

PEDESTRIAN DETECTION USING STEREO AND BIOMETRIC INFORMATION

Philip Kelly ^{*}, Eddie Cooke, Noel O'Connor and Alan Smeaton

Centre for Digital Video Processing, Adaptive Information Cluster, Dublin City
University, Ireland

Abstract. *A method for pedestrian detection from real world outdoor scenes is presented in this paper. The technique uses disparity information, ground plane estimation and biometric information based on the golden ratio. It can detect pedestrians even in the presence of severe occlusion or a lack of reliable disparity data. It also makes reliable choices in ambiguous areas since the pedestrian regions are initiated using the disparity of head regions. These are usually highly textured and unoccluded, and therefore more reliable in a disparity image than homogeneous or occluded regions.*

1 Introduction

Real-time pedestrian detection and tracking is valuable in many situations, such as security applications in airports, the study of crowd flow patterns for city planning, or automated driving assistants. Many computer vision based applications [1–3] depend on accurate detection and segmentation of pedestrians from a given background as a first step in their algorithmic process.

Various different techniques for segmenting individual pedestrians have been investigated using with 2D computer vision techniques. Some require a certain camera orientation which is difficult to achieve in outdoor scenarios [4], others search images for pedestrian shapes [3]. Multiple cues, such as skin color and face detection, have also be applied [5]. The results of many approaches are depreciated in unconstrained real-world environments due to dynamic conditions such as rapidly changing lighting conditions causing shadows, pedestrian occlusion and the large variability in a pedestrians local and global appearance due to pose, orientation and clothing. 3D stereo information has been proposed as a technique to guide pedestrian detection, as stereo and shape are more reliable and helpful cues than color and face detection in general situations. The use of stereo information carries with it some distinct advantages over conventional 2D techniques [6]: (a) it allows explicit occlusion analysis and is robust to illumination changes; (b) the real size of an object derived from the disparity map provides a more accurate classification metric than the image size of the object; (c) using stereo cameras can detect both stationary and moving objects.

^{*} This material is based on works supported by Science Foundation Ireland under Grant No. 03/IN.3/I361.

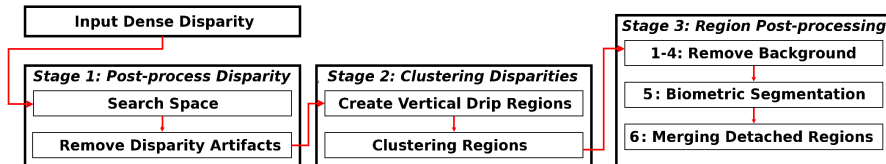


Fig. 1. System Overview

In this paper we define a pedestrian detection technique that can be applied using a stereo camera system that is located, as most surveillance cameras are, above human height and orientated at approximately a 45 degree angle to look down on pedestrians below, making this technique applicable to both indoor and outdoor scenarios. The proposed technique is based on the use of disparity information, ground plane estimation and human biometric information. It can detect pedestrians even in the presence of severe occlusion or a lack of reliable disparity data. It also makes reliable choices in ambiguous areas since the pedestrians are initiated using the disparity of head regions. These are usually highly textured and unoccluded, and therefore more reliable in a disparity image than homogeneous or occluded regions. As the technique uses disparity information, pedestrian characteristics, such as the variability in clothing colour, for example, does not affect the approach. In addition, biometric information, based on the *Golden Ratio* is used to remove regions that do not adhere to a pedestrians global shape.

This paper is organized as follows: Section 2 presents the details of the developed algorithmic approach. Firstly, an overview of post-processing the dense disparity estimation process; we then illustrate how individual pedestrians are segmented and post-processed. In Section 3 we present experimental results from a real world outdoor situation containing multiple pedestrians at various depths, some with severe occlusion, displaying a large variability in both local and global appearance. Finally, Section 4 details conclusions and future work.

2 Algorithm Details

Figure 1 illustrates an overview of the system. As a starting point, a dense disparity image is obtained using the technique defined in [7]. This technique can suffer, as can all other dense disparity estimation algorithms, from inaccuracies. These artifacts can be removed by post-processing the resultant disparity image, described in Section 2.2. This stage of the algorithmic process, as shown in Figure 1, is split into two separate steps. The first step removes disparity points that are outside a predefined search space. The second step removes artifacts that were generated during the disparity estimation process itself.

The resultant *foreground* disparities are then clustered together, as described in Section 2.3. Initially, so called *Vertical Drip Regions* are created which are regions of 1 pixel wide and 1 or more pixels in height. These regions are then

clustered together using knowledge of the camera position and biometric information.

Finally, in Section 2.4, the clustered regions are post-processed into individual pedestrians. As shown in Figure 1, this is a 6 step process. The first four steps remove background regions, which includes regions due to noise. The fifth step applies biometric information to segment multiple pedestrians that exist in an individual region. The final post-processing step merges, if necessary, two or more regions that belong to the same pedestrian.

2.1 Camera Position and Orientation

The technique described in this paper has been designed specifically for surveillance type applications. Most surveillance camera systems are located above human height and orientated at approximately a 45 degree angle to look down on activity below. This camera setup can be easily applied to both indoor and outdoor scenarios. Our stereo camera system emulates this setup and is attached to a crossroads traffic light at approximately 2.5 meters above the ground. Due to position of the camera, pedestrians can have a large range of disparities, an example of this can be seen for pedestrian **g** in Figure 4(e). Simply breaking the input image into disparity layers will not result in the detection of individual pedestrians. The camera setup, however, can be advantageous when detecting pedestrians using stereo information. For instance, depending on the distance of a pedestrian from the camera, the following observations can be made:

1. a pedestrian can have roughly the same disparity throughout its region;
2. the parts of a pedestrian which are higher above the groundplane, such as the head and shoulders, can have greater disparity than the rest of the pedestrians body;
3. if a pedestrian, ped_1 , is occluded by another pedestrian, ped_2 , then ped_2 must be closer to the camera than ped_1 and therefore, in general, ped_2 has a greater disparity than that of ped_1 .

2.2 Dense Disparity Estimation

The observations made in Section 2.1 assume that ideal dense disparity has been obtained and is used as the input to the algorithm. This assumption is not realistic, as every disparity estimation technique proposed in literature to date is subject to errors, especially in regions of homogeneous colour or occlusion. The technique used to obtain dense disparity information used in this paper suffers, as all others techniques do, from inaccuracies. These artifacts can be greatly reduced by integrating certain constraints into the dense disparity estimation algorithm and by post-processing the resultant disparity image.

The technique used to obtain dense disparity information in this paper is based on the work described in [7], that details a dynamic programming based stereo correspondence technique that has been specifically developed for pedestrian surveillance type applications. It employs two background models, a background disparity and an edge model. In addition, it introduces a technique for

obtaining a *dynamic* disparity limit constraint, whereby a separate disparity limit is defined for each scanline based on the background disparity model, the previous scanlines maximum disparity and Ground Control Points (GCPs). An advantage to this is that if an inaccuracy occurs within the dense disparity estimation, then the disparity will be *equal to or lower* to the previous scanlines maximum disparity unless there is an error in either the background model or the GCPs. This is an important constraint as although the disparity image may contain errors, we can still expect that, in general, the three observations set out in Section 2.1 to hold true. Figures 5(b) and (f) show results of the dense disparity estimation technique.

The technique described in this paper post-processes the dense disparity image for two reasons; (1) we are interested in detecting pedestrians within a certain depth range, therefore we can remove points that are outside a predefined search space; (2) to remove artifacts that were generated during the disparity estimation process.

Defining a Search Space The first post-processing step involves removing disparity points that are outside a predefined search space. However, in order to achieve this, the search space must be defined. In general, the scenes from which the detection of pedestrians is required are man-made scenes such as crossroads, streets, airports, railway stations, etc. A common factor to all these scenarios is a relatively flat ground plane. Information about this 3D plane can be advantageous as it allows us to gauge the position of a 3D point with respect to the *scene* in a more effective way than can be achieved using just its position with respect to the camera rig.

A geometric constraint called a *homography*, as described in [8], is used to map points on a plane in one image to the corresponding plane points in a second image of the same scene. Using a homography it is possible to obtain corresponding ground plane points between two different images of the same scene. From these correspondences disparity is obtained, and thus the 3D information of multiple ground plane points are determined. A Least Mean Squares (LMS) approach is then applied to obtain an equation of the 3D groundplane from the multiple 3D input points. This 3D ground plane is used in conjunction with pedestrians' biometric information and assumptions about pedestrian pose to define a predefined search space.

For each 2D point in the dense disparity image, the 3D information is determined from the rectified images using triangulation, as defined in [9]. The minimum Euclidean distance between the 3D ground plane and each 3D point is then determined via:

$$height_{euc} = \frac{A * x^p + B * y^p + C * z^p + D}{\sqrt{A^2 + B^2 + C^2}} \quad (1)$$

where $A * x + B * y + C * z + D = 0$ is the equation of the 3D ground plane and (x^p, y^p, z^p) is the 3D point. Each point is then defined as either *background* or *foreground*. A point is defined as background and removed from the disparity image if any of the following is true:

1. No disparity data exists at that point due to occlusion.
2. If $z^p > z_{max}$, where z_{max} is a threshold value that represents the maximum relevant z^p value. In our experiments, z_{max} is set to 8 meters, due to the image resolution and the degradation of accurate stereo information beyond this distance.
3. $height_{euc} < height_{min}$. In our experiments, $height_{min}$ is set to 0.9 meters.
4. $height_{euc} < height_{max}$. In our experiments, $height_{max}$ is set to 2.4 meters.

$height_{min}$ and $height_{max}$ are set to the minimum and maximum expected pedestrian height above the groundplane, but allowances are made for fluctuations in the groundplane and disparity inaccuracies.

Removing Dense Disparity Artifacts In unconstrained conditions inaccuracies can appear in the output of dense disparity estimation techniques. These are mainly due to lack of texture, occlusion or inefficiencies inherent in the chosen stereo correspondence algorithm. How these artifacts manifest themselves depends upon the stereo correspondence algorithm. In dynamic programming based algorithms they tend to take the form of horizontal streaks of inaccurate disparity data. A typical example of this can be seen in Figure 2(a), which shows a closeup of pedestrian **e** from Figure 4 (b). Note the lack texture in the midsection of the pedestrian due to the colour of the pedestrians coat being similar to the background colour on *both* sides. After the first post-processing step, a large streak can be seen through this area, see Figure 2(c). Also notice smaller streaks in the top left of the image and at the back of the pedestrians head. Fortunately, the streaks can be removed by searching the image for the characteristic vertical bar of constant disparity.

In order to remove streaks the image is traversed vertically. When a pixel, (x, y) , is a foreground point and the previous pixel, $(x, y - 1)$, is not, the y value is noted as the start of a strip, y_{start} . The value of y is incremented until (x, y) is a background point, whereupon the y value is noted as the end of the strip, y_{end} . The height of a strip is determined via $height_{strip} = y_{end} - y_{start}$. For a strip to be part of a streak, then:

- $(x + 1, y_{start} - 1), (x + 2, y_{start} - 1) \dots (x + width_{strip}, y_{start} - 1)$ must be a background point, where $width_{strip} = height_{strip} * \beta$, where β determines how wide a streak must be compared to its height for it to be retained. In our experiments $\beta = 1$.
- All points $(x + 1, y), (x + 2, y) \dots (x + width_{strip}, y)$ must be foreground points and have the same disparity value as (x, y) , where y is any value from y_{start} to y_{end} .
- $(x + 1, y_{end} + 1), (x + 2, y_{end} + 1) \dots (x + width_{strip}, y_{end} + 1)$ must be background points.

If a streak is found then all points within it are declared as background points. $width_{strip}$ is then incremented and the next line is tested. If the strip continues onto this next column and the above 3 properties hold then the strip on this line



Fig. 2. Remove Streaks; (a) Input Region; (b) Dense Disparity; (c) Foreground Disparity; (d) Found Streaks (in Red); (e) Post-processed Disparity

is also removed and $width_{strip}$ is incremented again. Figure 2(d) shows in red the regions that have been determined to be streaks in the foreground disparity image.

2.3 Clustering Regions

Vertical Drip Regions The clustering of foreground disparities is a two step process. In the first step, regions that are 1 or more pixels in height but *only* 1 pixel in width are created. These resemble drips of paint, and so we refer to them as *vertical drip regions*. To create vertical drip regions, each column y of the image is traversed vertically, from top to bottom. A new region is created at the pixel, (x, y) if; (a) $(x, y - 1)$ was a background point; (b) the disparity at (x, y) is greater than the disparity at $(x, y - 1)$. Otherwise (x, y) is clustered into the same region as $(x, y - 1)$; or (c) (x, y) is an edge. We create a new vertical dip region at every edge otherwise drips emanating from regions above pedestrians heads, such as high walls, will pass down the image covering pedestrians below. Results from this step can be seen in Figure 4(f), where each colour represents a different region.

Clustering Regions The second step involves merging these drip regions to form more coherent foreground objects. The merging of these drip regions needs to be achieved in a way that the final regions adhere to the three rules set out in Section 2.1. These rules can be summed up by the statement that, for a single object, as the y value *increases* the disparity remains *constant* or *decreases*. We will refer to this as the *Decreasing Disparity Rule*.

In order to merge two regions, r_1 and r_2 , and adhere to this rule, two separate tests must be made; (a) two neighbouring pixels, p_1 from r_1 , and p_2 from r_2 , must have similar disparities, and; (b) if r_1 has a lower minimum disparity than r_2 then r_1 must have a greater, or similar, disparity to r_2 , or vice versa.

Two disparities, $disp_1$ and $disp_2$, are declared to have similar disparities if the following is true:

$$\sqrt{(disp_1 - disp_2)^2} \leq \frac{disp_x * \theta}{\sigma} \quad (2)$$

where $disp_x$ is the minimum of $disp_1$ and $disp_2$, and θ is the maximum disparity difference allowed for every σ pixels of disparity that exists in $disp_x$. For example,

if $\theta = 1.5$ and $\sigma = 10$, this allows 1.5 pixels disparity difference when $disp_x = 10$, 3 pixels disparity difference when $disp_x = 20$, and so on. This allows pixels closer to the camera to cluster together easier, allowing for fluctuations in an objects surface.

In order to implement the second test, a value must be found for the disparity of a *region*. Using the maximum disparity of a region would make the assumption of a perfect input disparity image, but a single erroneous high disparity value may split a pedestrian into two separate regions that would never be allowed to merge. The average disparity of the *whole* region is also not adequate, due to the possibility of occlusion of regions. Figure 3 illustrates this point. Pedestrian **b** is occluded by pedestrian **a**. Due to this only pedestrian **b**'s head can be seen, which is the area of pedestrian **b** with the *highest* disparity. The *whole* of pedestrian **a** can be seen. However, due to the lower disparity of pedestrian **a**'s torso, the average disparity of the region of pedestrian **a** could drop below that of pedestrian **b**, thereby allowing the two regions to merge. This would then violate the *Decreasing Disparity Rule*. A solution to this problem is to use biometric information.

Applying Biometric Information The height above the groundplane, found using Equation 1 for defining a search space in Section 2.2, can also be used to define the proportions of a human body by applying the number Φ , ($\Phi = \sqrt{5} * 0.5 + 0.5 \simeq 1.618$). This number is known as the *Divine Proportion* or the *Golden Section/Ratio/Mean/Number* [10]. Figure 3 (a) shows how a body is segmented using Φ . Let $|aj|$ be the Euclidean distance between the horizontal lines a and j . Therefore, $|aj|$ is the height of a human body. Using Φ and $|aj|$ various other points on the human body can be defined. In Figures 3 (a) and (b) $|ai| = \frac{|aj|}{\Phi}$, $|ah| = \frac{|ai|}{\Phi} \dots |ab| = \frac{|ac|}{\Phi}$ [10] and $|mn|$ is equivalent to $|ae|$. Similarly $|lo| \equiv |ag|$ and $|kp| \equiv |ah|$. Distances of interest include; $|ac|$, which is the distance from the head to the forehead; $|ae|$ which is the width of the head; $|af|$ which is the the distance from the head to the base of the skull; $|ag|$ which is the width of the shoulders; and $|ah|$ which defines the distance from the head to the navel and the elbows.

This biometric information constitutes useful information for assisting region clustering. Statistics, such as a region's maximum height and bounding box, and therefore the value of $|ac|$ are built up about each region as they are created. The value $|ac|$ is chosen as, in our scenario, a pedestrian is assumed to be unoccluded by at least this amount in an image. The average disparity is then summed for all pixels that occur between the minimum y value of the region, y_{min} , and $y_{min} + |ac|$. In Figure 3 (f) the two regions are obtained using the proposed technique to merge vertical drip regions. Figures (h) and (i) show the biometric data of two regions overlaid onto the image. From this it can be seen that for each region the average disparity is calculated from region pixels between the lines **a** and **c**.

Merging is implemented in two stages, in the first stage, $\theta = 0$, and in the second stage $\theta = 1.5$ and $\sigma = 10$. The first stage is implemented to allow initial

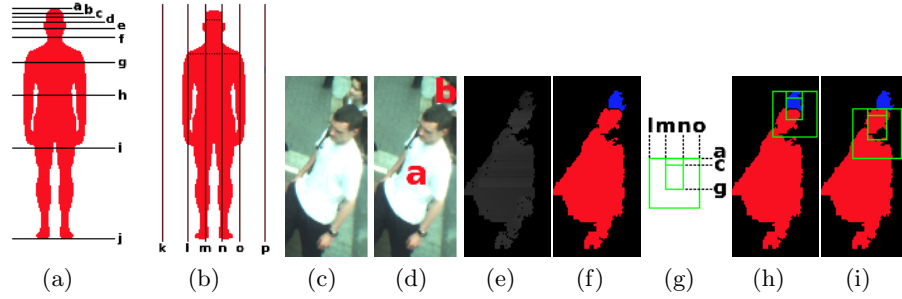


Fig. 3. Golden Ratio; (a) Vertical; (b) Horizontal; (a) Initial Image; (b) Tagged Image; (c) Disparity; (d) Regions; (e) Dense Disparity; (f) Final Regions; (g) Biometric Data with respect to (a); (e) Biometric Pedestrian **b**; (f) Biometric Pedestrian **a**;

regions to grow in order to obtain more accurate region statistics before the second stage, which allows pixels closer to the camera to cluster together easier, allowing for fluctuations in disparity in an objects surface.

2.4 Region Post-processing

The clustered regions are post-processed to obtain a single pedestrian in each region. The post-processing is split into six distinct steps. The first four steps removes background regions, which includes regions due to noise. The fifth step applies biometric information to segment multiple pedestrians that exist in an individual region. The final post-processing step merges, if necessary, two or more regions that belong to the same pedestrian.

The first step removes noise regions. These regions generally occur due to small inaccuracies in the disparity image. Regions that have a number of pixels less than a threshold, $pixels_{min}$, are removed. In addition, as the height of a pixel is known, if the absolute difference between the maximum and minimum height values appearing in a region is less than a threshold, $height_{min}$, then the region is also removed. In our experiments $pixels_{min}$ is set to 1000, assuming this to be the minimum size of a pedestrian in pixels, and $height_{min}$ is as defined in Section 2.2.

The second step is used to remove large background regions. This is achieved using the background edge model referred to in Section 2.2. Let the foreground image edges be referred to as *edge activity points*. All the edge activity points that are within the search space, as defined in Section 2.2, are clustered together in to *edge activity regions*, *EARs*. Then, as before, if the maximum height difference between all the pixels in the *EAR* is less than a threshold, $height_{min}$, then the *EAR* is removed. Figure 4(c) shows the edge activity points (in red) that pass this test. Notice how there is no *EAR* pixels on the background wall. Finally, all foreground regions are removed that do not have at least one *EAR* pixel contained in that region.

The third step is almost identical to Section 2.2, and removes streaking effects caused by the disparity estimation process. It differs from Section 2.2 as each region is taken in isolation and searched for streaks, whereas before all foreground points were taken all at once. This is necessary as a streak could easily join two separate pedestrians together.

The fourth post-processing step finds the bounding box of *EARs* in each region and then trims the region to remove any pixels outside this bounding box. It then checks each region to make sure it has more pixels than $pixels_{min}$. This removes any remaining background regions that exist due to a small number of *EAR* pixels that are locally spread in the region. The final two post-processing steps are biometric segmentation followed by merging of detached regions.

Biometric Segmentation Biometric information, introduced in Section 2.3, can be applied to segment multiple pedestrians that exist in an individual region. Each region is taken independently and its width is examined. If the width of the region on any given scanline is greater than the estimated width of the pedestrians shoulders, $|lo|$, then the possibility exists that the region is more than a single pedestrian.

If two pedestrians become merged into the same region, it is assumed that the pedestrians are at roughly the same depth, and therefore standing in a side by side orientation. Two such pedestrians **a** and **b** in Figure 4(g) have been merged. From the shape of this region the separation point of the two pedestrians is visible due to the rise and fall of height in the region. The region contains two maximum height points which are the two heads and a minimum height point which is where the two bodies meet. It is this characteristic rise and fall of height greater than $|ac|$ that is used to segment multiple pedestrians in a single region. In addition, at the point of a drop, there must be a distance of the width of a head, $|mn|$, within the region on both sides of the drop. This is necessary as otherwise areas of a single pedestrian, such as extended arms, may be incorrectly split into a separate region. If more than one possible drop and rise is found, then the one with the largest drop in height is selected, and the region is split at the mid-distance between the drop and the rise. The two separate regions are then independently tested to see if any more pedestrians are located within that region.

If, however, no drop and rise can be found within the region, the width of the region is again examined. If its width is greater than $|kp|$ then it is removed as it is an object that has not of the correct dimensions of a pedestrian, otherwise the region is left intact.

Merging Detached Regions The final post-processing step merges, if necessary, two or more regions that belong to the same pedestrian. Figure 4 (g) shows that, due to the lack of reliable disparity information, pedestrian **e** has been split in two. Although their positions and average disparities indicate that the two could merge, this has been impossible up until this point due to the lack of any foreground disparity information between the two regions.

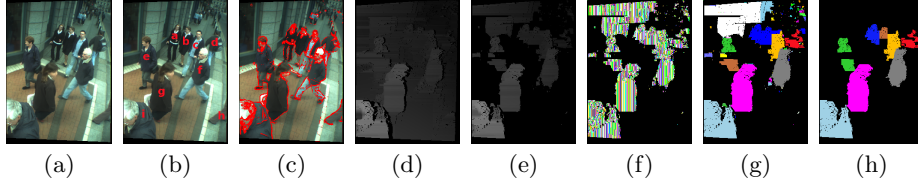


Fig. 4. Results; (a) Input Image; (b) Tagged Pedestrians; (c) Edge Activity; (d) Dense Disparity; (e) Foreground Disparity; (f) Vertical Drip Regions; (g) Merge Drip Regions; (h) Final Pedestrian Regions;

In order to merge two distinct regions, r_1 and r_2 , the center point, cp , of each region is obtained using the regions bounding box. If either: (a) cp_1 is within the x bounds of r_2 and cp_2 is within the x bounds of r_1 ; or (b) the x bounds of r_2 are contained within the x bounds of r_1 ; or (c) the x bounds of r_1 are contained within the x bounds of r_2 ; then it is possible that the two regions could be merged together. To determine if the two regions can merge then the region with the highest minimum y value for its cp is determined, for argument sake, let this be the region r_1 . The two regions are then merged if: (a) cp_1 is lower down in the image than r_2 's bounding box; (a) cp_2 is higher up in the image than r_1 's bounding box; (c) either Equation 2 holds for the average disparity of r_1 and the *maximum* disparity of r_2 , or r_1 's average disparity is greater than the *maximum* disparity of r_2 .

Finally, the first four post-processing steps are repeated on any new regions that have been created. Figures 4(h) and 5(d) and (h) show examples of the final post-processed regions, where each colour region represents an individual pedestrian.

3 Experimental Results

Figures 4 and 5 show results from various stages of the algorithmic process. These images are from a real world outdoor scenario with unconstrained conditions. They contain difficult scenarios where multiple pedestrians at various depths, some with severe occlusion, are detected. The pedestrians display a large variability in both local and global appearance. In some cases there is no distinct point of separation between pedestrians and other objects, whether they be other pedestrians or background. Inaccuracies in the dense disparity estimation process occurs in a number of these examples, but the inaccuracies are removed or overcome by the technique.

Table 1 displays an overview of results for 1000 images at various pedestrian densities. num_{ped} and num_{det} indicates the total number of pedestrians that exist and that are found in this search space respectively within these images. $\%_{found}$ defines the overall percentage of pedestrians correctly segmented. A correctly segmented pedestrian is defined as a region that contains at least the pedestrians head and no substantial area of a second pedestrian. The final

Image Pedestrian Numbers	num_{ped}	num_{det}	$\%_{found}$	$\%_{mult}$	$\%_{back}$	$\%_{disp}$
1-3	964	850	88.17	60.53	7.02	21.05
4-6	1186	954	80.44	66.38	14.66	6.47
>6	521	407	78.12	59.65	15.79	8.77
Total	2671	2211	82.78	63.26	13.04	10.65

Table 1. Results Overview

four columns of the table represents the percentage of pedestrians that were not segmented correctly for the following reasons: $\%_{mult}$ due to the region containing more than one pedestrian; $\%_{back}$ due to the pedestrian becoming merged into the background, which is caused by the pedestrian having almost an identical disparity as an area of the background model, such as a wall, and finally removed by biometric segmentation; $\%_{disp}$ due to a severe lack of reliable disparity and edge information for the pedestrian.

From these results it can be seen that 63.26% of the pedestrians that have *not* been segmented correctly can be attributed to the pedestrian becoming merged with another pedestrian. These results point could be improved by improving the biometric segmentation section. However, it is unrealistic to define and search for every possible orientation of a pedestrian that has become joined together, in a single region, with another object. However the use of temporal information could prove very useful as an additional pointer to pedestrian segmentation. Temporal information could also be employed as a feedback into the dense disparity estimation technique to further improve results, possibly helping to detect the 23.69% of missed pedestrians defined by $\%_{back}$ and $\%_{disp}$.

4 Conclusions and Future Work

This paper described a pedestrian detection technique based on disparity information, ground plane estimation and biometric information. The approach can detect pedestrians even in the presence of severe occlusion or a lack of reliable disparity data and will make reliable choices in ambiguous areas. As the technique uses disparity information, pedestrian characteristics, such as the variability in clothing colour, for example, does not affect the approach. In addition, biometric information, based on the *Golden Ratio* is used to remove regions that do not adhere to a pedestrians global shape.

In future work, the use of temporal data along with the improvement of the dense disparity algorithm to enforce better inter scanline consistency would increase the accuracy of both dense disparity estimation and pedestrian segmentation. In addition, the pedestrian model can be improved by taking into account the projective distortion of the image and the camera position relative to the groundplane. The biometric segmentation can also be improved to included other common cases when pedestrians get clustered together and there is no characteristic rise and fall of pedestrians heads. Other biometric informa-

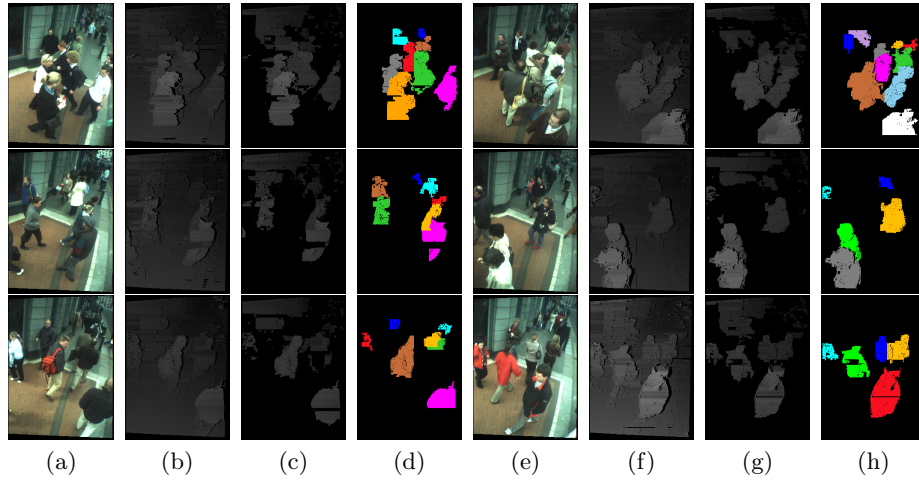


Fig. 5. Results; (a)/(e) Input Image; (b)/(f) Dense Disparity; (c)/(g) Foreground Disparity; (d)/(h) Final Pedestrian Regions;

tion such as skin colour could also be detected and used for the segmentation of possible heads within a region.

References

1. Uchida, K., Miura, J., Shirai, Y.: Tracking multiple pedestrians in crowd. In: Workshop of Machine Vision and its Applications. (2000) 533–536
2. Pai, C., Tyan, H., Liang, Y., Liao, H., Chen, S.: Pedestrian detection and tracking at crossroads. In: PR. Volume 37. (2004) 1025–1034
3. Gavrilu, D., Giebel, J.: Shape-based pedestrian detection and tracking. In: IEEE Intelligent Vehicles Symposium. (2002) 215–220
4. Terada, K., Yoshida, D., Oe, S., Yamaguchi, J.: A counting method of the number of passing people using a stereo camera. In: IEEE Conference on Industrial Electronics. Volume 3. (1999) 1318–1323
5. T. Darrell, G. Gordon, M.H., Woodfill, J.: Integrated person tracking using stereo, color and pattern detection. *International Journal of Computer Vision* **37**(2) (2000) 175–185
6. Zhao, L., Thorpe, C.: Stereo and neural network-based pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems* **1**(3) (2000) 148–154
7. Kelly, P., Cooke, E., O’Connor, N., Smeaton, A.: 3d image analysis for pedestrian detection. In: Image Analysis for Multimedia Interactive Services (WIAMIS). (2006) 177–180
8. Kelly, P., Beardsley, P., Cooke, E., O’Connor, N., Smeaton, A.: Detecting shadows and low-lying objects in indoor and outdoor scenes using homographies. In: The IEE International Conference on Visual Information Engineering. (2005) 393–400
9. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis and Machine Vision, Second Edition. PWS Publishing (1999)
10. : <http://www.goldennumber.net> (2005)