IMPROVED IMAGE SEGMENTATION USING PHOTONIC MIXER DEVICES

Frank Wallhoff, Martin Ruß[*]*, Gerhard Rigoll*

Technische Universität München
Human-Machine Communication
Theresienstr. 90, 80333 Munich
{wallhoff,rum,rigoll}@mmk.ei.tum.de

Johann Göbel, Hermann Diehl

EADS Corporate Research Center Germany
Depearment LG-ME
81663 München
{johann.goebel,hermann.diehl}@eads.net

ABSTRACT

Aiming at improving image segmentation and extending regular computer vision algorithms an image sensor, acquiring additional depth information is deployed. The novel Photonic Mixer Device (PMD) technology will be summarized, by which it becomes possible to measure the observed object's distance to the camera. Motivated by expanding and making existing image processing and segmentation algorithms more robust, the additional depth information is overlapped with the image map gained by a regular camera system. Due to the fact of the used cameras are displaced and have different image resolutions, a sophisticated calibration algorithm will be introduced. By applying improved algorithms to applications, such as people counting, gesture recognition, and barrier detection its effectiveness will be shown.

***Index Terms*—** Computer Vision, 3D Surveillance, Photonic Mixer Device, Segmentation, Gesture Recognition, People Counting, Barrier Detection

1. INTRODUCTION

Especially for object detection and image segmentation tasks, 3D information can aid numerous applications in the field of computer vision, i.e. surveillance tasks, face detection and many more. However, image processing has consolidated itself as a very important sector within the signal processing domain. Herein many elaborated and reliable algorithms have already been introduced. Most of them represent the real 3D environment by 2D images. Although several hard- and software approaches exist to also capture 3D information, it has turned out that most of them are either computationally too expensive, impractical due to their hardware constraints or both [1]. Furthermore most of them are lacking real-time capabilities, i.e. the acquisition of approx. 25 frames per second.

As a consequence, we have concentrated on a hybrid way to incorporate the good results from well-known 2D algorithms with a rather coarse but fast 3D representation of the observed scenery in a sophisticated way. To achieve this goal the outputs from an arbitrary image sensor and a spatially low resolution pixel range scanner are combined and interpreted. Since the herein employed Photonic Mixer Device (PMD) is capable to scan images with a frame rate of up to 25 Hertz by working in an autarkic manner, it can be foreseen and integrated even into applications with hard real-time constraints. Together with additional depth information an image segmentation task can become trivial, for example to isolate an image's foreground from its background.

To demonstrate the functionality of the presented approach, the rest of the paper is outlined as follows: after a brief theoretical introduction of the obeyed PMD technology, the setup for acquiring the desired image pairs and a calibration procedure are presented. The effectiveness of the presented approach is demonstrated on three applications, a head counting algorithm, a gesture recognizer and a barrier detection system. The paper closes with some conclusions and an outlook.

2. DEPTH INFORMATION ACQUISITION PRINCIPLE

The image and depth information acquisition principle of the Photonic Mixer Device (PMD) is basing on the run-time difference of a light impulse directly send to the detector and the reflected light from the surface of objects in the environment. In Figure 1 the simplified so-called time-of-flight measurement principle for smart pixel is shown.

With utmost precise counters, emitters and receivers the distance between the camera pixel and the object can be approximated by $d = \frac{t}{2} \cdot C$, where t represents the measured turnaround time between the start of a light impulse and its return to the receiver. The variable C represents the speed of light. The measurement of the flight-time is carried out using the phase shift of modulated infrared light pulses [2]. By combining several smart pixels in a two dimensional structure an image sensor with fully parallel operating cells arises, allowing the 3D surface reconstruction of the scene. Since this measurement paradigm is directly implemented in the detector's hardware there is no additional computational effort, such as that arising from stereo cameras [3]. The refresh rate for one measurement loop allows between 5 and 50 frames/second.

To overcome the problem of background illumination, which superposes the running pulse, various further techniques, such as optical filters and active circuits are implemented on the chip's cite. The sensor's usage of the suppression of background illumination makes it even possible to suppress the effects of bright ambient light [4] thus this measurement becomes independent from existing lighting conditions. The emitted infrared light has a wavelength of 870nm. By integrating the received light impulses over a certain interval the PMD camera could further serve as a NIR infrared camera.

However, there are still some drawbacks that the employed measurement technique is suffering from. Range measurement problems occur in conjunction with highly reflective surfaces that are too close to the sensor. By the mirroring effect of the infrared diodes on the material's surface pixel distances become too large. On light adsorbent materials the depth values are very noisy.

[*] now with: Autonomous Systems Technology, Bundeswehr University

1-4244-1437-7/07/$20.00 ©2007 IEEE

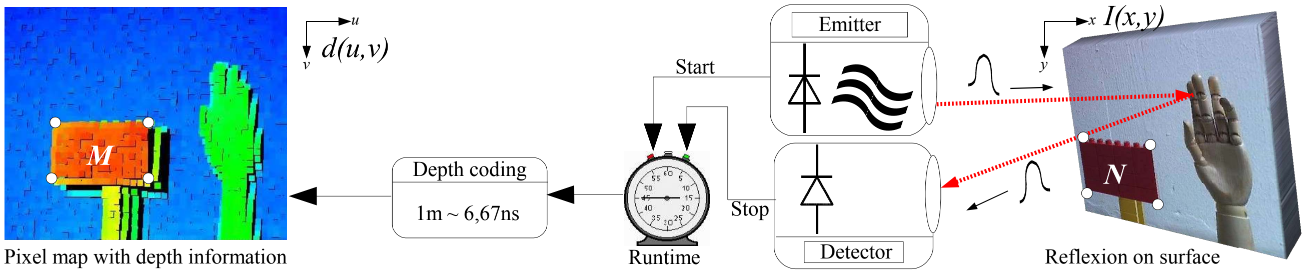

Fig. 1. Time-of-flight measurement principle and calibration body from both sensors' perspectives.

3. EXPERIMENTAL SETUP

To enable image overlaying algorithms for the introduced combination of depth maps with high resolution images, two image sensors are equipped with the same Cosmicar/Pentax lens having a focal length of 16mm resulting in a field-of-view of $\approx 40°$. For all further experiments the same hardware components and test conditions are satisfied. First a PMD[vision] model 3k-s 3D video range sensor [5] with 64×48 pixels running at 25 frames per second is deployed. The camera resolution in z-direction is specified by $z >$ 6mm. Secondly a 3CCD Sony XC003P with a resolution of 752×582 (PAL) in full frame mode is digitized using a Cinergy USB 200 capture device.

By mounting the CCD image sensor piggyback on the PMD device as shown in Figure 2 a parallel offset of the camera axis arises. Aiming at having block-wise overlapping results the camera setup as well as intrinsic camera parameters have to be calibrated.

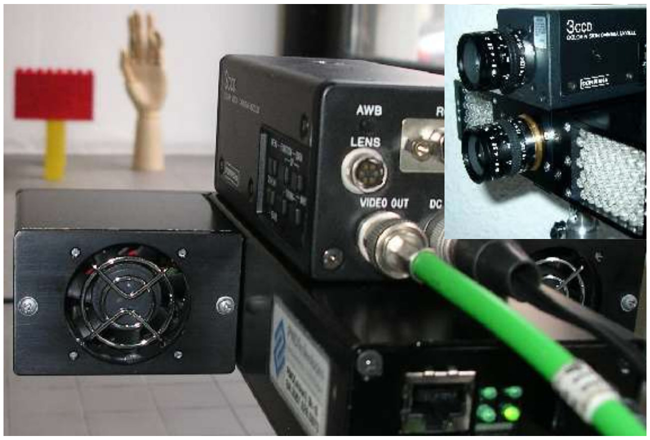

Fig. 2. Piggyback assembly of PMD and color CCD.

4. CAMERA CALIBRATION AND OVERLAY

As introduced, the fundamental idea bases on overlapping a higher resolution image with a coarse depth map. Thus both images can be carried over to each other by a coordinate translation followed by an expansion as shown in Figure 1. The variable $I(x, y)$ denotes a RGB color triple of the CCD camera in the x/y coordinate system, $d(u, v)$ is the distance matrix of the PMD in the displaced and scaled u/v system measured in meters.

The scaling parameters are denoted by s_x and s_y, the adjustment by a_x and a_y. Ideally the scaling is given by the fraction of both resolutions and a_x should be zero. However, due to the coarse resolution of the PMD and other external impacts all parameters have to be estimated automatically by finding characteristic landmarks in both images. In principle the presented calibration object may have an arbitrary shape and color as long as it can be distinguished from the scenery's background, here the nearest closed shape with distance $d_{\text{min}} = min(d(u, v))$. In our examples the calibration corpus consists of a LEGO model with a flat red facing and an depth of 15cm. Its surface map is denoted with M.

$$M(u, v) = \begin{cases} 1 & \text{if } d(u, v) \le d_{\text{min}} + 15\text{cm} \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

From M four calibration landmarks are derived by a bounding box surrounding this shape. In the high resolution image the corresponding four landmarks are found by a fitting box around the red object N, where its HSV triple falls into a pre-defined area:

$$N(x, y) = \begin{cases} 1 & \text{if } ((H \ge 319)|(H \le 31))\&(S \ge 5)\&(S \le 16)\& \\ & (V \ge 0)\&(V \le 16) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

The scale factors become:

$$s_x = \frac{max_x(N) - min_x(N)}{max_u(M) - min_u(M)} \text{ and } s_y = \frac{max_y(N) - min_y(N)}{max_v(M) - min_v(M)} \qquad (3)$$

After expanding $d(u, v)$ by s_x and s_y, a equally scaled binary map $D(x, y)$ gained by bicubic interpolation arises. This has to be displaced by a_x and a_y so that the color and depth box around the calibration body overlap. Thus an area results where the color image and the depth map are defined simultaneously.

$$a_x = min_x(N) - min_x(D) \text{ and } a_y = min_y(N) - min_y(D) \quad (4)$$

5. APPLICATIONS

5.1. Person Counting

The first approach demonstrates how the personnel flow through a door can be measured using a segmentation gained by depth information. Through the derivation of shape information it becomes possible to segment a head and distinguish a person passing the scene from other items. Due to the PMD's measurement principle there are no restrictions regarding external lighting conditions.

Human heads are assumed to bear some resemblance to egg shaped structures when seen from above. A typical image recorded with the PMD above the door is depicted in Figure 3.

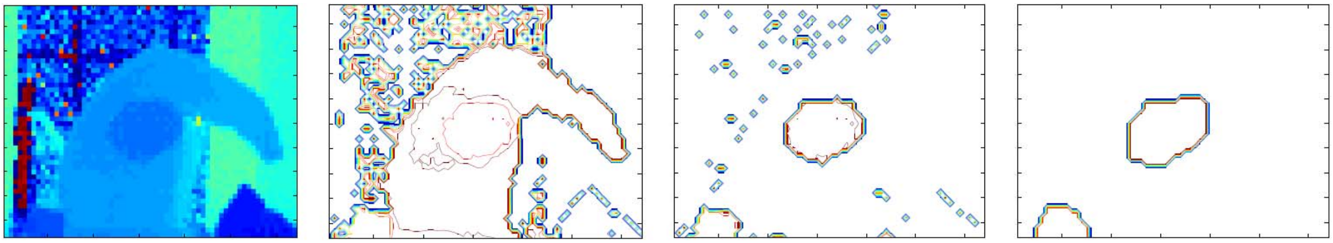

Fig. 3. Person counting (from left to right): original color coded top-down depth map, isolated depth information of one slice, preprocessed shape matrix and final edge map.

By cutting the observed depth map (Fig. 3, left) into slices with a height of $50cm$ a head inside this subspace will appear as a connected 3D object (Fig. 3, right).

After a connected component and thresholding operation the slice image can be transferred to a flat area (Fig. 3, lower left). Thus the outer curve or edge function of a head will appear as an ellipsis, which can be detected by a classical correlation with predefined head patterns or invariant moments [6]. To enable a robust detection of heads from persons with different body heights, this sampling procedure starts at the top of the door ($\approx 2m$) and ends at the height of a children ($\approx 0.8m$). In order to seek for heads in all heights, several overlapping slices with a step width of $25cm$, the typical skull height, are processed. Several overlapping head detection results are clustered by their mean values. A person can finally be counted if its trajectory can be tracked within a pre-defined region-of-interest, for example using a particle filter [7].

5.2. Gesture Recognition

The second use case is an action and gesture recognition system. Originally it was constituted on difference image based seven dimensional global motion feature vectors: the mass-center, the deviation and its change (each in x- and y direction) as well as the intensity of the motion [8]. An unknown feature sequence is dynamically classified by Hidden Markov Models. Self occlusions, changing lighting conditions and compression artefacts cause a high noise level within the feature stream, which could not be significantly removed even with a Kalman filter [9].

However, invariant moments have shown good classification results if the object to be classified can be isolated [10].Hence the region of interest in the foreground, i.e. the hand, is segmented by depth information. Due to the calibration, this mask can be overlapped with the CCD image pixel-wise, as shown in Figure 4. The hand position can further be extracted using skin color filters [11]. The advantage of the additional depth shape becomes obvious, since there are no interferences even with skin colored objects.

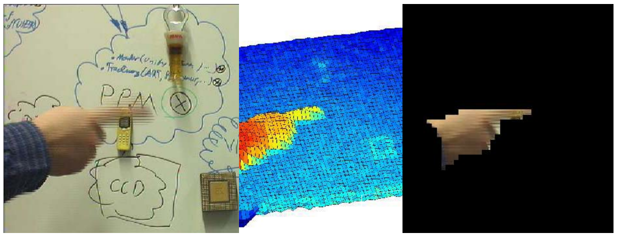

Fig. 4. Pointing gesture: CCD image (left), depth map (middle), and isolated hand (right).

5.3. Barrier detection

A barrier detection on color or greyscale images is usually mainly based on edge detection, as shown in Figure 5. But especially with a structured background it becomes hard or even impossible to decide, which edges belong to a barrier.

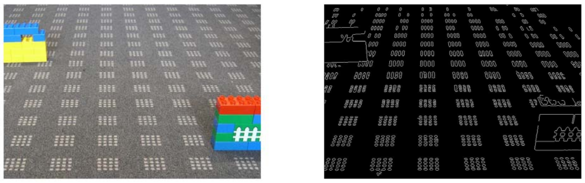

Fig. 5. CCD image (left), Canny edge detector

In a distance map, as shown for in the left two images in Figure 6 it is much easier to isolate an object. In order to strengthen the robustness and to reduce the computational effort, we concentrate on a certain region-of-interest (ROI), which makes it very promising for in-vehicle applications. Far pixels beyond the ROI, e.g. $d_{\mathrm{ROI}} \in [10; 75]$ cm, are floored to a corresponding threshold.

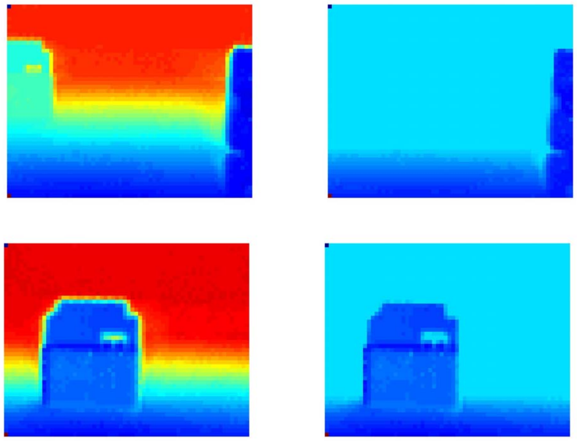

Fig. 6. Distance maps: two obstacle images (upper and lower left), fore side ROI (upper and lower right)

To be aware of vanishing and nearing obstacles, a difference map of two subsequent ROIs is build and median filtered, as depicted in

Figure 7, upper right. Thus it even becomes possible to decide if a obstacle is comes nearer or disappears. If a obstacle is approaching, its distance will be smaller. This means that the difference will be positive. On the other hand the vanishing barriers are represented by negative values.

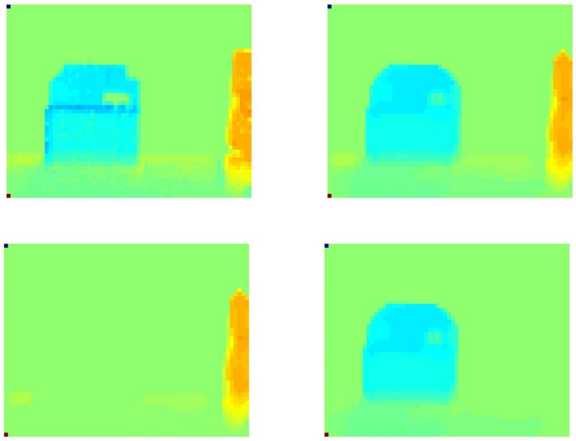

Fig. 7. Difference maps: two subsequent ROIs (upper left), median filtered (upper right), vanishing (lower left) and nearing obstacles (lower right)

The canny edge detector on the resulting image maps in Figure 7 (lower row) provides a very good segmentation of the barrier. The analysis of the edge maps in the left column of Figure 7 is based on finding the absolute and relative maximum in a histogram for each direction, e.g. vertical an horizontal.

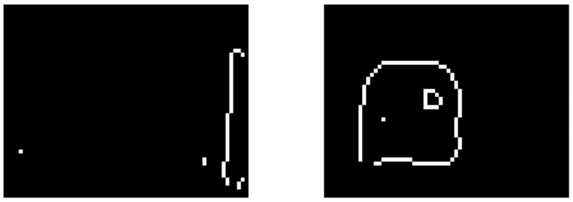

Fig. 8. Canny edge detector used for segmentation: vanishing (left) and nearing obstacle (right)

In addition to the detection if an object is approaching, a subsequent shape based classification of the obstacle can be used to detect pedestrians or cyclists.

6. SUMMARY AND CONCLUSIONS

Reliable image segmentation approaches have been introduced and integrated with very low computational efforts by making use of depth information from a PMD range sensor.

Existing algorithms were extended efficiently enabling possibilities to new image processing algorithms.

Three uses cases have been presented, which show very promising qualitative results. Therefore it is planned to measure the quantitative improvements that can be gained by integrating additional depth information into regular approaches in the future. Furthermore it is planned to concentrate on 3D surface reconstruction tasks for augmented reality applications as well as environmental exploration for autonomous navigation systems.

7. ACKNOWLEDGEMENT

The authors would like to express their thanks to Bianca Hagebeuker from PMDTechnologies [12] for her support and the short term provisioning of a sensor device.

8. REFERENCES

[1] Frank Forster, *Real-Time Range Imaging for Human-Machine Interfaces*, Ph.D. thesis, Technische Universität München, Lehrstuhl für Mensch-Maschine-Kommunikation, 2004.

[2] T. Kahlmann, F. Remondino, and H. Ingensand, “Calibration for increased accuracy of the range imaging camera swissranger,” in *Proceedings of the ISPRS Commission V Symposium Image Engineering and Vision Metrology*, D. Schneider Editors: H.-G. Maas, Ed., Dresden, Germany, 25-27 September 2006, vol. XXXVI, pp. 136–141.

[3] Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, and T. Ringbeck, “Smart pixel - photonic mixer device (pmd),” *Proc. M2VIP ’98 - International Conference on Mechatronics and Machine Vision in Practice, Nanjing*, pp. 259–264, 1998.

[4] T. Möller, H. Kraft, J. Frey, M. Albrecht, and R. Lange, “Robust 3d measurement with pmd sensors,” in *In: Proceedings of the 1st Range Imaging Research Day at ETH Zurich, Zurich, Switzerland, pp. ”upplement to the Proceedings”*, 2005.

[5] PMDTechnologies, “Data Sheet PMD(vision) 3k-s,” Online document *http://www.pmdtec.com/inhalt/download/documents/PMDvision 3k-S_000.pdf*.

[6] A. Chalechale, F. Safaei, F. Naghdy, and P. Premaratne, “Hand posture analysis for visual-based human-machine interface,” in *WDIC 2005 APRS Workshop on Digital Image Computing*, In B. Lovell & A. Meader (Eds.), Ed. Queensland: The Australian Pattern Recognition Society, 2005, pp. (pp. CD Rom 91–96).

[7] F. Wallhoff, M. Zobl, G. Rigoll, and I. Potucek, “Face tracking in meeting room scenarios using omnidirectional views,” *Proceedings Intern. Conference on Pattern Recognition (ICPR)*, Aug. 2004.

[8] F. Wallhoff, M. Zobl, and G. Rigoll, “Action segmentation and recognition in meeting room scenarios,” *Proc., IEEE Int. Conf. on Image Processing (ICIP)*, Oct. 2004.

[9] M. Zobl, A. Laika, F. Wallhoff, and G. Rigoll, “Recognition of partly occluded person actions in meeting scenarios,” *Proc., IEEE Int. Conf. on Image Processing (ICIP)*, Oct. 2004.

[10] M. K. Hu, “Visual pattern recognition by moment invariants,” *IRE Trans. Info. Theory*, vol. IT-8, pp. 179–187, 1962.

[11] Moritz Störring, *Computer Vision And Human Skin Colour*, Ph.D. thesis, Faculty of Engineering and Science, Aalborg University, 2004.

[12] PMDTechnologies, “Homepage,” *http://www.pmdtec.com/e_index.htm*, 2007.