

Graffiti Detection Using a Time-Of-Flight Camera

Federico Tombari^{1,2}, Luigi Di Stefano^{1,2}, Stefano Mattoccia^{1,2}, and Andrea Zanetti^{1,2}

¹ Department of Electronics Computer Science and Systems (DEIS)
University of Bologna

Viale Risorgimento 2, 40136 - Bologna, Italy

² Advanced Research Center on Electronic Systems (ARCES)
University of Bologna

Via Toffano 2/2, 40135 - Bologna, Italy

{federico.tombari, luigi.distefano, stefano.mattoccia}@unibo.it,
andrea.zanetti@studio.unibo.it
www.vision.deis.unibo.it

Abstract. Time-of-Flight (TOF) cameras relate to a very recent and growing technology which has already proved to be useful for computer vision tasks. In this paper we investigate on the use of a TOF camera to perform video-based graffiti detection, which can be thought of as a monitoring system able to detect acts of vandalism such as dirtying, etching and defacing walls and objects surfaces. Experimental results show promising capabilities of the proposed approach, with improvements expected as the technology gets more mature.

1 Introduction

Nowadays, countless acts of graffiti are committed daily against public and private properties all around the world. The costs caused by such damage are huge. These include obviously the direct costs of cleaning tagged surfaces as well as the loss of value of properties repeatedly damaged by graffiti. Besides, other significant indirect costs are to be imputed to the perceived uncleanness and insecurity associated with the presence of graffiti in a certain area. This results in decreases of revenues for commercial activities or services taking place in the area (e.g. shops, restaurants, rental of apartments). Similar considerations apply to public transport, for the uncleanness and perceived insecurity in railway/bus/underground stations lower passenger confidence in the transport system and consequently tend to decrease ridership.

Therefore, public bodies are trying to tackle the graffiti problem from many different perspectives. In some towns these include also attempts to establish forms of cooperation with the writers, so as to channel their needs for expression into a well-defined and socially acceptable framework (e.g. by giving them specific spaces for drawing and organizing exhibitions to show their works). Nonetheless, the need to protect properties clearly calls for a widespread adoption of surveillance cameras. Given the difficulty for security personnel to monitor simultaneously large arrays of cameras, there is also an increasing attention toward the development of intelligent video-surveillance systems specifically conceived to detect automatically the act of drawing graffiti on a surface.

As far as video analysis is concerned, in principle graffiti detection may be carried out according to two approaches: detection of the act of writing on a surface or detection

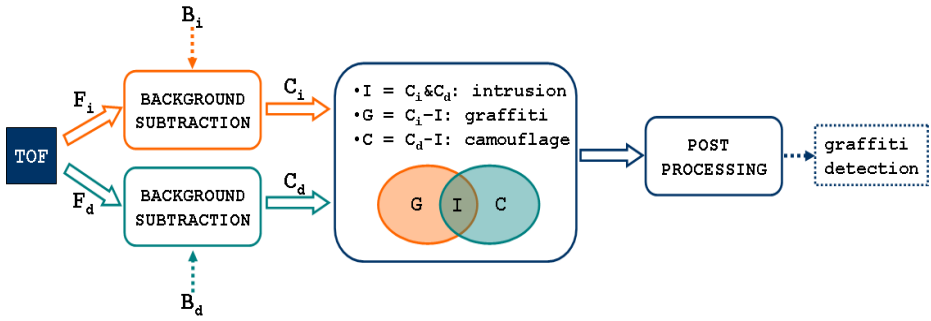


Fig. 1. Outline of the proposed algorithm

of a change on a monitored surface. The former deals with the deployment of behavior and gesture analysis techniques (e.g. [1]) for recognizing a spatio-temporal pattern corresponding to a person perpetrating the act of writing on a monitored surface. However, such techniques require accurate training of a classifier and generally perform much better when a certain degree of cooperation from the subject can be achieved, which is obviously not the case of graffiti detection. On the other hand, the latter approach consists simply in trying to detect as soon as possible the changes of appearance of the monitored surface caused by the drawings. Hence this approach may be accomplished based on standard change detection algorithms, which aim at detecting changes between the current frame and a background image and make up the primal processing step in most intelligent video-surveillance systems. In particular, as proposed in [2], graffiti detection may be carried out by detecting appearance changes with respect to a reference background which are stationary both in space and time. Yet, this method is prone to false positives in practical scenarios, since objects other than graffiti may exhibit the sought for pattern of spatio-temporal stationarity: people standing very still, parked vehicles (such as cars, motorcars, bicycles), abandoned objects.

In this paper we investigate on the use of a TOF (Time-of-Flight) camera to perform automatic graffiti detection. The rationale of this study is that since a TOF camera senses both brightness and depth at each pixel location, it may be deployed to detect graffiti by looking for stationary changes of brightness that do not correspond to changes in depth. It is clear that the use of a TOF camera holds the potential to overcome the false positives issue of the method in [2], for most still objects other than graffiti would yield both brightness and depth changes. Furthermore, it is worth noting that the same idea can be usefully employed to detect further events rather than only graffiti, such as modifications to the background surfaces. This can be useful, e.g., for cultural heritage environments or museums, to detect acts of vandalism such as painting, dirtying, etching or defacing of parts of an artwork. This will be investigated in the experimental part (Section 4).

2 TOF Cameras and Applications

Recently developed low-power TOF depth sensors [3,4,5,6] combine in single package a depth perception module and a CMOS camera. Although the first generation of these sensors had several limitations, TOF technology gained remarkable interest by virtue of

its potential suitability to many real applications. Several studies were aimed at assessing the performance of TOF technology [7, 8, 9] and compare it with other state of the art technologies for depth perception such as stereo vision [10, 11, 12].

Several improvements are expected by the forthcoming next generation of TOF cameras: increased invariance to objects reflectivity, increased spatial and depth resolution, increased signal to noise ratio for the 2D and 3D perception modules and the integration of higher resolution color CMOS cameras. Nevertheless, in spite of their limitations, current generation of TOF based depth sensors was successfully used in several applications such as video surveillance and tracking [13], facial features detection and tracking [14], classification of moving objects [15], gesture recognition [16], documentation of heritage objects [11], robot vision and obstacle detection [17, 18], safety [12]. Some authors combined TOF sensors with additional high resolution cameras [19, 20] in order to overcome the limitations of the embedded CMOS camera or stereo vision systems [21, 22] in order to improve effectiveness of depth perception. In order to improve depth and spatial resolution of depth sensors Yang *et al.* [23] recently proposed an iterative technique based on bilateral filtering [24]. Depth maps are iteratively improved applying bilateral filters on each slice of the Disparity Space Image (DSI) volume. The weights of each bilateral filter are obtained processing images provided by a high resolution color camera. This technique is currently not suited for real-time applications but provided excellent results with TOF and stereo-based depth sensors.

3 Graffiti Detection Algorithm

The proposed algorithm for graffiti detection jointly deploys depth and intensity information to detect events such as changes in the appearance of the visible surfaces in the monitored scene. The basic idea of our algorithm can be outlined as follows. First, by means of an intensity-based analysis visible changes can be detected by comparing the current intensity information with a model of the background of the scene. Then, the use of depth information can discriminate between changes occurring in the space between the background and the camera (e.g. intrusion) and those occurring directly on the background surface (e.g. graffiti). The outline of the proposed algorithm, described hereinafter, is shown in Fig. 1.

The proposed algorithm is based on background subtraction [25], that aims at detecting changes in a monitored scene by comparing each frame of the incoming 2D video stream with a model of the background of the scene. In order to build such a model, for each point of the background a statistical parameter can be computed by collecting the pixel values in that point along an initialization sequence. Typically this parameter is either the mean, the median or the mode of the pixel intensities along the initialization sequence. In our approach, we set each background pixel to the mean value computed over the initialization sequence. Then, in order to compare the background model with each frame, we adopt a basic background subtraction approach, i.e. we compute a change mask C_i by thresholding the absolute difference between each pixel intensity in the background B_i and in the current frame F_i :

$$C_i(x, y) = \begin{cases} false & |B_i(x, y) - F_i(x, y)| < T \\ true & elsewhere \end{cases} \quad (1)$$

Thus, C_i is a binary mask denoting all points which result *changed* after the comparison with the background. In the basic model, T is a fixed parameter of the algorithm. Nevertheless, when dealing with TOF cameras, the amount of noise perceived by each pixel is lower at the center of the image, where the power of the reflected light signal is higher, and increases as we get far from it. Hence, T should depend on the position (x, y) where the change mask is being evaluated. In particular, we assume that noise can be modeled as a zero-mean Gaussian distribution:

$$F_i(x, y) = F_i^T(x, y) + N_i(x, y) \tag{2}$$

with $F_i^T(x, y)$ being the noise-free version of $F_i(x, y)$ and

$$N_i(x, y) = \frac{1}{2\pi \cdot \sigma_i^2(x, y)} \cdot \exp\left\{-\frac{(x^2 + y^2)}{2 \cdot \sigma_i^2(x, y)}\right\} \tag{3}$$

Hence, parameter T in (1) is chosen to be proportional to the standard deviation of the Gaussian distribution of each pixel, which is estimated during the initialization sequence. This leads to

$$C_i(x, y) = \begin{cases} false & |B_i(x, y) - F_i(x, y)| < k_i \cdot \sigma_i(x, y) \\ true & elsewhere \end{cases} \tag{4}$$

with k_i typically ranging within $[1, \dots, 3]$ ($k_i = 1$ in our experiments).

The same approach can be carried out also for what concerns the depth information coming from the TOF sensor. In particular, a depth background model B_d can be built by averaging the depth value of each point of the depth map over an initialization sequence, assuring the background is static along that sequence. Moreover, for each point (x, y) the standard deviation $\sigma_d(x, y)$ of the depth values over the initialization sequence is also computed. Then, similarly to (4), the current depth map F_d can be compared at run-time to the depth background model B_d :

$$C_d(i, j) = \begin{cases} false & |B_d(i, j) - F_d(i, j)| < k_d \cdot \sigma_d(x, y) \\ true & elsewhere \end{cases} \tag{5}$$

with k_d typically ranging within $[1, \dots, 3]$ ($k_d = 1.5$ in our experiments).

Once C_i and C_d are computed, they are compared so to determine the presence of graffiti in the scene. In particular, the event of a point (x, y) resulting *changed* in either one of the two masks refers to one of the following three possible circumstances:

1. $C_i = true, C_d=true$: a change in intensity corresponds to a change in depth. This means that an intrusion by something/someone is currently going on.
2. $C_i = true, C_d=false$: a change in intensity does not correspond to a change in depth. Thus, a change of the appearance of the background surface has occurred: a graffiti event is triggered.
3. $C_i = false, C_d=true$: a change in depth does not correspond to a change in intensity. In this case, an intrusion has been performed by something/someone having an intensity similar to that of the background (i.e. *camouflage*).

Thus, graffiti are detected simply by choosing all points marked as *changed* on C_i and as *unchanged* on C_d .

In addition, we carry out a final post-processing stage in order to improve the reliability of our detector. Along this last stage, three different conditions are checked in order to eliminate false positives from the final graffiti mask. First, a stationarity check is performed, that is a point is detected as graffiti only if it was positively detected in the last t frames. This is a necessary measure against the high amount of noise of the camera sensor, which otherwise would produce a high number of flickering points in the final change mask. Successively, a labeling algorithm is applied on the detected graffiti regions. This allows to eliminate from the final change mask all graffiti whose area is less than a certain number of pixels, which are as well typically generated by noise. Finally, the last check eliminates all graffiti blobs having any of their 8-connected neighboring points detected as 3D intrusions in C_d , since they are most probably generated by the parts of an intruding object/person laying close to the background surface.

4 Experimental Results

In this section we show some preliminary results dealing with the application of our graffiti detector to real video sequences. Unfortunately, due to the current limits of the TOF technology and, above all, due to the characteristics of TOF camera (Canesta DP205, Field of View: 55 deg.) available to us for the experiments, which is not state-of-the-art, resolution is limited to 64×64 for both intensity and depth. Furthermore, the power of the infrared illuminator limited the maximum depth range during our tests, forcing the camera to stand not farther than 1.5~2 meters away from the background walls, otherwise the sensor is unable to detect the majority of details appearing on the background surface. In fact, at a farther distance, the intensity image tends to appear very dark.

Thus, we now show some footage dealing with some typical acts of vandalism which can be detected by our system. These acts include graffiti (e.g. writings on a surface, Video 1), object stealing (e.g. stealing a painting or a drawing hung up on the wall, Video 2), surface defacing or damaging (e.g. tearing apart a drawing on the wall, Video 3). For each sequence, we show some qualitative results by uniformly taking some snapshots of the outputs of the various stages of the algorithm along the whole sequence. In particular, for each snapshot we show the current intensity frame F_i , the current depth frame F_d , the intensity change mask C_i , the depth change mask C_d . Besides, we also superimpose on F_i the final output of our algorithm, which is a bounding box around each graffiti blob detected by our system.

Fig. 2 shows the results dealing with the graffiti sequence (Video 1). At the very beginning of the sequence, the scene is empty, with a white wall on the background (Frame #1). Then, a person enters the scene and starts making some drawings on the wall (Frames #100, 250, 500, 750, 1000). As soon as the graffiti start being visible on the scene (Frame #250), the system detects their presence and localizes them quite accurately. It is worth noting that in Frames #250 and #1000 2 false positives arise due to the fact that the person's arm, laying on the wall, is almost at the same depth as the background and is recognized as a graffiti. Then, when the person stands in front of the

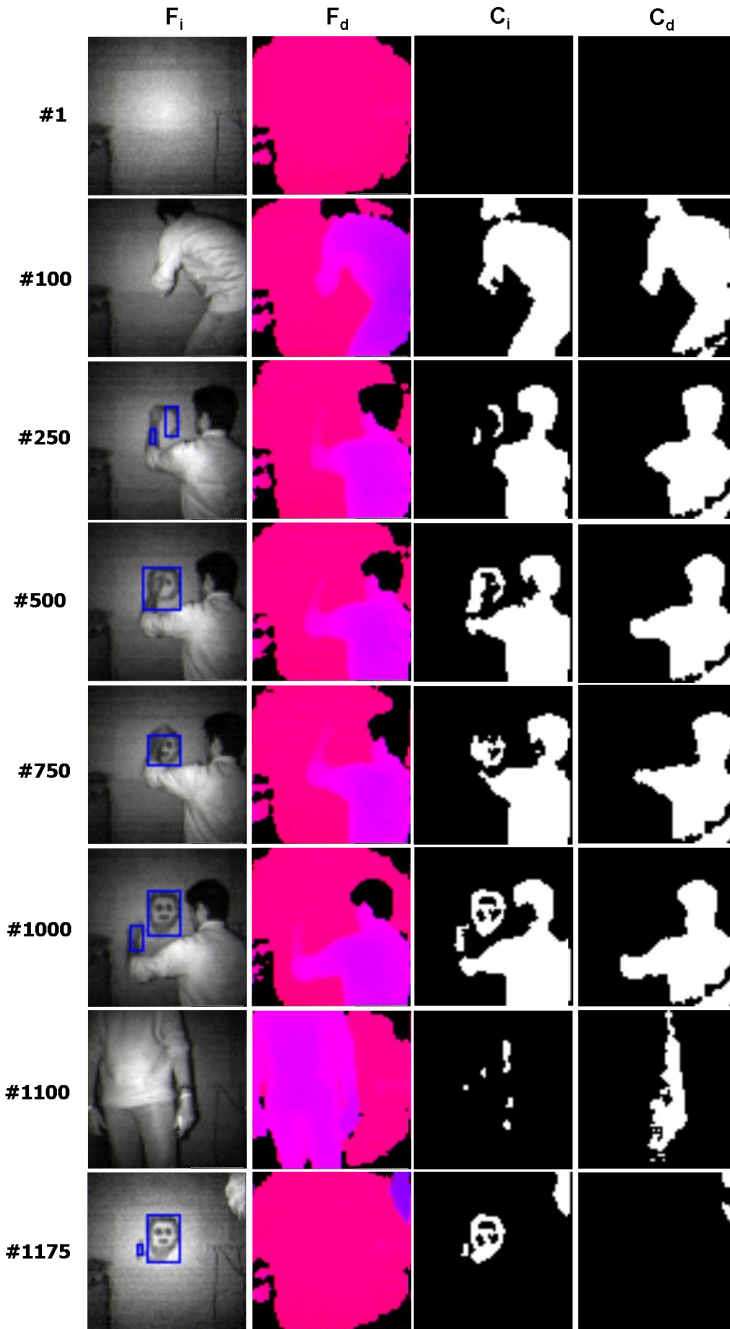


Fig. 2. Video 1: graffiti sequence

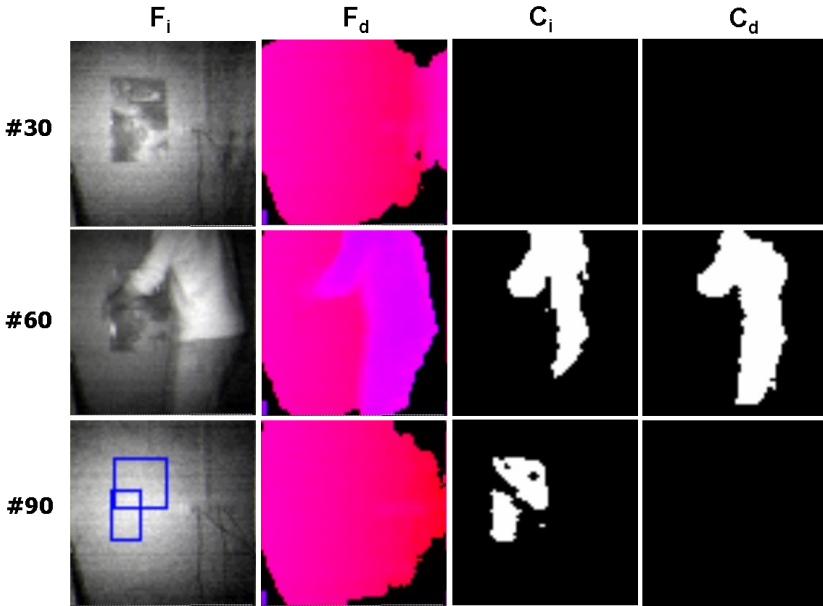


Fig. 3. Video 2: object stealing sequence

drawing, no graffiti is detected in the output (Frame #1100). Finally, Frame #1175 shows the output of the system at the very end of the sequence.

As said before, the proposed algorithm can be usefully deployed also to detect other events rather than just graffiti. In Fig. 3 results are shown concerning a sequence where a painting hanging on the wall is stolen (Video 2). The object appears at the beginning of the sequence (Frame #30). While a person is stealing the painting, no output is raised since an intrusion is present but the background has not structurally changed yet (Frame #60). Finally, when the object is removed the event is correctly detected by our system (Frame #90).

Finally, we show a sequence concerning the defacing of an object hanging on the wall (Video 3). In this last case, we also propose a slight modification to the output of the algorithm. It is easy to note, from the various depth frames F_d shown in Figs. 2, 3, that the depth map computed by the TOF sensor is rarely able to determine a good depth estimation of the scene on the regions around the 4 corners of the map. This is mainly due, as said in Section 3, to the amount of noise which increases as the distance of point from the image center increases, and it is maximum around the 4 corners, which are the points laying farthest from the center. This phenomenon is also evident for points belonging to the intensity frames. Hence, as a consequence, regions around the four corners of the image are highly unreliable, their depth and intensity variances being extremely high. In practice, this increases the chances of having false positives around those regions. Hence, we propose to use a binary mask which excludes the graffiti detection over these points, which can be regarded as peripheral regions where

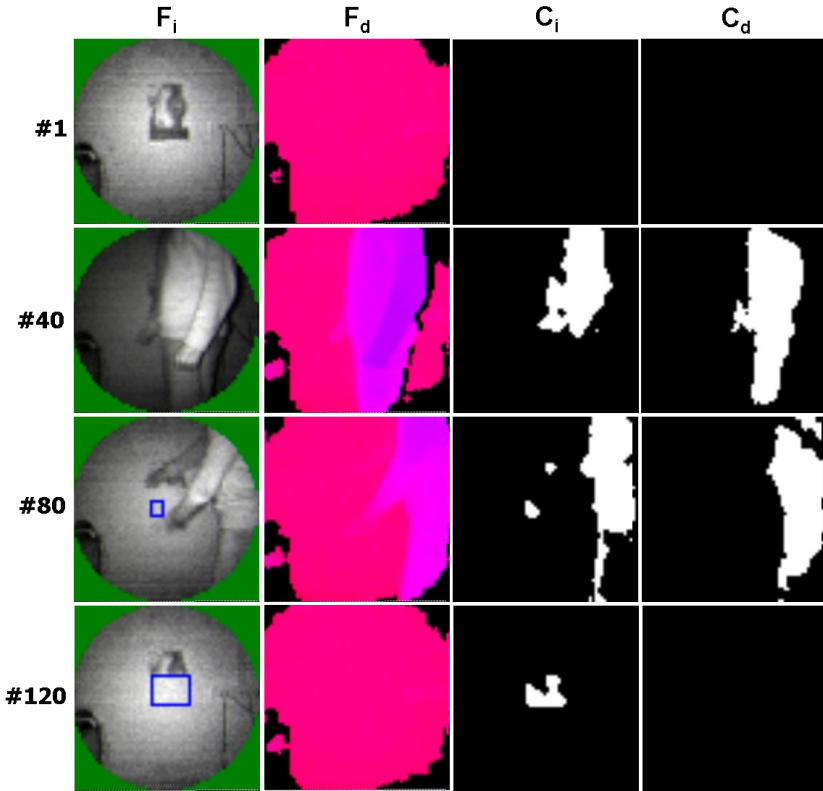


Fig. 4. Video 3: object defacing sequence

detection can not be performed. The output frames of Video 3 (Fig. 4, left column) show in green this mask.

The first frame of the sequence (#1) shows a painting hanging on a white background wall. Then a person enters the room (Frame #40) and starts tearing apart the painting (Frame #80). Correctly, only when defacing is being performed, the algorithm produces an alarm (Frame #80). At the very end of the sequence (Frame #120) defacing is correctly detected, as only the lower half of the painting (the part which has been torn apart) is being highlighted by the bounding box.

5 Conclusions

Overall, our study demonstrated promising capabilities to perform graffiti detection by means of a TOF camera. In particular, our system was able to effectively deploy intensity and depth information coming from a TOF sensor to detect acts of vandalism such as dirtying and painting on walls, object stealing and defacing. In our experiments, though the testing conditions were simplified due to the limits of the available TOF

sensor, our system showed a notable sensitivity, with "graffiti events" being correctly detected and their area being accurately estimated. Moreover, the system yielded a limited number of false positives. It is also worth pointing out that our system is able to operate in real-time.

One major improvement to the system developed so far deals with the capability to handle illumination changes, which were not explicitly taken into account in the experiments. In fact, in case of both slow and sudden illumination changes in the monitored scene, the basic background subtraction approach adopted for the intensity data would fail and illumination changes modifying the appearance of the background surface would easily be detected as graffiti. However, it is well known in literature that background subtraction can be made very robust with respect to illumination changes by deploying similarity/dissimilarity measures that are invariant to some transformation of the intensities. For example, the use of the Zero-mean Normalised Cross-Correlation (ZNCC) to obtain C_i would render the system invariant to illumination changes that can be modelled locally by affine transformations of intensities. As vouched by work concerning the evaluation of visual correspondence measures, e.g. [26], the ZNCC would notably improve the robustness of our system with respect to illumination changes. Other robust background subtraction methods, such as e.g. [27], can be adopted to achieve an ever higher degree of robustness (i.e. invariance to locally order preserving transformations).

Future work is aimed at render the system more robust with regards to illumination changes by means of the aforementioned methodology, and to improve the quality of the intensity images by integrating this information with that coming from a higher resolution color camera. Finally, we also look forward to test the performance of our system within more challenging environments by using a TOF sensor with improved resolution and depth range.

References

1. Sacchi, C., Regazzoni, C., Vernazza, G.: A neural network-based image processing system for detection of vandal acts in unmanned railway environments. In: Proc. IEEE Int. Conf. Image Analysis and Processing (ICIAP 2001), pp. 529–534 (2001)
2. Angiati, D., Gera, G., Piva, S., Regazzoni, C.: A novel method for graffiti detection using change detection algorithm. In: Proc. IEEE Int. Conf. Advanced Video and Signal-Based Surveillance (AVSS 2005), pp. 242–246 (2005)
3. Canesta inc., www.canesta.com
4. Pmd technologies, www.pmdtec.com
5. Mesa imaging, www.mesa-imaging.ch/index.php
6. 3dv systems, www.3dvsystems.com
7. Lottner, O., Sluiter, A., Hartmann, K., Weihs, W.: Movement artefacts in range images of time-of-flight cameras. In: International Symposium on Signals, Circuits and Systems (ISSCS 2007), vol. 1, pp. 117–120 (2007)
8. Göktürk, S., Yalcin, H., Bamji, C.: A time-of-flight depth sensor - system description, issues and solutions. In: Proc. Conf. Computer Vision and Pattern Recognition Workshop (CVPRW 2004), p. 35 (2004)
9. Falie, D., Buzuloiu, V.: Noise characteristics of 3d time-of-flight cameras. In: International Symposium on Signals, Circuits and Systems (ISSCS 2007) (2007)

10. Beder, C., Bartczak, B., Koch, R.: A comparison of pmd-cameras and stereo-vision for the task of surface reconstruction using patchlets. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1–8 (2007)
11. Rizzi, A., Voltolini, F., Remondino, F., Girardi, S., Gonzo, L.: Optical measurement techniques for the digital preservation, documentation and analysis of cultural heritage. In: VIII Conference on Optical 3D Measurement Techniques, vol. 2, pp. 16–24 (2007)
12. Hsu, S., Acharya, S., Rafii, A., New, R.: Performance of a time-of-flight range camera for intelligent vehicle safety applications. In: Advanced Microsystems for Automotive Applications 2006, pp. 205–219 (2006)
13. Grest, D., Kruger, V., Koch, R.: Single view motion tracking by depth and silhouette information. In: Ersbøll, B.K., Pedersen, K.S. (eds.) SCIA 2007. LNCS, vol. 4522, pp. 719–729. Springer, Heidelberg (2007)
14. Haker, M., Böhme, M., Martinetz, T., Barth, E.: Geometric invariants for facial feature tracking with 3D TOF cameras. In: Proc. IEEE International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, vol. 1, pp. 109–112 (2007)
15. Göktürk, S., Rafii, A.: An occupant classification system eigen shapes or knowledge-based features. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops, p. 57 (2005)
16. Holte, M., Moeslund, T., Fihl, P.: View invariant gesture recognition using the csem swiss-ranger sr-2 camera (2007)
17. Craighead, J., Day, B., Murphy, R.: Evaluation of canesta's range sensor technology for urban search and rescue and robot navigation. In: IEEE Workshop on Safety Security Rescue Robots, Gaithersburg, MD (2006)
18. Hong, T., Bostelman, R., Madhavan, R.: Obstacle detection using a tof range camera for indoor agv navigation. In: Performance Metrics for Intelligent Systems (PerMIS 2004), Gaithersburg, MD (2004)
19. Prasad, T., Hartmann, K., Wolfgang, W., Ghobadi, S., Sluiter, A.: First steps in enhancing 3d vision technique using 2d/3d sensors. In: Chum, O., Franc, V. (eds.) 11th Computer Vision Winter Workshop 2006, Czech Society for Cybernetics and Informatics, pp. 82–86 (2006)
20. Lindner, M., Kolb, A.: Data-fusion of pmd-based distance-information and high-resolution rgb-images. In: International Symposium on Signals, Circuits and Systems (ISSCS), vol. 1, pp. 121–124 (2007)
21. Hahne, U., Alexa, M.: Combining time-of-flight depth and stereo images without accurate extrinsic calibration. In: Dynamic 3D Imaging Workshop in Conjunction with DAGM (2007)
22. Knoop, S., Vacek, S., Dillmann, R.: Sensor fusion for 3d human body tracking with an articulated 3d body model. In: Proc. IEEE Int. Conf. on Robotics and Automation, Orlando, Florida (2006)
23. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1–8 (2007)
24. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. Int. Conf. Computer Vision (ICCV 1998), pp. 839–846 (1998)
25. Radke, R., Andra, S., Al-kofahi, O., Roysam, B.: Image change detection algorithms: a systematic survey. *IEEE Trans. Image Processing* 14, 294–307 (2005)
26. Tombari, F., Di Stefano, L., Mattoccia, S., Galanti, A.: Performance evaluation of robust matching measures. In: Proc. 3rd International Conference on Computer Vision Theory and Applications (VISAPP 2008) (2008)
27. Mittal, A., Ramesh, V.: An intensity-augmented ordinal measure for visual correspondence. In: Proc. Conf. Computer Vision and Pattern Recognition, pp. 849–856 (2006)