Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

A Comparison of Color and Infrared Stereo Approaches to Pedestrian Detection

Stephen J. Krotosky and Mohan M. Trivedi
Computer Vision and Robotics Research Laboratory
University of California, San Diego
La Jolla, CA 92093–0434
Email: {krotosky, mtrivedi}@ucsd.edu

Abstract— **This paper presents an analysis of color and infrared stereo approaches to pedestrian detection. We design a four camera experimental testbed consisting of two color and two infrared cameras that allows for synchronous capture and direct frame-by-frame comparison of pedestrian detection approaches. We incorporate this four camera system in a test vehicle and conduct comparative experiments of stereo-based approaches to obstacle detection using color and infrared imagery. A detailed analysis of these experiments shows the robustness of both color and infrared stereo imagery to generate the dense stereo maps necessary for robust object detection and motivates investigation of color and infrared features that can be used to further classify detected obstacles into pedestrian regions. The complementary nature of color and infrared features gives rise to a discussion of a feature fusion techniques, including a cross-spectral stereo solution to pedestrian detection.**

I. Introduction

Pedestrian safety is a problem of global significance. Naturally, such an important concern to public safety has received significant attention from all aspects of the research community. Specifically, ongoing computer vision research is making strides to detect and track pedestrians from both moving vehicles and the static transportation infrastructure. Typically, these approaches to pedestrian detection make use of visual or infrared imagery [1] in both monocular and stereo camera configurations.

The choice of visual or infrared imagery is significant, as each provides disparate, yet complementary information about a scene. Visual cameras capture the reflective light properties of objects in the scene, while infrared cameras are sensitive to the thermal emissivity properties of the same objects. Features extracted from each type of modality can be used to determine the presence of pedestrians in a scene. Additionally, binocular stereo systems have been incorporated into pedestrian detection approaches. The use of two cameras allows for the accurate depth estimates crucial to the task of pedestrian detection and collision mitigation. For color-based stereo systems, these estimates have been determined through dense stereo correspondence matching [2]. For infrared-based stereo systems for pedestrian detection, correspondence matching has been typically accomplished with sparser feature based matching techniques [3], [4].

This paper presents research toward the development of a stereo system that can extract the dense stereo depth and features necessary for robust pedestrian detection using either color or infrared stereo imagery. We design a four camera experimental testbed consisting of two color and two infrared cameras that allows for comparative experiments of stereo-based detection approaches using color and infrared imagery and demonstrates high obstacle detection rate achievable with such stereo imagery. From these comparative experiments, we provide a detailed analysis of the features and properties of color and infrared imagery that are used to classify detected obstacles into pedestrian regions. This analysis is used to motivate a discussion of feature fusion techniques, including a cross-spectral stereo solution to the pedestrian detection problem.

II. Stereo-based Pedestrian Detection

A fundamental step to analyzing pedestrians with stereo imagery is to detect obstacles in the scene and localize their position in 3D space from the disparity maps generated from stereo correspondence matching. The disparity images derived from stereo analysis can be used to generate a list of candidate pedestrian regions in the scene. We adapt a classical approach to obstacle detection in stereo imagery proposed by Labayrade *et al.* [5] that utilizes the concept of *v-disparity* to identify potential obstacles in the scene. Essentially, v-disparity is a histogram of the disparity image that counts the occurrence of disparity values for each row in the image and can be used to detect the ground plane in the scene and isolate regions that contain obstacles. Variations of this approach to detecting objects in stereo imagery have been implemented in [6], [7], [8]. However, we illustrate a generalized framework that is able to obtain dense stereo correspondences and robust ground plane estimates with both color and infrared-based stereo.

A. Disparity-based Obstacle Detection

Our goal is to provide a framework for a comparative analysis of color and infrared stereo imagery for pedestrian detection and have chosen to use the relatively simple v-disparity approach to obstacle detection so that it can be implemented for both color and infrared stereo imagery without modification or specialization. We examine the ability of each to generate stereo disparities and determine obstacle areas in the scene. This comparison of low-level detection accuracy will then lead to an evaluation of each camera type's potential

1-4244-1068-1/07/$25.00 ©2007 IEEE.

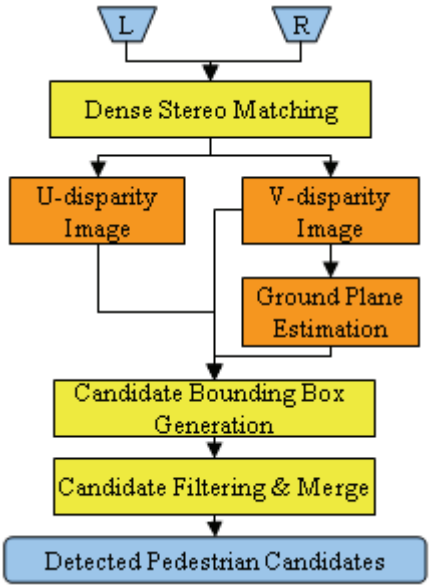
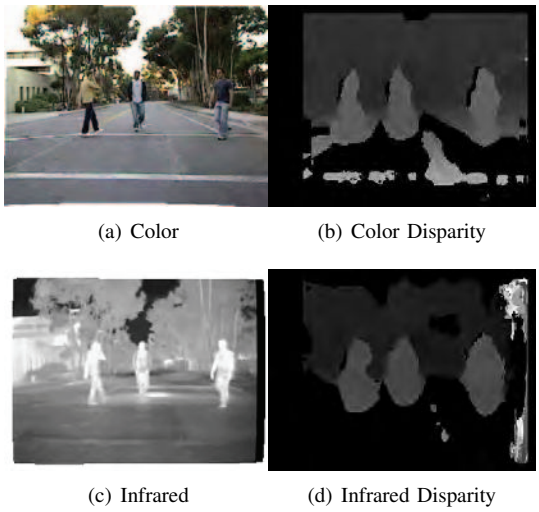


Fig. 1. Flowchart of stereo disparity-based obstacle detection algorithm.



(a) Color (b) Color Disparity

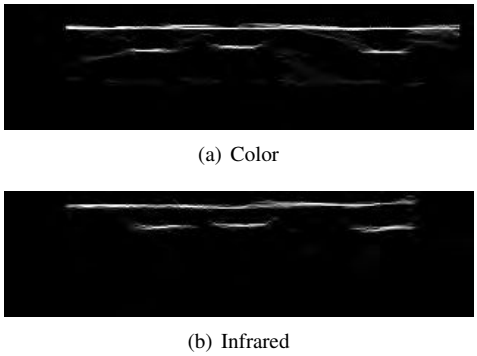
(c) Infrared (d) Infrared Disparity

Fig. 2. Example disparity images from color and infrared stereo input images.

for higher level obstacle classification and analysis. Fig. 1 shows a flowchart of the obstacle detection algorithm.

1) Dense Stereo Matching: As a first step, it is necessary to perform dense stereo matching to yield disparity estimates of the imaged scene. We elect to use the correspondence matching algorithm developed by Konolige [9] for its ease of use and reliable disparity generation with both color and infrared stereo imagery. Example disparity images generated using this approach are shown in Fig. 2.

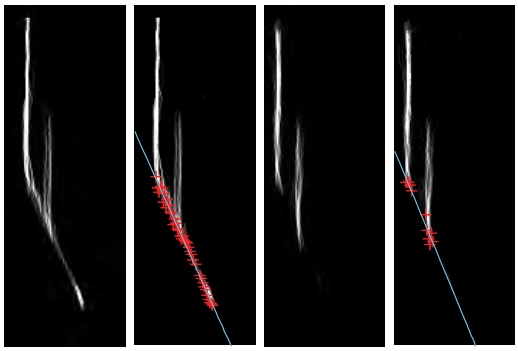
2) U- and V-Disparity Image Generation: The u- and v-disparity images are histograms that accumulate the number of pixels at a given disparity value, d, for each column or row in the image, respectively. For example, each row in the v-disparity image is the histogram of disparities in the corresponding row of the stereo disparity image D. The



(a) Color

(b) Infrared

Fig. 3. Example u-disparity images from color and infrared stereo input images.

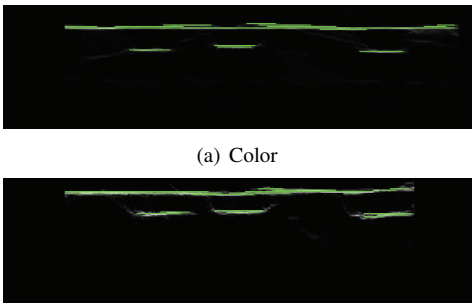


(a) Color (b) Color w/ Ground Plane (c) Infrared (d) Infrared w/ Ground Plane

Fig. 4. Example v-disparity images from color and infrared stereo input images along with the detected ground plane.

resulting v-disparity histogram image indicates the density of disparities for each image row v, while the u-disparity image shows the density of disparities for each image column u. Fig. 3 shows an example u-disparity image for color and infrared stereo imagery, and Fig. 4 shows the corresponding v-disparity images generated from the color- and infrared-based stereo disparity maps in Fig. 2.

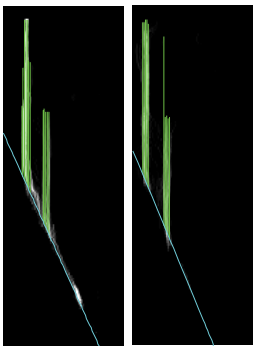
Notice how the u-disparity images in Fig. 3 show three distinct horizontal regions of high disparity density corresponding to the three pedestrians in the scene. These regions can be detected in order to help build candidate pedestrian areas. The image spanning high density region at the top of the u-disparity image indicates the background disparities of the image and can be detected and filtered from processing. Similarly the v-disparity images in Fig. 4 show vertical peaks of high density for both the background plane and the range of disparities in D containing pedestrians. These regions will also need to be detected to generate pedestrian candidates. Additionally, there is a distinct downward sloping trend for the lowest image point for each disparity in the v-disparity image. It has been shown that this phenomenon can be used to estimate the ground plane of the image [5] for color stereo imagery. We show this can also be extended to dense stereo estimates from infrared stereo imagery.



(a) Color

(b) Infrared

Fig. 5. Example region-of-interest generation in u-disparity images with color and infrared stereo input images.



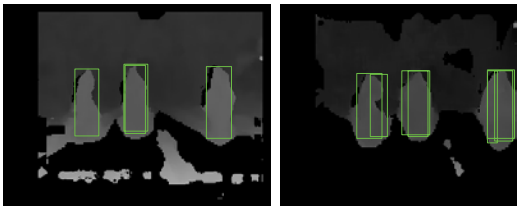
(a) Color (b) Infrared

Fig. 6. Example region-of-interest generation in v-disparity images with color and infrared stereo input images.

3) Ground Plane Estimation: To derive an estimate of the line indicating the ground plane, we must first extract candidate points on that line. For each column corresponding to a disparity d in the v-disparity image, we select the lowest pixel location whose value is above a given threshold as a candidate point in the ground plane. If there is no value that exceeds that threshold for a given disparity, then we do not consider that disparity point. The ground plane is estimated by fitting the candidate points to a line with a robust linear regression scheme that uses weighted least squares that iteratively reweights at each iteration using the bisquare weighting function. Figs. 4(b),(d) show the v-disparity images for color and infrared stereo imagery with the candidate ground plane points in red and the fitted ground plane estimate plotted in cyan. Because we are using a dense stereo correspondence algorithm with robust point candidate generation and linear least squares fitting, we are able to reliably estimate the ground plane with both color and infrared stereo imagery.

4) Candidate Bounding Box Generation: Bounding box candidates can be extracted by first identifying regions-of-interest in the u- and v-disparity images. Regions in the u-disparity image can be extracted by scanning along the rows of the image and identifying continuous spans along a row where the histogram values exceed a given threshold. Fig. 5 shows the extracted regions in green on the u-disparity image. Regions are also extracted in the v-disparity image by scanning each column and summing the histogram value above the ground plane. If this sum is greater than a threshold, then the region is selected that spans from the ground plane to the high point in the column where the histogram entry exceeds a given threshold. Fig. 6 shows the extracted regions in green on the v-disparity image.

Candidate bounding boxes are then determined by associating the regions-of-interest in the u- and v-disparity images based on their disparity values. For a given disparity d, the width of the bounding boxes at that disparity are determined by the regions found in the u-disparity image and the height is correspondingly derived from the regions in the v-disparity image. Bounding boxes associated with the background regions that are obviously too large are removed. The resulting bounding box candidates are shown in green in Fig. 7.



(a) Color (b) Infrared

Fig. 7. Example bounding box candidates with color and infrared stereo input images.

5) Candidate Filtering and Merging: As shown in Fig. 7, there are often multiple overlapping candidate bounding boxes generated in the previous step. This usually arises when disparities associated with a single pedestrian span a range of disparity values. We merge overlapping bounding box candidates if their overlap is significant and the disparities associated with the bounding boxes are close. The final selection of pedestrian candidate bounding boxes is shown in Fig 8. Notice how the multiple bounding box candidates have merged into three appropriate bounding boxes associated with the correct pedestrians in the scene.

B. Experimental Framework and Testbed

We establish a framework for experimenting and analyzing pedestrian detection approaches for color and infrared stereo imagery. This framework needs to facilitate a direct, frame-by-frame comparison of the data coming from color and infrared stereo imagery. To that end, we have designed a custom rig consisting of a matched color stereo pair and a matched infrared stereo pair. The two pairs have been arranged so that their imaged scenes are as consistent as possible. The two pairs have identical baselines and the corresponding cameras in the color and infrared pairs are positioned as close as possible so as to maintain the same approximate fields of view. Additionally, lenses for the color cameras were selected to best match the fixed zoom of the infrared cameras. All four cameras are arranged in a single row and care was taken in aligning the pitch, roll and yaw of the cameras to maximize the similarity in field of view.

(a) Color

(b) Infrared

Fig. 8. Example of the final selection of pedestrian candidates after bounding box merging with color and infrared stereo input images.

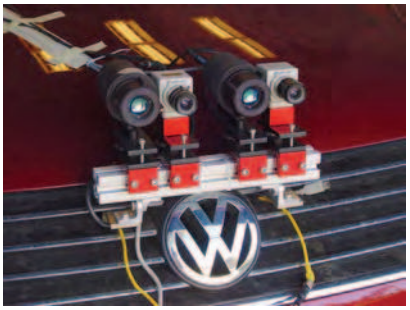


Fig. 9. Experimental testbed: Two color cameras and two infrared cameras arranged in stereo pairs and mounted to the front of the LISA-P testbed.

Once aligned the rig was mounted to the grill of the LISA-P testbed described in Trivedi *et al.* [10], [11]. The LISA-P is a Volkswagen Passat equipped with the computing, power, and cabling requirements necessary to synchronously capture and save the four simultaneous camera streams of our custom rig. Fig. 9 shows the four camera rig properly arranged and mounted on the LISA-P.

C. Experimental Analysis of Disparity-based Obstacle Detection in Color and Infrared Stereo Imagery

Experiments were conducted so pedestrians walk in front of the LISA-P testbed. The experiments included multiple pedestrians in the scene with varying degrees of depth, complexity and occlusion. The experimental data was captured simultaneously with the color and infrared stereo cameras to allow for direct comparison of the approaches. The captured data was analyzed using the disparity-based obstacle detection algorithm in Section II-A and detection was determined successful if a bounding box correctly overlaid a corresponding pedestrian region. If two candidate bounding boxes associated with two separate pedestrians merged into a single bounding box after the merge process, we still consider the detection correct, yet note it as a "merge error". We reason that errors associated with a lack of sophistication of our chosen merging algorithm should not adversely affect the detection rate, as our desire is to evaluate the effectiveness of color and infrared stereo disparities to identify pedestrian areas and not the robustness of the merging procedure. This is also a fair assessment when using pedestrian detection for collision mitigation, as finding all the critical areas in the scene is given priority over discerning merged bounding boxes. Therefore, false negatives were counted only when a bounding box did not properly identify a pedestrian region and false positives were counted when a bounding box enclosed an area where no pedestrian existed. Still, had we incorporated the merge errors, the total detection rate would decrease by only 1% for color and 1.4% for infrared. Table I shows the compiled results of the comparative experiments and Fig. 10 shows additional examples for both color and infrared stereo inputs.

TABLE I

RESULTS OF EXPERIMENTAL COMPARISON BETWEEN COLOR AND INFRARED STEREO IMAGERY FOR DISPARITY-BASED OBSTACLE DETECTION.

Modality	# Peds in Frame	Peds Correct	% Correct	False Positives	Merge Errors
Color	1	758	100.0%	0	0
	2	2376	99.5%	2	7
	3	1525	99.9%	0	35
	4	377	99.2%	1	6
	5	22	88.0%	0	0
	Total	5058	99.6%	3	48
Infrared	1	880	97.9%	1	0
	2	2257	98.7%	4	14
	3	1231	98.9%	0	43
	4	123	99.2%	1	10
	5	5	100.0%	0	0
	Total	4496	98.6%	6	67

III. ANALYSIS OF STEREO-BASED PEDESTRIAN DETECTION

Our comparative experiments with stereo-based pedestrian detection for color and infrared imagery show a very high level of detection accuracy and low false positive rate from the both modalities. However, a deeper analysis of the experiments is necessary to truly understand and evaluate the success of these experiments. The experiments yielded such a high rate of detection accuracy because our analysis equated low level obstacle detection with the higher level analysis of pedestrian determination. That is to say, since the experiments did not include non-pedestrian obstacles, a detection of any obstacle region is assumed to be a pedestrian. For the scope of our experiments, this sort of assumption is appropriate, as we are interested in evaluating the ability of color and infrared dense stereo correspondences to be used in low level pedestrian detection. In that respect, our experiments demonstrate that color and infrared stereo disparities both achieve high rates of low level obstacle detection, an imperative first step towards robust pedestrian detection and collision mitigation.

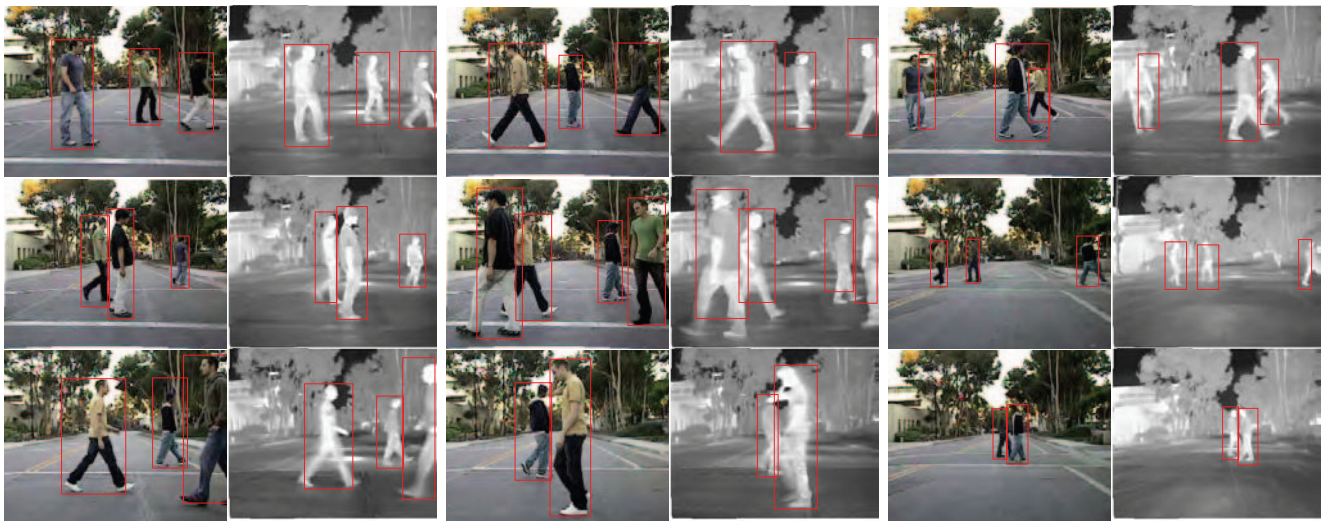


Fig. 10. Example of the final selection of pedestrian candidates with color and infrared stereo input images.

However, in real world driving scenarios, low level obstacle detection, while an imperative initial step, is not sufficient for pedestrian detection. Detected obstacles can include a variety of objects found in common driving scenes other than pedestrians, such as parked and moving vehicles, trees, buildings, parking meters and other spurious candidates in the scene. Additional processing is necessary to filter the detected obstacles into appropriate pedestrian and non-pedestrian regions.

In the disparity image domain, it is possible to filter some of the detected obstacles based on the bounding box features of typical pedestrian obstacles (e.g Bertozzi *et al.* [8]). Bounds on pedestrian bounding box features such as size, disparity and aspect ratio can be learned or heuristically selected to filter out bounding boxes associated with other objects in the scene. However, the success of such filtering techniques can prove unreliable, as it will not filter non-pedestrian bounding boxes that fall within the selected bounds of pedestrian candidates. Additionally, the selection of appropriately robust bounds is a challenging task, as bounding box sizes can vary with pedestrian pose and disparity fidelity. To achieve more reliable detection of pedestrian candidates, it is necessary to analyze the specific image features of the chosen modality.

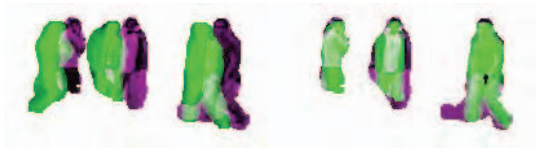
Color features that have been used for pedestrian classification attempt to identify the unique contours and shapes that discriminate pedestrians from other objects. Such features include Haar wavelet responses [12], Gabor filter response [13], Sobel edge responses [6], Implicit Shape Models with Chamfer distance matching [14], image countours with Mean Field models [15], and local receptive fields for support vector machine classification [16].

Infrared features for classification typically include features that identify the specific thermal characteristics of the scene, including hotspots [17], warm element and head template matching [4], body model templates [18], shape independent multidimensional histogram, inertial and contrast base features [19] and Histograms of Oriented Gradients [20]. Obstacle detection using stereo disparities derived from color or infrared imagery is highly accurate with low false positive rates. However, this level of detection is still too primitive to be used for real world pedestrian detection, as it can include obstacles not associated with pedestrians or other critical regions. To supplement and filter these obstacle candidates, specific features of color or infrared imagery can be extracted and analyzed to determine the true pedestrian regions in the scene. Although both color and infrared imagery have been used to identify pedestrians in a scene, it is unclear which camera system is preferred. However, our framework will allow for a direct comparison of these approaches and will give insight into how the disparate features in color and infrared imagery directly affect pedestrian detection accuracy.

Additionally, a more interesting proposition would be to use both modalities in concert to obtain all sets of available features in color and thermal imagery. Naturally a detection architecture that incorporates more features has a higher potential for detection accuracy than one with a lesser feature set. For example, the thermal "hotspots" of humans that often make pedestrians easily segmentable can be used together with the fine level of color image detail that has proven useful for tracking multiple people in a scene.

Although it is possible to incorporate the advantages of stereo color and infrared analyses by separately combining the two camera systems and pedestrian detections [8], it is costly and cumbersome to incorporate a four camera solution from both a computational and vehicle integration standpoint. A more economical and desirable solution would be to obtain the benefits of each using a cross-spectral stereo approach using a single color and single infrared camera. The challenge is achieving the accurate and dense stereo correspondences of unimodal stereo systems with cross-spectral stereo, where conventional assumptions for matching do not hold. Because of the disparate nature of color and infrared imagery, conven-

(a) Color (b) Infrared



(c) Unaligned (d) Aligned

Fig. 11. Cross-spectral stereo approach [22] for pedestrian detection. Pedestrian pixels can be associated across color and infrared imagery using this approach.

tional and state-of-the-art stereo correspondence algorithms for unimodal imagery are unsuccessful in providing any reliable matches in multimodal imagery. As an advancement towards a similar dense stereo algorithm for cross-spectral stereo imagery, we have proposed a stereo registration algorithm that can accurately align multiple people in a scene [21], [22]. Fig. 11 shows a typical result of how multiple pedestrians can be aligned in the cross-spectral framework.

IV. Discussion and Conclusion Remarks

The use of stereo imagery has helped researchers take large steps towards achieving accurate and robust pedestrian detection. The depth estimates obtainable from vehicle mounted stereo imagery give a straightforward approach to extracting obstacle regions from the scene. We have outlined a general algorithm for obstacle detection in either color or infrared stereo imagery and have provided comparative experiments to gauge the detection rates achievable with each. Our analysis indicates that color and infrared-based stereo disparities are capable of highly accurate pedestrian detection ($> 98\%$) with low false positives ($\ll 1\%$).

Given the high detection rates obtainable from color and infrared stereo imagery, the selection of an appropriate camera system for pedestrian detection turns to the consideration of each modality's ability to further classify detected obstacles into pedestrian and non-pedestrian regions. Because the physical processes that give rise to color and thermal imagery are disparate, the extractable features from color and infrared imagery are also very different and largely unique to each modality. Previous approaches have demonstrated the usefulness of features from both color and infrared imagery for classifying pedestrian regions, and we have laid the groundwork for a direct comparison of those features for future work. Additionally, a complementary system that utilizes all the available features of color and infrared imagery is most desirable. Specifically, we suggest moving towards a two camera, cross-spectral stereo solution to obtain the depth, color and thermal features desirable for a pedestrian detection system.

References

[1] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, and I. Masaki, "Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection," in *IEEE Intelligent Vehicles Symposium*, 2003.

[2] D. Scharstein and R. Szeliski. (2005) Middlebury college stereo vision research page. [Online]. Available: http://cat.middlebury.edu/stereo/

[3] X. Lie and K. Fujimura, "Pedestrian detection using stereo night vision," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1657–1665, Nov. 2004.

[4] M. Bertozzi, A. Broggi, C. Caraffi, M. D. Rose, M. Felisa, and G. Vezzoni, "Pedestrian detection by means of far-infrared stereo vision," *Computer Vision and Image Understanding*, 2007, 10.1016/j.cviu.2006.07.016.

[5] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation." in *IEEE Conference on Intelligent Vehicles*, 2002.

[6] G. Grubb, A. Zelinsky, L. Nilsson, and M. Rilbe, "3D vision sensing for improved pedestrian safety," in *IEEE Conference on Intelligent Vehicles*, 2004.

[7] M. A. Sotelo, I. Parra, D. Fernandez, and E. Naranjo, "Pedestrian detection using svm and multi-feature combination," in *IEEE Conference on Intelligent Transportation Systems*, 2006.

[8] M. Bertozzi, A. Broggi, M. Felias, G. Vezzoni, and M. D. Rose, "Low-level pedestrian detection by means of visible and far infra-red tetra-vision," in *IEEE Conference on Intelligent Vehicles*, 2006.

[9] K. Konolige, "Small vision systems: hardware and implementation," in *Eighth International Symposium on Robotics Research*, 1997.

[10] M. M. Trivedi, S. Y. Cheng, E. M. C. Childers, and S. J. Krotosky, "Occupant posture analysis with stereo and thermal infrared video: Algorithms and experimental evaluation," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1968–1712, Nov. 2004.

[11] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *IEEE Trans. Intell. Transport. Syst.*, Mar. 2007.

[12] L. Andreone, F. Bellotti, A. D. Gloria, and R. Laulette, "Svm-based pedestrian recognition on near-infrared images," in *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, 2005.

[13] H. Cheng, N. Zheng, and J. Qini, "Pedestrian detection using sparse gabor filters and support vector machine," in *IEEE Conference on Intelligent Vehicles*, 2005.

[14] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *Computer Vision and Pattern Recognition*, 2005.

[15] Y. Wi, T. Yu, and G. Hua, "A statistical field model for pedestrian detection," in *Computer Vision and Pattern Recognition*, 2005.

[16] S. Munder and D. Gavrila, "An experimental study on pedestrian classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 11, pp. 1863–1868, Nov. 2006.

[17] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," *IEEE Trans. Intell. Transport. Syst.*, vol. 6, no. 1, pp. 63–71, Mar. 2005.

[18] A. Broggi, A. Fascioli, P. Grisleri, T. Graf, and M. Meinecke, "Model-based validation approaches and matching techniques for automotive vision based pedestrian detection," in *Computer Vision and Pattern Recognition*, 2005.

[19] Y. Fang, K. Yamada, Y. Ninomiya, B. K. P. Horn, and I. Masaki, "A shape-independent method for pedestrian detection with far-infrared images," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1679–1697, Nov. 2004.

[20] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian detection using infrared images and histograms of oriented gradients," in *IEEE Conference on Intelligent Vehicles*, 2006.

[21] S. J. Krotosky and M. M. Trivedi, "Multimodal stereo image registration for predestrian detection," in *IEEE Conference on Intelligent Transportation Systems*, 2006.

[22] ——, "Mutual information based registration of multimodal stereo videos for person tracking," *Computer Vision and Image Understanding*, 2007, doi:10.1016/j.cviu.2006.10.008.