Annals of Emerging Technologies in Computing (AETiC)
Vol. 2, No. 2, 2018

Article

Stereoscopic Human Detection in a Natural Environment

Ross Davies[1], Ian Wilson[1], Andrew Ware[1,*]

[1]Faculty of Computing, Engineering and Science, University of South Wales, UK
ross.davies@southwales.ac.uk, ian.wilson@southwales.ac.uk, andrew.ware@southwales.ac.uk
*Correspondence: andrew.ware@southwales.ac.uk

Received: 14th February 2018; Accepted: 15th March 2018; Published: 1st April 2018

Abstract: The algorithm presented in this paper is designed to detect people in real-time from 3D footage for use in Augmented Reality applications. Techniques are discussed that hold potential for a detection system when combined with stereoscopic video capture using the extra depth included in the footage. This information allows for the production of a robust and reliable system. To utilise stereoscopic imagery, two separate images are analysed, combined and the human region detected and extracted. The greatest benefit of this system is the second image, which contains additional information to which conventional systems do not have access, such as the depth perception in the overlapping field of view from the cameras. We describe the motivation behind using 3D footage and the technical complexity of human detection. The system is analysed for both indoor and outdoor usage, when detecting human regions. The developed system has further uses in the field of motion capture, computer gaming and augmented reality. Novelty comes from the camera not being fixed to a single point. Instead, the camera is subject to six degrees of freedom (DOF). In addition, the algorithm is designed to be used as a first filter to extract feature points in input video frames faster than real-time.

Keywords: 3D Image; Human Detection; Human Tracking; Foreground Detection

---

1. Introduction

Computer vision is a challenging field of computing where the ability of an algorithm to produce a valid output is often not the only measure of success. Often, one of the biggest problems in computer vision is the computation cost to run the algorithm in real-time. The most common and difficult task is real-time human tracking, which has seen many advances in recent years. Problems associated with algorithms created to tackle issues are discussed. The majority of current systems utilise a single camera with no depth perception. The goal of the system presented is to take advantage of the depth perception provided by adding a second camera spaced just under that of the intraocular distance. Recent advances in 3D media create an industrial requirement for future innovations to keep up with the demand that followed.

Multiple camera human tracking is not a new area of research with many researchers and companies trying to find robust and easy to set-up cameras. Typically, such systems consist of multiple cameras that can see the human from different viewpoints [1]. A larger number of these systems are starting to focus on using stereoscopic cameras in fixed locations, utilising background subtraction techniques for creating disparity mapping on the resultant image.

Ross Davies, Ian Wilson and Andrew Ware, "Stereoscopic Human Detection in a Natural Environment", Annals of Emerging Technologies in Computing (AETiC), Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 15-23, Vol. 2, No. 2, 1st April 2018, Published by International Association of Educators and Researchers (IAER), DOI: 10.33166/AETiC.2018.02.002, Available: http://aetic.theiaer.org/archive/v2n2/p2.html.

Figure 1 shows how using two cameras gives an overlaid region where a three-dimensional view exists, assuming webcams with a 90° range. All objects within the 3D region have a different parallax. Closer objects have larger parallax than distant objects. This knowledge is used in the creation of depth mapping. Here, a system is presented that uses this knowledge of differing parallax to detect a person who is close to the camera.

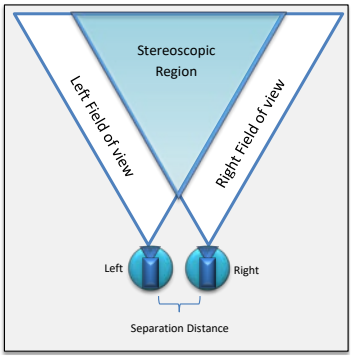

Figure 1. Image showing the stereoscopic region created by using two webcams

Human tracking is a complex vision problem for computers to simulate. Computers lack depth perception without use of specialist equipment such as the Microsoft Kinect. The goal of the system is to give the computer depth perception in the way humans have by having two cameras at eye distance apart. This system will be used in further work in augmented reality where 3D analysis of a scene has the potential to create more robust systems than are currently available. In this scenario, users of the system can be exposed to uniquely generated frames for each individual eye.

2. Related Work

Numerous solutions to recognising humans in images have been devised [1-6]. These typically make use of offline processing. The ones that are not have limited scope of use, which is discussed in the following section.

Algorithms such as Pfinder ("people finder") [2] records multiple frames of unoccupied background taking one or more seconds to generate a background model. This model is subtracted from an image before processing occurs. The only details remaining after background subtraction are the "moving objects" or changes such as people in the scene. Pfinder has limitations in its ability to deal with scene movement. The scene is expected to be significantly less dynamic than the target person, meaning the scene should not update as often as the target.

The benefit over similar systems like player tracking and stroke recognition [3] is that Pfinder processes in real-time. This does not produce clear models of the person in question. Skeleton structures are generated from the images that include the shadow as part of the human. In that system only top body movement was analysed meaning this did not cause a problem. Alternative systems for the same task exist, such as Players Tracking and Ball Detection for an Automatic Tennis Video Annotation [4]. This algorithm works in real-time and is able to detect and recognise tennis strokes although the detail of human movement is limited.

People tracking systems conceived for surveillance applications already work in real-time without the need to pre-initialise the background model [5]. Their system constructs a background model based on checking the frame-to-frame differences. The abilities of the previous algorithm surpass many competitors providing the benefit of human tracking. It appears as though systems currently developed work in real-time with little accuracy or with accuracy but offline. Scope for improvement still exists in the ability to develop an algorithm that works in real-time that does not require long background initialisation and has the detail required for gesture recognition. These significant advances made by researchers in the past use single lens cameras, which does not benefit from depth perception.

The work presented in [7] was capable of more than human detection based upon Discriminatively Trained Part Based Models of a certain type of object. This work produces a system that is highly accurate detecting the correct object in most of the test cases presented in the paper. The unfortunate downside, as with most vision based algorithms, is the processing requirement. Here, each new object requires a three-hour training session (on a single CPU) and each frame of an image requires two seconds to be analysed. Although this is a significant improvement in vision based object detection, it is not at the point where it can be used for applications such as augmented reality due to high processing costs.

A relatively new field has emerged in computer vision utilising different cameras providing various viewpoints of a scene. Stereoscopic systems such as [1] provide the ability for human tracking in natural environments. This system uses conventional difference checking techniques to determine where motion has occurred in a scene. Motion of both cameras combined generates a location of a human (including their limbs) within a scene. This project produced a robust system capable of tracking multiple people, but with a pre-setup requirement. Multiple camera human detection has also been used in a museum environment [8]. People could come to an exhibit and interact with the environment though their movement alone. The system was capable of handling a large number of guests successfully but required multiple cameras meaning there was a problem with lack of portability.

Multi-lens imagery when set up correctly can have more than the advantage of viewing different viewpoints. Two cameras set-up at a distance close to that of the intraocular distance facing towards the same focal point allows for stereoscopic imagery with the ability to extract a perception of depth. Finding out the displacement between matching pixels in the two images allows creation of a disparity map. This includes the depth information for each pixel viewable by both cameras. It is possible to extract and reconstruct 3-D surfaces from the depth map [9-10]. Work conducted into depth mapping has improved the clarity of the result [11]. In [12], disparity estimation was improved by repairing occlusion. This allows for a more realistic depth map as occluded pixels are approximated from surrounding data. Processing requirements remains the fundamental problem that needs to be addressed for successful application in dynamic space in real-time. Generation of depth maps for the entire image is not currently possible in real-time. Research needs to be directed into subtracting regions from an image to give a smaller image to use for depth map generation.

Early work that utilises stereo cameras for human detection also adds in other characteristics such as height, face pattern and colour to improve the tracking performance, labelled Integrated Person Tracking (IPT) [13]. IPT uses stereo range finding and foreground segmentation passed onto systems that estimate head location based upon the silhouette. This information is passed onto face recognition.

Recent work on stereoscopic human tracking includes set-up of camera distributed within an environment to gather information from different angles. There is a large amount of information held in just a short distance between cameras, which is evidenced in the subtraction stereo algorithm [14]. Using conventional techniques for background subtraction on both the right and left image, only the regions of "movement" remain. It is possible to generate a disparity map for only the relevant section of the image instead of the whole image when comparing movement in both images. The disparity then allows the extraction of data such as size and location of the object detected, which is not available in single view cameras. Although this is an improvement on single vision, the original proposed algorithm also extracted shadows [15]. In detection of pedestrians using subtraction-stereo [6], the algorithm was expanded to exclude shadow information and a test case was put forward for the use of this algorithm in video surveillance. A further expansion of this work provided a robust system for tracking motion of individual persons between frames [15].

In [16] a dense depth map, generated through a hardware implementation of Semi Global Matching (SGM), was used to reduce inaccuracies in the matching procedure. Estimation is possible through the highly detailed dense stereo map in addition to the standard dynamic ability to estimate camera height, pitch and road imperfections that a sparse map can provide. The road surface is recognised using B-spline estimates. Feeding the road estimates into a Kalman filter increases the

accuracy. Combining the ability to detect people with a general ability to control a car's key systems the algorithm was later improved to both stop before hitting a person or steer to avoid contact [17]. In summary, previous work successfully shows how computer vision can have practical implications for not only streamlining people's lives and creating a safer environment.

3. Our Work

With all the previous work that has been analysed the significant advantages and improvements made always come at an increase in computation. The system presented here is designed to be a novel Natural Feature Tracking (NFT) system that allows the camera to move freely and still process at a high frame rate. The system makes the following assumption: only one person is tracked in the scene and the person being tracked is going to be prominent on the camera and not just another distant object. The tracking in this paper is designed for augmentation of the person in the frame not for video surveillance. Differences in both images will be considered as 'real' objects rather than background noise.

The most common techniques for human detection are background subtraction and motion detectors. Both have significant disadvantages and limit the ability of any system. Background subtraction techniques require knowledge of the scene and objects without the human. Once set up they suffer from noise issues and lighting variations but otherwise are robust and allow detection of numerous different objects (people). Motion detectors are affected by lighting variations showing motion is occurring when lighting levels in the room change. However, they only require a couple of frames set-up so they have faster initialisation than background subtraction. Although differing on implementation both techniques work on a similar principle, the camera has to be stationary.

Our system is designed to be better than previous systems but work in similar ways. The conventional way of motion detection is to check for difference in pixels. Only outlines of foreground objects remain when this stereoscopic vision algorithm is performed. Through different filters and grouping techniques the most prominent object in the scene is detected. The person is detected when the assumptions are valid. Our system requires no initial setup and is not affected by light variation between frames. Unlike traditional systems, the one presented here runs off a single frame comparison between left and right images allowing for camera movement and change in environment.

The remainder of this paper is organised as follows: section 2 gives a description of the algorithm development process, section 3 shows the algorithm in use, section 4 provides discussion and future uses and section 5 documents the conclusions.

4. Methods

The algorithm we developed is interesting in its simplicity. The first attempt used just an XOR filter in order to find the difference in the image. This highlighted lighting variations in the images and output an interesting pattern of colour with the useful information lost amongst the noise. To improve upon this, the next step was to test a variety of filters including: difference, minimum and arc tan filters. Out of the three, the minimum at first appeared to produce the best output removing a lot of the noise with the side effect of slightly eroding the desired result. When filter use alone was discovered to be ineffective, a Gaussian filter was applied over the both inputs to remove minor noise. Even though this did remove minor noise, large patches of lighting variation noise remained largely unaffected. Thresholding was then applied to remove everything but the brightest changes. The problem with this was that even though the displacement between closer objects was larger than that of distant objects it was not necessarily bright. Valuable data was lost once again. A breakthrough was made by checking each pixel against its horizontal and vertical neighbours. Noise was almost eliminated and only slightly affected the required information.

$$\sum_{r=0}^{h}\sum_{c=0}^{w} left[r][c] \oplus right[r][c] \quad (1)$$

$$\sum_{r=0}^{h}\sum_{c=0}^{w} |left[r][c] - right[r][c]| \quad (2)$$

$$\sum_{r=0}^{h}\sum_{c=0}^{w} MAX(left[r][c] - right[r][c], 0) \quad (3)$$

$$\sum_{r=0}^{h}\sum_{c=0}^{w} |tan^{-1} left[r][c] - tan^{-1} right[r][c]| \quad (4)$$

h is the *height* of the input images.
w is the width of the input images.
y is the current row being evaluated.
x is the current column being evaluated.
left is the left camera lens input image.
right is the right camera lens input image.

The first filter attempted was XOR (1). This filter was initially used with the expectation that only the areas on the image that were displaced would remain. Unexpectedly, lighting variations between the left and right image produced interesting output images. The output did include all of the information expected with a lot of added lighting noise. This prompted the effort to find a filter that would be more resistant to lighting variation between the left and right frame. The conventional difference filter (2) is performed on each channel of the image, which produces results that were anticipated from the XOR filter. This form of filter is slightly slower than straight bitwise operations. Although the results of this filter are as good as could have originally been expected there was still need to investigate further filters. The subtraction filter (3) is similar to the differential filter but filters out parts of the results that would otherwise remain. Unfortunately, the tests proved the filter to be indiscriminate given the elimination of valid parts of the result data. The final filter (4) followed on from research by [1]. The filter was designed to eliminate lighting variation in frame-by-frame comparisons for motion detection. Even though different in terms of program use, in principle the idea is similar. Although proving effective in the images with high and low contrast between person of interest and the scene background the filter failed to be effective in the other sample groups. The extra computational expense proved to be wasteful, providing output that in some cases eliminated the useful information with background remaining.

TABLE I
Filter results

Image Description	**XOR (1)**	**Diff (2)**	**Sub (3)**	**ATan (4)**
Ideal conditions	0	3	2	0
Low contrast	2	3	3	3
High contrast	0	3	2	3
Close Target	2	3	3	1
Distant but prominent	2	2	2	2
Neither distant nor prominent	0	2	2	1
Results	**1.00**	**2.67**	**2.33**	**1.67**

Key:
0: Indicates the filter failed to produce any suitable results.
1: Indicates the detection of the person in question with significant background noise.
2: Specifies the detection of the person with slight background, which preferably should have been eliminated.
3: Three indicates a complete success with the region detected including the target person with all reasonable background noise eliminated.

The algorithm is dependent upon the operation of the orphan filter passed through the data, which filters out any pixel that does not have a significantly strong bond connection (set by a threshold) to any of their horizontal or vertical neighbours.

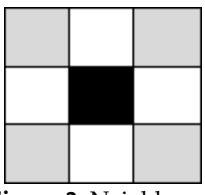

Figure 2. Neighbours

Fig. 2 shows the selection of neighbours of a pixel (black), where white shows a valid neighbour and grey shows a pixel that is not going to be analysed. Thresholding creates a scenario were the best-fit needs to be found in order for an algorithm to be developed that works in the widest range of environments as possible. Lower thresholds are preferable as data kept in the scene provides a larger number of reference objects. Table 2 shows the way in which the optimum threshold value was calculated. A selection of images were analysed plotting their lowest and highest working thresholds. Unfortunately, no single threshold exists, but when excluding the final image the values of best fit are the highest of the lowest thresholds to the lowest of the highest. The best-fit threshold of one hundred and nine will be the default in the program as a lower value is preferable to keep as much detail in the output as possible.

$$\sum_{y=1}^{h}\sum_{x=1}^{w} image[y][x] = {}^{1\ if}_{else\ 0}(\forall x \in B > t) \quad (5)$$

A is a set of all pixels

$B = \{A \mid A \text{ is a neighbour}\}$

t is the threshold

image is the output result from the difference filter

The threshold was determined by calculating the best fit. In a number of test images the best matching threshold range was 109 to 110. Due to lower thresholds keeping in more useful information, 109 is the threshold used. Table 2 shows the valid threshold range for a number of images.

TABLE 2
Threshold

Image Description	Lower Thresh	Upper Thresh
Ideal conditions	105	110
Low contrast	109	174
High contrast	68	203
Close Target	97	178
Distant but prominent	83	127
Neither distant or prominent	164	211
Average	**104**	**167**

The next step now the stereoscopic imagery has been converted into a single image of useful information is to extract a region that contains the most prominent change. This will always be the human in the scene when the previously described assumptions are valid. Pixels of interest are grouped together into the appropriate small region of interest on a grid 16 in width by 16 in height. When there is a sufficient amount of change in the smaller region, it is considered a region of interest. The largest bulk of these regions of interest are then expanded into a single region of best fit. This region encompasses the person in the scene successfully in all tested environments, even where the assumptions do not quite hold true.

5. Results

Detection of the user is performed by grouping the small-detected regions of interest into a larger group around their average distribution point. Regions are considered to be connected via either horizontal, vertical or diagonal neighbours with no gap between them.

In both Figure 3 and Figure 4 only the left image is shown. The images are labelled as follows, (a) top-left, (b) top-right, (c) bottom-left and (d) bottom-right. Figure 3 shows the major region detected and minor detections are also shown. The algorithm was tested indoors:

a. The region includes the subject's complete body despite the fact that the arms are spread out.
b. In this image, the colour of the clothing being worn is similar to the background image. This image is one of the ones in which the algorithm was expected to experience difficulty. Instead, the subject is picked up accurately with the smallest possible analysis region being generated.

c. He is still detected as the most likely region to hold a human in this image despite the fact that the subject is either further away and at different distances to other potential objects of interest.
d. Despite the assumptions not holding true, the algorithm deals well with generating the human region. The overall region detection picks up the area of the image including the person but has extra image that we would rather not analyse. The size of the image that requires analysis is far smaller than the original.

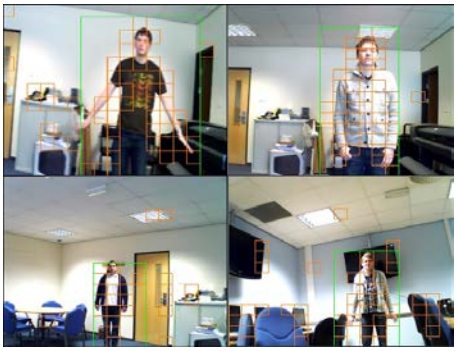

Figure 3. Sample Images

The results of the indoor testing were quite promising; even in cluttered environments. Indoor is likely to be the main environment in which this system would be utilised for motion capture. Augmented reality applications on the other hand would be typically aimed at outdoor activities meaning the algorithm must also function well outdoors. This is the environment we anticipate the system will thrive in comparison to infrared systems which can be problematic on sunny days (for example Microsoft Kinect designed to tackle human detection indoors does not work well in this kind of environment). Figure 4 shows the abilities to use the system outdoors.

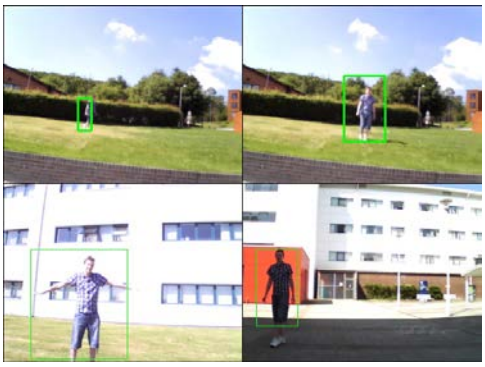

Figure 4. Outdoor Images

The results of this experiment prove that it is possible to detect a human in stereoscopic imagery where the camera is subject to movement. Utilising the parallax effect allows detection of the foreground region of the images. The algorithm developed works in a similar way to disparity mapping. The algorithm successfully detects a small region for further investigation by existing techniques such as disparity when the assumption of the person being the forward focal object in the scene is true. Most importantly the tests prove that the system has three major advantages: the camera is not fixed as it is not subject to background model initialization; the system is unaffected by lighting variations between frames; and the system works well outdoors.

6. Discussion

The scope of the system developed reaches into multiple fields of computer vision such as motion capture and augmented reality. The basic principle has been proven, allowing further development of systems that are dependent upon knowing the difference between background and objects in a scene efficiently. Systems can be developed from this that are able to track the objects

detected between frame to frame without compromising upon the flexibility of the camera movement.

One of the major issues associated with this problem is computational expense. Real-time human region detection is possible with our system with VGA resolution images being analysed up to 120fps, which is twice that our camera was capable of recording. The two separate images are combined and analysed, and the human motion detected and extracted. The system makes use of the parallax effect of objects in a way that conventional stereoscopic systems do not. Traditional systems use the data to reconstruct a depth map of the scene. In this research, the parallax data was filtered and grouped causing closer objects to be extracted quickly. The result is a system that produces extremely quick approximations of a person's location.

Although the system is designed to detect people in a scene when the assumptions are valid, the output has far more potential. This system quickly identifies regions of an image that include objects. The algorithm could have abilities in the future to be used for tasks such as the Mars automated missions. Alternate systems have been designed using conventional means such as disparity mapping [18]. The system presented has the ability to accompany such systems for an initial processing option pointing out regions that could potentially obscure the route of the rover. Through improving the algorithm to provide the outline of the detected person as well as the region, the system can be used for augmented reality applications. Although augmented reality was the primary motivation, the system shows potential for recognising and modelling human movement. This will allow for an effective motion capture piece of software that could be used in small rendering companies due to the low system cost.

7. Conclusion

This paper documents a system for use in stereoscopic vision that has industrial application potential. Although the system is not as accurate as some of its predecessors, there are some significant advantages, specifically: in that the camera can move freely without any initialisation between location changes; the processing is unaffected by light variation between frames; the system works well in both indoor and outdoor environments; and the system can process extremely quickly. Each frame is independent of previous frames due to the comparison of the left and right imagery. The algorithm developed has significant possibilities for enhancement in the future to develop a system that has all the capability of its predecessors while maintaining the advantage of speed and camera mobility.

The algorithm is designed to be used as a first filter to extract feature points in input video frames faster than real-time. The system presented here is designed to be a novel natural feature tracking system that works with a dynamic background due to the camera not being bound to a single fixed point. Instead, the camera is subject to 6DOF allowing it to move freely and still process at a high frame rate. In further work, the algorithm is going to be enhanced by having a lower level representation of the scene by grouping pixels together in smaller collections. These collections can then be compared between frames generating a representation of the world through spatial world mapping. The next phase is to allow for multiple grouping and recognition of different objects such as in the figure where the feature has been collected along with the person. Giving the algorithm the ability to detect change in type of object e.g. colour variation has the potential to stop false groupings.

Acknowledgements: This work received funding from the European Social Fund (ESF) through the European Union's Convergence programme administered by the Welsh Government.

References

[1] J. Amat, A. Casals and M. Frigola, "Stereoscopic System for Human Body Tracking in Natural Scenes," IEEE International Workshop on Modelling People, Kerkyra, Greece, 1999.

[2] C. R. Wren, A. Azarbayejani, T. Darrell and A. P. Pentland, "Pfinder: real-time tracking of the human body," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780 - 785, 1997.

[3] T. Bloom and A. P. Bradley, "Player Tracking and Stroke Recognition in Tennis Video," in Proceedings of the APRS Workshop on Digital Image Computing. APRS Workshop on Digital Image Computing (WDIC'03), Brisbane, Australia, pp. 93-97, 2003.

[4] K. Teachabarikiti, T. H. Chalidabhongse and A. Thammano, "Players Tracking and Ball Detection for an Automatic Tennis Video Annotation," in: Proceedings of 11th International Conference on Control, Automation, Robotics and Vision, Singapore, 2010.

[5] L. M. Fuentes and S. A. Velastin, "People tracking in surveillance applications," Journal of Image and Vision Computing, vol. 24, pp. 1165–1171, 2006.

[6] Y. Hashimoto, Y. Matsuki, T. Nakanishi, K. Umeda, K. Suzuki and Takashio, "Detection of pedestrians using subtraction stereo," in: Proceedings of International Symposium on Applications and the Internet, Turku, 2008.

[7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," IEEE Trans. PAMI, vol. 32, no. 9, 2010.

[8] X. Zabulis, D. Grammenos, T. Sarmis, K. Tzevanidis, P. Padeleris, P. Koutlemanis and A. A. Argyros, "Multicamera human detection and tracking supporting natural interaction with large-scale displays," Journal of Machine Vision and Applications, pp. 1-18, 2012.

[9] R. Koch, "3-D surface reconstruction from stereoscopic image sequences," in: Proceedings of Fifth International Conference on Computer Vision, Cambridge, MA, USA, 1995.

[10] F. Devernay and 0. D. Faugeras, "Computing differential properties of 3-D shapes from stereoscopic images without 3-D models," in: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 1994.

[11] L. Falkenhagen, "Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuous Surfaces," Image Processing for Broadcast and Video Production, pp. 115–127, 1994.

[12] W.-S. Jang and Y.-S. Ho, "Efficient Disparity Map Estimation Using Occlusion Handling for Various 3D Multimedia Applications," IEEE Transactions on Consumer Electronics, vol. 57, no. 4, pp. 1937-1945, 2011.

[13] T. Darrell, G. Gordon, M. Harville and J. Woodfill, "Integrated Person Tracking Using Stereo, Color, and Pattern Detection," International Journal of Computer Vision, vol. 37, no. 2, pp. 175-185, 2000.

[14] K. Umedaa, Y. Hashimotob, T. Nakanishib, K. Irieb and K. Terabayashia, "Subtraction stereo - A stereo camera system that focuses on moving regions," in: Proceedings of Three-Dimensional Imaging Metrology, San Jose, CA, USA, 2009.

[15] K. Terabayashi, Y. Hoshikawa, A. Moro and K. Umeda, "Improvement of Human Tracking in Stereoscopic Environment Using Subtraction Stereo with Shadow Detection," International Journal of Automation Technology, vol. 5, no. 6, pp. 924-931, 2011.

[16] C. G. Keller, M. Enzweiler, M. Rohrbach, D. F. Llorca, C. Schnörr and D. M. Gavrila, "The Benefits of Dense Stereo for Pedestrian Detection," IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1096-1106, 2011.

[17] C. G. Keller, T. Dang, H. Fritz, A. Joos, C. Rabe and D. M. Gavrila, "Active Pedestrian Safety by Automatic Braking and Evasive Steering," IEEE Transactions on Intelligent Transportation Systems, vol. 99, pp. 1-13, 2011.

[18] S. B. Goldberg, M. W. Maimone and L. Matthies, "Stereo Vision and Rover Navigation Software for Planetary Exploration," in: Proceedings of IEEE Aerospace Conference Proceedings, Big Sky, Montana, USA, 2002.