Shape improvement of traffic pedestrian hypotheses by means of stereo-vision and superpixels

Ion Giosan, Sergiu Nedevschi, Ciprian Pocol
Computer Science Department
Technical University of Cluj-Napoca, Romania
{Ion.Giosan, Sergiu.Nedevschi, Ciprian.Pocol}@cs.utcluj.ro

***Abstract*—Shape is a powerful descriptor frequently used in pedestrian detection process. This paper presents a novel stereo and superpixel-based approach for extracting high quality shapes of pedestrian hypotheses from urban traffic scenarios. Gray-levels stereo-vision images of traffic scenes are acquired, high quality stereo-reconstruction and optical flow algorithms are used for computing the depth and motion information. Superpixels are extracted using the intensity images and clustered in different obstacles by a novel paradigm that fuses intensity, depth and motion information. Pedestrian hypotheses are defined as a subset of the scene obstacles obtained by imposing human-specific geometric constraints. A contour tracing algorithm is used for extracting a continuous contour that defines the shape of each pedestrian hypothesis. A comparison between the contours quality of pedestrian hypotheses obtained by this stereo and superpixel approach and another approach based only on stereo-reconstructed points grouping shows improvements in both object shape description and area coverage. Improvements in shape description will increase the accuracy of any further pedestrian detection processes that use pattern matching techniques.**

Keywords—stereo-vision; intensity; depth; motion; superpixels; pedestrian hypotheses; shape description

I. Introduction

Nowadays, the number of modern cars that integrate on-board advanced driving assistance systems (ADAS) is continuously growing. In each ADAS, functions like adaptive cruise control, automatic parking, collision avoidance, forward collision warning, lane departure warning, lane change assistance, traffic sign recognition are usually present. In urban traffic scenarios, pedestrian warning and protection functions are very important, being necessary to be integrated in every ADAS. This motivates the researchers to provide robust solutions for pedestrian protection which consist in both vehicle protection accessories integration and robust software algorithms development for pedestrian detection. There are vehicles that have special accessories for reducing the impact with pedestrians when the collision avoidance becomes impossible. These special frontal accessories must be triggered only in these special situations, meaning that the pedestrian detection algorithms must be very accurate. Although highway traffic scenes are relatively simple mainly containing obstacles such as cars, fences and poles, urban traffic scenarios are much more complex. In urban traffic scenarios several problems occur due to the crowded scenes in which occlusions occur frequently and makes the obstacle detection problem very difficult.

Pedestrian detection is a quite simple problem for humans but it is much more difficult for an ADAS due to their different body poses, clothing or accessories they are carrying and due to the different traffic scene backgrounds in which they appear. These factors may frequently cause weak detections. Pedestrian detection algorithms must reduce the rate of false alarms almost to zero while not losing many pedestrians from observation. They also must perform real-time, otherwise the ADAS cannot take a collision avoidance or protection measure in time. Pedestrians' shape contour can eliminate most of the problems previously presented, being invariant to them. Different sensors and technologies like laser scanners, radars, ultrasound sensors and video cameras are frequently used for obstacle detection and specifically pedestrian detection. Stereo-cameras are frequently used in computer vision at the expense of other sensors.

In this paper, we propose a low cost setup consisted in a pair of gray level stereo-cameras that take advantage from both the 2D intensity and the associated 3D points' information. We compute the superpixels and detect the traffic obstacles by clustering them based on a novel approach that fusion intensity, depth and motion information. Specific geometric constraints are applied for selecting the pedestrian hypotheses from the traffic obstacles. We use a contour tracing algorithm for extracting the superpixel pedestrian hypotheses contours. The main goal is to obtain high quality shapes of the pedestrian hypotheses that may further improve the pedestrian detection accuracy by pattern matching. We make a comparison between the contours quality of superpixels pedestrian hypotheses and another approach based on stereo-reconstructed points grouping in order to evaluate the possible improvements in both shape description and area coverage.

II. Related Work

Although monocular vision is still widely used for processing the traffic scenes, stereo-vision can clearly acquire much more and accurate information that leads to further better obstacle detection and classification. Stereovision is usually used first for traffic scenes obstacle segmentation [1] and last for validating the classification results [2]. A high quality of the stereo-reconstruction process [3] is absolutely necessary for computing a dense and accurate 3D reconstructed points' map.

Analyzing this map, several algorithms like points grouping [4] or density map analysis [5] are usually used for obstacle segmentation. In comparison with other monocular vision based techniques that uses symmetry [6], edges [7] and textures [8] information, stereovision based obstacle segmentation approaches [9-11] offer better results.

Features extraction is then employed for characterizing each segmented obstacle and classifiers are trained. 2D features like color or gray intensities [12], symmetry [6], edges [7], shadows [13] and textures [8] are usually extracted from the obstacles appearance images and used further for obstacle classification. These features can also integrate depth information [14] and optical flow motion information [15] for increasing the segmentation and classification power. The classifiers may be trained directly on the extracted feature vectors or on a subset of relevant features for improving the classification speed and usually also the accuracy [16].

The 2D pedestrian hypothesis contour is a strong feature that can be used for pedestrian validation [17]. It describes the shape of the pedestrian being invariant to clothing and illumination variations. Pedestrian hypotheses are usually matched against a database containing several pedestrian contour templates in order to validate them as being pedestrians. Object detection can be achieved by using a hierarchy of contour templates and pattern matching score computation with distance transform on edges [18]. A similar approach is also used in a pedestrian detection system based on edges and shapes [19]. In order to speed-up the pattern matching process, a hierarchy of pedestrian contours can decrease the matching time by reducing the search space. These kind of approaches that perform template matching are described in [5, 20, 21]. In [22, 23] a pedestrian detection system that integrates stereo-based depth segmentation, Chamfer template matching for hypotheses shapes, texture classification for verification using neural network, stereo-based verification, and tracking is used for robustly detecting pedestrians.

The extracted obstacle shapes and contours may be quite far from the reality even if stereo-cameras are used for acquiring the traffic scene information and superior stereo-reconstruction and obstacle segmentation algorithms are applied. This may decrease the accuracy of pedestrian detection when using pattern matching of these shapes against a set of pedestrian contour templates. In this paper, we manage to obtain better pedestrian hypotheses shapes closer to the reality that can clearly improve the pedestrian detection by pattern matching techniques.

Our approach divides first the image pixels into regions having the properties that all pixels from a separate region are similar with respect to a chosen similarity metric like in [24, 25]. These methods represent the basis of superpixels based image segmentation. In [26] we proposed a novel obstacle detection method based on the original scene segmentation in superpixels. The method combines the intensity, depth and motion information within the SLIC superpixels. A novel algorithm was proposed for superpixels clustering into scene obstacles. A method for very close obstacles separation was developed based on the motion vectors analysis of their compound superpixels. The results showed a very good obstacles detection with precise segmentation of their surfaces which is particularly useful for subsequent processes like pedestrian detection. Continuing our previous work, in this paper we propose a method for stereo and superpixels-based pedestrian hypotheses high-quality contour extraction. The superpixels benefits in defining the shape of the pedestrian hypotheses are clearly highlighted through an quantitative evaluation of the obtained contours quality by a comparison with the ground truth and also with another approach that uses only stereo-vision and points grouping [4].

III. SYSTEM ARCHITECTURE

In this section we present the entire system architecture for pedestrian hypotheses shape extraction and validation by using stereo and superpixels information. All the component modules and the data flow between them are depicted in Figure 1.

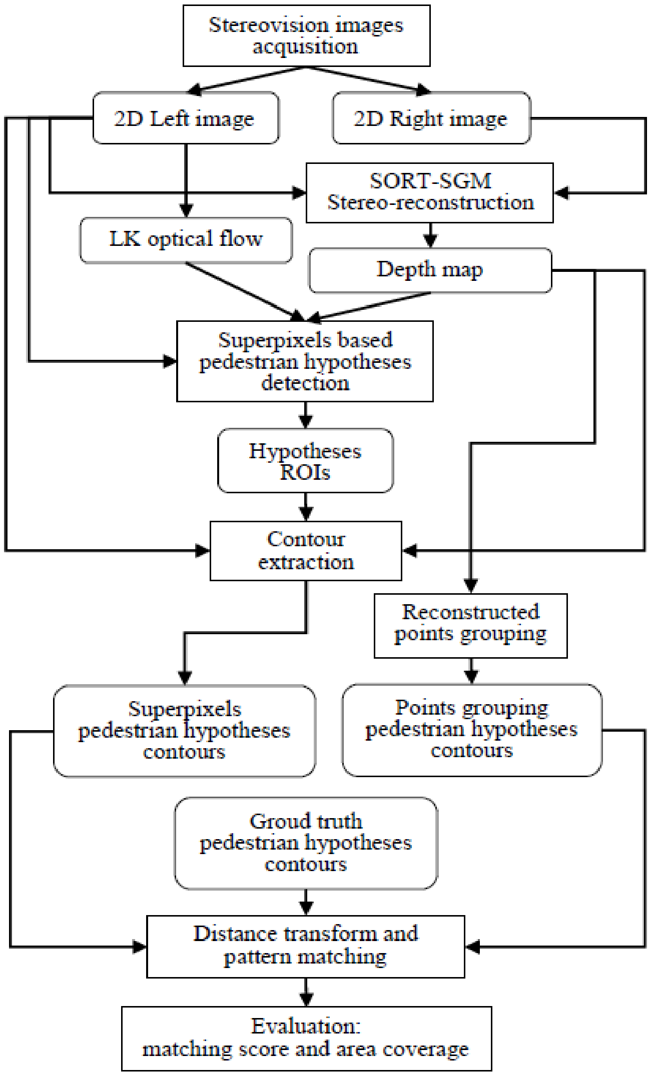


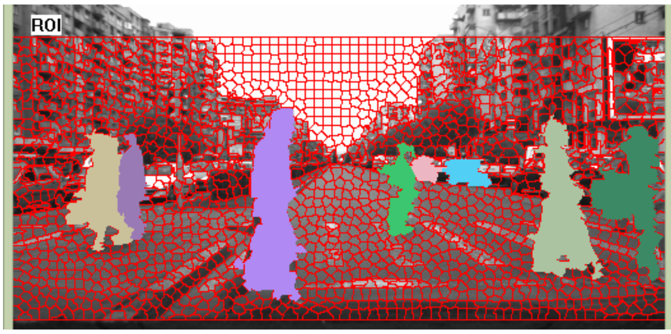
Figure 1. Superpixels and stereo based pedestrian hypotheses contour extraction and validation system architecture

Stereo-vision traffic scene grayscale images with resolution of 512x383 pixels are acquired first with a stereovision cameras setup. The SORT-SGM algorithm [3] is used for stereo-reconstruction process, offering an accurate and dense depth map which is very important for further processing steps. Sparse optical flow is computed by using the Lukas-Kanade algorithm applied on good features to track [27]. Our multi-paradigm that fuses intensity, depth and optical flow information in superpixels is used for obstacles detection [26]. Specific geometrical constraints are applied for selecting the pedestrian hypotheses from the previously detected obstacles. Our previously work [28] is used for extracting a continuous chain of contour pixels for each pedestrian hypothesis segmented region.

In order to assess the quality of the proposed stereo and superpixels pedestrian hypotheses extracted contours and their coverage area, we compare them with the contours of the same hypotheses obtained by grouping the 3D reconstructed points into separate entities [4]. This is achieved via pattern matching performed between the extracted contours and the ground truth pedestrian hypotheses contours. The matching score is obtained through Distance Transform (DT) computation [29]. This offers a fast processing while approximating a high quality matching score.

In the final evaluation, we compare both the pattern matching contour scores and also the area coverage of the extracted pedestrian hypotheses. This will highlight the quality of both extracted contours and area coverage when using stereo and superpixels proposed approach for pedestrian hypotheses generation.

a)



b)

Figure 2. Pedestrian hypotheses generation: a) intensity image ROI; b) stereo and superpixels-based pedestrian hypotheses (random color for each hypothesis)

IV. SHAPE IMPROVEMENT OF PEDESTRIAN HYPOTHESES

This section presents the entire methodology proposed for stereo and superpixels based pedestrian hypotheses contour extraction, area coverage definition and their evaluation against ground truth images.

First, the pedestrian hypotheses are generated by using stereo and superpixels. This is done by continuing our previous work [26] for superpixels based obstacle detection. Fast SLIC superpixels are computed in a rectangular region of interest (ROI) (see Figure 2a) in the intensity image, defined by the values: (*left*=0, *top*=100, *right*=512, *bottom*=320). We uniformly spread a number of *N*=2000 superpixels seeds in the intensity image ROI. A novel paradigm that combines intensity, depth and motion information is used for clustering the superpixels in different scene obstacles.

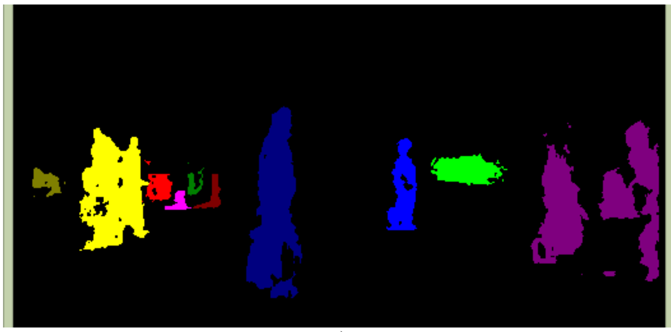
The main steps of the algorithm are: gray-levels SLIC superpixels computation; superpixels specific features extraction; superpixels clustering in obstacles hypotheses; obstacles hypotheses validation and refinement; pedestrian hypotheses generation. Pedestrian hypotheses (see Figure 2b) are selected from the scene obstacles that have the following human specific properties within the specified ranges:

$$3D \rightarrow Height \in [1.5\text{m}, 2.2\text{m}]$$

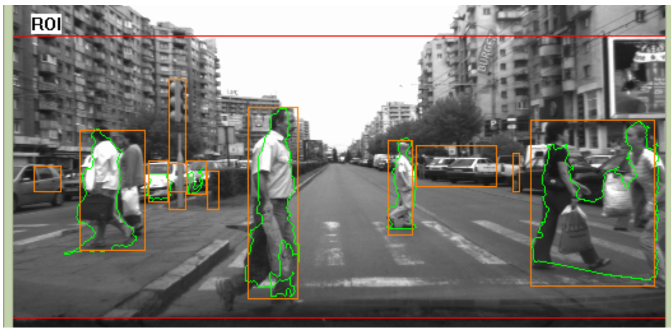
$$3D \rightarrow Width \in [0.3\text{m}, 0.9\text{m}]$$

$$3D \rightarrow Length \in [0.3\text{m}, 0.9\text{m}]$$

$$2D \rightarrow AspectRatio \in [1.0, 4.0]$$



a)



b)

Figure 3. Contour extraction: a) 3D reconstructed points (random color for each hypothesis); b) corresponding 2D image bounding rectangle and extracted contour for each hypothesis based only on 3D reconstructed points

A contour extraction algorithm proposed by us in [28] is used for tracing a continuous contour, defining the shape of each pedestrian hypothesis previously generated (see Figure 2b). The same algorithm is applied directly on the 3D reconstructed points of each pedestrian hypotheses (see Figure 3b) for comparison reasons. A qualitative comparison between the contours extracted by both approaches shows that stereo and superpixels based proposed approach better defines the exterior shape of the pedestrian hypotheses (see Figure 4).

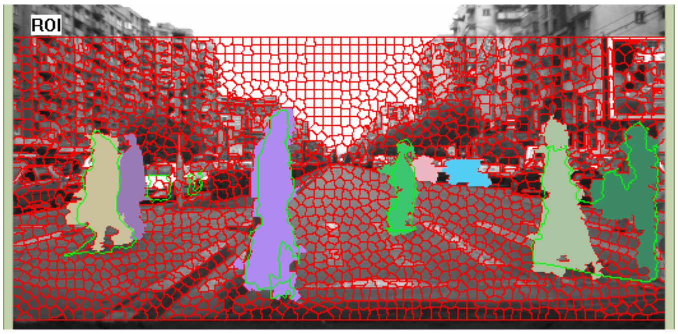
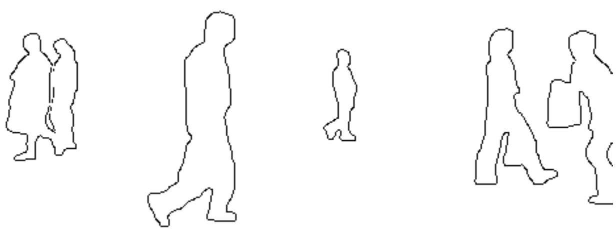


Figure 4. Superimposing the extracted contour (with green color) based on the 3D points over the superpixels regions of pedestrian hypotheses

In order to quantitatively evaluate the superpixels based and 3D points based pedestrian hypotheses shape confidence, ground truth images (see Figure 5) and a distance metric are required. We define the distance between the ground truth pedestrian hypotheses and the extracted contours by the Chamfer distance. This is computed by computing first the distance transformation (DT) [29] (see Figure 6c, Figure 7c) of the contours that will be evaluated (see Figure 6b, Figure 7b) and then by superimposing the ground truth contour templates (see Figure 5b) and summing the values from DT images that lies under the templates [30].

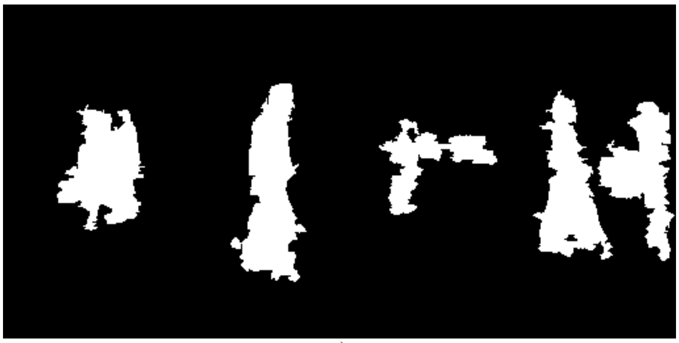
a)



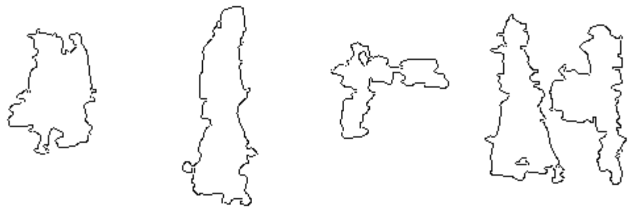
b)

Figure 5. Ground truth pedestrian hypotheses: a) area coverage of hypotheses b) corresponding extracted contour of each hypothesis

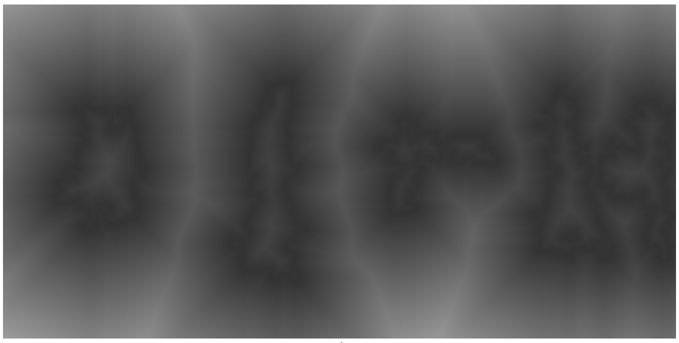
Area coverage is also evaluated by finding the direct and inverse inclusion of the ground truth pedestrian hypotheses areas to the corresponding generated pedestrian hypotheses areas. The errors are computed by summing the missing and exceeding correspondent surfaces.



a)



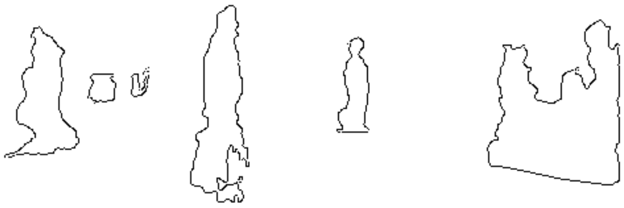
b)



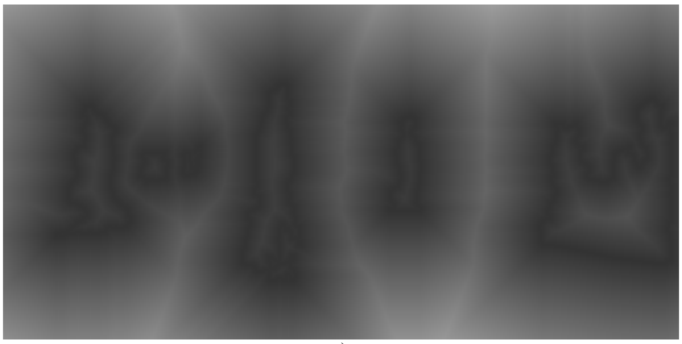
c)

Figure 6. Stereo and superpixel based pedestrian hypotheses generation: a) area coverage; b) extracted contour; c) DT image

a)



b)



c)

Figure 7. 3D points grouping based pedestrian hypotheses generation: a) area coverage; b) extracted contour; c) DT image

V. Experimental Results

In this section we present the quantitative shape and area coverage evaluation of the generated stereo and superpixels based pedestrian hypotheses. We also evaluate the same indicators on 3D points grouping based pedestrian hypotheses and highlight the better quality of the proposed stereo and superpixels approach. The evaluation was done on several sequences containing thousands stereo-vision grayscale images with 512x383 pixels acquired from different urban traffic scenarios.

We extract the pedestrian hypotheses, both with the proposed stereo and superpixels approach and with the 3D points grouping approach, from urban traffic image sequences where we manually labeled the segmentation of ground truth hypotheses. We determine the area coverage by checking the obtained pedestrian hypotheses surfaces against the ground truth surfaces. The intersection (common surface) of detected hypotheses over the ground truth must be as large as possible. In Table I, the missing and exceeding area percentages of the hypotheses to the ground truth are presented. Better hypotheses area coverage is obtained when these percentages are lower. Superior results in area coverage are achieved for pedestrian hypotheses detection by means of proposed stereo and superpixels approach when errors are decreasing by about 45% in missing area and 36% in exceeding area.

TABLE I. Errors in Pedestrian Hypotheses Area Coverage with Respect to Ground Truth

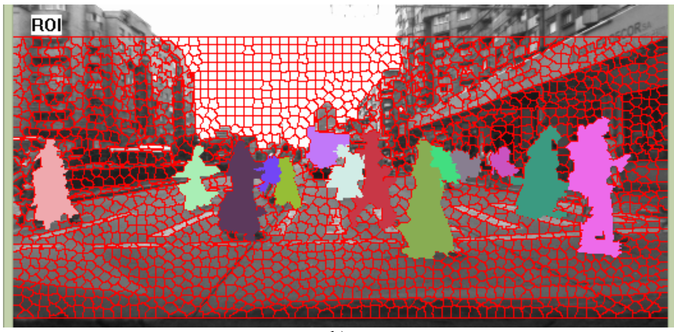
	Missing area	**Exceeding area**
Stereo and superpixels approach	16.23 %	17.96 %
3D points grouping approach	29.22 %	28.03 %

We also determine the pattern matching score between the generated pedestrian hypotheses contours and the ground truth contours. The matching score is represented by the Chamfer distance, which is computed via distance transformation of the contours. In Table II, the matching scores of the pedestrian hypotheses contours against the ground truth contours are presented. Better hypotheses contour definition is obtained when the matching score (Chamfer distance) is lower. Superior results in shape description are achieved for pedestrian hypotheses generated by means of stereo and superpixels when the matching score decreases with about 31%.

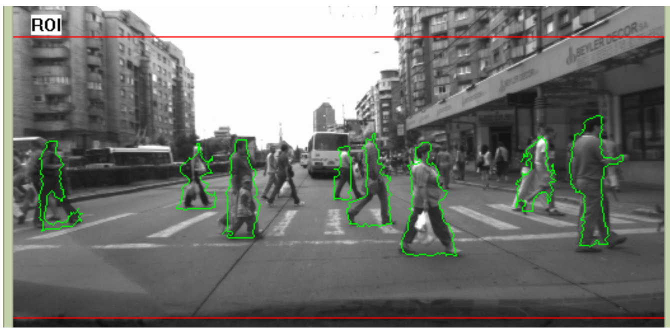
TABLE II. Pedestrian Hypotheses Contours Matching Score with Respect to Ground Truth

	Matching score (Chamfer distance)
Stereo and superpixels approach	18.14
3D points grouping approach	26.45

a)



b)



c)

Figure 8. Pedestrian hypotheses shapes: a) 3D points grouping (random color for each hypothesis); b) segmentation based on stereo and superpixels; c) contours based on the 3D points groups

Sample pedestrian hypotheses shape description results obtained by the proposed stereo and superpixels based method in comparison with the other approach based on 3D points grouping are depicted in Figure 8. Our proposed approach shows clearly superior (qualitative and quantitative) results in both area coverage and shape description of the generated pedestrian hypotheses.

VI. CONCLUSIONS

A novel stereo and superpixel based approach for extracting high quality shapes of pedestrian hypotheses from urban traffic scenarios was presented. Superpixels were extracted using the intensity images and grouped into separate obstacles by a novel paradigm that fused intensity, depth and motion information. Pedestrian hypotheses were defined as being the scene obstacles that satisfy some human-specific geometric constraints. A contour tracing algorithm was used for defining the shape of the pedestrian hypotheses.

A comparison between the pedestrian hypotheses segmentation obtained by the proposed superpixel approach and another approach based on stereo-reconstructed points grouping showed significant improvements in both object shape description and area coverage with respect to ground truth. This may further improve the accuracy of subsequent pedestrian detection modules that are using contour pattern matching techniques.

ACKNOWLEDGMENT

This work has been supported by UEFISCDI (Romanian National Research Agency) in the national research project Multi-scale Multi-modal Perception of Dynamic 3D Environments Based on the Fusion of Dense Stereo, Dense Optical Flow and Visual Odometry Information (MultiSens), project code PN-II-ID-PCE-2011-3-1086.

REFERENCES

[1] F. Oniga and S. Nedevschi, "Processing Dense Stereo Data Using Elevation Maps: Road Surface, Traffic Isle, and Obstacle Detection," *IEEE Transactions on Vehicular Technology,* vol. 59, pp. 1172-1182, 2010.

[2] M. Bertozzi, L. Bombini, P. Cerri, P. Medici, P. C. Antonello, and M. Miglietta, "Obstacle detection and classification fusing radar and vision," in *IEEE Intelligent Vehicles Symposium*, 2008, pp. 608-613.

[3] C. D. Pantilie and S. Nedevschi, "SORT-SGM: Subpixel Optimized Real-Time Semiglobal Matching for Intelligent Vehicles," *IEEE Transactions on Vehicular Technology,* vol. 61, pp. 1032-1042, 2012.

[4] C. Pocol, S. Nedevschi, and M. M. Meinecke, "Obstacle Detection Based on Dense Stereovision for Urban ACC Systems," in *International Workshop on Intelligent Transportation (WIT)*, Hamburg, Germany, 2008, pp. 13-18.

[5] S. Nedevschi, S. Bota, and C. Tomiuc, "Stereo-Based Pedestrian Detection for Collision-Avoidance Applications," *IEEE Transactions on Intelligent Transportation Systems,* vol. 10, pp. 380-391, 2009.

[6] M. Bertozzi, A. Broggi, A. Fascioli, and S. Nichele, "Stereo vision-based vehicle detection," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2000, pp. 39-44.

[7] M. Bertozzi and A. Broggi, "GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing,* vol. 7, pp. 62-81, 1998.

[8] M. Heikkila and M. Pietikainen, "A Texture-Based Method for Modeling the Background and Detecting Moving Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 657-662, 2006.

[9] A. Broggi, M. Buzzoni, M. Felisa, and P. Zani, "Stereo obstacle detection in challenging environments: The VIAC experience," in *International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 1599-1604.

[10] S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, *et al.*, "High accuracy stereo vision system for far distance obstacle detection," in *IEEE Intelligent Vehicles Symposium*, 2004, pp. 292-297.

[11] D. F. Llorca, M. A. Sotelo, A. M. Hellín, A. Orellana, M. Gavilan, I. G. Daza, *et al.*, "Stereo regions-of-interest selection for pedestrian protection: A survey," *Transportation research part C: emerging technologies,* vol. 25, pp. 226-237, 2012.

[12] G. Dong, T. Fraichard, X. Ming, and C. Laugier, "Color modeling by spherical influence field in sensing driving environment," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2000, pp. 249-254.

[13] H. Mori and N. M. Charkari, "Shadow and rhythm as sign patterns of obstacle detection," in *IEEE International Symposium on Industrial Electronics*, 1993, pp. 271-277.

[14] S. Nedevschi, R. Danescu, T. Marita, F. Oniga, C. Pocol, S. Sobol, *et al.*, "A Sensor for Urban Driving Assistance Systems Based on Dense Stereovision," in *IEEE Intelligent Vehicles Symposium*, 2007, pp. 276-283.

[15] N. S. Boroujeni, S. A. Etemad, and A. Whitehead, "Fast obstacle detection using targeted optical flow," in *IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 65-68.

[16] L. You and Y. Ruichek, "Moving objects detection and recognition using sparse spatial information in urban environments," in *IEEE Intelligent Vehicles Symposium*, 2012, pp. 1060-1065.

[17] C. Hilario, J. M. Collado, J. M. Armingol, and A. d. l. Escalera, "Pedestrian detection for intelligent vehicles based on active contour models and stereo vision," *Proceedings of the 10th international conference on Computer Aided Systems Theory,* pp. 537-542, 2005.

[18] D. M. Gavrila and V. Philomin, "Real-time object detection for "smart" vehicles," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, pp. 87-93 vol.1.

[19] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *IEEE Intelligent Vehicles Symposium*, 2000, pp. 215-220.

[20] D. Gavrila, "Pedestrian Detection from a Moving Vehicle," *Proceedings of the 6th European Conference on Computer Vision-Part II,* pp. 37-49, 2000.

[21] C. Tomiuc and S. Nedevschi, "Real-time pedestrian classification exploiting 2D and 3D information," *Intelligent Transport Systems, IET,* vol. 2, pp. 201-210, 2008.

[22] D. M. Gavrila and S. Munder, "Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle," *International Journal of Computer Vision,* vol. 73, pp. 41-59, 2007.

[23] D. M. Gavrila, J. Giebel, and S. Munder, "Vision-based pedestrian detection: the PROTECTOR system," in *IEEE Intelligent Vehicles Symposium*, 2004, pp. 13-18.

[24] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision,* vol. 59, pp. 167-181, 2004/09/01 2004.

[25] R. Xiaofeng and J. Malik, "Learning a classification model for segmentation," in *Proceedings of IEEE International Conference on Computer Vision*, 2003, pp. 10-17 vol.1.

[26] I. Giosan and S. Nedevschi, "Superpixel-based obstacle segmentation from dense stereo urban traffic scenarios using intensity, depth and optical flow information," in *IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, 2014, pp. 1662-1668.

[27] J. Shi and C. Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593-600.

[28] I. Giosan and S. Nedevschi, "Building Pedestrian Contour Hierarchies for Improving Detection in Traffic Scenes," *Computer Vision and Graphics - Lecture Notes in Computer Science,* pp. 154-163, 2009.

[29] R. Fabbri, L. D. F. Costa, J. C. Torelli, and O. M. Bruno, "2D Euclidean distance transform algorithms: A comparative survey," *ACM Computing Surveys (CSUR),* vol. 40, pp. 1-44, 2008.

[30] I. Giosan, S. Nedevschi, and S. Bota, "Real time stereo vision based pedestrian detection using full body contours," in *IEEE 5th International Conference on Intelligent Computer Communication and Processing*, 2009, pp. 79-86.