On improving IED object detection using stereo by exploiting scene geometry processing

van de Wouw, D.W.J.M.; Dubbelman, G.; de With, P.H.N.

Published: 01/01/2015

On improving IED object detection by exploiting scene geometry using stereo processing

Dennis W.J.M. van de Wouw
Eindhoven University of Technology
ViNotion B.V., Eindhoven
d.w.j.m.v.d.wouw@tue.nl

Gijs Dubbelman, Peter H.N. de With
Eindhoven University of Technology
vca.ele.tue.nl

1. SUMMARY

Detecting changes in the environment with respect to an earlier data acquisition is important for several applications, such as finding Improvised Explosive Devices (IEDs) and reliable detection of traffic participants. We explore and evaluate the benefit of depth sensing in the context of automatic change detection, where we extend an existing monocular system with a second camera in a fixed stereo setup. We propose an alternative frame registration that exploits scene geometry, in particular the ground plane. Additional feature selection techniques are applied to improve feature distribution on the ground plane, thereby increasing the reliability of the registration. Furthermore, change characterization is applied to localized depth maps to distinguish between physical changes and shadows, which is one of the main challenges of a monocular system. An extensive evaluation on real-world acquisitions containing geo-tagged test objects of 10x10x10 cm is presented in the full paper. Preliminary results show that false positives are reduced by more than 30%.

2. INTRODUCTION AND PROBLEM STATEMENT

In order to reduce casualties by IEDs, reliable detection of changes along roads is of primary importance. Change detection on aerial and satellite images is widely researched and already applied to Countering IEDs. However, analysis of aerial footage has its limitations, such as resolution because of the large capturing distance and the uniform viewing point. Therefore, the interest for mobile change detection systems for ground-based surveillance is growing. This is a challenging task[4], where images from different viewpoints and time instants, have to be accurately registered and compared in real time. This paper aims at improving the robustness of such a change detection system using scene geometry acquired by a stereo camera. The proposed solution addresses the main challenges of the monocular system and aims at a more reliable system, especially in urban environments, and a higher robustness to strong shadows.

3. RELATED WORK AND CONTRIBUTIONS

Ground-based change detection is a challenging task. Some challenges, such as inaccurate image registration in poorly textured road scenes and false positives due to local shadows, are difficult to solve with a monocular setup[4]. Haberdar and Shah tackle the first problem by applying depth sensing[3]. They assume changes are located on the ground surface and register the ground planes instead of the entire image. In their work, the ground plane is found by segmenting the disparity map. Wathen et al. employ a LIDAR sensor combined with a color camera to populate a point cloud[5]. Change detection is performed point-by-point on the 3D-point cloud, hence reducing the effect of shadows which do not affect the depth.

This paper extends the work of van de Wouw et al [4] with depth sensing through the addition of a second camera, yielding a fixed stereo setup. Similar to Haberdar and Shah[3], we register the ground plane of the live and historic recording. However, instead of using image segmentation, the ground plane is extracted by fitting a spline through the 3D point cloud. This allows for multi-planar warping in the case of non-planar ground planes. Our approach improves image registration, which results in more accurate change detection. Next, we exploit depth information to improve the change detection accuracy. We propose a post-filtering metric to reduce sensitivity to shadows in the scene, by validating the detected changes from the preceding stage. Localized 3D point clouds are used to distinguish real objects from shadows, exploiting the knowledge that shadows do not affect the depth of the scene. The resulting system is extensively tested on manually annotated real-world videos, similar to the earlier referred approach[4].

4. DEPTH EXTENSIONS

We extend the original change detection system with depth sensing, where the most significant changes w.r.t. the monocular system are now described.

4.1. Stereo capturing

Images are captured by a stereo camera consisting of two state-of-the-art cameras with an adjustable baseline of up to 150 cm. At full-HD resolution, an object of $10 \times 10 \times 10$ cm will have a disparity difference of 5.7 and 3.8 pixels w.r.t. its background at a distance of 40 and 60 m, respectively. Taking into account that the disparity map has a sub-pixel resolution of 1/16th pixel, this is more than sufficient to accurately distinguish objects from their background.

4.2. Image registration

The matched features between historic and live images are the basis for image registration. In this case, only those features residing on the ground plane are used. This ground plane can be accurately determined by fitting a spline through the y-z histogram of the 3D point cloud obtained from the stereo camera[2]. The ground plane is then obtained by finding all triangulated points whose (y, z)-values satisfy the spline. This method easily extends to multi-planar warping in the case of non-planar ground surfaces, by finding piecewise linear segments in the spline.

Instead of merely selecting all features that lie on the ground plane, new features are extracted within the ground plane region. Together with feature selection techniques, such as Adaptive Non Maximal Suppression[1], the feature distribution over the ground plane is significantly improved. We have found that this positively affects the robustness of the image registration.

4.3. Depth filtering

To reduce the sensitivity to shadows, an additional change characterization module is added. For each change blob found by the monocular change detection system, the depth map is (locally) compared to the depth map of the historic scene. A change blob representing a shadow will not lead to any differences in the depth map, while a physical change, e.g. an object, will show clear changes and discontinuities. The challenge lies in coping with different viewpoints (caused by driving a different path), which result in a difference in the relative depth between the live and historic scene. Therefore, the localized 3D point clouds are first co-registered and then compared. If the resulting clouds differ significantly, the change blob is accepted as a true change.

5. EXPERIMENTS AND RESULTS

Our extended change detection system is extensively evaluated similar to that of the earlier work[4]. The system is mounted on a vehicle and videos are acquired while driving through urban and rural environments. After the first recording, wooden test objects of $10 \times 10 \times 10$ cm in multiple colors are placed in the environment at predefined locations. These objects are manually annotated and assigned the correct GPS location, after which the system can be automatically validated. After placement of the objects, a second video is captured by driving the same route now containing the manually placed test objects. This process will be explained in more detail in the final paper. The proposed depth-sensing solutions shows promising results, where false positives are reduced by more than 30% and that reliability improves in urban scenes. An extensive quantitative evaluation will be presented in the final paper.

6. ACKNOWLEDGEMENT

This research has received funding from the D-SenS project in the European Unions FP7 program, managed by REA - Research Executive Agency under grant agreement FP7-SME-2012.

References

[1] M. Brown, R. Szeliski, and S. Winder. Multi-Image Matching Using Multi-Scale Oriented Patches. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:510–517.

[2] D. Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis*, 54(4):1167–1178, Apr. 2010.

[3] H. Haberdar and S. Shah. Disparity map refinement for video based scene change detection using a mobile stereo camera platform. In *20th International Conference on Pattern Recognition (ICPR)*, pages 3890–3893, Aug 2010.

[4] D. W. J. M. Van de Wouw, K. van Rens, H. van Lint, E. G. T. Jaspers, and P. H. N. de With. Real-time change detection for countering improvised explosive devices. volume 9026, pages 90260T–90260T–15, 2014.

[5] M. Wathen, N. Link, P. Iles, J. Jinkerson, P. Mrstik, K. Kusevic, and D. Kovats. Real-time 3d change detection of ieds. volume 8379, pages 837908–837908–10, 2012.