Motion Detection and Tracking Using the 3D-camera

Xiang Yin*, Noboru Noguchi**

* Bio-production Engineering. Graduate School of Agriculture, Hokkaido University, Kita-ku, Sapporo, CO 060-8589, Japan (e-mail: yin@bpe.agr.hokudai.ac.jp).
** Bio-production Engineering. Graduate School of Agriculture, Hokkaido University, Kita-ku, Sapporo, CO 060-8589, Japan (e-mail:noguchi@bpe.agr.hokudai.ac.jp)

Abstract: This paper presents how to detect and track the real-time motion using the PMD 3D-camera. Different modalities, including intensity image, range image and amplitude image, can be provided by the 3D-camera. Both the intensity and the range images are used in the motion detection and tracking process. First, intensity and range backgrounds are separately modeled. During the motion detection and tracking, the segmentation results based on intensity and range images are combined, which results in a robust segmentation of the current scene. And such a segmented image has to be filtered in real-time applications because the NIR light emitted by the 3D-camera will be affected by the objects too close to the 3D-camera, resulting in intensity distortion, and low resolution of the intensity image constitutes another reason. After segmentation, edge and contour detection is used to try to figure out the number and positions of mobile objects. By experiments, the process of real-time motion detection and tracking is demonstrated. Finally, some existing problems and proposed solutions such as shadow removal are discussed.

Keywords: background modelling, motion detection and tracking, image segmentation, image filtering.

1. INTRODUCTION

Motion detection and tracking is an important research area of mobile systems, which is widely used in obstacle avoidance (T. Hong et al., 2004), real-time navigation (J. Weingarten et al., 2004; A. Prusak et al., 2008), augmented reality (C. Portales et al., 2010), and human-machine interaction (such as the gesture recognition in Z. Li and R. Jarvis, 2009). For the purposes mentioned above, a three-dimensional perception of the current environment is really preferred to two-dimensional information. With the rapid development of computer vision, there are many conventional approaches to acquiring 3D information, such as stereo photogrammetric systems (P. Cerveri et al., 1998; S.M. Seitz et al., 2006), laser range scanners (Y. Alshawabkeh, 2005) and optical/scene flow approaches (R.A. Newcombe et al., 2010), most of which are based on 2D-images from CCD- or CMOS-sensors to recovery 3D-information and also need laborious calibration process and relatively complex and time-consuming algorithms. The development of Photonic Mixer Devices (PMD) technology enables 3D-imaging system with pixel-level distance information beyond grayscale values. And many existing problems find their new solutions such as edge detection (X. Jiang and H. Bunke, 1999) and environment reconstruction (B. Huhle et al., 2008; A. Sappa et al., 2001).

The objective of this research is to detect moving objects and tracking them in real-time based on both range images and intensity (grayscale) images from the PMD 3D-camera using background subtraction and image processing techniques. One task in real-time motion detection and tracking is to segment targets from background. Previous studies presented effective methods to perform this task, including region splitting/merging, thresholding, morphological methods, statistical models, and so on. Traditional segmentation of objects from background is based mainly on color or grayscale in 2D images. But the segmentation becomes impossible under conditions of low illumination or when the objects and background look similar to each other, as J. Leens (2009) mentioned. Fortunately, range images from the 3D-camera are available and work very well even in those cases mentioned above. However, there are systematic errors and distortions caused by environment conditions in range images as mentioned in M. Lindner and A. Kolb (2006). Besides, one drawback in practical applications is that range images are useless for segmentation in case that the objects are too close to the background. So, intensity and range images can compensate for disadvantages of each other.

The used 3D-camera is shown in Fig.1: a ToF 3D-camera with an optical sensor of 204 × 204 pixels, 40° × 40° field of view and 0.3-7m measurement range. Integrated SBI (Suppression of Background Illumination) technology enables working well both indoors and outdoors. A visual application is shown in Fig.2, which combines each pixel’s grayscale value with its 3D-coordinate.

Fig. 1. The 3D-camera: PMD[vision]® CamCube 2.0
2. IMAGE SEGMENTING

2.1 Related Work of Image Segmenting

Traditional image segmenting algorithms such as region splitting and/or merging and thresholding methods are based mainly on 2D information, colour or grayscale. With the PMD technology application, the distance information, in the form of real-time range images, can be obtained beyond intensity information, providing an extended perception for image segmentation. With the real-time distance acquisition of every pixel, the computational complexity is greatly reduced and many applications and new algorithms came into being. A novel edge detection algorithm is presented in X. Jiang and H. Bunke (1999) for range images based on a scan line approximation technique, which supports a classification of edge points into several subtypes such as crease edge and jump edge and achieve a complete segmentation of current scenes into regions with both good segmentation quality and high computational efficiency. The range information is introduced to achieve a reliable segmentation in (F. Wallhoff et al., 2007) for surveillance and gesture recognition. For 3D environment reconstruction, it proved that the 3D imaging system outperformed the stereo system in terms of achievable accuracy for distance measurements (C. Beder et al., 2007).

For image segmenting, effective methods are reviewed about how to interpret the content of a video scene by combining color, depth and motion information (S. Piérand et al., 2009). Considering limited resolution and insufficient gray-level information of the 3D-camera, high-resolution RGB-images are introduced by mounting a standard 2D camera on top of the PMD camera. Based on calibration and correct mapping between two camera views, a robust segmentation is achieved (M. Lindner et al., 2007a).

In this section, we use the statistical modelling method, presented as SACON in H. Wang and D. Suter (2007), to simultaneously model the pixel-level intensity and distance backgrounds in order to improve the precision and robustness in real-time image segmentation.

2.2 Background Modelling Using Intensity and Range Information

Background modelling is an effective approach to helping segment the foreground objects from the background. Using the modelling method SACON, We establish a statistical model of the background according to the consensus of background samples. For each pixel, its intensity and range values are both kept as long as N frames, and the averages of each pixel are calculated as the values of intensity and range backgrounds, respectively:

\[
B_I(m) = \frac{\sum_{i=1}^{N} x_I(m, I)}{N} \quad (1)
\]

\[
B_R(m) = \frac{\sum_{i=1}^{N} x_R(m, R)}{N} \quad (2)
\]

Where \(x_I(m, I)\) and \(x_R(m, R)\) are the intensity value and range value at pixel \(m\) at frame \(i\), respectively, and \(B_I(m)\) and \(B_R(m)\) are the average values of intensity and range at pixel \(m\), and \(N\) is the number of background sample frames, which is a constant and determined empirically.

Because of both the relatively low resolution in intensity images and noises in range images, pre-processing is applied to every background sample frame. In our case, the intensity image is equalized in terms of its histogram to achieve proper brightness and contrast and smoothed using medial filter. For the range image, a noise removal operation is performed according to the amplitude of each pixel and then medial filter is applied. After the pre-processing, \(B_I(m)\) and \(B_R(m)\) are calculated according to (1) and (2), respectively. After the intensity and range backgrounds are modelled, we can calculate the differences \(D_I(m)\) and \(D_R(m)\) between real-time scenes and the modelled backgrounds as in (3) and (4).

\[
D_I(m) = \begin{cases} 
1 & \left| C_I(m) - B_I(m) \right| \geq T_I(m) \\
0 & \text{otherwise}
\end{cases} \quad (3)
\]

\[
D_R(m) = \begin{cases} 
1 & \left| C_R(m) - B_R(m) \right| \geq T_R(m) \\
0 & \text{otherwise}
\end{cases} \quad (4)
\]

Where \(T_I(m)\) and \(T_R(m)\) are the values to threshold the intensity and range differences at pixel \(m\) between the current scene and background models, respectively. \(C_I(m)\) and \(C_R(m)\) are the intensity and range values at pixel \(m\) at the current scene. \(D_I(m)\) and \(D_R(m)\) are both binary values with “one” for differences larger than the corresponding thresholding values and “zero” for smaller differences.

The thresholding values \(T_I(m)\) and \(T_R(m)\) are in relationships with intensity and range errors, respectively, and will be carefully determined by the camera calibration in section 2.3.

According to the binary values \(D_I(m)\) and \(D_R(m)\) of every pixel, two binary images are obtained, which are processed pixel by pixel with “or” operation:

\[
D(m) = D_R(m) \| D_I(m) \quad (5)
\]

And a binary image can be composed of values of \(D(m)\). The resulting image is that the pixels with “one” belong to the foreground objects and pixels with “zero” stand for the background.

2.3 Motion Detection and Tracking
2.3.1 Application of Intensity Images

For some simple applications based on a background modelling technique, an intensity image can provide sufficient information for segmenting, detecting and tracking objects.

However, some intrinsic limitations exist in the application of intensity images. When the scene has low illumination or small contrast or when the colour differences between the background and the foreground objects are not large enough, it is impossible to use the intensity image to achieve an effective and precise segmentation. And other problems for intensity images of the 3D-camera are its relatively low resolution and only grayscale information, compared with commonly used digital cameras.

2.3.2 Application of Range Images

For range images, there are no limitations of scene illumination/contrast or colours because the 3D-camera uses its own light sources and can be used in complete darkness or outdoor environment with the help of background light suppression, the SBI technology. Unfortunately, the range image still has its own intrinsic problem. That is, if the foreground objects are physically too close to the background, it becomes impossible to discriminate between the foreground objects and the background because of the limitation of grayscale level and the determination of parameters related to the existing inherent errors in range images mentioned in many previous contributions, as shown in Fig.3.

![Fig. 3. Segmentation based on a range image: a man with his left hand on the wall. Only from the range image (a), it is difficult to recognize his complete hand from the background, as (b) shows.](image1)

And additionally, range images are greatly affected by the highly reflective material like glass, metal and glossy plastic (M. Wiedemann et al., 2008) or by intensity distribution (M. Lindner and A. Kolb, 2007b). But it is feasible to correct measurements by using amplitude-related filtering methods to filter out the bad pixels, such as amplitude mean or amplitude median, as M. Wiedemann et al. (2008) presented.

2.3.3 Background Error Modelling

The thresholding values $T_d(m)$ and $T_b(m)$ are very important parameters in classifying the corresponding pixel into objects and the background. So, a basic camera calibration and careful error estimation is of great necessity to determine the values $T_d(m)$ and $T_b(m)$ of each pixel.

The purpose of the camera calibration is mainly to determine the range errors of each pixel at different distances. In our camera calibration, we use checkerboards of different pattern size as calibration models for different distances varying from 1.2m to 6.0m.

For each range image, there is an intensity image taken in parallel. The Intel’s OpenCV library is used to mark all the reference points in the intensity image and provide the position of each pixel in the image corresponding to the marked reference point, as shown in Fig.4 (a) for example.

![Fig. 4. (a) An example of the reference points marked by OpenCV; (b) 3D reconstruction of range image pixels by OpenGL.](image2)

And according to the pixel positions in the image, the range value of each reference point can be obtained. It is true that the intensity image includes distortion because of the intrinsic properties of pinhole camera model. But we do not need to correct it since what we want is the corresponding range values that have been automatically corrected by the 3D-camera itself. Fig.4 (b) shows the range values obtained directly from the 3D-camera and reconstructed in three-dimension using the OpenGL library.

As basic references, the true distance value from the 3D-camera to each reference point is measured by the total station with an accuracy of 2mm±2ppm. By comparing the difference between the measured and true values in each reference point, we know deviations at each pixel at distances from 1.2m to 6.0m. For each non-reference pixel in one range image, its distance deviation is acquired by the linear interpolation of its two nearest reference points. And a curve of deviation as the function of nominal distance is made for each pixel using the least square method (LMS). By now, a range deviation model has been established. However, such a model is time-consuming, computationally complex, and unsuitable for real-time applications. For simplicity, the maximum average deviation of five shots at each distance is used to set up the background error model. One resulting problem is that segmented images based on range images always contain noises because the thresholding values $T_d(m)$ and $T_b(m)$ are not so precise for every pixel. But at the same time we can use filtering methods to remove these noises. And filtering is also needed in the intensity image segmentation because of low resolution and properties of the PMD sensor.
2.3.4 The Complete Framework of the Proposed Detection and Tracking Method

The main steps of the proposed method are shown in Fig.5. Important components in the framework are the background modelling, pixel-level comparison, object recognition and tracking, and image processing through the whole process. The quality of pixel-level comparison depends mainly on the background error model. During the whole process, image processing methods like median filtering, dilating and eroding, grayscale normalizing and edge/contour detection are of great significance and should be properly used.

Fig. 5. Block diagram of the complete framework.

3. EXPERIMENTAL RESULTS

During the whole experiment, camera parameters are set to constant values with the integration time 2.5ms, field of view 40°×40° and lens correction on. The proposed detection and tracking process is realized in C based on OpenCV and OpenGL libraries.

In one test, we use the proposed method to detect and track the human being with a white wall background as a relatively optimal environment, as shown in Fig.6. From the segmentation results, it can be seen that the intensity-based segmentation image is not a complete segmentation with inside holes and contains some noise around the scene while there is little noise in the range-based segmentation image. And a series of video frames captured in this test are shown in the left column of Appendix A.

Fig. 6. Result with a white wall as the background: (a) the intensity image and segmentation; (b) the distance image and segmentation; (c) the final result of segmentation (left) and edge detection, marking and tracking (right).

In the following test, a glass door as a noise source is located at bottom-right of the background and causes noise in segmentation images. Compared with results in the above test, the range-based segmentation image is greatly affected. Although the glass door causes unavoidable noise in range images, a good segmentation still can be achieved using filtering methods like median filtering, dilating and eroding.
Fig. 7. Result with a noise source in the background: (a) the intensity image and segmentation; (b) the distance image and segmentation; (c) the final result of segmentation (left) and edge detection, marking and tracking (right).

In another test we try to figure out the number of objects and tracking them. A series of video frames are shown in the right column of Appendix A. From these video frames we can see that the number of mobile human beings is known when they have no overlap with each other while the two persons will be taken as only one person when overlapping with each other. Obviously some noise exists during the process, which might be recognized as objects. A feasible solution for separating objects from noise areas is to record their positions. The noise areas remain at constant positions while mobile objects keep moving.

As shown in the above tests, the image segmentation will fail by the separate use of the range and intensity images while a proper segmentation can be achieved by combining them. And we can find that the range-based segmentation always works much better than the intensity-based segmentation. Another drawback of intensity images is that, the shadow around the human will appear when the human is too close to the 3D-camera, which is difficult to be removed only from limited grayscale information.

4. CONCLUSIONS

In this paper we use a statistical background modelling method SACON to help segment the real-time scene and then detect and track objects based on intensity and range images. To achieve precise segmentation we establish the background error model that is a very important part to determine the error tolerance of every pixel. And during the whole process filtering methods play a significant role in suppressing the noise and removing bad pixels.

Further work should be done to make a more suitable background error model according to the specific application environment. Since the inaccuracy of final segmentation results comes mainly from intensity images, it is much better to integrate other colour spaces like RGB into the source information so that we can obtain RGBZ-data for advanced processing (M. Lindner et al., 2007a). In this way the shadow caused by too close objects can be recognized and removed from the final segment result. Future work should also solve the problem that different objects can be separated and tracked even when they are overlapping with each other. And a proposed solution of this problem is to perform the object 3D-analysis like volume calculation and position tracking and further reconstruct them.

REFERENCES


Appendix A. RESULTS OF MOBILE OBJECT DETECTION AND TRACKING

The left column shows a simple test result with one person in the scene and relatively optimal environment. Another test is performed with two persons and the result is presented in right column.