Proceedings of the Technical Sessions, 26 (2010) 17-24
Institute of Physics – Sri Lanka

Pedestrian Collision Detection through Monocular Vision

C.T. Wannige and D.U.J. Sonnadara
Department of Physics, University of Colombo, Colombo 3

ABSTRACT

In this paper, a depth detection based on monocular vision coupled with the motion was used to predict the possible collisions on pedestrian crossings. A video taken while in motion is extracted into frames and processed with skin color detection methods to identify the face blobs of pedestrians. The distance to the pedestrian was determined using the difference in size of the face blobs in the consecutive image frames. The field tests show that the size variation does not depend on the speed of the vehicle but fewer steps are available for processing. The breaking distance which depends on the speed increases with the speed of the vehicle. A trigger for the breaking signal can be fired depending on the face blobs' size variation. Reasonable results are observed from the field tests, when theoretical and experimental breaking distances are compared.

1. INTRODUCTION

Automatic detection and verification of pedestrian movement through image processing is a central challenge in computer vision because of the large variation in pedestrian appearances and environmental outlooks. Many interesting approaches can be found in literature for pedestrian detection. Normally, sensors are used for detecting pedestrians, such as Cameras, Near IRs, Thermal IRs, RADAR and Laser scanners. Among them, RADAR is having a higher detection rate [1] but they are expensive compared to cameras. Imaging sensors capture detailed description of the scene but involves high amount of processing. On the other hand, sensors like RADAR and LASER scanners give the information about object distance, their resolution is limited [1].

Stereo vision is a powerful method for depth perception which can be used as a cue to avoid the collision between pedestrians and vehicles. However, the stereo vision based methods are complex due to two cameras which require high processing power. Thus, they are computationally expensive [2]. In this paper, we are trying to see to what degree one can relay on monocular vision for depth detection. To our knowledge limited number of researches are carried out for pedestrian detection using monocular vision [2, 3].

Shape based approaches for pedestrian detection uses characteristic features from the images to be used as its trained classifier. However, the training needs a large number of image sets and time as well as huge processing power. Moreover, there are many algorithms for object detection and tracking such as correlation based methods, optical flow based methods, motion based methods, frame difference based methods, model based methods, and contour based methods [2, 3]. Processing speed is important for pedestrian detection and it should predict the breaking distance simultaneously by tracking the pedestrian movement. The frame difference technique is widely used to detect moving objects due to the low computational complexity [2].

Color appearance based obstacle detection methods are used in monocular vision where each individual pixel is classified as belonging to either to an obstacle or to the background based on its color appearance. Although there are many methods to choose from, the tracking method for pedestrians should be independent of the background to adapt to any location and work efficiently as images should be processed and decision has to be taken within few seconds. Since color space based methods work efficiently with reasonable speeds, in this work, color space method was chosen to detect the pedestrians. The head was identified by fitting a circle inside each skin color region and filtering wrong candidates based on radius values.

Depth perception is done frequently using binocular vision. Binocular stereo can be used for depth perception based on disparity analysis [2, 3]. However, few researches have used monocular vision for depth perception using feature based methods, such as size/ scaling, vertical-horizontal projection, scaling of supervised regions [2, 3]. Simple feature extraction algorithms should be used in pedestrian detection for avoiding collision. In this work a feature based size/scaling method for prediction of collision was used. The main scope of this work is to identify a pedestrian from monocular vision simultaneously with the video taken while in motion and generate a trigger to the vehicle’s breaking system to stop the vehicle at the required breaking distance.

2. PEDESTRIAN DETECTION

2.1 Skin color segmentation

The first objective is to identify the pedestrians using an efficient and simple technique. The method used in this work was based on visual tracking. The video taken during the movement of the vehicle is extracted into image frames simultaneously. The extracted image frames are processed using image processing techniques to detect the pedestrians efficiently.

Classifying pedestrian from other objects in a complex background has many challenges. First, the background region should be removed from the image. Among background removal methods, dynamic background creation is done using a series of images by finding the mean and standard deviation of each and every pixel [4]. The dynamic background is created using the calculated mean and the standard deviation values. However, this method needs extra time for processing as well as a series of background images. On the other hand static background subtraction requires a good background image [5]. Although there are many methods to choose from, the tracking method for pedestrians should be independent of the background to adapt to any location and work efficiently as images should be processed and decision has to be taken simultaneously while the vehicle is moving. Since color space based methods work efficiently with reasonable speeds, it was decided to use skin colour to identify pedestrians which can be applied to even single image.

There are two main methods of skin detection using color spaces, pixel based and region based. Pixel based methods check each and every individual pixel to classify as skin or non skin using color. Region based methods consider the spatial arrangement of the skin pixels. However, additional knowledge such as texture is needed to apply this technique. Many researches have developed pixel based skin classifiers with high detection rates at high speeds. Thus, in this work, a pixel based, color space method for skin detection was adapted.

There are different color models which can be used for skin detection such as RGB, Normalized RGB, HIS, HSV, HSL, TSL, YcrCb etc. The work done by Zarit et. al. [6] has shown that HS and HSV give the best results followed by normalized RGB. In the RGB space, the red, green and blue components represent color and luminance. Luminance may vary across a person's face due to the ambient lighting and it is not a reliable measure for separating skin from non-skin regions. We selected HSV color space for the skin detection as it is having a higher detection rate. The HSV stands for Hue, Saturation and Value or Brightness. The Value component from the HSV color space was ignored and the HS color space was considered in this work. Figure 1 shows the distribution of extracted HS color components (histograms) for different skin samples selected for the training set.

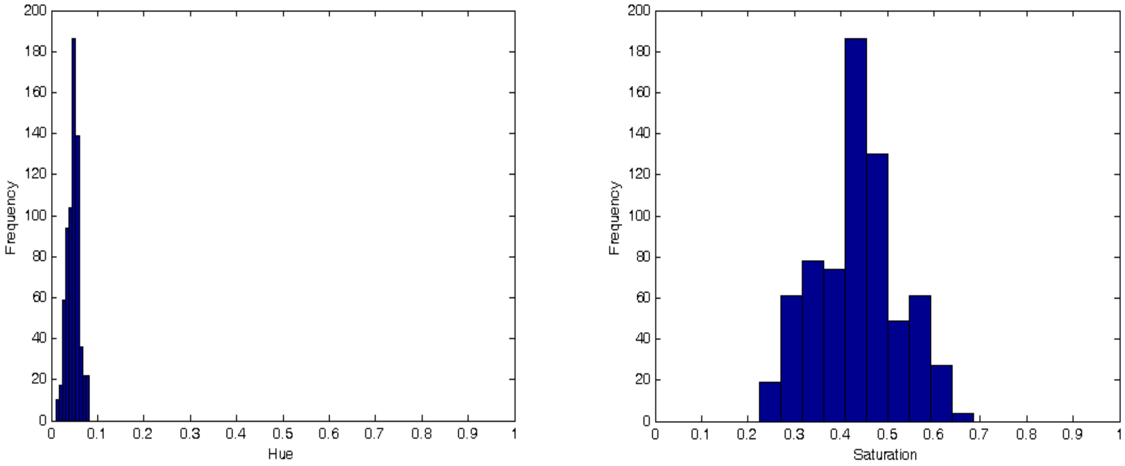

Figure 1: Distribution of skin colors of the training set in HS color space.

From the selected skin color pixel set, the mean and the standard deviation was calculated and values in the range of 3 standard deviations, where 99.8% of the skin color pixels are belong was considered as the true values for skin color. For each input image, HS color components of every pixel was checked and pixels which are not in the range of 3 standard deviations were turned into black and others were allowed to remain in their original color.

Figure 2 shows one of the test images subjected to the above procedure. Then the images were turned into binary images using a fixed threshold value.

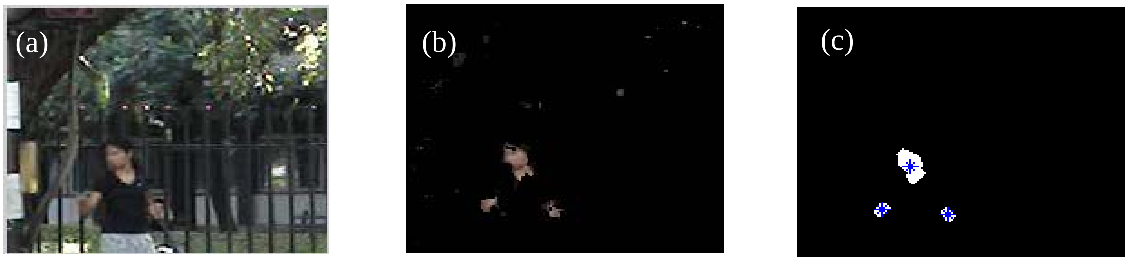

Figure 2: The resultant images after the skin color segmentation and processing (a) Original image (b) Skin colour segmentation using HS colour space (c) Region removal and morphological processing.

2.2 Morphological processing

The resultant binary image contains falsely detected pixels as well as missed pixels. Filling the holes and removing unwanted pixels was done through morphological processing.

Dilation causes objects to grow in size or dilate while erosion causes objects to shrink. The structuring element determines the amount and the way the objects grow or shrink. Dilation and erosion are combined to create two morphological processes called morphological opening and morphological closing. In morphological opening process, erosion is applied first followed by the dilation causes individual objects that are connected in a binary image to separate. In morphological closing where dilation is applied first followed by erosion causes small holes and gaps to be filled. The resulted image after morphological opening and morphological closing is shown in the Figure 2(c).

2.3 Face recognition from skin color

The main goal of the skin colour segmentation was to find the pedestrian from the detected skin regions. There are faces, hands and legs among the segmented skin regions. Face detection was applied to detect the people in an image since faces are having unique cues such as two eyes, a mouth and a nose. There are many techniques to find the face or head from images such as elliptical head tracking, neural network methods, contour mapping, template matching etc. However, all these techniques require high processing time and unsuitable for this work.

In this work, a new but simple method was applied to identify the face blobs. In general, faces when compared with hands or legs have round shapes. This property was used to identify faces from legs and hands. First, the centroid of each blob is computed and the distance from the centroid to each point at the border of the blob was calculated. When the average radius of each blob is considered, the face blobs have higher average radius than other shapes. Within each image, all the blobs passing a minimum threshold radius but having a radius higher than 0.5 of the maximum radius was considered as true face blobs.

2.4 Depth Perception

The binocular vision system captures two different views of a scene. Depth estimation from these two images involves three steps. First step is establishing the correspondences between two images, and then calculating the relative displacements between the features in each image. Determination of the 3D depth of the feature relative to the cameras, using the camera geometry is the third step.

Monocular vision as opposed by binocular vision has an increased field of view, while depth perception is limited. Humans use monocular cues such as texture variations, texture gradients, interposition, occlusion, known object sizes, light and shading, defocus, haze etc. Texture gradients which capture the distribution of the direction of edges also help to indicate depth. Haze is caused by atmospheric light scattering.

When an observer moves, the apparent relative motion of several stationary objects against a background gives hints about their relative distance. Thus, if the information about the direction and the velocity is known, motion parallax can provide absolute depth information. As an example, when driving a car, nearby objects pass quickly while far away objects appear stationary. Reader may refer the review by Gandhi et al [1] for further information. Enzweiler et al., [2] have obtained good results using motion parallax for monocular pedestrian detection. However, high computer processing power is required for this methodology.

In our approach, we used the above discussed monocular cue of object size variation with distance for depth perception because of its simplicity and high processing speed. Figure 3 shows the image frames extracted from a video while a person is crossing the road perpendicular to the motion direction of the vehicle. The face size (or the object) increases when the person is nearer to the camera.

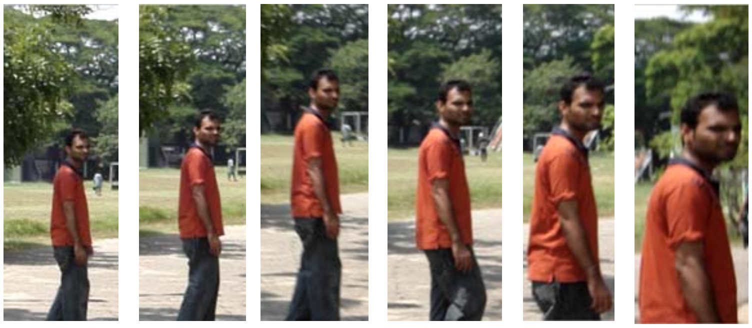

Figure 3: The size variation of an image of a pedestrian when the frames are extracted from a video.

Figure 4 shows the increase in the size of the face blob with the decrease in the distance to the pedestrian (or vise versa). The pixel difference of two consecutive face size increase when the person is nearer than in far. This information, i.e. the size difference between two consecutive image frames was used to decide whether the person is within a risk leading to a collision with the vehicle.

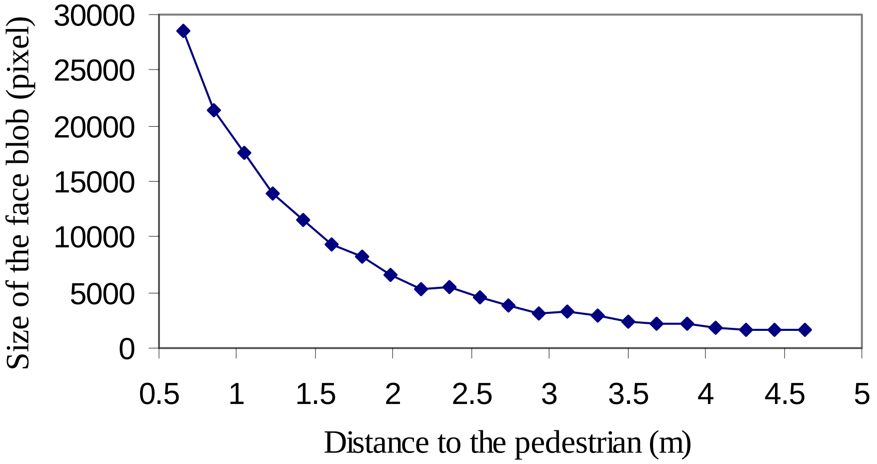

Figure 4: The variation of the distance to the pedestrian vs. the size variation of the face blob when the camera moves with a low average speed of 2 kmh^{-1}.

2.5 Breaking Distance

The breaking system should be invoked early within the correct distance between the pedestrian and the vehicle to avoid collision. This distance is defined in transportation engineering as the breaking distance.

Assume a flat ground and straight road to travel on. Then the breaking distance is given by [7],

$$BD = \frac{V^{2}}{254[\frac{(a)}{(g)} \pm G]} \tag{1}$$

where, *BD* is the braking distance in meters, *V* is the speed of the vehicle in kmh^{-1}, *a* is the deceleration rate (ms^{-2}), *G* is the grade (decimal) and *g* is the acceleration due to gravity (9.18ms^{-2}).

Considering the capability of drivers to stay within their lane and control the vehicle when breaking on wet surfaces, the deceleration comfortable for most drivers is 3.2 ms^{-2} [7].

The breaking distance was used to determine the break invoking point and the size variation of two consecutive face blobs was used to decide whether the vehicle is in the limit of its breaking distance to avoid collision.

3. RESULTS AND DISCUSSION

Field tests were carried out to determine whether the object size variation with distance depend on the speed of a vehicle. Tests were carried out with a moving camera towards a stationary object in different speeds. The resultant plots are shown in figure 5. It can be seen that the size variation with distance is not effected by the speed of the vehicle.

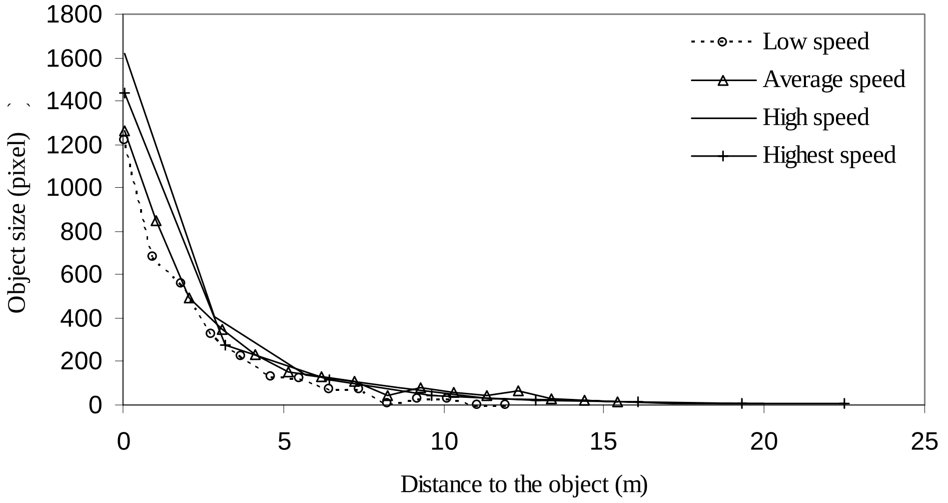

Figure 5: The size variation of the object with distance when the camera is moving with different speeds.

The tests were carried out to check whether the vehicle can be stopped at the breaking distance with a moving camera when a person is crossing the road. Different events were experimented such as the pedestrian has crossed the road when the camera moves nearer, pedestrian has stopped in the middle of the road when the camera comes nearer and pedestrian is crossing the road in different speeds.

The tests were also carried out to compare the variation of theoretical breaking distance and the observed breaking distance with velocity. The results are shown in Figure 6.

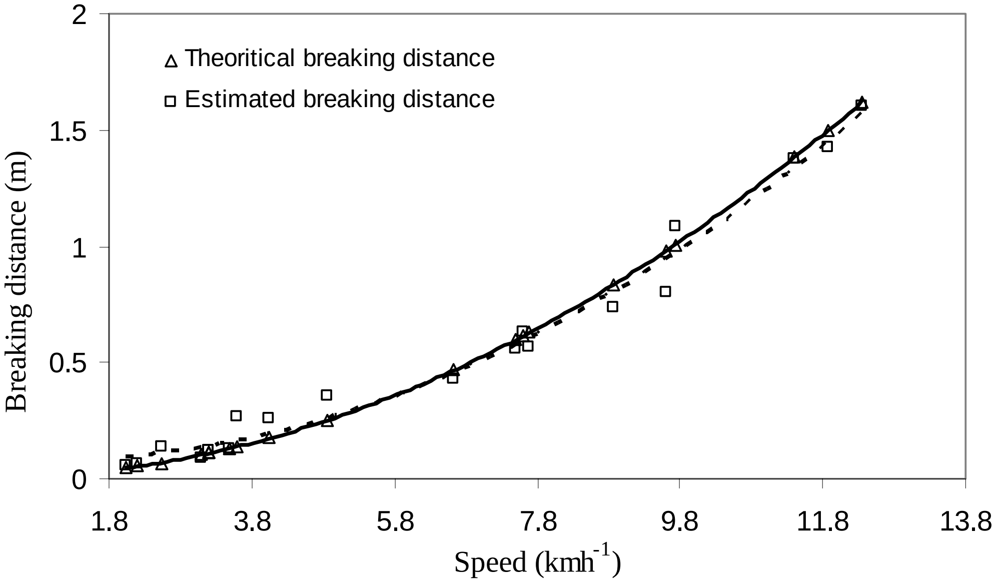

Figure 6: Variation of theoretical and observed breaking distance with velocity. Continuous line: The polynomial fit for the theoretical values. The dotted line: The polynomial fit for the theoretical values.

It can be seen from the Figure 6 that the breaking distance increases with the speed. When the estimated breaking distance is compared with the theoretical breaking distance, it can be seen that the estimated values are slightly different than the theoretical values. This may be because of the different lighting conditions and facial expressions leading to fluctuations of the face blob size. However, when the data points are fitted to a polynomial of the degree of 2, the two trend lines (solid line and the dotted line) appear closely.

4. CONCLUSIONS

This paper presents preliminary results of a simple technique for pedestrian collision avoidance which can be used in vehicles. The system consists of a camera that is fixed to a vehicle. Whenever the camera detects a pedestrian via skin color detection methods, the face size will be examined and the system will check the speed of the vehicle and calculate the breaking distance according to the speed. The vehicle decides whether the person is in its breaking distance using size difference change of two consecutive frames. If the person is in the corresponding breaking distance, the system can generate a trigger.

Acknowledgements: This research work was supported by the National Science Foundation (NSF) grant RG/2007/E/03 and International Science Program (ISP) grant SRI 01/3.

5. REFERENCES

1. T. Gandhi, M .M. Trivedi, Pedestrian collision avoidance systems: A survey of computer vision based recent studies, IEEE International Transportation systems conference (2006) 976-981.
2. M. Enzweiler, P. Kanter, D. M. Gavrila, Monocular pedestrian recognition using motion parallax, IEEE intelligent vehicle symposium (2008) 792-797.
3. A. Wedel, U. Franke, J. Klappstein, T. Brox, D. Cremers, Real-time depth estimation and obstacle detection from monocular video, DAGM symposium on pattern recognition (2006) 475-484.
4. J. Lin, K.Y. Kuo, Application of fuzzy set theory on the change intervals at a signalized intersection, Applied Soft Computing (2001) 161-177.
5. S. Birchfield, Elliptical Head Tracking Using Intensity Gradients and color Histograms, Proc. IEEE Conference on Computer Vision and Pattern Recognition (1998) 232-237.
6. C. Chou, J. Teng, A fuzzy logic controller for traffic junction signals. Information Science, 143 (1-4) (2002) 73-97.
7. D. Levinson, Stopping Distance, http://nexus.umn.edu/Courses/ce3201/CE3201-L3-02.pdf