# Real Time 3D-perception by TOF-echoing 3D-Video Cameras

Rudolf Schwarte

INV Inst. für Nachrichtenverarbeitung and ZESS Center of Sensor Systems, University of Siegen, 57076 Siegen, Hölderlinstr. 3, Germany, Tel. 0271/7403330,
Email schwarte@ZESS.uni-siegen.de,
in cooperation with PMDTechnologies GmbH, 57076 Siegen, Am Eichenhang 50.

**Abstract.** We are faced with urgent needs of fast 3D-perception of moving sceneries in various fields, like

- Human-Safety, Industrial Automation, 3D-Man Machine Interfacing,
- Autonomous Mobile Robots, e.g. for Service, Assembling, Exploration & Rescue,
- Accident-free Driving, Precrash, Automatic Stop&Go, Traffic Control etc.

Evolution created ingenious concepts and systems for fast 3D-perception and evaluation for the sake of orientation and goal-directed reaction in dynamic 3D-environments. So far our understanding and still more our technical solutions seem to be miles behind the evolutionary state-of-the-art, e.g., of the 3D-sonar TOF (time-of-flight) systems of dolphins or of the 3D-stereo vision systems of advanced mammals and of human beings.

In our technical evolution we are pursuing similar concepts, but neither our two-camera stereo vision systems nor our echo TOF (time-of-flight)-based ultrasonic-, microwave- and laser-radar systems are able to provide comparable performance to the natural standard examples in real-time 3D-environment perception.

Finally an innovation to be described gives us the chance to solve the technical challenge of high speed 3D-capturing dynamic environments. After two decades of laser-radar research in 1996 a key component was found [1], providing an amazing way to reduce the volume of the laser-radar receiver by a factor of more than 1 : 100.000, including a dramatic improvement of all measurement specifications.

The purpose of this paper is to describe this proved multi-channel laser-radar technology which enables thousands of high-sensitive laser-radar receivers to be integrated on a fingernail-sized CMOS-chip for fast 3D-perception. Furthermore, the consequences of this innovation have to be illustrated with respect to the increasing autonomy of technical systems, to the huge number of applications and to the substantial scientific, economic and social impact.

These extraordinary capabilities rely on a smart photodiode-inherent mixing-process. We called this specialized photo diode with two controllable photo-current outputs **PMD (Photonic Mixer Device).** That's why the opto-electronic mixing process of optical and modulation signals is accomplished directly in the photonic state **before** the photo-current is read out, followed by an integration process to get the OE-correlation function which stands for the optical echo delay. As outlined, PMD enables to realize 3D-cameras respectively **3D-video cameras** for high speed 3D-perception in dynamic environments, today using up to 20.000 PMD-pixel.

## 1 Motivation for real-time 3D-machine vision

At present the ability to capture the surrounding area at high speed in three dimensions is one of the most challenging tasks in industrial automation, in production and automotive safety, in autonomous mobile robotics, and in artificial intelligence. Real-time technical 3D-perception would bring a huge number of new opportunities and various applications in various fields.
Human beings are habited to “see” spatially at relatively high speed. They are able to model and to interpret – more or less unconsciously - our three dimensional and naturally moving environments at a glance. Precisely adapted to the world coordinate system by means of our sense of equilibrium, which temporarily is retained by means of our vestibular system, we evaluate the static and dynamic 2D-projection information of light emitting surfaces in a complex and sophistic way, among other things, with respect to 3D-triangulation for 3D-perception, orientation and navigation. These fast 3D-perception sensing and cognition abilities are absolutely essential if we want to move around, for instance, to play basketball, tennis or only to walk.
Same is valid if we want our cooperating technical systems to do so and to help us e.g. performing autonomous transport and handling activities. They also have to provide a minimum of integrated performance of cooperating **sensor intelligence, actuator intelligence** and **goal-controlled computer intelligence.** It’s 3D-percetion which has risen to the most important sensor intelligence system in our evolution, seen for example at advanced mammals and at human beings. Obviously this development to be observed in the biological evolution can also be observed in the young development of our enhanced technical **Intelligent Sensor-Actuator-Controlled-Computer (S-A-CC)-Systems**. Only under these sensing preconditions e.g. our robots are able to find their way autonomously, and to do their job of servicing.
But what is the state-of-the-art of fast technical 3D-vision in this context?
Fact is that so far we have no satisfying solution for these tasks and no equivalent technical 3D-perception system being able to capture (first inevitable step! [2]), to recognize objects, to orientate, and to navigate within the entire dynamic environment three-dimensionally as fast as our biological visual system operates – neither mentioning the dynamic 3D-scene simulation and prediction abilities nor additional multi-sensor fusion and evaluation of colors, textures, sound, inertial-system information, gravity, and sensing actuator intelligence.
The state-of-the-art presently is described on one hand by **3D-stereo vision** systems. They suffer from cost and time consuming burden in dynamic and natural environments by a lack of reliable and sufficient corresponding pixel groups. On the other hand there are **rotating laser-radar scanners** for plane-scanning or for volume-scanning. Autonomous vehicles in unstructured scenes at least require 3D-laser scanners. They are well suited for motionless 3D-scenes. But due to arbitrary motion over ground during scanning they are not able to model the 3D-surrounding and to fulfill the SLAM-conditions for “Simultaneous Localization And Mapping”. So far they are condemned to stop&go-operation.
In the following we propose a multi-thousand channel laser-radar concept towards human-like 3D-vision capabilities, which enables to fulfill the inevitable first step: **High speed capturing** of the **environmental 3D-data.**

## 2. PMD-principle of operation and 3D-video-camera application

For high speed capturing of the environmental 3D-data we have to measure as fast and as many 3D-surface coordinates as possible. Now exactly this can be performed via revolutionary improvements of the laser-radar receiver. The PMD-principle provides an unbelievable simplification, size-reduction and improvement by a smart inherently-mixing photodiode using two photo-gates for controlling the photo-current read-out electrodes. **Fig. 2b** below depicts the equivalent circuit of this specialized double read-out photodiode. It enables to reduce the receiver size by a factor of more than **1:100.000**, involving a dramatic improvement of all measurement specifications and the opportunity to realize 100.000-channel laser-radars for fast parallel 3D-perception on fingernail-sized CMOS-chips. We call it **PMD (Photonic Mixer Device)** because the opto-electronic mixing process is accomplished directly in the photonic state, followed by an integration process to get the OE-correlation which stands for the delay of the optical echo against the modulation signal.

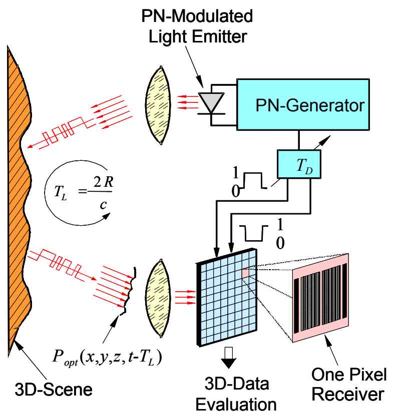



Fig. 1. In this PMD-camera each PMD-receiver operates as a 1D-laser radar but here in a PMD-matrix in parallel and at the same time.

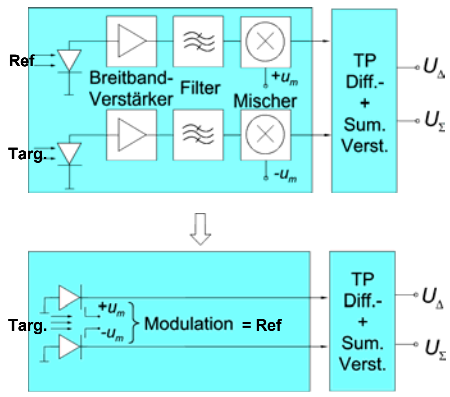



Fig. 2. Comparison of a) the conventional and b) the new PMD-receiver concept of LED/Laser-Ranging by echo TOF (shown without optical sender)

**Fig. 1** schematically illustrates a multi-channel laser-radar respectively a 3D-PMD camera with a PMD-matrix and with one PMD-pixel out of the matrix being enlarged and representing one independent laser radar receiver. In **Fig. 1** the whole 3D-scene is illuminated by means of a modulated optical sender, e.g., a laser diode or a LED, and mapped onto the PMD-array in the focal plane of the receiving optics for high-parallel demodulation of the echo signals, for echo time-of-flight TOF and distance evaluation. Since so many receivers operate independently in parallel, they deliver high speed 3D-snapshots in the ms-range and high 3D-frame rates, thus representing a novel **3D-video camera** without scanner.

The two block diagrams in **Fig. 2** illustrate the quantum gap of simplification and progress of the PMD-laser radar receiver in **Fig. 2b** – replaced by the PMD-circuit symbol - against the conventional laser-radar receiver in **Fig. 2a.** The latter requires a big expenditure of two relatively expensive identical RF-receiving channels, one for the target and the other for reference channel with a known internal reference delay to reduce the well known high mean-value drift from the nanosecond to the 10ps region.
The conventional troublesome and elaborate way in **Fig. 2a** depends on the cumbersome way to firstly transform the optical RF (radio frequency)-signal to the electrical state by the photodiode. This requires critical and costly RF-amplification including high phase drift and interference sensitivity with low dynamics and finally an electric/electric-mixing and correlation process, thus producing increased noise, jitter, drift, dynamic range reduction, and temperature sensitivity.
The PMD-laser radar receiver in **Fig. 2b** reduces the elaborately, voluminous erroneous, and expensive waste to a photo diode sized, directly mixing and correlating semiconductor pixel of about 50x50 $\mu m^2$ without any further RF-circuitry.
To understand this progress we consider the Photo-Gate PMD-principle of operation in **Fig. 3**. It schematically describes the PMD-semiconductor structure performing the process of inherent opto-electronic mixing and correlation which can be understood as a balanced and correlated photo-charge sampling process [3, 4, 5].
This specialized photo diode is equipped with two symmetrical read-out cathodes $K_a$ and $K_b$ and with the common anode on mass potential.

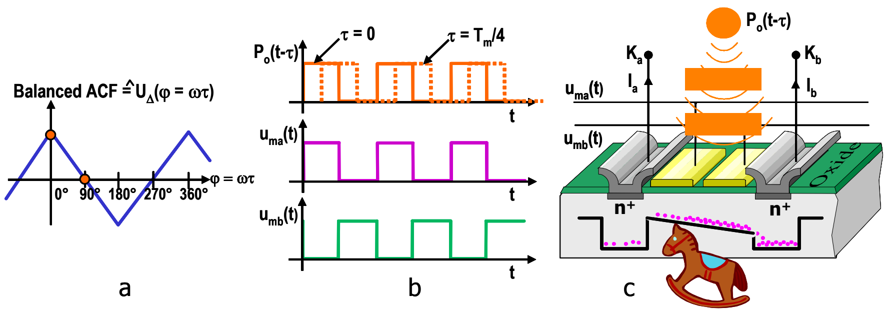



Fig. 3. c) PMD operates like a voltage seesaw for photo electrons generated below the balanced modulated photo-gates thus mixing by anti-symmetrical sampling to the left and to the right. b) At a light-wave phase zero the total charge drifts to the positive right side with maximum charge difference; at 90°: equally distributed charge results in zero difference. a) The charge or voltage difference described by the balanced Auto-Correlation-Function is a measure for delay.

Photo charges are generated by the incident modulated light signal $P_o(t)$ in the central photo sensitive region below two modulated photo gates which control the current flow direction by a seesaw effect via the applied balanced modulation voltage $\Delta u_m = u_{am} - u_{bm}$. This is based on an offset voltage $U_o$. The efficiency of the percentaged current flow distribution to the left $I_a$ and to the right $I_b$ depending on $\Delta u_m$ is depicted by the simulated and measured two pairs of modulation characteristics $\eta_{Ta,b}$ in **Fig. 4.** The percentaged currents are drawn against the only relevant odd part of the modulation voltage $\Delta u_m$, one of an older PMD1 and another of a newer PMD2 with higher modulation sensitivity from 10% to 90% within only +400 mV. These results indicate an excellent common mode rejection ratio of the PMD. Considering **Fig. 3** it's obvious that without any modulation there is a symmetrical photo current distribution of 50% while the sum is always 100%. $I_a + I_b = I_\Sigma$ give the total light intensity of both, modulated and background illumination.

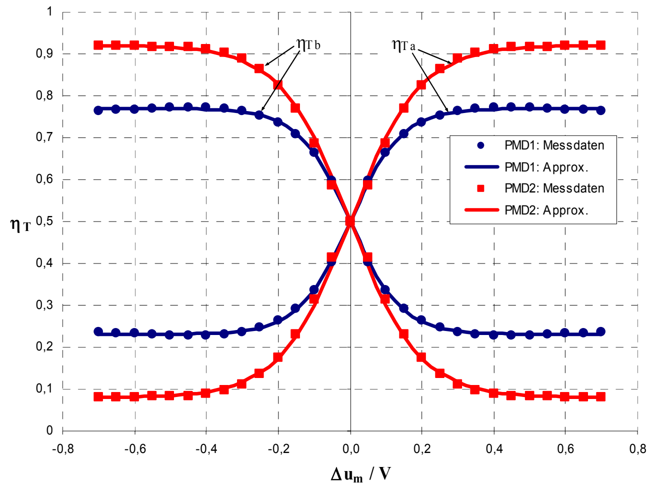



Fig. 4. Modulation characteristic describing the photo current distribution over the odd-mode modulation voltage $\Delta u_m$ [8]

On the further way to distance evaluation the photo currents $I_a$ and $I_b$ are integrated on capacities over an integration time $T_i$ to $U_a$ and $U_b$. Apart from overflow effects the voltage difference $U_a - U_b = \Delta U_{ab} = U_\Delta$ is not affected by the background illumination due to the non-correlated rectangular modulation signal depicted in **Fig. 3b**. $U_\Delta$ stands directly for the wanted correlated differential/balanced autocorrelation function DACF as shown in **Fig. 3a**. It contains the optical phase delay and modulation intensity and can be evaluated by means of the first harmonic coefficients in the following way:
Two measurements of $U_\Delta$ are made at 90° phase difference and taken as real and imaginary parts DACF ($\omega\tau_d = 0^o$) and DACF ($\omega\tau_d = 90^o$). Thus time-of-flight TOF=$\tau_d$ or distance $D = \tau_d \cdot c / 2$ can be calculated directly by the phase delay $\varphi_d = \omega \cdot \tau_d = \arctan (U_{\Delta Im} / U_{\Delta Re})$.
As seen here the balanced operation of PMD provides a lot of superior features, in particular to filter non correlated background light $P_{onc}(t)$ from the wanted and correlated modulated light $P_{om}(t - \tau_d)$ by means of even-odd selection as demonstrated in **Fig 3a** and **Fig. 4**.

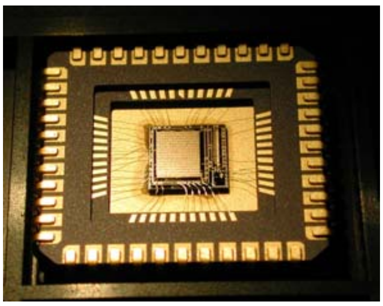
The substantial advantages of the PMD-solution may be summarized with respect to the high demands of variable real 3D-measuring conditions and of the aspired range resolution of mm to cm respectively 10 ps region:

(1) The PMD-laser radar receiver is reduces to a photo diode sized, directly mixing and correlating semiconductor pixel of about 50x50 $\mu m^2$.
(2) No electromagnetic interference problems of dense packaged PMD-matrices.
(3) 3D-PMD-cameras are using simple global illuminations and don't need pixel-wise laser pointing, no scanning. Sender and receiver are microelectronic parts.
(4) No reference channel and no broadband RF-amplifier are needed.
(5) The balanced sampling mixer provides a differential photo current integration result $U_\Delta$ which is equivalent to the balanced auto-correlation function DACF and directly delivers the wanted TOF
(6) The PMD-receiver provides the huge photodiode dynamics of bigger $1 : 10^6$
(7) Balanced mixing and correlation eliminates a lot of errors produced by temperature, aging, nonlinearities, RF-interference, distortions and background light,
(8) The measuring range is about 0.1 to 10 m with mm to some cm resolution depending on optics and modulated illumination power. Additionally using burst modulation and image intensifying or inherent amplification later on could achieve the some km-range.
(9) A newer realization of PMD-technology applying Schottky-technology provides sub-mm control using modulation frequencies up to 10 GHz with a nearly proportional enhancement of distance resolution [7, 9, 10].]

## 3 Some experimental results of 3D-PMD-video camera

To demonstrate the potential of the PMD-camera concept for future high speed 3D-perception we consider some of the first PMD-video cameras results. Under these exquisite conditions it was worth to aim for PMD-lines and PMD matrices. After external trials finally the Siegen cooperating teams of PMDTechnologies and the INV succeeded in 2003 with a PMD-matrix of 256 parallel laser radar receivers with high mixing efficiency or modulation contrast, high pixel uniformity and synchronisation, and low temperature sensitivity.

**Fig. 5a** presents the 16x16 PMD-pixel chip and **Fig. 5b** the 256-pixel 3D-PMD-camera. The LED-sender comprises about 100 LED's with 1 watt infrared radiation power and is modulated with about 20 MHz.



a)

Fig. 5a. 16x16 PMD-pixel in CMOS-Technology

b)

Fig. 5b. 3D-camera with 256 PMD-pixels and 100 LED

**Fig. 6** demonstrates four 3D-snap-shots selected out of a 50 Hz 3D-video sequence. The integration and evaluation time for each of 256 distances is 20 ms, proving the PMD-camera to be applied as a 3D-videocamera. On the theft is seen a scientist carrying a pale board towards the PMD-camera, starting from a spatial reference – a blue or dark grey ground-stripe staying in place on the left side during the board arises. The 3D-snap-shots show the ascending motion in coarse steps of 50 to 100 not shown snap-shots between.

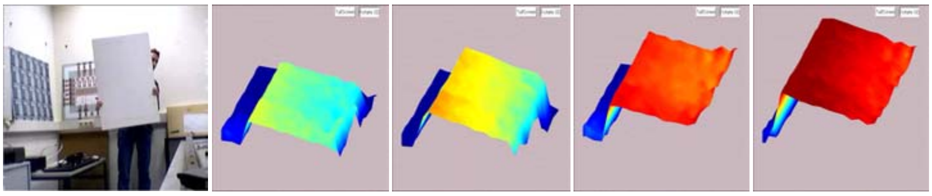


Fig. 6. Four 3D-snapshots out of a 50Hz video sequence show the upwards moving pale board in color coded height

Compared to three or five-beam microwave radars this PMD-camera has already realized the high number of 256 “beams” to be appropriate for a lot of applications, e.g., like precrash. So far we see no technical limitation on the way to higher pixel numbers. However grading up to the next PMD-matrix is a very complex, time consuming and laborious way.
A 3D-image of a human hand in **Fig. 7** has been taken with the same camera. The finger resolution of a few mm or about 10 to 20 ps is worth to be mentioned. But still more amazing is the homogeneity of time synchronization of the 256 separate PMD-receiver delays.
**Fig. 8** shows a 3D-snap-shot taken with a 1024 PMD-pixel camera equivalent to 1024 laser radar receivers and beams. The relating 64x16 pixel PMD-chip is seen in the midst of **Fig. 9** with a pixel area of about 11mm x 3.2mm.

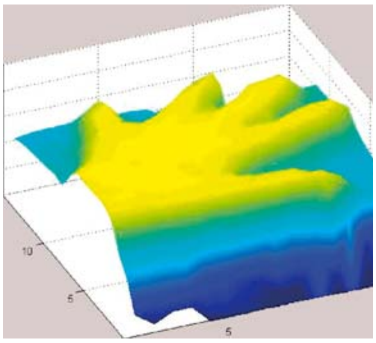



Fig. 7. A 3D-photo of a hand, resolves some mm distance

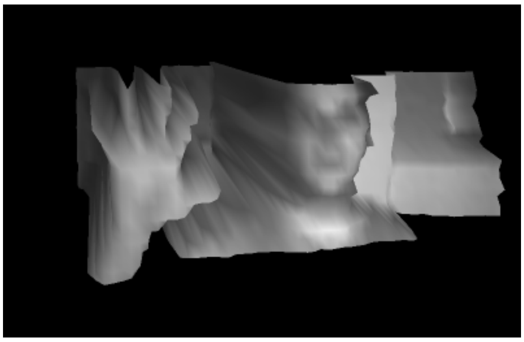


Fig. 8. This 3D-snapshot taken from a 3D-video sequence shows a co-worker greeting with his hand

It gives an idea of a “3D-photo” representing only one sculpture surface out of a 3D-video sequence. We see the greeting hand and the head of a co-worker and a little of the 3D-background taken by this 64x16 pixel 3D-Videocamera. Instead of the original snap-shot perspective the sculpture should be considered from another spatial viewpoint or perspective to see the real 3D-contour. That’s why the shown perspective of this 3D-image has been changed from the original face-to-face position more to the left side and upwards in order to better recognize the three dimensions at the special snap-shot moment.

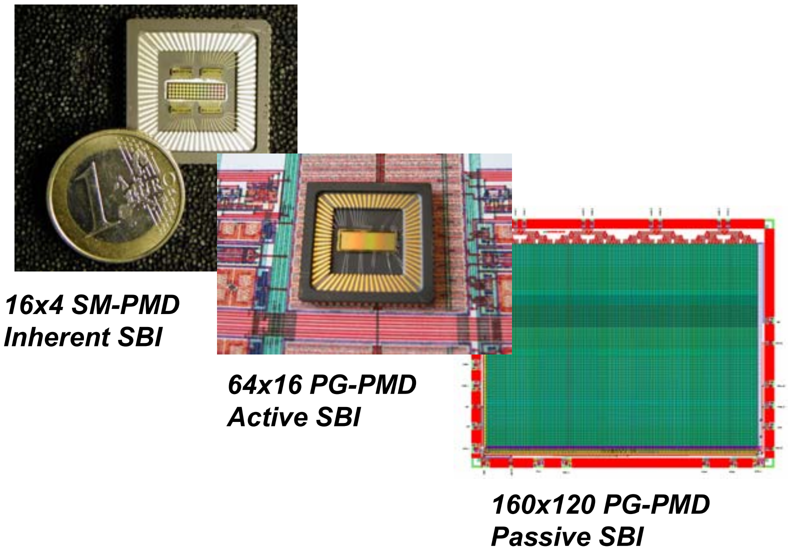



Fig. 9. Three PMD-matrices: a) 64 SM-MSM / OEP-PMD-pixels, b) 1024 Photo-Gate PMD-pixels and c) Layout of a 19200 Photo-Gate PMD-pixel array.

It should be mentioned that the measured 3D-shape principally does not depend from the distance and position – within 3D-camera specifications. Consequently overlapping contour parts of different snapshots have to be consistent and enable a substantially faster surface matching procedure than known from 2D-image matching in photogrammetry. Additionally the 3D-environment can be captured, modeled and mapped using time multiplex with one 3D-camera going around a sculpture or using space multiplex by means of several 3D-cameras positioned around an object of interest. This is essential for man-machine-interfacing to get the full freedom of wanted stereo vision perspectives by the tele-operator and als to robot teams for optimized 3D-perception. Very helpful for navigation, in particular of robot teams and for common safety will be an 3D-PMD camera with a mirror optimized for omni-directional 3D-vision and looking around e.g. by means of a rotationally symmetrical mirror.
On the right of **Fig. 9** we see the layout of the next PMD-matrix generation providing 19.200 TOF-receivers in a 160 x 120 matrix using a meanwhile proven CMOS-chip

design and technology with excellent properties. Before pursuing extremely high PMD-pixel numbers of e.g. more than 200.000 PMDs a combination of one 3D-camera and one or two 2D-cameras with optimized spatial synchronization and data fusion has to be taken into account. Simulations show a great advantage in object recognition and tracking by synchronizing the perception process in space and time using combined information of the 3D-PMD-videocamera of e.g. 20.000 to 160.000 PMD-pixels and a high resolution 2D-RGB-videocamera of 800.000 to 2 Million CCD or CMOS-pixels, observing preferably the same field-of-view.

## 4. Retrospective and outlook

Looking back to more than two decades of laser-radar research at the universities of Aachen and Siegen and in a smaller firm of Dr. Bölkow* finally in 1996 this surprising simple key component PMD was found [1].
Already in 1999/2000 [6] Dr. Lange realized the world-wide first 3D-camera with 1600 none balanced CCD/CMOS-pixels. For my opinion it's the beginning of a fascinating 3D-imaging area, in particular thanks to the competence and merit of the PMDTechnologies GmbH and their leaders and co-workers in cooperation with the INV of the University of Siegen, since they brought forward the Photo-Gate PMD technology in big steps to high reliability and large pixel numbers and towards products. Today about 20.000 receivers in a 120x160 PMD-matrix provide homogenous and exquisite specifications like very low and constant mean value, low standard deviation and very low cost compared with conventional radar receivers. This perspective opens an exciting view to future applications like

- Accident-free driving, precrash, intelligent airbag, automatic Stop&Go,
- Optimized automotive safety system including Differential-GPS, Omnidirectional 3D-PMD-vision and inertial system cooperation,
- Autonomous mobile robotics, 3D-PMD robotic-hand camera, robotic teamwork with cooperating PMD-cameras for optimized 3D-perception
- Industrial and agriculture production automation, human safety,
- 3D-man-machine interface, virtual and augmented reality, tele-operation, 3D-PMD-web camera for enhanced information exchange a.s.o.

in order to name only some examples.
Another perspective not treated here are various PMD-technologies, e.g. Self Modulation of Schottky and PN-diodes, mixer based spectral enhancement of existing PMD-cameras, PMD-technology with inherent amplification etc.
For example, in the upper left corner of **Fig. 9** we see a 16x4-SM(self-modulation)-PMD-chip based on MSM-Schottky technology. This new emerging PMD-technology based needs no photo gates and provides a lot of promising new features like very high modulation speed in the 10 GHz-range for sub-mm distance resolution and an inherent background light elimination. A minimum of two MSM-PMDs with common current read-out is combined to a new OE-component, called **OEP** for **Opto Electronic Processor**. This OEP may advance to a key-component in high-speed digital and analog opto-electronic signal processing technologies [7, 9, 10].

## 5. Acknowledgements

Here presented results in fast technical 3D-vision are mainly based on the engagement and competence of the cooperating teams of PMDTechnologies and S-TEC GmbH and the University Institute of Information and Signal Processing INV. Thanks also to the collaborating members of the Interdisciplinary Centre of sensor systems ZESS and to the helpful partly funding by the German BMBF-ministry.

## References

1. Schwarte, R.: Verfahren und Vorrichtung zur Bestimmung der Phasen- und Amplitudeninformation einer elektromagnetischen Welle. DE 19635 932.5; DE 19704 496.4
2. Schwarte, R. et. al.: Principles of Three-Dimensional Imaging Techniques. In Handbook of Computer Vision and Applications, Volume 1, Sensors and Imaging; Academic Press; Boston 1999
3. Z., Xu, Investigation of 3D-Imaging Systems Based on Modulated Light and Optical RF-Interferometry ORFI, ZESS Forschungsberichte, Shaker Verlag, ISBN 3-8265-6726-6, Aachen 1999
4. B. Buxbaum, Optische Laufzeitentfernungsmessung und CDMA auf Basis der PMD-Technologie mittels Phasenvariabler PN-Modulation, ZESS Forschungsberichte Bd. 17. ISBN 3-8265-9805-9
5. T. Ringbeck, Untersuchung eines analogen optoelektronischen Phasenregelkreises hinsichtlich dessen Anwendungspotenzials für die Photonik, Dissertation, INV, Universität Siegen, Siegen 2001
6. Lange, R.: 3D Time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. Dissertation, Universität Siegen, 2000
7. Kraft, H., Frey, J., Moeller, T., Albrecht, M., Grothof, M., Schink, B., Hess, H., Buxbaum, B.: 3D-Camera of High 3D-Frame Rate, Depth-Resolution and Background Light Elimination Based on improved PMD (Photonic Mixer Device)-Technologies. OPTO 2004, Nürnberg
8. Zhang, Z.: Untersuchung und Charakterisierung von PMD (Photomischdetektor)-Strukturen und ihren Grundschaltungen. Dissertation, Universität Siegen, 2004
9. Schwarte, R., Zhang, Z., Kraft, H., Moeller, T., Frey, J., Grothof, M., Buxbaum, B., Ringbeck, T.: OEP for Extremely Fast Multi-Channel Analog and Digital OE-Signal Processing. OPTO 2004, Nürnberg.
10. Schwarte, R.: Ultra-fast Analog and Digital OE-Multi-channel Signal Processing using PMD/OEP-Technologies. $10^{th}$ Microoptics Conference MOC 04, Jena, Sept. 1-3, 2004