ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

Volume 3, Special Issue 3, March 2014

2014 International Conference on Innovations in Engineering and Technology (ICIET'14)
On 21st & 22nd March Organized by

K.L.N. College of Engineering, Madurai, Tamil Nadu, India

Human Tracking and Counting In Range Images by Using Mean Shift and Pylon Grid Algorithm

P. Karpagavalli[#1], A.Vinoth Nelson[*2], A.V. Ramprasad[#3]

[#1]Department of Electronics and communication Engineering, K.L.N College of Engineering, Madurai, Tamil Nadu, India.

[*2]Department of Electronics and communication Engineering, K.L.N College of Engineering, Madurai, Tamil Nadu, India.

[#3]Department of Electronics and communication Engineering, K.L.N College of Engineering, Madurai, Tamil Nadu, India.

ABSTRACT— Pylon grid method used for automatic tracking and counting of people on consumer path in a shopping mall, a subway and an airport in order to collect statistical data about the consumer behavior. In vision-based algorithms for the people counting and tracking has failed when the background changes gradually over time. The contour segmentation method has failed due to multistage threshold limits. If the observed area is so crowded, an occlusion problem has occurred. So we introduce the new algorithm, Pylon Grid, which is used to eliminate all these problems by using a static grid of measure points – pylons. A modified human body tracking system based on the Discrete Wavelet Transform and Mean-shift algorithm is proposed. The proposed algorithm of motion object detection and segmentation is applied to the video for a crowd of fast moving people. Tracking multiple people accurately in cluttered and crowded scenes has failed in a single view. We present multi view approach to solving this problem. This can be done with a stereo camera. Pylon grid method is the efficient and fast method for human head detection in range images captured by a stereo camera that is positioned vertically, pointing from the roof to the ground. The human head is detected using segmentation of the images. The segmentation process helps to identify the heads of the humans. The objects in the video are obtained using normal Background subtraction. The identified objects are then marked. By using the number of blobs identified the number of people in the video. An essential component of any tracking system is the detector, which determines the positions of the people in each frame of the image sequence. While the vertical camera orientation limits the size of the observed area, it provides a clear view with minimal occlusions among the people. We show how a fast median filter can be used for effective preprocessing of the range data. The person can be detected by using the method only two threshold values. The algorithm is guaranteed to find all local minima and its speed only depends on the resolution of the grid.

KEYWORDS—Range Image, Head Detection, Stereoscopy, Computer Vision, People Counting, Local Minimum Search, Spatial Information, Discrete Wavelet Transform, Mean-Shift Algorithm, ARM, Human Body Tracking.

I. Introduction

There is a class of applications that make use of automatic tracking and counting of people (crowd) in a scene. One example is consumer path tracking in a shopping mall in order to collect statistical data about the consumer behavior. Another one application is people counting and pedestrian flow analysis in public areas.

An essential component of any tracking system is the detector, which determines the positions of the people in each frame of the image sequences. In the context of stereo vision, the problem of precise detection can be attacked effectively with a vertical (stereo) camera orientation.

While the vertical camera orientation limits the size of the observed area, but it provides a clear view, with minimal occlusions among the people. It enables a reliable segmentation of the people from the ground, with accurate head detection, even in crowded scenes with high people density. Another one advantage of the vertical camera orientation is that human heads appear as local minima in the range image. Tracked objects in video sequences can be used for many applications such as video surveillance and monitoring, traffic monitoring, and object based coding.

II. System Overview

The human head has detected by using segmentation of the images. The segmentation process helps to identify the heads of the humans. The objects in the video are obtained using normal Background subtraction. The identified objects are then marked. Using the number of blobs identified the number of people in the video are found out. Discrete wavelet transform (DWT) has been widely used in the field of motion detection and object tracking.

If the observed area is so crowded, an occlusion problem has occurred. Occlusion and lack of visibility in crowded and cluttered scenes make it difficult to track individual people correctly and consistently, particularly in a single view. We present multi view approach to solving this problem. This can be done with a stereo camera. Stereo camera is a type of camera with two (or) more lenses to separate image sensor (or) film frame for each lens.

In the context of stereo vision, the problem of precise detection can be attacked effectively with a vertical camera orientation. While the vertical camera orientation the limits size of the observed area, it provides a clear view, with minimal occlusions among the people. This enables a reliable segmentation of the people from the ground, with accurate, head detection even in crowded scenes.

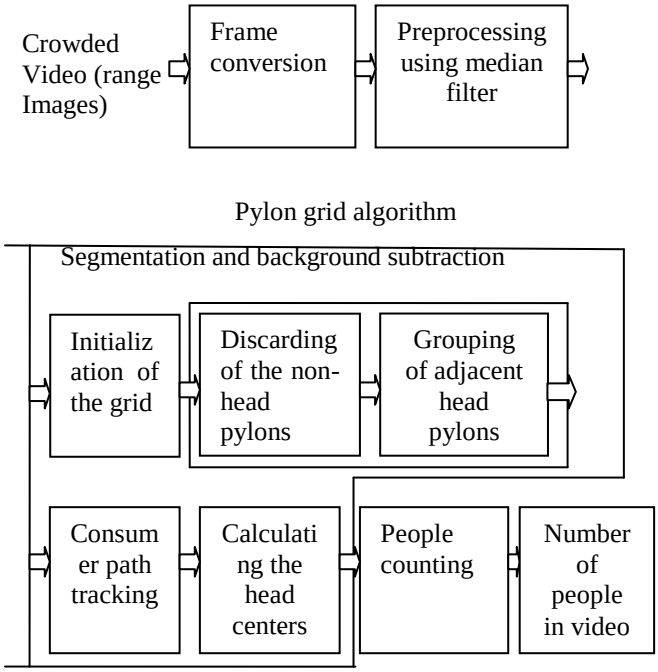


Fig. 1 Block diagram of pylon grid algorithm

III. Related Work

Vision-based algorithms for the people counting and tracking problem may be divided into two main classes: the direct and the indirect methods [2]. The direct methods [3–6] are based on the detection of each person in the scene, by using some form of segmentation and object detection algorithm. The counting, in turn, is performed as a second step. The indirect methods [2, 7–9] instead perform the people counting by using the measurement of some features that do not require the separate detection of each person in the environment.

A method for counting moving people in video surveillance, video approach proposes a novel method for counting moving people in a video surveillance scene. And also this indirect method employed the people counting by using the measurement of some features such as the amount of moving pixels, fractal dimension.

This work presents a new algorithm, the Pylon Grid, which is used to detect efficiently human heads in range images. This algorithm lies at the heart of a novel direct vision-based approach to estimate people flow in both indoor and outdoor environments.

IV. Proposed Algorithm

A. Segmentation of the People

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image

segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

B. Stereo Model for Humans

Stereo camera is a type of camera with two (or) more lenses to separate image sensor (or) film frame for each lens. In the context of stereo vision, the problem of precise detection can be attacked effectively with a vertical camera orientation. While the vertical camera orientation the limits size of the observed area, it provides a clear view, with minimal occlusions among the people.

C. Tracking - Background Subtraction

Following step defines the approach to perform the background subtraction operation:

- Subtract the estimated background of the input frame
- Apply a threshold 'T' to the absolute difference to get the foreground information. A sample background subtracted image

$$| F – G | > T \quad (1)$$

Where, F is the Image at time‘t’, G is the Background Image and T is Threshold

Various methods are used for estimating the background of an image in a video sequence. In Frame Differencing method, the background is considered to be the previous frame in the sequence of frames.

$$G(x, y, t) = F(x, y, t - 1) \quad (2)$$

$$|F(x, y, t) - F(x, y, t - 1)| > T \quad (3)$$

$$|F(x, y, t) - G(x, y, t)| > T \quad (4)$$

Where, F is the Image at time ‘t’, G is the Background Image and T is Threshold value.

In Mean Filter approach, the background is the mean of the previous 'n' frames.

$$B(x, y, t) = \text{mean } \{F(x, y, t - i)\} \quad (5)$$

$$F(x, y, t) - B(x, y, t)| > T \quad (6)$$

In Median Filter approach, the background is the median of the previous 'n' frames.

$$B (x, y, t) = \text{median } \{F (x, y, t - I)\} \quad (7)$$

$$|F(x, y, t)-B(x, y, t)| > T \quad (8)$$

D. Mean Shift Algorithm

This algorithm is a robust method of finding local extrema in the density distribution of a data set. Robust means that it ignores data points that are far away from peaks in the data. It runs as the following below conditions:

1. Choose a search window and mention the type.
2. Compute the window’s center of mass.
3. Center the window at center of mass.
4. Return to step 2 until the window stops moving.

E. The Pylon Grid Algorithm

The input video is preprocessed to remove the noises in the video. The videos are first converted into frames. For Preprocessing Median filter is used. Before preprocessing Gaussian noise is added to the image. After preprocessing the frames are segmented to identify the objects in the video. After Segmentation the background subtraction extracts the foreground objects.

step1—Initialization of the grid: The Video is divided into grids for identification of the objects. The grid formation helps in the identification of the objects and it also helps in segmentation. Grids are like small sections of the images. Grids are applied to every frame of the object.

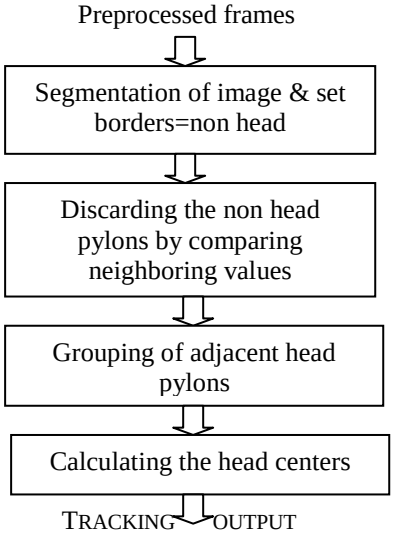


Fig. 2 Flow diagram of pylon grid algorithm

Step2—Discarding of the non-head pylons: After preprocessing the frames are segmented to identify the objects in the video. After Segmentation the background subtraction extracts the foreground objects. The extracted foreground is then marked. The head region is identified and is marked using a bounding box.

Step3—Grouping of adjacent head pylons: Background subtraction is one of techniques for segmenting out objects of interest in a scene for application such as surveillance, human-component interaction and traffic monitoring. In this method; we are going to remove all the non significant components in the background from a reference frame in a video.

Step4—Calculating the head centers: By counting the number of Blobs identified the number of people in the video can be calculated. The estimation of the number of people present in an area can be extremely useful information both for security safety reasons

V. Results And Discussion

A. Preprocessing of the Range Data

The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve the results of later processing. Median filtering is very widely used in digital

image processing, because under certain conditions, it preserves edges while removing noise.

B. Handling ''Mud Stains'' Noise

The typical noise that we encountered was a kind of dark mud stains that appear frequently on the borders of the range silhouettes of the people. The median filter turned out to be especially well suited for removing this kind of noise.

C. Preprocessing by a Fast Median Filter

From a qualitative viewpoint, the median filter performed very well for filtering the mud stains. Unfortunately, while the noise removal ability of the median filter is excellent, it is also considered slow, because of its non-linear computational complexity. It exploits the fact that the range of all possible values is known. For example, if we sort byte values (which is very common in grayscale images), we know that all values will be in the range.

D. Robustness and Speed

Here the orthographic projection, followed by convexization and noise filtering of the range data, the humans can be distinguished by one very stable feature the local minimum, which becomes unique for every head. From this point, the pylon grid algorithm can be used to find all local minima accurately and quickly.

E. Fast Moving People

In a situation with fast moving people, there will be two additional effects:

1. The positions of the people will be shifted in subsequent frames;
2. Range data might be less accurate and might contain more noise.

Simulation Results

1. DATASET1:

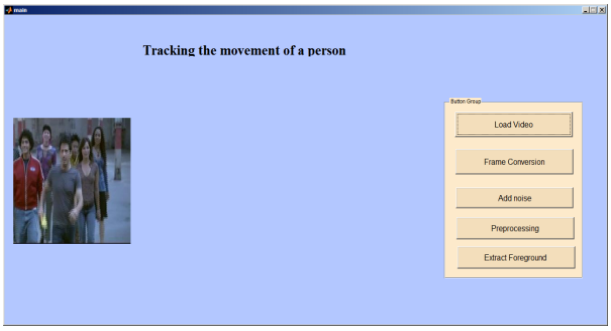


Fig. 3Video read output

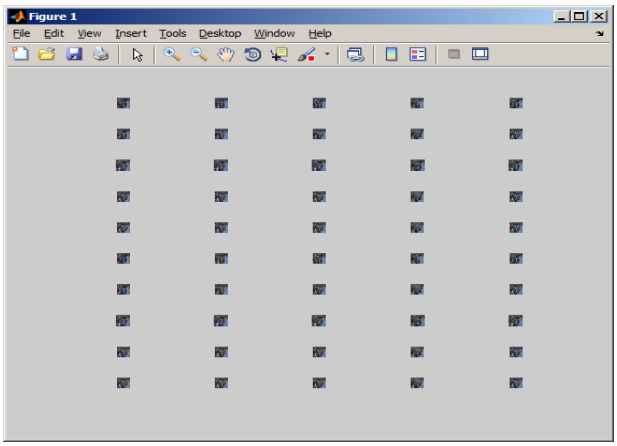
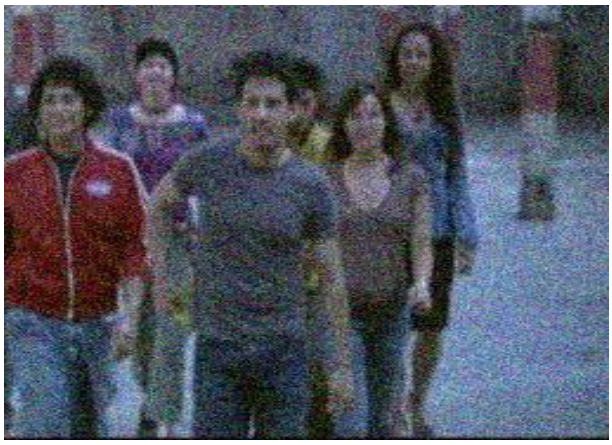


Fig. 4Video to frame conversion window



(a)

…..

(b)

Fig. 5 (a) and (b) Noised frames

Fig. 7 Grid formation output

(a)

….

(b)

Fig. 6 (a) and (b) Preprocessed frames

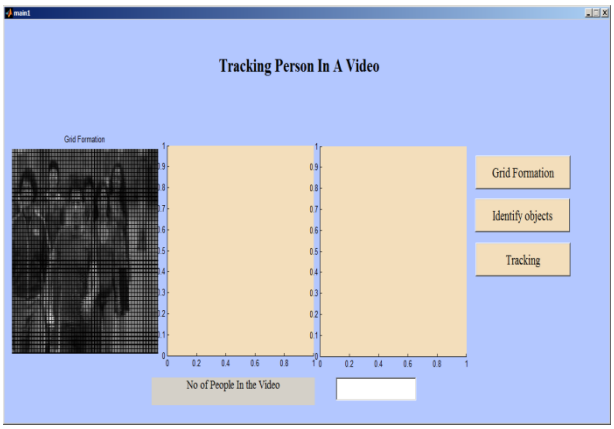
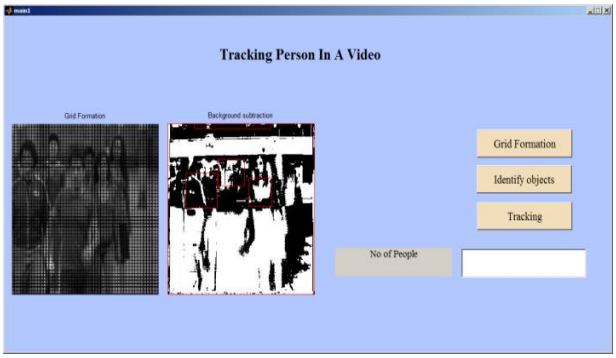


Fig. 8 Segmentation output

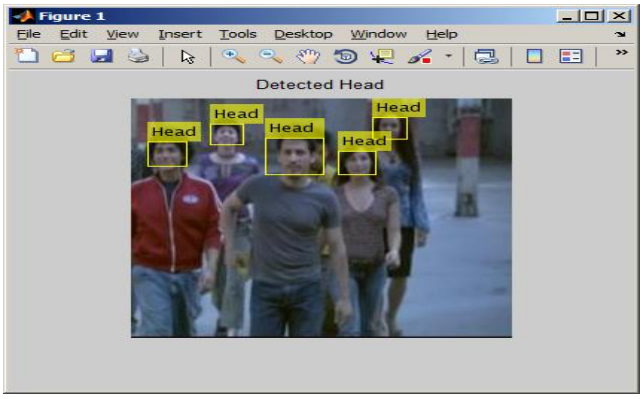
(a)

(b)

Fig. 9 (a) and (b) Background subtraction output



(a)



(b)

Fig. 10 (a) and (b) Tracking output

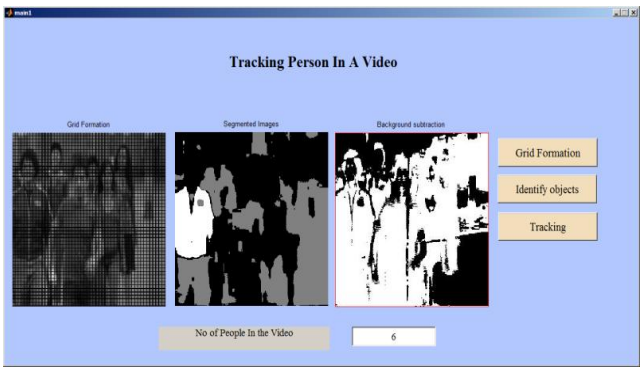


Fig. 11 Person counting output

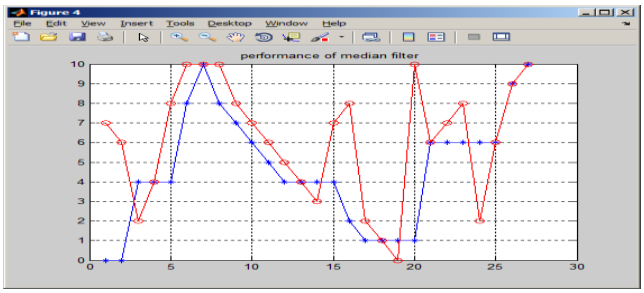


Fig. 12 Performance of median filter

2. DATASET 2:

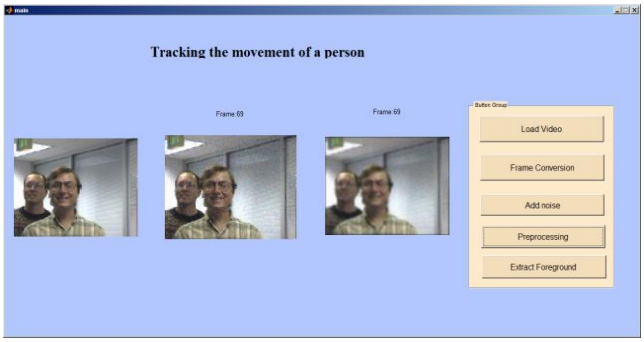


Fig. 13 Video read and preprocessed output

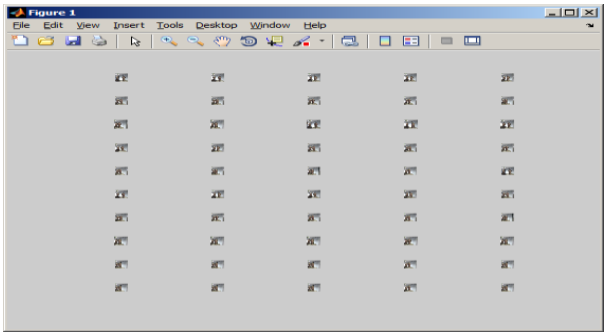


Fig. 14 Video to frame conversion window

Fig. 15 Background subtraction output

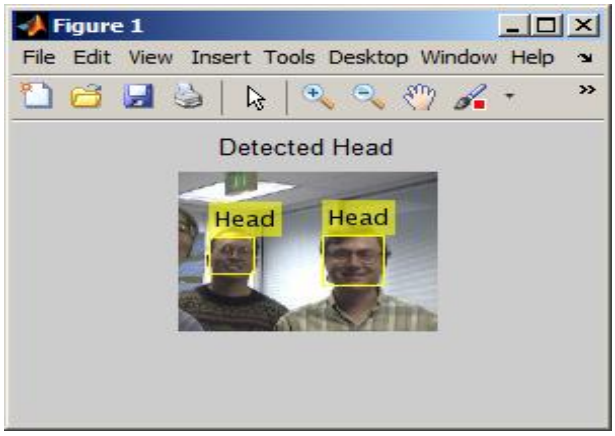


Fig. 16 Head tracking output

Fig. 17 Person counting output

VI. Conclusion

Here I propose a mechanism to track humans in a video. After the preprocessing stage, the videos are supported into grids for segmentation and object identification. By segmenting the objects in the video are identified. Using Background subtraction the objects are extracted from the video and they are tracked. Using the number of blobs identified the total number of people in the video are found out. Then the improved multiple human body tracking system with minimum computation time. And also propose the algorithm to detect motion and track the human body.

References

[1] A.Rusi Antonov Filipov, B.Flavio Luis Cardeal Padua, and B.Marco Aurelio Buono Carone, "pylon grid: A fast method for human head detection in range images," ELSEVIER, Neurocomputing, 2012.

[2] G.Rakate "Human body tracking system based on DWT and Mean-shift algorithm on ARM-Linux platform," IEEE Trans. Pattern Anal. Mach. Intell. , 715–720, 2012.

[3] D.Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, "A method for counting moving people in video surveillance videos," EURASIP J. Adv. Signal Process. , Vol. 5, pp: 1–10, 2010.

[4] S.M.Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," IEEE Trans. Pattern Anal. Mach. Intell. Vol. 3, pp 505–519,2009

[5] T. Zhao, R. Nevatia, and B. Wu,"Segmentation and tracking of multiple humans in crowded environments," IEEE Trans. Pattern Anal. Mach. Intell. Vol.30, pp. 1198–1211,2008.

[6] D. Ramanan, D.A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," IEEE Trans. Pattern Anal. Mach. Intell., Vol.29, pp. 65–81, 2007.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B*, vol. 39, no. 1–38, pp. 1–1, 1977.

[8] D.B. Yang, H.H. Gonzalez-Banos, and L.J. Guibas, "Counting people in crowds with a real-time network of simple image sensors," in: Proceedings of the Ninth IEEE International Conference on Computer Vision, vol. 2, ICCV '03, IEEE Computer Society, Washington, DC, USA, pp. 122, 2003.

[9] Donald E. Knuth, "The Art of Computer Programming," Sorting and Searching, 2nd edition, Addison-Wesley, vol. 3,1998.

[10] A. Albiol, M.J. Silla, A. Albiol, and J.M. Mossi, "Video analysis using corner motion statistics," in: Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 31–38, 2009..

[11] O. Masoud, N.P. Papanikolopoulos, and S. Member,"A novel method for tracking and counting pedestrians in real-time using a single camera," IEEE Trans. Veh. Technol., Vol. 50 pp. 1267–1278, 2001.

[12] D.S. Jang, H.I. Choi, Active models for tracking moving objects, Pattern Recogn , vol.33, pp. 1135–1146, 2000.

[13] D. Beymer, "Person counting using stereo," in: Proceedings of the Workshop on Human Motion (HUMO'00), HUMO '00, IEEE Computer Society, Washington,DC, USA, pp. 127, 2000.

[14] H. Ning, T. Tan, L. Wang, and W. Hu, "Kinematics-based tracking of human walking in monocular video sequences," Image and Vision Computing, Vol. 22, pp. 429-441, May2004.

[15] C. Rambabu and I. Chakrabarti, "An efficient hill climbing-based watershed algorithm and its prototype hardware architecture," J. Signal Process. Syst, Vol. 52, pp. 281–295, 2008.

[16] S. Stiene, K. Lingemann, A. Nuchter, and J. Hertzberg, "Contour-based object detection in range images," in: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), 3DPVT '06, IEEE Computer Society, Washington, DC, USA, pp. 168–175, 2006.