsystems for Automotive Applications 2000, Berlin 2000

[Bux05] B. Buxbaum, B. Hagebeuker, "Dreidimensionale Umfeld-Erfassung", Elektronik Automotive, Nr. 5, Sep. 2005, page(s) 77

[Kro04] S. J. Krotosky, S. Y. Cheng, M. M. Trivedi: Face Detection and Head Tracking using Stereo and Thermal Infrared Cameras for 'Smart' Airbags: A Comparative Analysis, in The 7th International IEEE Conference on Intelligent Transportation Systems, page(s) 17-22, October 2004.

[Kal05] B. Kallelis: Occupant Sensing from IEE, aei (aei-online.org), page(s) 33-34, Dec 2005.

[Ale04] B. Alefs, M. Clabian, H. Bischof, W. Kropatsch and F. Khairallah: Robust Occupancy Detection from Stereo Images, Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems, page(s) 1-6 Oct. 2004.

[Dev05] P.R.R.Devarakota, B. Mirbach, M. Castillo-Franco, B. Ottersten: Classification of vehicle occupants using 3D image sequences, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume 4, page(s): 18-23, Mar 2005.