PEDESTRIAN DETECTION BY RANGE IMAGING

Heinz Hügli, Thierry Zamofing
Institute of Microtechnology, University of Neuchâtel, Rue Breguet 2,
CH-2000 Neuchâtel, Switzerland
heinz.hugli@unine.ch, thierry.zamofing@csem.ch

Keywords: Change detection, 3D camera, range imaging, stereo, time-of-flight, pedestrian detection, shadow effects, changing illumination

Abstract: Remote detection by camera offers a versatile means for recording people activities. Relying principally on changes in video images, the method tends to fail in presence of shadows and illumination changes. This paper explores a possible remedy to these problems by using range cameras instead of conventional video cameras. As range is an intrinsic measure of object geometry, it is basically not affected by illumination. The study described in this paper considers range detection by two state-of-the art cameras, namely a stereo and a time-of-flight camera. Performed investigations consider typical situations of pedestrian detection. The presented results are analyzed and compared in performance with conventional results. The study shows the effective potential of range camera to get rid of light change problems like shadow effects but also presents some current limitations of range cameras.

1 INTRODUCTION

Pedestrian detection plays a central role in many applications. An overview of different pedestrian detection sensors such as passive infrared, ultrasonic, microwave radar, video imaging and piezometric is presented in reference [1][2]. This paper concentrates on pedestrian detection by a fixed camera. Various systems based on monocular vision to detect and track pedestrians are extensively described in reference[2]. Basically the detection process tries to model the background and to detect the presence of persons or objects from the difference between the modeled background and the current scene. A major difficulty of background modeling with 2-D cameras arises in presence of changing illumination and shadows. Therefore shadow suppression algorithms have been designed to deal with this problem [3][4][5]. Other interesting and robust background modeling algorithms use kernel-density model [6], hidden markov models, adaptive color mixture models, weighted match filtering or a Cauchy statistical model [7].

As alternative to above efforts, this study evaluates new detection systems based on range image measurements, analyses their efficiency and compares them with video systems operating in difficult conditions. The usage of range (3D) cameras instead of conventional video (2D) cameras is expected to improve the robustness of detection and to make the system insensitive to illumination and shadow perturbations. Two range camera systems are considered in this paper: stereo cameras and time-of-flight cameras [8].

Next section presents a change detection procedure suited for range. Then, two range imaging technologies are presented and compared: stereo and time-of-flight (TOF) imaging. Finally, a section is devoted to the application of these range cameras for pedestrian detection.

2 PRESENCE DETECTION BY VIDEO AND RANGE

Persons or objects are detected where changes with respect to a background model occur.

2.1 Change detection from video

Figure 1 presents the basic processing scheme for detecting the presence of persons or objects. Central to it is change detection, which consists mainly is detecting differences between the current image I and the background B, which is a representation of the static scene. Foreground modeling segments and labels the change image based on a priori available knowledge in order to provide a best estimate of the objects or persons in presence.

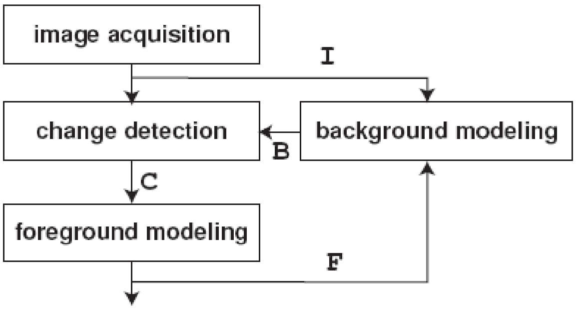


Figure 1: Detecting persons from change.

Background modeling schemes are numerous. Let us mention, in order of increasing complexity, fixed or adaptive models, scalar, Gaussian, mixture of Gaussian models and other advanced models [6]. As all models can be applied to video as well as to range, we limit this paper to the presentation of the simpler adaptive scalar background model and rather stress differences in video and range processing.

In this simple context, the adaptation of the background B_{t-1} is performed according to:

$$B_t = \begin{cases} B_{t-1} & if\ F_{t-1} \\ (1-\alpha)B_{t-1} + \alpha\, I_t & else \end{cases} \quad (1)$$

i.e. only pixels not belonging to the foreground F_{t-1} (line 1) are processed by recursively substituting in them a small part ($0<\alpha<<1$) of the current image I_t (line 2)

Then, image values which differ from the background by more than a given threshold value ΔI constitute the boolean change image C_t

$$C_t = \begin{cases} true & if\ |I_t - B_{t-1}| > \Delta I \\ false & else \end{cases} \quad (2)$$

Finally, in the simplest way, the foreground is set equivalent to the change image

$F_t = C_t$

while more generally, foreground modeling performs an interpretation of the change image C_t in order to provide a best possible estimate of the foreground F_t.

2.2 Change detection from range

A specific aspect of range images lies in their domain of definition:

$$Z_t \in \{[0 ... z_{\max}], nil\} \quad (3)$$

i.e. they take values in a range of positive z values and can possibly take the value *nil* that encodes all situations where the range camera delivers undefined values. Such undefined range values appear for instance in stereo cameras in absence of texture, and in TOF cameras when the modulated reflected signal is weak.

In this context, classical background modeling must be adapted to the presence of *nil* values. In addition, it can take into account that, unlike intensity in video, presence in the range domain always decreases the Z value with respect to the background. A suited means for the updating of the range background is:

$$B_t = \begin{cases} B_{t-1} & if (Z_t = nil)\ or\ F_{t-1} \\ Z_t & if (B_{t-1} = nil) \\ (1-\alpha)B_{t-1} + \alpha\, Z_t & else \end{cases} \quad (4)$$

where the first line says there is no update in presence of a foreground pixel or with a nil Z value; the second line says that the background starts with the first non-nil Z value, and the last line expresses the standard recursive update.

Regarding the change detection, it must also consider the nature of possibly undefined signals. In the following definition of the change image:

$$C_t = \begin{cases} false\ if\,(Z_t = nil)\ or\,(B_{t-1} = nil) \\ false \quad else\,if\,(B_t - Z_{t-1}) < \Delta Z \\ true \qquad\qquad else \end{cases} \tag{5}$$

only pixels with valid Z and B values are considered (line 1) and a change is not detected (line 2) unless the decrease in range surpasses a threshold ΔZ (line 3).

Note that the presence of *nil* values in range images can be partially compensated by so-called hole filling algorithms, and multi-scale methods are well suited to do so [9]. This possibility can be introduced at several places in above procedure, but is not discussed further here.

3 RANGE CAMERAS

Two range imaging technologies are considered in this study: stereo camera and time-of-flight (TOF) cameras

3.1 TOF cameras

TOF cameras measure the time needed for light to travels from the camera to the object and back again. Typically, the phase shift between sent and received modulated signal is measured and converted into a range value.

3.2 Stereo cameras

Stereo cameras record sequences of image pairs. The images of a pair are recorded at the same time and represent images of the scene viewed from two neighboring location. Stereo interpretation consists in computing the disparity of corresponding pixels in an image pair, and the Z range is then simply derived. Disparity computation is quite tedious. It is usually not performed directly in the camera but requires a powerful computer to reach real-time performance.

Some basic differences of the two technologies considered are compared in table 1 below.

Table 1: Comparison of TOF and stereo ranging.

	TOF	Stereo
range calculation method	phase shift of sent and received light	disparity computation of stereo pairs
range resolution over Z range	constant over Z	decreases with increasing Z
range accuracy	decreases with increasing Z	depends on surface texture
sensitive to ambient light	yes	no
need of own light source	yes	no
sensitive to bad surface structure	no	yes
additional processing needed	no	yes

4 PEDESTRIAN DETECTION EXPERIMENTS

Practical pedestrian detection experiments are performed in order to evaluate the performance of range detection per se, but also in comparison to video detection.

The TOF camera is the SwissRanger [10] SR-02 which delivers 16 bit range images (160x128 pixels) at a rate of 30 Hz or less, together with an intensity image of same size. Range is derived locally by the camera, from the measured phase shift between sent and received modulated light. The maximum range is limited, and set to 7.5 m in the device used.

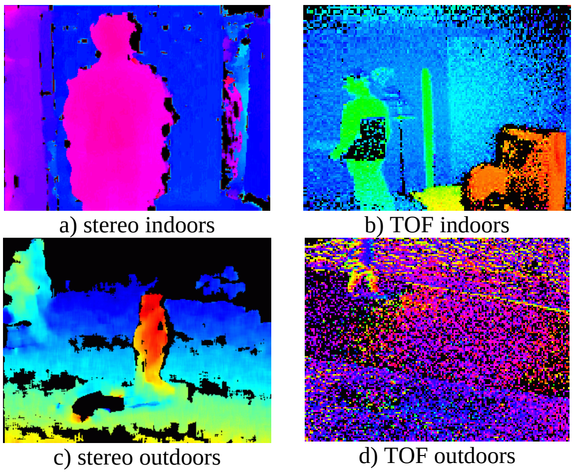
The stereo camera is the Bumblebee [11]. It delivers pairs of images (1024x768) from two cameras located on a 12 cm long baseline, at a rate of about 7 Hz. Disparity computation is performed on a fast PC. Because of the processing complexity, there is a tradeoff between high resolution and high speed. A typical range images size is 320x240.

Three different situations are considered successively.

Indoor versus outdoor site: Figure 2 provides a comparison of range imaging by stereo and TOF in two different sites, namely an indoor and outdoor site. Indoors, pedestrians walk along a corridor. The range of interest is 1 to 3 m (fig. 2a and b). Both stereo and TOF work fine.

Outdoors, the pedestrians walk along a pathway and the distance ranges from 4 to 8 m (fig. 2c and d). Here, only stereo works fine, because TOF is strongly affected by sunlight illumination that surpasses by far the camera own illumination. On the other hand, because operated with IR light, TOF operates also invisibly during the night, both indoors or outdoors.

Therefore, both stereo and TOP ranging systems are suited for pedestrian detection, each method has specific advantages. Among main advantages of stereo for pedestrian detection is the capability to work indoor as well as outdoor, the availability of a registered high-resolution video image. Among main advantages of TOF cameras are the locally embedded range processing, the capacity to work at night and good object independence regarding texture.



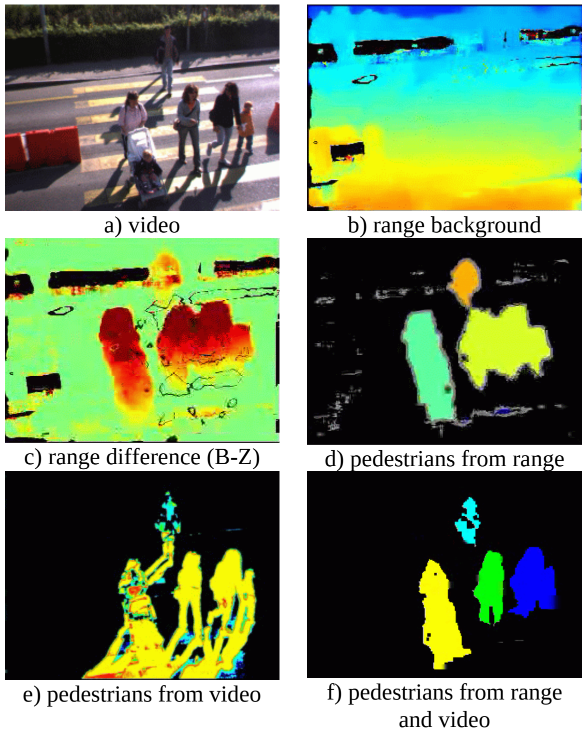
a) stereo indoors | b) TOF indoors | c) stereo outdoors | d) TOF outdoors

Figure 2: Range from stereo and TOF.

Road crossing site: This outdoor road crossing is about 8 m long and the Z range of interest reaches up to 12 m. TOF cannot be used outdoor and stereo, given the fixed baseline, is at its practical resolution limit at about 10 m. Stereo images are recorded in situ and processed off-line.

Of major interest is the pedestrian detection illustrated in figure 3, where the scene is strongly affected by the pedestrian shadows. Video detection (fig. 3e) labels shadows as pedestrians which then, cannot be correctly segmented. In contrast, range detection (fig. 3d) is not affected by shadows and provides a correct segmentation.

Note that a combination of video detection and range detection provides the best result (fig. 3f)

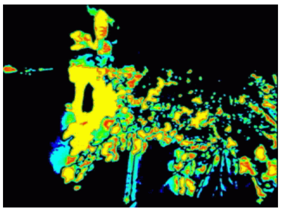


a) video | b) range background | c) range difference (B-Z) | d) pedestrians from range | e) pedestrians from video | f) pedestrians from range and video

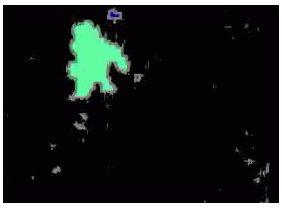
Figure 3: Results from the road crossing site.

Pathway site: A pathway for pedestrians is affected by strong illumination changes. In one situation, fast traveling clouds in the sky produce fast illumination changes. In another situation, shadows from moving trees produce even stronger and faster illumination changes. While video detection (fig. 4a) is completely unable to distinguish even the presence of groups of pedestrians, stereo range detection (fig. 4b) performs correctly and detects the pedestrians walking along the pathway.

These results confirm the capacity of range detection to perform well in presence of illumination changes and show therefore its robustness for people detection. Given other weaknesses of range imaging compared to video, like a poorer resolution, it is suggested that optimal performance will result from a suitable combination of both methods.



a) pedestrian from video



b) pedestrian from range

Figure 4: Results from the pathway site.

5 CONCLUSIONS

The paper considers range cameras for presence detection, specifically for pedestrian detection where conventional video detection systems perform poorly due their sensitivity to shadows and illumination changes. A first part was devoted to the presentation of a change detection scheme that suits the specificities of range detection, specifically by considering the presence of undefined range values and the property of range measurements to always decrease in presence of objects or persons.

A second part was devoted to two ranging systems, namely stereo and time-of-flight (TOF). Among main advantages of stereo for pedestrian detection is the capability to work indoor as well as outdoor, the availability of a registered high-resolution video image. Among main advantages of TOF cameras are the locally embedded range processing, the capacity to work at night and good object independence regarding texture.

A final part was devoted to practical pedestrian detection experiments, in particular in difficult situations. For indoor pedestrian detection, both stereo and TOF are suited, the later with the advantage to be operated also by night. For outdoor pedestrian detection, TOF is not (yet) suited and only stereo can be used. The capability of range detection to get rid of shadow and illumination changes affecting strongly the video detection was demonstrated on two sites. On the road crossing site, range detection is not affected by the strong pedestrian shadows cast on the road. On the pathway site, where cast shadows from moving trees make video detection completely hopeless, range detection performs correctly.

Finally, using together video and range for presence detection performs optimally, as it combines the advantages of both worlds, essentially good resolution for the first and good robustness for the second.

ACKNOWLEDGEMENTS

This project was partially supported by the Swiss Federal KTI/CTI Innovation Promotion Agency. Fruitful collaboration with MIS Institute and ACET is kindly acknowledged.

REFERENCES

[1]D. M. Beckwith and K. M. Hunter-Zaworski. Passive pedestrian detection at unsignalized crossings. Transportation research records, Paper No. 98-0725, 1996. URL http://www.enhancements.org/trb\1636-016.pdf

[2] I. Haritaoglu, D. Harwood, and L. Davis. Who, when, where, what: A real time system for detecting and tracking people. Proceedings of the Third Face and Gesture Recognition Conference, pages 222--227, 1998.

[3] G. Finlayson, S. Hordley, and M. Drew. Removing shadows from images. 2002. URL http://citeseer.ist.psu.edu/finlayson02removing.html.

[4] H. Jiang and M. S. Drew. Shadow-resistant tracking in video. 2003
URL http://citeseer.ist.psu.edu/jiang03shadowresistant.html.

[5] L. Jianguang & al. An Illumination Invariant Change Detection Algorithm", Proc. ACCV, 2002

[6] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis. Background and foreground modeling using nonparametric kernel density for visual surveillance. Proc. of the IEEE, 90(7), pp. 1151-1163, 2002.

[7] J. M. Ying Ming, Jingjue Jiang. Background modeling and subtraction using a local-linear-dependence-based cauchy statistical model, 2003. URL www.cmis.csiro.au/Hugues.Talbot/dicta2003/cdrom/pdf/0469.pdf

[8] P. Seitz. Solid-state time-of-flight range camera. IEEE Journals of Quantum Electronics, 2003

[9] T. Zamofing, H. Hügli. Range Image Filtering Using Reliability Information. Proc. SPIE Vol. 5606-16, 2004

[10] SwissRanger. URL http://www.swissranger.ch/

[11] P. G. Research. Bumblebee 3D camera. URL http://www.ptgrey.com/products/bumblebee/.