Combining Time-Of-Flight depth and stereo images without accurate extrinsic calibration

Uwe Hahne* and Marc Alexa

Computer Graphics Group,
Faculty of Electrical Engineering and Computer Science,
Technische Universität Berlin, Germany
E-mail: hahne@cs.tu-berlin.de
E-mail: marc@cs.tu-berlin.de
*Corresponding author

Abstract: We combine a low resolution Time-Of-Flight (TOF) depth image camera based on Photonic Mixer Devices with two standard cameras in a stereo configuration. We show that this approach is useful even without accurate calibration. In a graph cut approach, we use depth information from the low resolution TOF camera to initialise the domain, and colour information for accurate depth discontinuities in the high resolution depth image. The system is promising as it is low cost, and naturally extends to the setting of dynamic scenes, providing high frame rates.

Keywords: depth imaging; Photonic Mixer Device; PMD; sensor fusion; stereo; Time-Of-Flight; TOF.

Reference to this paper should be made as follows: Hahne, U. and Alexa, M. (2008) 'Combining Time-Of-Flight depth and stereo images without accurate extrinsic calibration', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 5, Nos. 3/4, pp.325–333.

Biographical notes: Uwe Hahne is a PhD Student and Teaching Assistant at Technical University Berlin, Germany. He received his diploma (MS) in Media Systems at Bauhaus University in Weimar, Germany. His research interests vary from stereo vision over mesh editing, display technologies, digital holography, augmented reality to fusion of 3D imaging technologies. His teaching interests include computer graphics, computer vision and functional programming as well as projects about 3D reconstruction and multi-touch displays.

Marc Alexa is a Professor in the Faculty of Electrical Engineering and Computer Science at the Technical University of Berlin and heads the Computer Graphics group. He is primarily interested in representing and modelling shapes, using point sampled geometry, implicit surfaces and explicit representations. For his earlier work on morphing he received a PhD in Computer Science from Darmstadt University of Technology. He has presented and lectured on topics related to shape modelling at SIGGRAPH and other conferences. He has been a Co-chair and has served as a member of several committees of major graphics conferences.

1 Introduction

Reliable and fast depth imaging for real world scenes is a difficult problem. A variety of methods are available for the task, however, each of them has its strengths and weaknesses. In particular, passive methods rely on feature matching, which is time-consuming and might fail if no features are present. Active techniques overcome this problem, yet at the price of being sensitive to the reflection properties of the elements in the scene and typically higher cost of the devices.

Recently, a new type of Time-Of-Flight (TOF) sensor has extended this spectrum: Photonic Mixer Devices (PMD) (Moeller et al., 2005; Xu, 2005) can be used to measure the phase of a modulated light source relative to its reflection in the scene. The phase directly relates to the distance. These devices are low-cost and very fast, but their spatial resolution is low and they suffer from noise, especially around depth discontinuities.

It is interesting to compare this new approach to the dominating low cost depth approach, namely stereo. For planar patches, Beder, Bartczak and Koch (2007) find that the PMD approach is more accurate than stereo. On the other hand, details and discontinuities in intensity and/or depth decrease the performance of PMD depth measurement, while they typically increase the performance of stereo. Also, traditional cameras have a much higher resolution and stereo setups have a larger working range. A natural conclusion is to combine the PMD approach with standard stereo photography.

Few approaches consider the fusion of PMD generated depth (and possibly intensity) images with standard photography. Reulke (2006) combines a single high resolution intensity image with the PMD depth image. Intensity information is exploited to steer the re-sampling of depth data into a higher resolution depth image. Kuhnert and Stommel (2006) propose combining the PMD approach with an additional pair of stereo cameras—this is identical to our set-up. They take the depth image of the PMD camera and a depth reconstruction from the stereo pair as independent measures of the depth in the scene. For pixel with low confidence in the computed disparities (i.e. no significant matching for the windows) they fill in the data gathered with the PMD camera.

We are using a similar approach of combining PMD and stereo depth. The stereo pair is mounted symmetrically to the PMD camera, with the three centres of projection (roughly) co-linear and parallel image planes (see Section 2 for details). This yields three images, one depth image and two (colour) intensity images. The three images are calibrated based on their intensity images using standard approaches (Heikkilae and Silven, 1997; Zhang, 1999, 2000). Because of the low resolution and the fact that the PMD sensor rather measures intensity differences than absolute values, this calibration turns out to be not very accurate.

However, even the data based on the inaccurate calibration can be fused. We are using similarity of intensity values (over windows) from the stereo pair together with the phase information from the PMD camera, which is new compared with prior approaches. While this idea could be used in any stereo setting, we are here using a global approach based on graph cuts (Kolmogorov, Zabih and Gortler, 2003; Paris, Sillion and Quan, 2004, 2006; Hornung and Kobbelt, 2006; Tran and Davis, 2006; Yu, Ahuja and Chen, 2006) for the reconstruction of a depth image. Typically, these approaches are computationally demanding, yet in our setting we can exploit the PMD generated depth image for restricting the domain of the volumetric grid. Section 3 explains how we use graph cut in our setting.

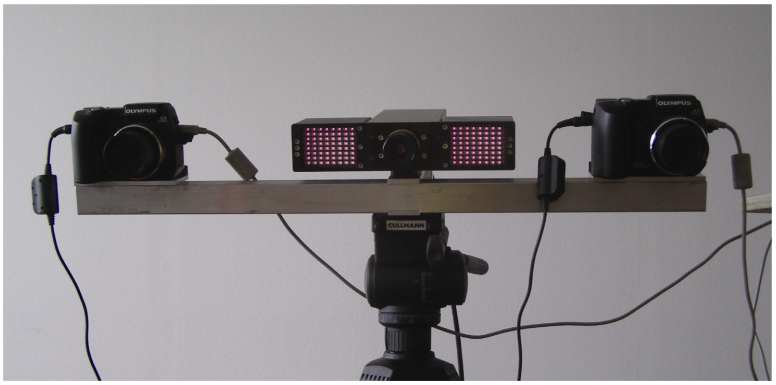
The results of this procedure allow increasing the resolution of the PMD depth data, while keeping sharp depth discontinuities based on intensity discontinuities.

An important feature of the set-up is that we could record a video sequence, and then recompute the depth on a per frame basis. This yields a system with dramatically improved depth reconstruction for dynamic scenes.

2 Set-up

We are using a PMD camera, type 19k, with a resolution of 160 × 120 pixels. It is centred between two standard photo cameras. As an experiment, we have used consumer grade cameras of type Olympus SP-500 UZ. These cameras could be interchanged by standard firewire cameras in order to capture a video sequence for reconstructing dynamic scenes. All three cameras are mounted on an aluminium bar (see Figure 1 for an image of the cameras). The stereo pair has a base line of roughly 50 cm. The cases are mounted so that their image planes are parallel.

Figure 1 All three cameras mounted on an aluminium bar



For calibration of this set-up, we need to pay special attention to the specifics of the PMD camera. It turns out that intrinsic calibration is not very accurate and extrinsic calibration in the classical sense would not be sufficient: extrinsic calibration usually tries to align the optical systems of the cameras, but it is unclear if zero phase shift in the PMD sensor would exactly match the optical centre.

2.1 Intrinsic calibration

Due to manufacturing mechanics, the intrinsic parametres of a consumer camera given by the producer are very inaccurate. Therefore, calibration is necessary. We use the algorithm of Zhang (1999, 2000) for the standard cameras.

The intrinsic calibration of the PMD camera is more difficult, mostly because of its sensitivity to reflections of the self emitted IR light and low resolution. Another problem is that the PMDTec 19k camera only measures differences in capacities. Hence, there is no true intensity image available. The intensity image can be simulated by weighting the amplitudes according to their squared distances. In order to recognise feature points

(e.g. corners of a chessboard) automatically, the calibration sheet has to be placed quite directly in front of the camera. Apart from these details, we follow the ideas of Reulke (2006) as well as Lindner and Kolb (2006).

Note that the intrinsic calibration of the PMD camera already yields Euclidean coordinates for each pixel: the coordinates of the pixel identify a ray by means of the intrinsic transformation. The pixel contains a distance value, which identifies a point on the ray. Thus, the intrinsic transformation allows computing a set of points in $\mathbb{R}^3$ from the depth image.

2.2 Extrinsic calibration

Our main idea is to use well-known techniques to perform an extrinsic calibration for all three cameras. We decided to use the OpenCV calibration methods as well as the Camera Calibration Toolbox for MATLAB® that are based upon the same composition of algorithms (Heikkilae and Silven, 1997; Zhang, 1999) and are written by Jean-Yves Bouguet. We have found, however, that the low resolution and relatively high noise level lead to very inaccurate and alternating calibration results. In addition, the plane of values with zero phase shift, which defines the plane with zero depth, is not necessarily coinciding with the image plane of the camera. We model this effect based on experiments by moving the camera plane along the depth axis so that depth values obtained with the camera coincide with stereo depth values.

2.3 Imaging

For taking the intensity/colour images, we use the application programmers interface provided by Olympus Corporation (2004). This allows setting similar parameters for both cameras.

In order to get a representative depth image, we take on the order of 20 images with the PMD camera. We have observed that the data values are sometimes instable. Working with the mean or the median image from several images not only significantly reduces noise in the depth image of the PMD camera, but also it provides useful information on the confidence in the depth values: we use the variation of the depth values around the median value as a measure of the confidence. In particular, we use the variance of each depth value as a weight for taking into account the depth values of the PMD camera as well as for defining the domain for the graph cut algorithm. This is explained in detail in Section 3.

3 Algorithm

In the following, we describe the details of the depth computation. We follow the graph cut reconstruction by Paris, Sillion and Quan (2006). Our goal is to reconstruct the surface S in the object space (x, y, z) defined parametrically by the function f: $S(u, v, f(u, v))$, which can be interpreted as a depth function $z = f(x, y)$. The surface S minimises the functional

$$\int\int\left(c(S) + \left(\alpha_u(u,v)\frac{\partial S}{\partial u} + \alpha_v(u,v)\frac{\partial S}{\partial v} \right) \mathrm{d}u\mathrm{d}v \right) \tag{1}$$

containing a consistency term c and two smoothing terms α_u and α_v. In the classic stereo graph cut, c is defined by stereo correspondences and, α_u and α_v by discontinuities in x and y direction in both images. The data obtained with the PMD camera allows us to refine the definition of c and α – we will describe our enhanced definition in Section 3.2.

3.1 Defining the volume

The graph cut approach works on a discrete volume. The number of voxels in x and y direction determine the resolution of the resulting depth map. The number of voxels in z direction defines the quantisation of depth values. In our experiments, we have used grids of dimensions 400 × 300 × 100 in width, height and depth.

For all of these voxels, we build a mapping function M: $N^3 \rightarrow R^3$ that maps a voxel id to its 3D position in the PMD camera coordinate system (x, y, z). For this mapping, we use the intrinsic values of the PMD camera to find the x and y positions of the voxel (see previous section). All depth steps, from the minimal depth value to the maximum depth of the PMD camera image, are filled with corresponding z values. We receive the minimal and maximal depth in the scene from the set of median depth values. In order to reduce the computational costs, a Domain Of Interest (DOI) is defined inside this volume. The DOI is determined as follows: First, we find the voxel which is closest to the PMD depth values. Hence, the DOI contains one voxel in each depth column. If this depth value has a large variance, voxels before and after (in z-direction) are added to the domain, as well. Then, these voxels are connected in x and y direction. Therefore, we add voxels from the neighbouring depth columns until we reach a connected set of voxels. Thus, the DOI can be interpreted as the voxelisation of the variance controlled blurred PMD image. This is a significant reduction of computation cost for stereo graph cut algorithms, which usually have to start from the overlap of the viewing frustums of the left and the right camera.

3.2 Constructing a graph

After defining the volume and the DOI, a graph is constructed according to Paris, Sillion and Quan (2006). This graph is connected to a source node which is placed in front of our object space (minimal depth) and a sink node behind the object at maximal depth. Basically, graph cut algorithms aim at setting up the capacities in such a way that the minimal cut describes the demanded surface. The graph contains two types of edges: consistency edges and smoothing edges. The consistency edges are inside one voxel and their capacity is determined by the measure of probability that this voxel belongs to the surface. We compute this measure as a linear combination between commonly used stereo consistency terms, here denoted as c_{stereo}, and a consistency value c_{pmd} computed from the difference of the depths of the voxel and the corresponding PMD depth. The stereo consistency is either derived from the normalised cross correlation or sum of squared distances between the corresponding regions in the left and right image. The depth difference d can then be mapped to a consistency value c_{pmd} using a convex or concave function. However, the convex term $c_{\text{pmd}} = 1/d^2 + 1$ and the concave term $c_{\text{pmd}} = \max[0, 1-d^2]$ were indistinguishable in our experiments.

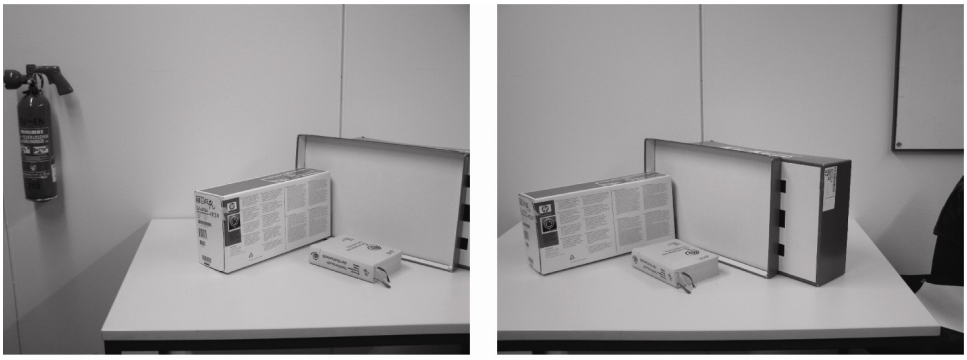
Computing the smoothing edges capacities is quite similar. The stereo smoothing value α_{stereo} depends upon colour discontinuities in both stereo images. For the PMD smoothing term α_{pmd}, we use depth discontinuities in the PMD median image. The terms are combined linearly, as well.

Based on the topology of the graph and edge weights, the minimal cut computed with standard graph cut algorithms determines the surface. As with other depth reconstruction approaches based on graph cut, this is the dominating factor in the computation time, and in our set-up it requires a few minutes at typical input.

4 Results

Figure 2 shows our stereo input images. We chose this arrangement of cardboard boxes in order to clearly see the gradient of depth in our results. Supplementary, the untextured cardboard is a typical case where classic stereo algorithm may fail, because there are problems in identifying disparities on large uniform coloured areas in the stereo image pairs.

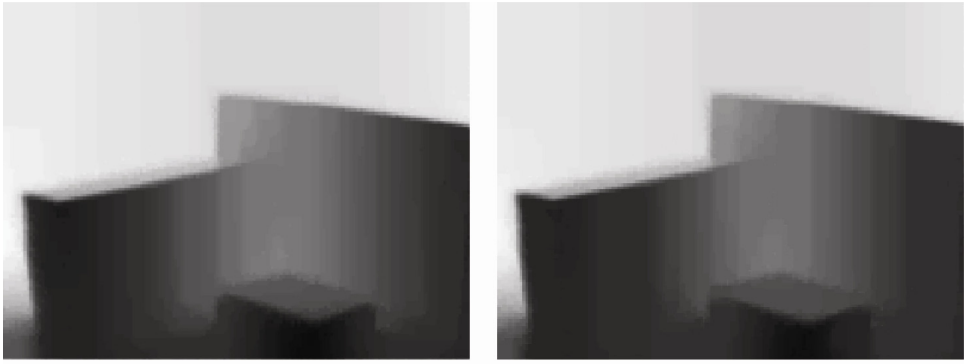
Figure 2 The stereo pair



(a) Left stereo image (b) Right stereo image

As mentioned in Section 3.2, the reconstruction of the surface is performed by finding a minimal cut for the constructed graph. This cut minimises the functional given in Equation (1). As expected, computation times are greatly reduced by using a tighter DOI. In addition, if we extend the DOI by a dilation, the results in Figure 3 show that using this larger DOI does not enhance the result. However, it takes much more time to compute.

Figure 3 Increasing the DOI leads to blurred results and longer computation time

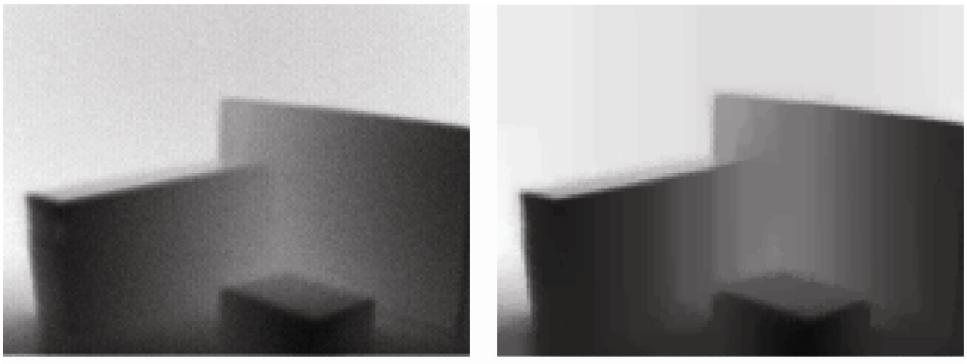


(a) normal DOI (b) dilated DOI

In addition to the computational savings, we were interested in whether the accuracy of the reconstructed surface also improves on using PMD depth data or stereo reconstruction alone. The characteristics of the PMD depth data would typically lead to either noisy or overly smoothed results, especially when re-sampled to higher resolution: note that each depth element is measured independently of its neighbours, resulting in noise with salt and pepper distribution. Removing this noise results in smoothing. In addition, increasing the resolution of the image also results in smoothed edges.

The stereo information and the smoothing term used in the graph cut algorithm greatly reduce this noise (see Figure 4). In all flat surface areas inside the scene, the grainy artefacts are removed. Depth estimation is enhanced as well: in contrast to the PMD median image, the cardboard boxes are reconstructed without distortion. Their depth increases smoothly along the surface and stays constant on the vertical axis. Also, depth discontinuities are more accurately modelled by exploiting the colour discontinuities in the higher resolution intensity images. This improvement can be recognised clearly. In the PMD image at the borders of the book lying in front of the cardboard boxes, we can see a slightly brighter area at the depth discontinuities – generated from inaccurate depth values, while in our resulting graph cut image, these effects are removed.

Figure 4 The salt and pepper noise is reduced and the depth information is enhanced



(a) PMD median image (b) Graph cut smoothing

5 Discussion

We have shown that the fusion of PMD data with stereo images enhances depth reconstruction in low-cost sensor systems. Our experiments are based on only roughly calibrated systems, and nevertheless, we have found the results to be better than with either system alone.

In the future, we want to calibrate the system accurately. This is more complicated than for usual camera based systems, as the sensing technology in the PMD camera appears to be not accurately modelled with a perspective transformation alone. Furthermore, the calibration of consumer grade zoom cameras has to be explored carefully, as well. An accurate calibration would allow us to exploit the two types of information for each voxel in a more systematic way. In addition, an accurate calibration will be necessary, if we want to start an exact evaluation about the accuracy of our system.

Using the video cameras for the stereo pair results in a system capable of recording depth data at high frame rates and for very low cost. The characteristics of such a system would make it very attractive for a variety of applications, perhaps most prominently 3DTV.

References

Beder, C., Bartczak, B. and Koch, R. (2007) 'A comparison of PMD-cameras and stereo-vision for the task of surface reconstruction using patchlets', Paper presented in the Proceedings of the *Second International ISPRS Workshop BenCOS*.

Heikkilae, J. and Silven, O. (1997) 'A four-step camera calibration procedure with implicit image correction', Paper presented in the Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)* (pp.1106–1112). San Juan, Puerto Rico.

Hornung, A. and Kobbelt, L. (2006) 'Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding', Paper presented in the Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2006)* (Vol. 1, pp.503–510).

Kolmogorov, V., Zabih, R. and Gortler, S.J. (2003) 'Generalized multi-camera scene reconstruction using graph cuts', Paper presented in the Proceedings of the *Fourth International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, July .

Kuhnert, K-D. and Stommel, M. (2006) 'Fusion of stereo-camera and PMD-camera data for real-time suited precise 3d environment reconstruction', Paper presented in the Proceedings of the *IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp.4780–4785), October.

Lindner, M. and Kolb, A. (2006) 'Lateral and depth calibration of PMD-distance sensors', *International Symposium on Visual Computing* Vol. 2, pp.524–533.

Moeller, T., Kraft, H., Frey, J., Albrecht, M. and Lange, R. (2005) 'Robust 3d measurement with PMD sensors', *Technical Report, PMDTec*.

Olympus Corporation (2004) *Olympus Camedia SDK* 3.4.

Paris, S., Sillion, F. and Quan, L. (2004) 'A surface reconstruction method using global graph cut optimization', *Asian Conference of Computer Vision*, January.

Paris, S., Sillion, F.X. and Quan, L. (2006) 'A surface reconstruction method using global graph cut optimization', *Int. J. Computer Vision*, Vol. 66, pp.141–161.

Reulke, R. (2006) 'Combination of distance data with high resolution images', *ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*.

Tran, S. and Davis, L. (2006) '3D surface reconstruction using graph cuts with surface constraints', *Computer Vision ECCV 2006, Volume 3952/2006 of Lecture Notes in Computer Science*. Berlin, Germany; Heidelberg, Germany: Springer.

Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B. and Ringbeck, T. (2005) 'Smart pixel – photonic mixer device (PMD) new system concept of a 3d-imaging camera-on-a-chip', *Technical report, PMDTec*.

Yu, T., Ahuja, N. and Chen, W-C. (2006) 'SDG cut: 3D reconstruction of non-lambertian objects using graph cuts on surface distance grid', Paper presented in the Proceedings of the *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp.2269–2276). Washington, DC: IEEE Computer Society.

Zhang, Z. (1999) 'A flexible new technique for camera calibration', *Technical report, Microsoft Research*.

Zhang, Z. (2000) 'A flexible new technique for camera calibration', *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 22, pp.1330–1334.