

A Multi-Baseline Stereo System for Scene Segmentation in Natural Environments

Annalisa Milella
Institute of Intelligent Systems
for Automation (ISSIA)
National Research Council (CNR)
Bari, Italy
Email: milella@ba.issia.cnr.it

Giulio Reina
Department of Engineering
for Innovation
University of Salento
Lecce, Italy
Email: giulio.reina@unisalento.it

Mario Massimo Foglia
Department of Mechanics,
Mathematics and Management
Polytechnic of Bari
Bari, Italy
Email: mm.foglia@poliba.it

Abstract—A long range visual perception system is presented based on a multi-baseline stereo frame. The system is intended to be used onboard an autonomous vehicle operating in natural settings, such as an agricultural environment, to perform 3D scene reconstruction and segmentation tasks. First, the multi-baseline stereo sensor and the associated processing algorithms are described; then, a self-learning ground classifier is applied to segment the scene into ground and non-ground regions, using geometric features, without any *a priori* assumption on the terrain characteristics. Experimental results obtained with an off-road vehicle operating in an agricultural test field are presented to validate the proposed approach. It is shown that the use of a multi-baseline stereo frame allows for accurate reconstruction and scene segmentation at a wide range of viewing distances, thus increasing the overall flexibility and reliability of the perception system.

I. INTRODUCTION

Stereovision is a widely adopted sensorial input for outdoor autonomous vehicles, as it provides an effective technique to extract range information and perform complex scene understanding tasks [1]-[5]. Nevertheless, the accuracy of stereo reconstruction is generally affected by various design parameters, such as the baseline [6]-[8]. A larger baseline decreases the common field of view of the two cameras, but it leads to higher accuracy at each visible distance. A long baseline also requires a larger disparity search range, which implies a greater possibility of false matches. Hence, the choice of the optimal baseline results from the balance of opposing although equally important factors.

In this paper, a multi-baseline stereo frame is proposed, which allows an autonomous vehicle operating in natural settings, such as an agricultural environment, to perform accurate 3D scene reconstruction and segmentation in a wide range of distances. The system was implemented within the project Ambient Awareness for Autonomous Agricultural Vehicles (QUAD-AV) funded by the ERA-NET ICT-AGRI action, aimed to enable safe autonomous navigation in high-vegetated, off-road terrain [9]. The developed stereo frame is shown in Figure 1. It is composed by two trinocular heads, one featuring a short baseline system and the other one featuring a long baseline system. By employing the narrow baseline to reconstruct nearby points and the wide baseline for more distant points, this system takes the advantage of the small minimum range of the narrow baseline, while preserving the higher accuracy and maximum range of the wide baseline

configuration. The two trinocular cameras can be either used simultaneously to widen the overall perception range of the vehicle, or alternately depending on the vehicle travel conditions. For instance, the narrow baseline configuration is useful in low-speed operations, where less noisy measurements are needed, while the wide baseline is suitable when the vehicle travels at higher speed, enabling it to perceive far away obstacles [8]. In addition, the wide baseline can improve the quality of the stereo range data for distant terrain mapping [10]. Therefore, the use of a multi-baseline stereo frame allows one to get good results at a wide range of viewing distances, and to increase the overall flexibility and reliability of the system. The 3D point cloud returned by either trinocular camera provides a rich source of information for the vehicle to perform key navigation tasks, such as terrain identification and scene segmentation. In this investigation, stereo reconstructed points were used as input data to a geometry-based classifier that segments the scene into ground and non-ground regions. This classifier features a self-learning framework [11], where the ground model is automatically built during an initial bootstrapping stage and is continuously updated to incorporate changes in the ground appearance. During the training stage, the classifier learns to associate the geometric appearance of data with class labels. Then, it makes predictions based on past observations classifying new acquired data. It is worth to note that, in the context of this work, the ground class denotes points from traversable terrain, whereas the non-ground class corresponds to all other data, including points from non-traversable ground, above ground objects (i.e., obstacles) or occluded areas, and poor stereo reconstructions. While most of the algorithms proposed in the literature rely on ground plane estimation [12]-[15] and perform obstacle detection by identifying objects that "stick out" of the ground, in this work,

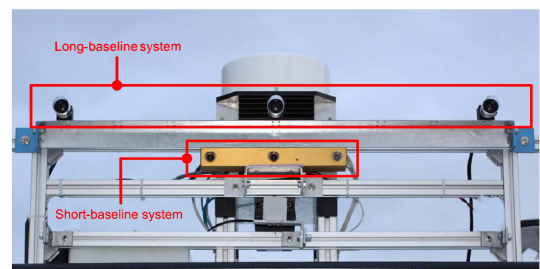


Fig. 1: The multi-baseline stereo vision system.



Fig. 2: Experimental test bed provided by IRSTEA and used for field validation in the QUAD-AV project.

ground plane reasoning is not explicitly needed and the system automatically adapts to the changing geometry of the terrain. Furthermore, in contrast to previous works that attempt to explicitly identify obstacles [1], [16], the proposed approach aims at detecting scene regions that are traversable-safe for the vehicle. This is a subtle, but significant difference; only those regions where there is evidence that it is safe are labeled as traversable, thereby avoiding both positive and negative obstacles and non-traversable terrains without explicitly detecting them.

The multi-baseline stereo system was integrated with an off-road vehicle (see Figure 2) that was made available by the partner IRSTEA at the Montoldre farm facility [17], during an experimental campaign in October 2012. The vehicle’s sensor suite included as well a 3D SICK laser rangefinder, a frequency modulated continuous wave (FMCW) radar, and a thermal infrared camera.

The remainder of this paper is organized as follows. Section II provides details about the implementation of the multi-baseline stereo system, including design and calibration issues, and stereo processing algorithms. In Section III, the ground detection approach is introduced. Experimental results are presented in Section IV. Section V concludes this paper.

II. MULTI-BASELINE STEREO VISION

The baseline, i.e., the distance between the optical centers of two cameras in a stereo head, is a critical parameter in the design of a stereo system. A short baseline makes the point of view closer, thus widening the common field of view of the two cameras. A long baseline improves the reconstruction accuracy and the range resolution (i.e., the smallest change in range that is discernible by the stereo system) at all visible distances, at the cost of a reduction of the field of view. Furthermore, a wide baseline needs wider disparity search range, thus leading to an increased possibility of false matches.

Another critical parameter is the lens focal length: a short focal length increases the angular field of view, but induces higher distortion. It also increases (i.e., makes worse) the range resolution. Lenses with larger focal lengths produce images that are zoomed in farther, allowing for the detection of distant objects. Nevertheless, the greater the focal length, the narrower the field of view. Different combinations of baselines and optics should be, therefore, used in different operational conditions. For instance, a stereo pair with longer baseline and focal length is suitable to detect far objects and should be used at high travel speed. At lower speeds, instead, a short baseline can be effectively adopted.

TABLE I: Specifications of the Bumblebee XB3.

Sensor	Three Sony 1/3" progressive scan CCD, Color
Baseline	12 cm and 24 cm
Resolution and FPS	1280 × 960 pixels at 15 FPS
Focal Length	3.8 mm
Field of View	66deg(H) × 50deg(V)

TABLE II: Specifications of the custom-built trinocular system.

Sensor	Three Sony ICX267 1/2" CCD, Color
Baseline	40 cm and 80 cm
Resolution and FPS	1384 × 1032 pixels at 16 FPS
Focal Length	12 mm
Field of View	23deg(H) × 17deg(V)

In this work, a multi-baseline system is employed to combine the advantages coming from different baselines and optics, thus improving the system flexibility. In the rest of this section, first, details concerning the implementation and calibration of the multi-baseline stereo frame used for data acquisition are provided, then the stereo processing algorithms for 3D scene reconstruction are presented.

A. Description of the System

The system comprises two trinocular cameras, featuring four baselines, two for each of them, covering the short range and the medium-long range, respectively. The system is shown in Figure 1. The short range camera is the Bumblebee XB3 by Point Grey. It consists of a trinocular stereo head with 3.8 mm focal length lenses, featuring two stereo configurations: a narrow stereo pair with a baseline of 0.12 m (*XB3-Narrow*) using the right and middle cameras, and a wide stereo pair with a baseline of 0.24 m (*XB3-Wide*) using the left and right cameras. The second trinocular system is custom-built. It comprises three identical Flea3 cameras by Point Grey with 12 mm focal length lenses, disposed in line on an aluminum bar to form two baselines: a narrow baseline of 0.40 m (*Flea3-Narrow*) using the left and middle cameras and a wide baseline of 0.80 m (*Flea3-Wide*), using the left and right cameras. Additional technical details are provided in Table I and Table II for the XB3 and the Flea3, respectively.

B. Reconstruction Error

The theoretical percentage error $E_{z_c}(\%)$ on the reconstruction in the direction of the camera optical axis (z_c -axis) can be calculated as:

$$E_{z_c}(\%) = \frac{E_{z_c}}{z_c} \times 100 \quad (1)$$

where

$$E_{z_c} = \frac{z_c^2}{B \cdot StereoF} \cdot corrAcc \quad (2)$$

being B the baseline, $StereoF$ the focal length in pixels, $corrAcc$ the correlation accuracy (i.e., the matching error in pixels) [7]. The percentage reconstruction error expressed by Eq. (1) and Eq. (2) assuming an image resolution of 640×480 pixels and a correlation accuracy of 0.2 pixels is shown in the graph of Figure 3, for each stereo pair of the system. It can be observed that the reconstruction accuracy decreases with the range and improves at higher baseline and focal

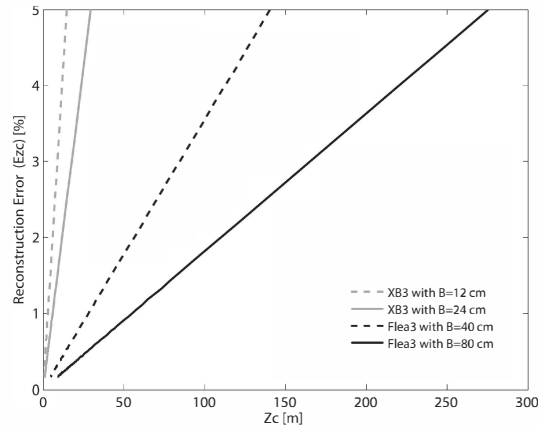


Fig. 3: Percentage reconstruction error for the multi-baseline system.

length. Therefore, by combining different baselines and optics, it is possible to keep the reconstruction error within a given percentage by setting different cutoff distance thresholds for each baseline. It should be also noted that for a disparity search range of 128 pixels the closest distance along the camera optical axis that can be reconstructed by the XB3 camera is of about 0.5 m, while it is of approximately 5 m for the Flea3 system. Hence, using a wide baseline, information in the close range is lost, while it can be preserved by adopting a short baseline. In addition, a longer focal length moves farther the viewpoint and restricts the angular field of view of the system, thus making even more significant the loss of information in the short range.

C. Calibration

Each stereo pair was calibrated using the OpenCV calibration functions [18]. Both intrinsic and extrinsic parameters were estimated for each stereo system based on a set of images of a planar checkerboard. The calibration functions also returned the rectification matrices to rectify the images as a preliminary step before applying the stereo matching algorithm. Since the four stereo pairs were calibrated separately, an additional calibration step was successively performed to align all the systems with respect to a common reference frame attached to the vehicle. To this aim, the calibration pattern was positioned at a known location with respect to the vehicle. Then, the position and orientation of each stereo pair



Fig. 4: Sample images acquired during experimentation by the Flea3 system (a) and the XB3 system (b). It can be noticed that the angular field of view of the XB3 is wider than the one of the Flea3 due to the use of a shorter focal length. However, the Flea3 image displays more clearly far away objects.

with respect to the pattern was estimated using the calibration functions. Finally, the position and orientation of each stereo pair with respect to the vehicle was inferred.

D. Stereo Processing Algorithms

Since the two trinocular systems have very different fields of view also due to the use of different lenses (see Figure 4 as an example), we integrate separately the wide and narrow baseline of each trinocular camera, so that, in the end, two point clouds are obtained: one from the Flea3 system, to be used to get accurate information in the long range (approximately up to 60 m from the vehicle), and the other one from the XB3 system to get accurate information in the short range (approximately up to 30 m from the vehicle).

For each stereo pair, the stereo processing algorithm includes the following steps:

- *Rectification*: each image plane is transformed so that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes. Using rectified images, the problem of computing correspondences is reduced from a 2-D to a 1-D search problem, typically along the horizontal raster lines of the rectified images. Rectification matrices are computed in the calibration step as described in Section II-C.
- *Disparity map computation*: to compute the disparity map a stereo block matching algorithm is used that finds corresponding points by a sliding Sum of Absolute Difference (SAD) window [18].
- *3D point cloud generation in the reference camera frame*: being the stereo pair calibrated both intrinsically and extrinsically, disparity values can be converted in depth values and 3D coordinates can be computed in the reference camera frame for all matched points.
- *Transformation from the reference camera frame to the vehicle reference frame*: in this stage 3D points are transformed from the camera frame to the vehicle frame.
- *Statistical filtering*: a statistical filter is applied to reduce noise and remove outlying points.
- *Voxelization*: in order to decrease the computational burden the number of points is reduced using a voxelized grid approach. A 3D voxel grid with a leaf size of 10 cm is created over the input point cloud space. Then, all the points in each voxel are approximated with their centroid.

The point clouds reconstructed by the narrow pair and by the wide pair of each trinocular sensor are fused in a unique point cloud: if a point of the scene has been reconstructed by both the wide baseline pair and the narrow baseline pair, only information coming from the wide baseline is retained, since a wider baseline generally assures better accuracy at every distance. Figure 5 shows the results obtained for a sample scene in the far range (a, b) acquired by the Flea3 system, and in the short range (c, d) acquired by the XB3 system.

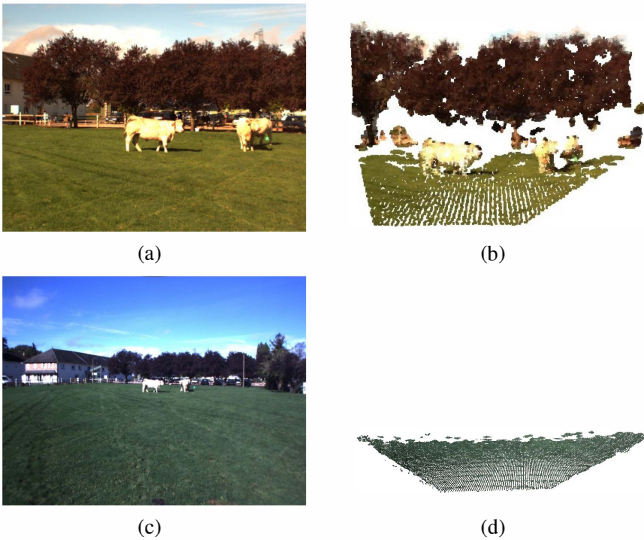


Fig. 5: Results of the stereo processing algorithm with selected cutoff thresholds in the long range (a, b) for the Flea3 system and in the short range (c, d) for the XB3 system.

III. GEOMETRY-BASED GROUND CLASSIFICATION

Geometry-based ground classification is a method for labeling observations based on their geometric properties. Specifically, the ground model is constructed upon a set of geometric features that can be extracted from stereo reconstruction. The proposed approach takes advantage of a self-learning scheme where the training instances are automatically produced using a rolling training set.

Self-learning methods have been recently introduced in the literature [19]-[21] to provide automatically trained classifiers that reduce or eliminate human intervention. In our system, the training set is initialized at the beginning of the robot's operation via a bootstrapping approach and progressively updated. The only underlying assumption to initialize the training set is that the vehicle starts its operation from an area free of obstacles, so that the trinocular system initially "looks" at ground only. Then, geometric features can be extracted from the 3D point cloud and associated with the ground class. When sufficient data is accumulated, the geometry-based ground classifier can be trained, and the ground class is related with point cloud properties. This allows the system to predict the presence of ground in successive scenes, based on past observations. The model (i.e., the training set) is continuously updated using the most recent acquisitions to account for variations in ground characteristics during the vehicle travel. The single steps of the classification system are described in the rest of this section. The reader may also refer to [5] for additional information.

A. Geometric Features

The raw point cloud obtained by the stereo system is divided into a grid of 0.4×0.4 m terrain patches projected onto a horizontal plane. Geometric features are statistics obtained from the point coordinates associated with each terrain patch. The first element of the geometric feature vector is the average slope of the terrain patch, i.e., the angle θ between the least-squares-fit plane and the horizontal plane. The second

component is the goodness of fit, E , measured as the mean-squared deviation of the points from the least-squares plane along its normal. The third element is the variance in the height of the range data points with respect to the horizontal plane, σ_h^2 . The fourth component is the mean height of the range data points, \bar{h} . Thus, the geometric properties of each patch is represented as a 4-element vector

$$f = [\theta, E, \sigma_h^2, \bar{h}] \quad (3)$$

B. Ground Modeling and Classification

The ground modeling problem is formulated as a one-class classifier [22]. One-class classification methods are generally useful in two-class classification problems, where one class, referred to as the target class, is relatively well-sampled, while the other class, referred to as the outlier class, is relatively under-sampled or difficult to model. This is the case for our application where most of the patches belong to the ground with sparse instances of non-ground. To model the ground, we adopt a multivariate Gaussian distribution to model positive ground samples, and we implement a Mahalanobis distance classifier [19].

Let us consider N_G ground patterns. The ground pattern i is represented by its m -dimensional row feature vector f_G^i , with m being the number of feature variables (4 in our case). These vectors constitute the training set X , expressed in the form of a $N_G \times m$ matrix. If we compute the sample mean μ and the sample covariance Σ of the data in X , we can denote the ground model as $M(\mu, \Sigma)$. Then, given a new pattern f_{new} , the squared Mahalanobis distance between f_{new} and $M(\mu, \Sigma)$ is defined as:

$$d^2 = (f_{new} - \mu)\Sigma^{-1}(f_{new} - \mu)^T \quad (4)$$

The pattern is an outlier, i.e. it is defined as a non-ground sample, if d^2 is greater than a threshold. The latter is computed as the α -quantile $\chi_{m;\alpha}^2$ of the *chi-square* distribution with m degrees of freedom. Note that, in order to update the ground class during the vehicle motion, the model $M(\mu, \Sigma)$ is continuously rebuilt, always using the ground feature vectors obtained by the most recent acquisitions.

IV. EXPERIMENTAL RESULTS

In this section, experimental results are presented to validate the proposed approach. The multi-baseline stereo frame was mounted on the off-road vehicle shown in Figure 2. During the experiments, the vehicle was driven by a human operator with a travel speed ranging between 10 and 20 km/h, as the on-board sensors acquired data from the surrounding environment. Various scenarios were analyzed including positive obstacles (trees, crops, metallic poles, buildings, agricultural equipment), negative obstacles (holes, ditches), moving obstacles (vehicles, people and animals), and difficult terrain (steep slopes, highly-irregular terrain, etc.). Then, the proposed classification framework was applied offline. For each data set, the vehicle started its operations from an area that was clear of obstacles in order to initialize the ground model by acquiring a few frames. After the training stage, the stereo classifier was able to predict the presence of ground in successive acquisitions.

A. Stereo Reconstruction

The results of scene reconstruction performed using XB3 and Flea3 data for two different scenarios are shown in Figure 6 and Figure 7. In order to evaluate the reconstruction capability of each system, no cutoff threshold on the range was used in these tests. In Figure 6, the scenario presents relatively flat ground and a building on the left in the vicinity of the vehicle, and buildings and people in the far range. Specifically, Figure 6a and Figure 6b show the reference images acquired by the Flea3 and XB3, respectively. A 3D view of the point clouds returned by each system is shown in Figure 6c and Figure 6d. It can be observed that while the Flea3 is able to reconstruct also the farthest building located at approximately 100 m from the vehicle, this building is filtered out in the XB3 reconstruction. This can be better seen in the close up of the far range reported in Figure 6e and Figure 6f for the

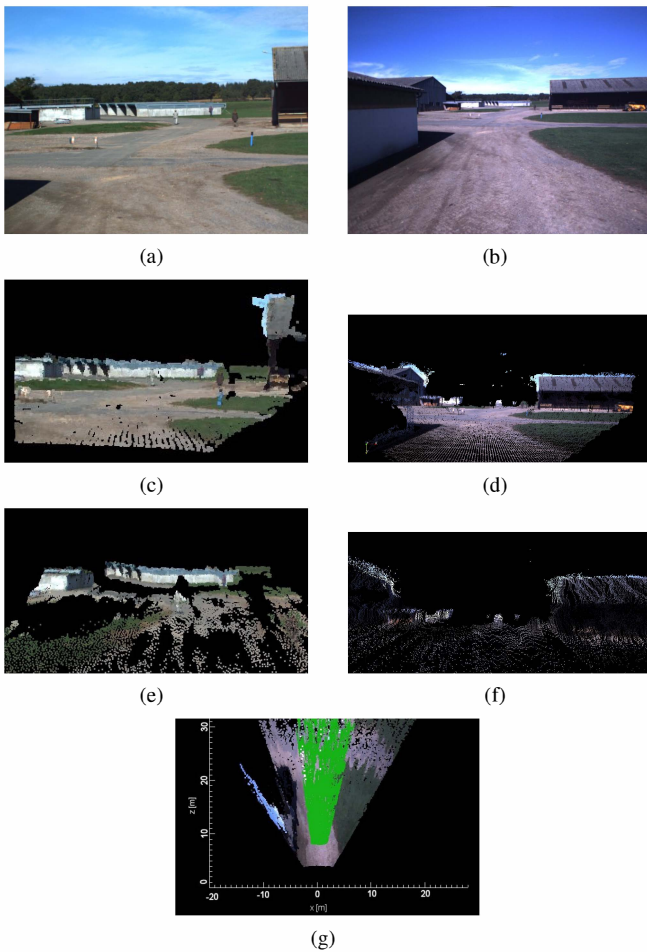


Fig. 6: Sample scenario acquired during field experiments: (a) reference image of the Flea3 system; (b) reference image of the XB3 system; (c) point cloud obtained by Flea3 for the whole scene; (d) point cloud obtained by XB3 for the whole scene; (e) close up of the long range in the Flea3 reconstruction; (f) close up of the long range in the XB3 reconstruction; (g) upper view of the close range: RGB points are used for XB3 and green points for Flea3 data. It is shown that the Flea3 system provides accurate information in the long range while losing information in the short range compared to the XB3 camera.

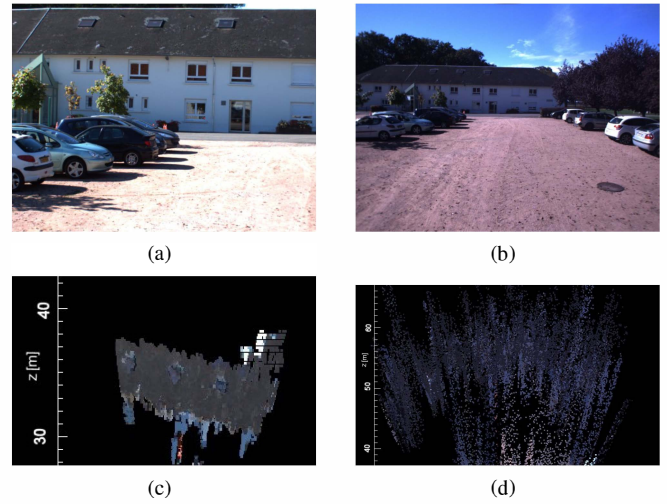


Fig. 7: Sample scenario acquired during experimentation on field: (a) reference image of the Flea3 system; (b) reference image of the XB3 system; (c) upper view of the 3D reconstruction of the building obtained by using Flea3 data; (d) upper view of the 3D reconstruction of the building obtained by using XB3 data. It can be observed that in the long range the XB3 produces noisier results with respect to the Flea3 camera.

Flea3 and the XB3, respectively. Finally, a close up of the short range with the two point clouds overlapped is shown in Figure 6g, with green points representing the Flea3 point cloud and RGB points representing the XB3 point cloud. This figure shows that the Flea3 is not able to detect nearby regions as its point of view is located farther than the one of the XB3. In addition, it has a narrower angular field of view that causes the loss of important information on the building on the left of the vehicle, which is detected by the XB3 system, instead. These considerations justify the need for combining both systems.

Similar observations can be done for the sample case reported in Figure 7. This figure is referred to a scenario with even ground in the vicinity of the vehicle, and cars, trees, and a building in the medium-far range (see Figure 7a for the Flea3 and Figure 7b for the XB3). The better accuracy of the Flea3 system in the medium-far range with respect to the XB3 can be seen by looking at the results obtained for the reconstruction of the building, which was located at a distance of approximately 35 m from the vehicle. Specifically, Figure 7c and Figure 7d show an upper view of the 3D reconstruction of the building as obtained by the two stereo systems. As expected, the results provided by the XB3 in the long range are noisier and have lower range accuracy than those produced by the Flea3 system (using the XB3, points belonging to the building are reconstructed with a wide range span between 40 and 60 m).

B. Ground Detection

Some typical results obtained from the geometry-based classifier during field experiments are shown in Figure 8 and Figure 9. A scenario with relatively even ground in the vicinity of the vehicle and some buildings in the far range is shown in Figure 8. Figure 8a refers to the Flea3 system, while Figure 8b refers to the XB3 camera. In these figures, the results obtained from the geometry classifier are projected

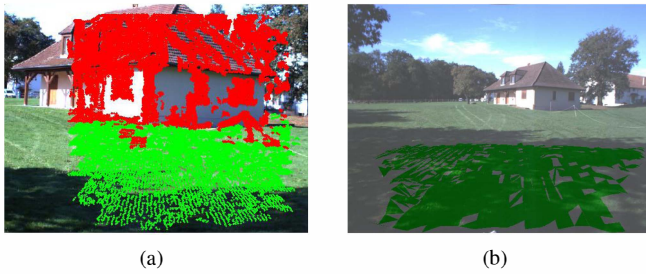


Fig. 8: Results of geometry-based classifier for a scene with low grass, trees and buildings: (a) long range classification with Flea3 data; (b) short range classification with XB3 data.

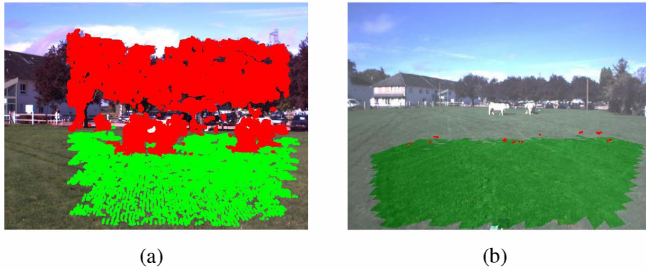


Fig. 9: Results of geometry-based classifier for a scene with low grass, trees, buildings, and animals: (a) long range classification with Flea3 data; (b) short range classification with XB3 data.

over the image plane of the reference camera of each system. Points that belong to a cell labeled as ground are denoted by green dots, whereas points falling into cells marked as non-ground are denoted by red dots. As can be seen, the classifier correctly segment ground and non-ground regions. The same considerations apply to Figure 9 where animals are also present in the scene.

V. CONCLUSION

In this paper, a multi-baseline stereovision system for autonomous navigation in unstructured environments was introduced. The system features a short baseline and a long baseline trinocular camera. The former is used to reconstruct nearby points, while the latter is employed to reconstruct distant points, thus allowing for accurate reconstruction both in the near range and in the medium-long range. A self-learning framework using geometric features extracted by the 3D stereo data returned by the multi-baseline camera was also described. It enables the vehicle to detect traversable ground based on a ground model, which is automatically built at the beginning of the robot operation and updated during the robot travel. Experimental results obtained using a test platform in agricultural scenarios were presented to validate the proposed system.

ACKNOWLEDGMENT

The financial support of the ERA-NET ICT-AGRI through the grant Ambient Awareness for Autonomous Agricultural Vehicles (QUAD-AV) is gratefully acknowledged.

REFERENCES

- [1] A. Rankin, A. Huertas, L. Matthies, *Evaluation of Stereo Vision Obstacle Detection Algorithms for Off-Road Autonomous Navigation*, In AUVSI Symp. on Unmanned Systems, 2005.
- [2] F. Rovira-Más, Q. Zhang, J.F. Reid, *Stereo vision three-dimensional terrain maps for precision agriculture*, Computer and Electronics in Agriculture, vol. 60, no. 2, pp. 133-143, March 2008.
- [3] K. Konolige, M. Agrawal, R.C. Bolles, C. Cowan, M. Fischler, B. Gerkey, *Outdoor Mapping and Navigation Using Stereo Vision*, Experimental Robotics, Springer Tracts in Advanced Robotics, vol. 39, pp. 179-190, 2008.
- [4] A. Milella, G. Reina, R. Siegwart, *Computer Vision Methods for Improved Mobile Robot State Estimation in Challenging Terrains*, Journal of Multimedia, vol. 1, no. 7, pp. 49-61, Nov./ Dec. 2006.
- [5] G. Reina, A. Milella, *Towards Autonomous Agriculture: Automatic Ground Detection Using Trinocular Stereovision*, Sensors, vol. 12, no. 9, pp. 12405-12423, doi:10.3390/s120912405, 2012.
- [6] M. Okutomi, T. Kanade, *A Multiple-baseline stereo*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 4, 1993.
- [7] D. Gallup, J.M. Frahm, P. Mordohai, M. Pollefeys, *Variable baseline/resolution stereo*, IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008, 23-28 June 2008.
- [8] A. Broggi, C. Caraffi, S. Cattani, S. Ghidoni, P. Porta, P. Zani, *TerraMax Vision at the Urban Challenge 2007*, IEEE Transactions On Intelligent Transportation Systems, vol. 11, no. 1, March 2010.
- [9] R. Rouveure, M. Nielsen, A. Petersen, G. Reina, M. Foglia, R. Worst, S. Seyed-Sadri, M. Blas, P. Faure, A. Milella, K. Lykkegard, *The QUAD-AV Project: Multi-Sensory Approach for Obstacle Detection in Agricultural Autonomous Robotics*, In Proc. of 2012 International Conference of Agricultural Engineering CIGR-Ageng, Valencia, Spain, 8-12 July 2012.
- [10] C.F. Olson, H. Abi-Rached, *Wide-baseline stereo vision for terrain mapping*, Machine Vision and Applications, vol. 21, pp. 713-725, 2010.
- [11] G. Reina, A. Milella, J. Underwood, *Self-learning classification of radar features for scene understanding*, Robotics and Autonomous Systems, vol. 60, no. 11, pp. 1377-1388, November 2012.
- [12] R. Hadsell, P. Sermanet, J. Ben, A. Erkan, M. Scoffier, K. Kavukcuoglu, U. Muller, Y. LeCun, *Learning long-range vision for autonomous off-road driving*, J. Field Robot., vol. 26, pp. 120-144, 2009.
- [13] K. Konolige, M. Agrawal, M.R. Blas, R.C. Bolles, B.P. Gerkey, J. Solà, A. Sundaresan, *Mapping, navigation, and learning for off-road traversal*, J. Field Robot., vol. 26, pp. 88-113, 2009.
- [14] P. Vernaza, B. Taskar, D.D. Lee, *Online, self-supervised terrain classification via discriminatively trained submodular Markov random fields*, IEEE International Conference on Robotics and Automation ICRA 2008, pp. 2750-2757.
- [15] G. Reina, G. Ishigami, K. Nagatani, K. Yoshida, *Odometry correction using visual slip-angle estimation for planetary exploration rovers*, Advanced Robotics, vol. 24, no. 3, pp. 359-385, 2010.
- [16] R. Manduchi, A. Castano, A. Talukder, L. Matthies, *Obstacle detection and terrain classification for autonomous off-road navigation*, Auton. Robot., vol. 18, pp. 81-102, 2003.
- [17] Available on line: <http://www.irstea.fr/la-recherche/th%C3%A8mes-de-recherche/motive/station-de-montoldre> (Accessed on 21 January, 2013)
- [18] G. Bradski, A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, October 1, 2008, ISBN-10: 0596516134.
- [19] A. Milella, G. Reina, J. Underwood, B. Douillard, *Combining Radar and Vision for Self-Supervised Ground Segmentation in Outdoor Environments*, In Proc. of the 2011 IEEE/RSJ IROS, San Francisco, CA, USA, 2011, pp. 255-260.
- [20] D. Stavens, S. Thrun, S., *Self-Supervised Terrain Roughness Estimator for Offroad Autonomous Driving*, In Proc. of the 22nd Conference on Uncertainty in Artificial Intelligence, USA, July 2006, pp. 255-260.
- [21] S. Zhou, J. Xi, M.W. McDaniel, T. Nishihata, P. Salesses, K. Iagnemma, *Self-supervised learning to visually detect terrain surfaces for autonomous robots operating in forested terrain*, J. Field Robot., March 2012, vol. 29, no. 2, pp. 277-297.
- [22] D. Tax, *One-Class Classification. Concept Learning in the Absence of Counter Examples*, Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2001.