A Comparison of PMD-Cameras and Stereo-Vision for the Task of Surface Reconstruction using Patchlets

Christian Beder, Bogumil Bartczak and Reinhard Koch
Computer Science Department
Universiy of Kiel
Hermann-Rodewald-Str.3
24118 Kiel, Germany
{beder,bartczak,rk}@mip.informatik.uni-kiel.de

Abstract

Recently real-time active 3D range cameras based on time-of-flight technology (PMD) have become available. Those cameras can be considered as a competing technique for stereo-vision based surface reconstruction. Since those systems directly yield accurate 3d measurements, they can be used for benchmarking vision based approaches, especially in highly dynamic environments. Therefore, a comparative study of the two approaches is relevant.

In this work the achievable accuracy of the two techniques, PMD and stereo, is compared on the basis of patchlet estimation. As patchlet we define an oriented small planar 3d patch with associated surface normal. Least-squares estimation schemes for estimating patchlets from PMD range images as well as from a pair of stereo images are derived. It is shown, how the achivable accuracy can be estimated for both systems.

Experiments under optimal conditions for both systems are performed and the achievable accuracies are compared. It has been found that the PMD system outperformed the stereo system in terms of achievable accuracy for distance measurements, while the estimation of normal direction is comparable for both systems.

1. Introduction

One of the main goals of computer vision and photogrammetry is the accurate and contactless measurement of surfaces, for which many approaches including stereo systems (cf. [17]) and active systems such as structured light or laser scanners have been used (cf. [6]). However, the application domain of all this systems is restricted. For instance the algorithmic complexity of stereo systems is quite high and they are not applicable in case of weakly textured surfaces. Laser scanners and structured light approaches on the other hand cannot cope with moving objects, because capturing is not instantaneous for this systems.

A new promising development for the area of surface reconstruction, that is able to cope with those caveats of the existing techniques, is the Photonic Mixer Device (PMD), which measures distances directly for a two dimensional field of pixels based on the time of flight of incoherent, modulated infrared light. Recently PMD cameras have been developed that are capable of capturing reliable depth images directly in real-time. Those cameras are compact and affordable, which makes them attractive for versatile applications including surveillance and computer vision (cf. [10]). The successful application of this technology in Structure from Motion [18], motion capturing [4] and face tracking[3] have been demonstrated.

In [8] the complementary nature of stereo vision based systems and PMD cameras is discussed qualitatively and a simple method for fusing the information gathered from those two system is proposed. Yet a quantitave comparison of both systems is not done. The main contribution of this work is such a quantitative comparison based on the estimation of patchlets (cf. [14]), i.e. small planar surface patches with an associated surface normal, being a very useful surface representation for tasks such as segmentation (cf. [13]) and visualization (cf. [19]).

First a short introduction to the technology underlying the PMD image formation will be given in section 2 in order to provide some background information.

Then in section 3 two least-squares estimation schemes for the PMD images as well as for the stereo images will be presented and it will be shown, how optimal estimates for the patchlets together with their covariance matrices can be obtained. While the PMD camera provides direct geometric measurements, which are used in the estimation of the patchlet, the stereo matching is based on estimating a local homography between the images (cf. [12, 15, 16]),

which optimally aligns the image intensities between the stereo image pair (cf. [1, 11, 20]).

The primary scope of this paper lies on benchmarking both systems against each other. Quality criteria of the surface reconstruction based on the estimated accuracy of the patchlet estimation are derived and the performance of both the stereo system as well as the PMD camera will be compared in section 4 based on those criteria. It will be shown, that the surface distance is estimated by an order of magnitude more accurate using the PMD camera than with the stereo system, implying the usability of such a combined stereo-PMD-system for the task of benchmarking the accuracy of surface reconstruction from stereo images under realistic conditions.

2. The PMD-Camera

We will first give some background information on the Photonic Mixer Device (PMD), which is a semiconductor structure based on CCD- or CMOS-technology [21]. Integrated in an image sensor array it is capable of modulating the current that is generated by received light intensities in every single pixel. This can be utilized to build affordable cameras that are able to measure depth with high precision. One such camera is shown in figure 1. It mainly consists of a camera with the PMD sensor and light emitter arrays that are used to send out modulated light. The light is reflected by 3D scene points and received by the PMD image sensor.

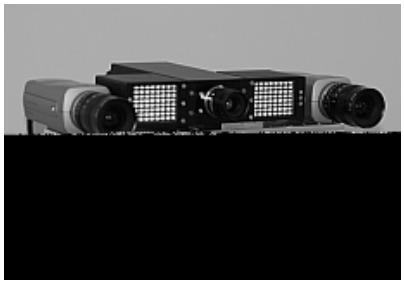

Figure 1. The Rig used to conduct the benchmarking. It consists of two color cameras that frame the PMD-Camera.

The depth measurement performed with a detector using the Photonic Mixer Device is based on the time of flight principle. This principle measures the time a signal with well defined speed spends, when traveling from the signal's transmitter to a receiver. Different approaches for measuring the time-of-flight with light exist [9]. One method suitable for the use with the PMD is to modulate the emitted light intensity with a periodic pattern. Depending on the "time of flight" a phaseshift of the periodic pattern is observable. The PMD-Camera is able to extract this phase shift in every pixel.

Though different intensity modulations using square waves or pseudo noise coding are possible, the use of a sinusoidal signal is technically well realizable [22]. Generally an intensity wave $I(x,t) = I_0 + I_A \cos(2\pi\nu_m(t + \frac{x}{c}) + \varphi_0)$ with modulation frequency ν_{m}, propagation speed c and initial phase φ_0 is sent out. At two points x_0 and x_1 of the wave the phase shift

$$\Delta\varphi = 2\pi\nu_m(x_1 - x_0)/c \tag{1}$$

is observable. Extracting this shift from the wave therefore delivers the distance between x_0 and x_1. Due to the repetitive nature of the wave the non-ambiguous wavelength of the measurement is $\lambda_{max} = c/\nu_m$. The effective measurement range is $\frac{\lambda_{max}}{2}$ because light wave has to return to its source to be detected. Typically a modulation frequency of $\nu_m = 20$MHz is used which gives the camera 7.5 meters of unambiguouse depth range. Reflections from distances beyond this range might cause measurement errors due to phase wrapping, however usually the reflected light intensity is too small to cause such errors. For very small distances below 2m, the reflected strong light intensity might cause nonlinear saturation effects which might limit the accuracy and cause bias [10]. Therefore, the usable range was chosen between $2 - 7.5$ meters.

The phase difference is measured by cross correlation between the sent and received modulated signal by the PMD chip. Since the resolution of the phase difference measurement is independent from distance, the achievable depth resolution is independent from scene depth. This is in contrast to stereo triangulation where depth accuracy is proportional to inverse depth.

After taking depth calibration and lens distortions [7, 10] into account, the model of a central perspective projection for the PMD-Camera can be used to compute a depth value for each PMD pixel.

3. Estimating Patchlets

For the purpose of benchmarking the PMD-Camera and the stereo-vision system it is investigated how well both systems perform on a region based non-linear estimation of the normal and position of a planar target. The estimation is done locally and in that equals the extraction of a patchlet proposed by Murray [14]. Patchlets are small planar elements that have a confidence for their normal and their position. They have been developed for segmentation and extraction of higher level surfaces from dense disparity maps, generated by correlation based stereo-vision. Since the PMD-Camera is directly delivering depth maps and no analysis of the underlying disparity estimation algorithm shall be performed, Murrays extraction algorithm is not used.

In order to evaluate both systems when they are expected to perform best, the tests are conducted under more or less optimal conditions. Therefore the estimation schemes presented in the following sections do not cover robustness.

The uncertainty of a patchlet which is delivered next to a position and a normal is proposed to be the benchmark for the individual system. In section 4 this uncertainty is used to evaluate the behaviour of the systems under varying conditions.

3.1. Estimation of patchlets from PMD images

The PMD-camera determines for each ray direction corresponding to pixel $\mathbf{x}$ its distance λ to the optical center. If the camera geometry is given by a calibration matrix (cf. [5, p.141f]) as

$$\mathsf{P} = \mathsf{K}(I_3|\mathbf{0}_3) \tag{2}$$

then the corresponding 3d point is obtained directly from the distance λ as

$$\boldsymbol{X} = \frac{\lambda \mathsf{K}^{-1}\mathbf{x}}{\sqrt{\mathbf{x}^T\mathsf{K}^{-T}\mathsf{K}^{-1}\mathbf{x}}} \tag{3}$$

This 3d point lies on the plane $(\boldsymbol{n}^T, d)^T$, if

$$\boldsymbol{X}^T\boldsymbol{n} + d = 0 \tag{4}$$

or equivalently

$$g(\boldsymbol{n}', \lambda, \mathbf{x}) = \lambda\mathbf{x}^T\mathsf{K}^{-T}\boldsymbol{n}' + \sqrt{\mathbf{x}^T\mathsf{K}^{-T}\mathsf{K}^{-1}\mathbf{x}} = 0 \tag{5}$$

using the substitution $\boldsymbol{n}' = \frac{\boldsymbol{n}}{d}$ to parameterize the plane. This expression is linear in the plane parameters $\boldsymbol{n}'$ so that initial values are easily computed. For an optimal estimation and derivation of the uncertainties the Gauss-Helmert-Model (cf. [2, p.80ff]) must be applied.

Therefore the partial derivatives are calculated as

$$\boldsymbol{a}^T(\lambda, \mathbf{x}) = \frac{\partial g(\boldsymbol{n}', \lambda, \mathbf{x})}{\partial \boldsymbol{n}'} = \lambda\mathbf{x}^T\mathsf{K}^{-T} \tag{6}$$

$$b_\lambda(\boldsymbol{n}', \mathbf{x}) = \frac{\partial g(\boldsymbol{n}', \lambda, \mathbf{x})}{\partial \lambda} = \mathbf{x}^T\mathsf{K}^{-T}\boldsymbol{n}' \tag{7}$$

and

$$\begin{aligned} \boldsymbol{b}_\mathbf{x}^T(\boldsymbol{n}', \lambda, \mathbf{x}) &= \frac{\partial g(\boldsymbol{n}', \lambda, \mathbf{x})}{\partial \mathbf{x}} && (8)\\ &= \lambda\boldsymbol{n}'^T\mathsf{K}^{-1} + \frac{\mathbf{x}^T\mathsf{K}^{-T}\mathsf{K}^{-1}}{\sqrt{\mathbf{x}^T\mathsf{K}^{-T}\mathsf{K}^{-1}\mathbf{x}}} && (9)\end{aligned}$$

so that the Taylor expansion of equation (5) yields

$$\begin{aligned} g(\boldsymbol{n}', \hat{\lambda}, \hat{\mathbf{x}}) \approx\ & g(\boldsymbol{n}'_0, \lambda_0, \mathbf{x}_0) && (10)\\ & +\boldsymbol{a}^T(\lambda_0, \mathbf{x}_0)(\boldsymbol{n}' - \boldsymbol{n}'_0)\\ & +\boldsymbol{b}_\mathbf{x}^T(\boldsymbol{n}'_0, \lambda_0, \mathbf{x}_0)(\hat{\mathbf{x}} - \mathbf{x}_0)\\ & +b_\lambda(\boldsymbol{n}'_0, \mathbf{x}_0)(\hat{\lambda} - \lambda_0)\\ =\ & g(\boldsymbol{n}'_0, \lambda_0, \mathbf{x}_0) && (11)\\ & +\boldsymbol{a}^T(\lambda_0, \mathbf{x}_0)(\boldsymbol{n}' - \boldsymbol{n}'_0)\\ & +\boldsymbol{b}_\mathbf{x}^T(\boldsymbol{n}'_0, \lambda_0, \mathbf{x}_0)(\hat{\mathbf{x}} - \mathbf{x} + \mathbf{x} - \mathbf{x}_0)\\ & +b_\lambda(\boldsymbol{n}'_0, \mathbf{x}_0)(\hat{\lambda} - \lambda + \lambda - \lambda_0)\end{aligned}$$

Setting the latter equation equal to zero yields the linearized model for each point

$$\boldsymbol{a}^T\Delta\boldsymbol{n}' + \boldsymbol{b}_\mathbf{x}^T\Delta\mathbf{x} + b_\lambda\Delta\lambda = c_g \tag{12}$$

with the substitutions

$$\Delta\boldsymbol{n}' = \boldsymbol{n}' - \boldsymbol{n}'_0 \tag{13}$$

$$\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x} \tag{14}$$

$$\Delta\lambda = \hat{\lambda} - \lambda \tag{15}$$

and

$$\begin{aligned} c_g =\ & -g(\boldsymbol{n}'_0, \lambda_0, \mathbf{x}_0) && (16)\\ & -\boldsymbol{b}_\mathbf{x}^T(\boldsymbol{n}'_0, \lambda_0, \mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\\ & -b_\lambda(\boldsymbol{n}'_0, \mathbf{x}_0)(\lambda - \lambda_0)\end{aligned}$$

Using a window containting more than three points, the plane parameter updates may be estimated iterativly as

$$\Delta\boldsymbol{n}' = \left(\sum_{i=1}^N \omega_i\boldsymbol{a}_i\boldsymbol{a}_i^T\right)^{-1}\sum_{i=1}^N \omega_i\boldsymbol{a}_i c_{g_i} \tag{17}$$

using the substitution

$$\omega_i = \frac{1}{\boldsymbol{b}_{\mathbf{x}_i}^T\boldsymbol{\Sigma}_{\mathbf{x}_i\mathbf{x}_i}\boldsymbol{b}_{\mathbf{x}_i} + \sigma_{\lambda_i}^2 b_{\lambda_i}^2} \tag{18}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{x}_i\mathbf{x}_i} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{x}_i\boldsymbol{x}_i} & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{pmatrix} \tag{19}$$

is the pixel accuracy of the PMD camera and $\sigma_{\lambda_i}^2$ is the accuracy of the distance measurement.

The residuals are computed as

$$\Delta\mathbf{x}_i = \boldsymbol{\Sigma}_{\mathbf{x}_i\mathbf{x}_i}\boldsymbol{b}_{\mathbf{x}_i}\omega_i(c_{g_i} - \boldsymbol{a}_i^T\Delta\boldsymbol{n}') \tag{20}$$

$$\Delta\lambda_i = \sigma_{\lambda_i}^2 b_{\lambda_i}\omega_i(c_{g_i} - \boldsymbol{a}_i^T\Delta\boldsymbol{n}') \tag{21}$$

allowing an iterative estimation until $\boldsymbol{n}'$ is converged.

Finally the expected covariance is obtained as

$$\boldsymbol{\Sigma}_{\boldsymbol{n}'\boldsymbol{n}'} = \left(\sum_{i=1}^N \omega_i\boldsymbol{a}_i\boldsymbol{a}_i^T\right)^{-1} \tag{22}$$

From the residuals the variance factor may be estimated as

$$\sigma_0^2 = \frac{1}{N-3}\sum_{i=1}^N\left(\frac{\Delta\mathbf{x}_i^T\boldsymbol{b}_{\mathbf{x}_i}\boldsymbol{b}_{\mathbf{x}_i}^T\Delta\mathbf{x}_i}{\boldsymbol{b}_{\mathbf{x}_i}^T\boldsymbol{\Sigma}_{\mathbf{x}_i\mathbf{x}_i}\boldsymbol{b}_{\mathbf{x}_i}} + \frac{\Delta\lambda_i^2}{\sigma_{\lambda_i}^2}\right) \tag{23}$$

yielding the estimate for the covariance matrix

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}'\boldsymbol{n}'} = \sigma_0^2\boldsymbol{\Sigma}_{\boldsymbol{n}'\boldsymbol{n}'} \tag{24}$$

The same quantity together with its covariance matrix may be estimated using a stereo system, which will be presented in the following section.

3.2. Estimation of patchlets from stereo images

Given a calibrated stereo system with the first camera being

$$\mathsf{P}_1 = \mathsf{K}_1(I_3|\mathbf{0}_3) \tag{25}$$

and the second camera being

$$\mathsf{P}_2 = \mathsf{K}_2(R|\boldsymbol{t}) \tag{26}$$

with the known rotation matrix R and translation vector t, the homography relating points on the plane $(\boldsymbol{n}^T, d)^T$ from the first into the second camera is given by (cf. [5, p.314])

$$\mathsf{H} = \mathsf{K}_2\left(\boldsymbol{R} - \boldsymbol{t}\frac{\boldsymbol{n}^T}{d}\right)\mathsf{K}_1^{-1} \tag{27}$$

which is linear in the vector $\boldsymbol{n}' = \frac{\boldsymbol{n}}{d}$. Hence, points on the plane are transformed according to

$$\begin{aligned}\mathbf{x}_2 &= \mathsf{H}(\boldsymbol{n}')\mathbf{x}_1 &(28)\\ &= \mathsf{K}_2R\mathsf{K}_1^{-1}\mathbf{x}_1 - ((\mathbf{x}_1^T\mathsf{K}_1^{-T}) \otimes (\mathsf{K}_2\boldsymbol{t}))\boldsymbol{n}' &(29)\end{aligned}$$

where $\otimes$ denotes the Kronecker product.

We now assume, that the grey value of corresponding points is equal in the two images. Using the Euclidean normalization function

$$\boldsymbol{h}(\mathbf{x}) = \boldsymbol{h}\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \frac{1}{w}\begin{pmatrix} u \\ v \end{pmatrix} \tag{30}$$

this is expressed in terms of the two images I_1 and I_2 as

$$I_1(\boldsymbol{h}(\mathbf{x}_1)) = I_2(\boldsymbol{h}(\mathbf{x}_2)) \tag{31}$$

or substituting equation (29)

$$I_1(\boldsymbol{h}(\mathbf{x}_1)) = I_2(\boldsymbol{h}(\mathsf{K}_2R\mathsf{K}_1^{-1}\mathbf{x}_1 - ((\mathbf{x}_1^T\mathsf{K}_1^{-T}) \otimes (\mathsf{K}_2\boldsymbol{t}))\boldsymbol{n}')) \tag{32}$$

Applying chain rule, the partial derivatives of this expression are given by

$$\begin{aligned}&\boldsymbol{a}^T(\mathbf{x}_1, \boldsymbol{n}') &(33)\\ &= \frac{\partial}{\partial \boldsymbol{n}'} I_2(\boldsymbol{h}(\mathsf{K}_2R\mathsf{K}_1^{-1}\mathbf{x}_1 - ((\mathbf{x}_1^T\mathsf{K}_1^{-T}) \otimes (\mathsf{K}_2\boldsymbol{t}))\boldsymbol{n}'))\\ &= -(\nabla I_2)(\boldsymbol{h}(\mathsf{H}(\boldsymbol{n}')\mathbf{x}_1))\mathsf{J}(\mathsf{H}(\boldsymbol{n}')\mathbf{x}_1)((\mathbf{x}_1^T\mathsf{K}_1^{-T}) \otimes (\mathsf{K}_2\boldsymbol{t})) &(34)\end{aligned}$$

with the Jacobian of the normalization function being

$$\mathsf{J} = \frac{\partial}{\partial \mathbf{x}}\boldsymbol{h} = \begin{pmatrix} \frac{1}{w} & 0 & -\frac{u}{w^2} \\ 0 & \frac{1}{w} & -\frac{v}{w^2} \end{pmatrix} \tag{35}$$

Hence, the Taylor expansion of equation (32) yields for every point on the plane

$$\underbrace{\boldsymbol{a}^T(\mathbf{x}_1, \boldsymbol{n}_0')}_{\boldsymbol{a}_i^T}\underbrace{(\boldsymbol{n}' - \boldsymbol{n}_0')}_{\Delta\boldsymbol{n}'} \approx \underbrace{I_1(\boldsymbol{h}(\mathbf{x}_1)) - I_2(\boldsymbol{h}(\mathsf{H}(\boldsymbol{n}_0')\mathbf{x}_1))}_{\Delta l_i} \tag{36}$$

Using again a window containing more that three points, the plane parameter updates may be estimated iteratively as

$$\Delta\boldsymbol{n}' = \left(\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{a}_i\boldsymbol{a}_i^T\right)^{-1}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{a}_i\Delta l_i \tag{37}$$

having the expected covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{n}'\boldsymbol{n}'} = \left(\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{a}_i\boldsymbol{a}_i^T\right)^{-1} \tag{38}$$

where σ_i^2 is the variance of the image noise.

From the residuals the variance factor may be estimated finally as

$$\sigma_0^2 = \frac{1}{N-3}\sum_{i=1}^{N}\sigma_i^{-2}(\Delta l_i - \boldsymbol{a}_i^T\Delta\boldsymbol{n}')^2 \tag{39}$$

yielding the estimate for the covariance matrix

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}'\boldsymbol{n}'} = \sigma_0^2\boldsymbol{\Sigma}_{\boldsymbol{n}'\boldsymbol{n}'} \tag{40}$$

The resulting plane parameterization together with its covariance matrix are equivalent to the result for the PMD camera allowing a direct comparison, which will be carried out in the following.

4. Benchmarking

4.1. Derivation of bechmarking quantities

It was shown in the previous sections, how for each of the two systems a local plane together with its uncertainity can be derived. The direct interpretation of those plane parameters is difficult, so that we derive some more meaningful quantities.

The first quantity is the distance of the patchlet from the camera, which is in both cases obtained by (cf. equation (5))

$$\lambda = -\frac{\sqrt{\mathbf{x}^T\mathsf{K}^{-T}\mathsf{K}^{-1}\mathbf{x}}}{\mathbf{x}^T\mathsf{K}^{-T}\boldsymbol{n}'} \tag{41}$$

having the standard deviation

$$\sigma_\lambda = \sqrt{\left(\frac{\partial\lambda}{\partial\boldsymbol{n}'}\right)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}'\boldsymbol{n}'}\left(\frac{\partial\lambda}{\partial\boldsymbol{n}'}\right)^T} \tag{42}$$

using the Jacobian

$$\frac{\partial\lambda}{\partial\boldsymbol{n}'} = \frac{\sqrt{\mathbf{x}^T\mathsf{K}^{-T}\mathsf{K}^{-1}\mathbf{x}}}{\left(\mathbf{x}^T\mathsf{K}^{-T}\boldsymbol{n}'\right)^2}\mathbf{x}^T\mathsf{K}^{-T} \tag{43}$$

The second quantity is the normal direction of the patchlet, which is obtained by spherical normalization

$$\overline{\boldsymbol{n}} = \frac{\boldsymbol{n}'}{\sqrt{\boldsymbol{n}'^T\boldsymbol{n}'}} \tag{44}$$

having the covariance matrix

$$\boldsymbol{\Sigma}_{\overline{\boldsymbol{n}}\overline{\boldsymbol{n}}} = \frac{1}{\boldsymbol{n}'^T\boldsymbol{n}'}\left(I_3 - \frac{\boldsymbol{n}'\boldsymbol{n}'^T}{\boldsymbol{n}'^T\boldsymbol{n}'}\right)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}'\boldsymbol{n}'}\left(I_3 - \frac{\boldsymbol{n}'\boldsymbol{n}'^T}{\boldsymbol{n}'^T\boldsymbol{n}'}\right) \tag{45}$$

This covariance matrix is singular and its two non-zero singular values are the variances along the two principal directions tangential to the unit sphere. Hence, computing the singular value decomposition

$$\boldsymbol{\Sigma}_{\overline{\boldsymbol{n}}\overline{\boldsymbol{n}}} = U\begin{pmatrix}\sigma_1 & & \\ & \sigma_2 & \\ & & 0\end{pmatrix}V^T \tag{46}$$

we obtain the angular standard deviations of the normal in the two principal directions

$$\alpha_1 = \arctan\sqrt{\sigma_1} \quad \text{and} \quad \alpha_2 = \arctan\sqrt{\sigma_2} \tag{47}$$

In the following those two quantities together with the standard deviation of the distance σ_λ will be used to characterize and compare the performance of the two systems.

4.2. Results

For benchmarking both systems a stereo-rig was used, which is shown in image 1. It consists of two Sony cameras which deliver images with a resolution of 1024×768 pixels and a field of view of $40° \times 32°$. The PMD camera in the middle is PMDtech's model 3K-s with a resolution of 64×48 pixels over a viewing angle of $22° \times 17°$. The rig was calibrated using a calibration pattern so that the internal parameters of each of the three camera are known. In order to compare or transfer measurements from the PMD camera to the stereo system and vice versa, the external calibration of all cameras with respect to the left stereo camera was also computed. The stereo system had a baseline of approximately 30 cm and the orientation was close to standard stereo geometry. The relatively small baseline compared to the object distance results from the requirement, that the system is more or less compact and is mountable on a robotic platform.

Three series of images of a planar poster shown in figure 2 at a distance of 3m from the rig were taken. In each series, 14 pictures for each camera were made. For each picture the poster was rotated around a fixed tilt axis, equidistantly sampling an angle interval of $70°$. Between the series the rig was rolled around the common optical axis so that the angle β between the poster rotation axis and the rig baseline was $\beta = 0°, 45°$ and $90°$.

In the images from the first camera and in the depth images of the PMD camera the image area covered by the movie poster was sampled at 10 randomly chosen points. The local plane parameters and their uncertainties were iteratively estimated using the estimation schemes described

Figure 2. The setup for the three experiments. The first column shows the first and last image from the series taken with a roll angle of $\beta = 0°$, the second column shows the first and last image for a roll angle of $\beta = 45°$ and the last column shows the first and last image for a roll angle of $\beta = 90°$. Only the image area covered by the movie poster was used in the experiments.

in the previous sections from a small region around each sample-point. An equal region size was chosen for both systems, such that an opening angle of approximately $1° \times 1.3°$ corresponding to 20×20 pixels for the stereo system and 3×3 for the PMD system was used. From those samples the standard deviations of the distances from the camera σ_λ for the stereo-system as well as for the PMD camera were derived. The mean over all samples together with the standard deviation were computed and plotted against the plane tilt angle. Because σ_λ is the deviation of the reprojection of the image sample point onto the estimated patchlet, it is correlated with the patchlets normal and its magnitude is expected to increase with growing plane tilt. While the uncertainty of the PMD-Camera's depth measurement is mainly influenced by constant quantization, the stereo system relies on the proper matching of texture.

From the results (cf. figures 3, 4 and 5) it can be seen, that the accuracy of the PMD system is in the range of $0.5mm - 8mm$ while the accuracy of the stereo system is in the range of $30mm - 100mm$ (cf. figures 6, 7 and (cf. figures 7), which is about one order of magnitude worse. This could be expected from our setup, because one pixel disparity in the stereo images corresponds approximately to $20mm$ of distance for the baseline-distance-ratio used, and the quantization interval of the used PMD-Camera is $3.6mm$ wide, which leads to an expected position deviation for fronto-parallel patchlets of $1.3mm$. Further observe that the reliability of distance estimation degrades with the tilt angle of the surface in both systems as expected. The plots for the different roll angles show that there is no significant effect of the roll angle on the experiments.

Also the mean and standard deviation of the angular deviation of the normal α_1 and α_2 (cf. figures 10,9 and 11) were determined as described in the previous section. The estimation of the normal direction depends on the ability to locate 3D points on the patchlet. For the stereo system

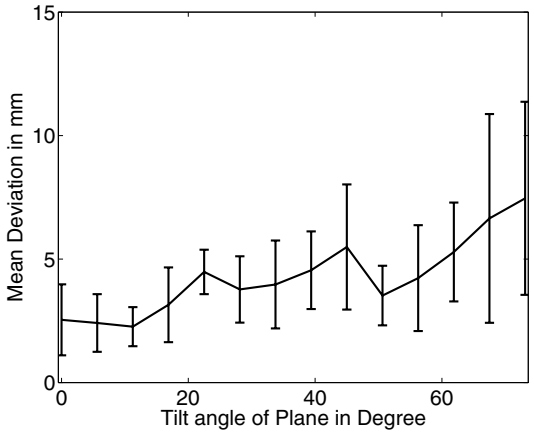

Figure 3. The standard deviation of the distance uncertainty σ_λ deviation for the PMD system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 0°.

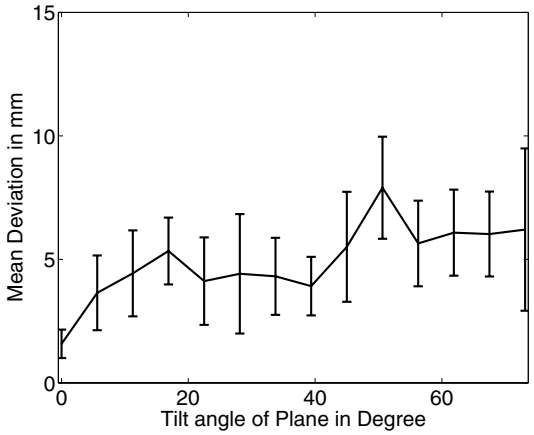

Figure 4. The standard deviation of the distance uncertainty σ_λ for the PMD system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 45°.

the uncertainty of 3d points in direction of the optical axis depends on the length of the baseline while the uncertainty parallel to the image plane depends on the image resolution. In our case the baseline is short and the resolution is high, so that we expect the angular accuracy of the normal direction of the patchlets to increase with the plane tilt angle. The situation is reversed for the PMD system, where the resolution is very low and the distance measurement is very accurate.

The angular uncertainty in the major principal direction is about $5^\circ - 15^\circ$ for the PMD system and in the range of $2^\circ - 8^\circ$ for the stereo system. Unlike the distance uncertainty, the stereo system performed slightly better here although the difference is not as significant. The reason for this is the very low resolution of the PMD camera allowing only very small window sizes for reasonable opening angles. However, for planar objects one might increase the window size yielding the effect shown in figure 12. The

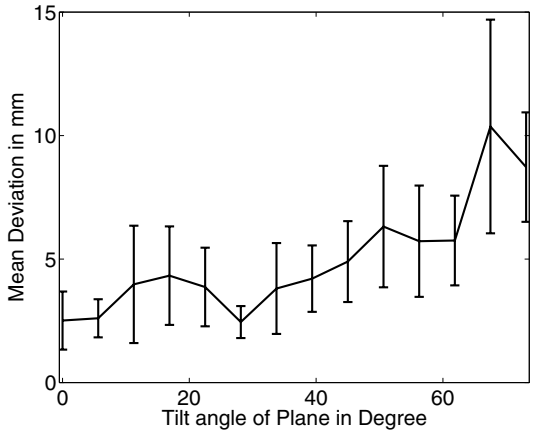

Figure 5. The standard deviation of the distance uncertainty σ_λ for the PMD system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 90°.

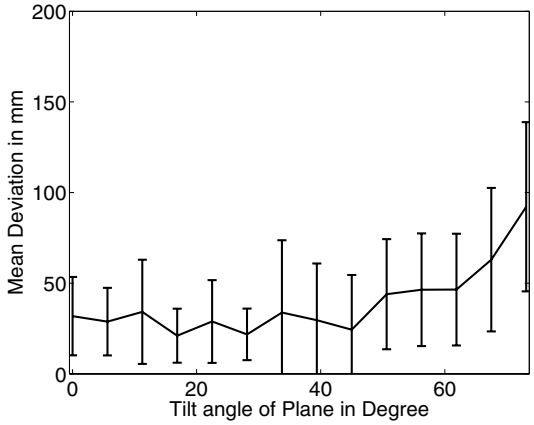

Figure 6. The standard deviation of the distance uncertainty σ_λ deviation for the stereo system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 0°.

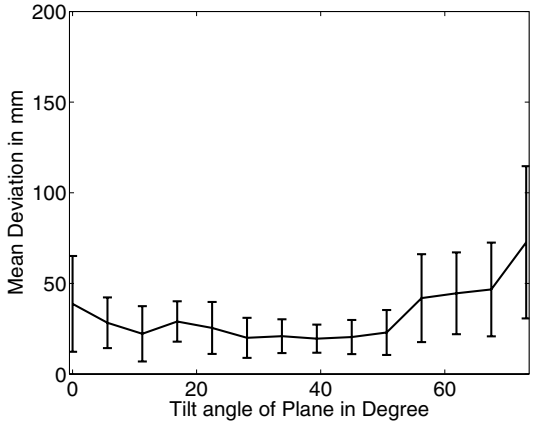

Figure 7. The standard deviation of the distance uncertainty σ_λ for the stereo system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 45°.

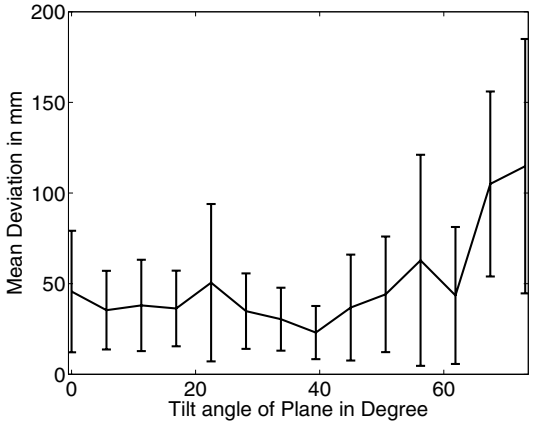

Figure 8. The standard deviation of the distance uncertainty σ_λ for the stereo system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 90°.

angular error drops to approximately $1^\circ - 2^\circ$ implying the potential of increasing the resolution of PMD systems for the task of normal estimation.

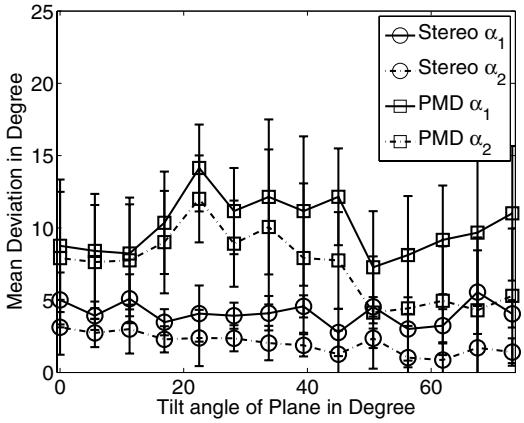

Figure 9. The standard deviation of the angular uncertainties α_1 and α_2 for the stereo system and the PMD system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 0°.

5. Conclusion

We have compared the achievable accuracies of estimated patchlets for stereo systems on the one hand and PMD range images on the other hand. It has been found, that the PMD system outperformed the stereo system in terms of accuracy. The distance of the patchlet from the camera was estimated by an order of magnitude better using the PMD system for the given setup, while the accuracy of the estimated normal direction was found to be comparable for both systems.

However, the major drawback of current PMD cameras is image resolution, so that a much lower spatial resolution

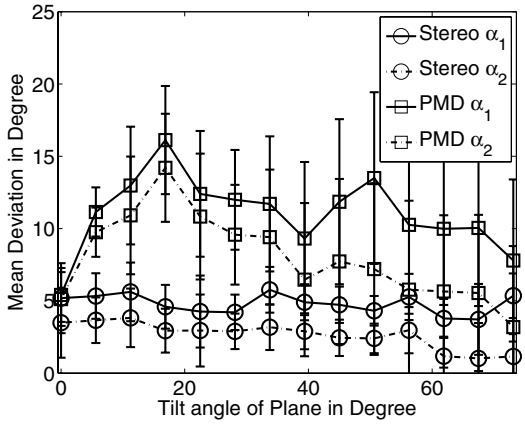

Figure 10. The standard deviation of the angular uncertainties α_1 and α_2 for the stereo system and the PMD system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 45°.

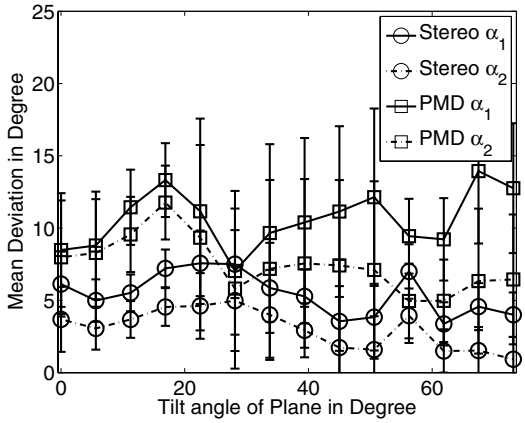

Figure 11. The standard deviation of the angular uncertainties α_1 and α_2 deviations for the stereo system and the PMD system plotted against the plane tilt angle for the experiment corresponding to the roll angle of 90°.

can be achieved. Hence, for the purpose of dense surface reconstruction a fusion of both systems would be desirable, as both systems are complemetary in terms of resolution and accuracy. Also, the PMD camera suffers from phase wrap of distant objects, and from saturation bias for nearby objects. Since stereo performs very well for nearby objects and has no distance ambiguity, these properties can also be combined in a joint setup.

Finally, an interesting aspect of a combined stereo-PMD-system is the potential for benchmarking vision based surface reconstructions. Because the distance measurement accuracy has been found to be so much better for the PMD system, it directly delivers a set of control points, which can be used for benchmarking any vision based algorithm running on the stereo system in a real environment.

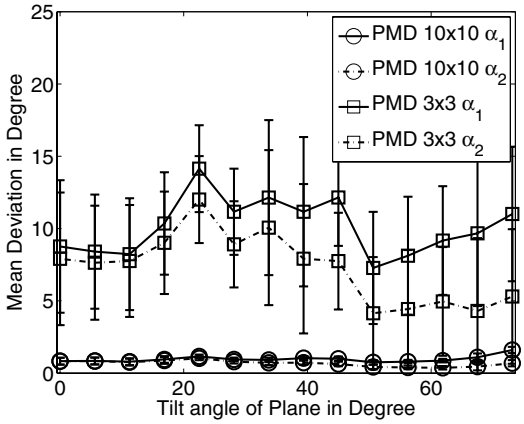

Figure 12. The standard deviation of the angular uncertainties α_1 and α_2 for the PMD system using two different window sizes (3×3 and 10×10) plotted against the plane tilt angle for the experiment corresponding to the roll angle of 0°.

Acknowledgements

The PMD camera used in the experiments is courtesy of Alexander Prusak and Hubert Roth, University of Siegen, Germany.

This work was supported by the German Research Foundation (DFG), KO-2044/3-1.

References

[1] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework: Part, 2002.

[2] W. Förstner and B. Wrobel. Mathematical concepts in photogrammetry. In J.C.McGlone, E.M.Mikhail, and J.Bethel, editors, *Manual of Photogrammetry*, pages 15–180. ASPRS, 2004.

[3] S. Gokturk and C. Tomasi. 3d head tracking based on recognition and interpolation using a time-of-flight depth sensor. In *Proc. CVPR*, pages 211–217, 2004.

[4] D. Grest and R. Koch. Single view motion tracking by depth and silhouette information. In *Scandinavian Conference on Image Analysis (SCIA07)*, 2007.

[5] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[6] H. Hoppe, T. DeRose, T. Duchamp, M. Halstead, H. Jin, J. McDonald, J. Schweitzer, and W. Stuetzle. Piecewise smooth surface reconstruction. *Computer Graphics*, 28(Annual Conference Series):295–302, 1994.

[7] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration for increased accuracy of the range imaging camera swissrangertm. In *IEVM06*, 2006.

[8] K. Kuhnert and M. Stommel. Fusion of stereo-camera and pmd-camera data for real-time suited precise 3d environment reconstruction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2006.

[9] R. Lange, P. Seitz, A. Biber, and R. Schwarte. Time-of-flight range imaging with a custom solid state image sensor. *Proc. SPIE Vol. 3823, p. 180-191, Laser Metrology and Inspection, Hans J. Tiziani; Pramod K. Rastogi; Eds.*, 1999(Tiziani, H. J. and Rastogi, P. K.), sep.

[10] M. Lindner and A. Kolb. Lateral and depth calibration of pmd-distance sensors. In *International Symposium on Visual Computing (ISVC06)*, volume 2, pages 524–533. Springer, 2006.

[11] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81*, pages 674–679, 1981.

[12] N. D. Molton, A. J. Davison, and I. D. Reid. Locally planar patch features for real-time structure from motion. In *Proc. British Machine Vision Conference*. BMVC, Sept. 2004.

[13] D. Murray and J. J. Little. Segmenting correlation stereo range images using surface elements. In *3DPVT '04: Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium on (3DPVT'04)*, pages 656–663, Washington, DC, USA, 2004. IEEE Computer Society.

[14] D. R. Murray. *Patchlets: a method of interpreting correlation stereo 3D data*. PhD thesis, The Univeristy of British Columbia, Vancouver, Canada, 2004.

[15] T. Pietzsch and A. Grossmann. A method of estimating oriented surface elements from stereo images. In *Proc. British Machine Vision Conference*, 2005.

[16] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proceedings of the International Conference on Computer Vision*, pages 754–760, 1998.

[17] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, 2001.

[18] B. Streckel, B. Bartczak, R. Koch, and A. Kolb. Supporting structure from motion with a 3d-range-camera. In *Scandinavian Conference on Image Analysis (SCIA07)*, June 2007.

[19] R. Szeliski and D. Tonnesen. Surface modeling with oriented particle systems. *Computer Graphics*, 26(2):185–194, 1992.

[20] B. Triggs. Detecting keypoints with stable position, orientation, and scale under illumination changes. In *Proceedings of European Conference on Computer Vision*, pages 100–113, 2004.

[21] 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.

[22] Z. Zhang. *Untersuchung und Charakterisierung von Photomischdetektor-Strukturen und ihren Grundschaltungen*. PhD thesis, Department of Electrical Engineering And Computer Science at Univeristy of Siegen, December 2003.