Detection of Moving Objects with Removal of Cast Shadows and Periodic Changes Using Stereo Vision

A. Moro
University of Trieste / CREST, JST
alessandro.moro@stud.units.it

K.Terabayashi
Chuo University / CREST, JST
terabayashi@mech.chuo-u.ac.jp

K. Umeda
Chuo University / CREST,JST
umeda@mech.chuo-u.ac.jp

Abstract

In this paper we present a method for the detection of moving objects for unknown and generic environments under cast shadow and *periodic movements of non relevant objects (like waving leaves), using a combination of non-parametric thresholding algorithms and local cast shadow analysis with stereo camera information. Good detection rates were achieved in several environments under different lighting conditions, and objects could be detected independently of scene illumination, shadow, and periodic changes.*

1. Introduction

The detection of moving objects and people from a streaming video is a fundamental problem in many vision systems including robotics applications, surveillance and objects and humans detection. In fact, much research has been devoted to human motion detection mainly due to the increasing number of potential applications [7]. However, changes in the illumination conditions, changes in the environment or shadow changes can limit the detector performance.

We propose a cascade of algorithms in order to improve the detection of moving objects. Main contributions have been given in the detection of periodic movements and shadow detection. The whole method demonstrated robustness and flexibility, achieving high detection accuracy. It does not require a training phase, and it is fast and easily extensible.

The remainder of paper is organized as follows: Section 2 is a report of related works. Section 3 contains a description of the proposed algorithm. Section 4 shows the results obtained analyzing different environments. Section 5 is the conclusion, a report of future work, and possible applications.

Fig. 1. Detection of changes. In blue, detected changes; in green, shadow detected; in yellow, previous memorized layers

2. Related Work

In order to detect relevant elements on the scene, several algorithms have been proposed. In [9] a review of change detection algorithms in video sequence is reported. In [1, 4] approached based on an *a-priori* knowledge of the objects are described. In the latter case, the presence of unknown objects can lead to detection failures. With change detection algorithm, Su et al. [10] proposed a non-parametric method for computing an adaptive threshold, however, is very sensitive to high variation of the scene contrast.

Periodic changes, such as those due to trees moving in the wind, can generate false detection. Mixture of Gaussian (MoG) model or Codebook [3] have been proposed to solve this problem in complex environments. Codebook performs faster and requires a smaller amount of memory than MoG; however for specific cases, it is less accurate. Both require a training phase. Also shadow can cause misdetection. Martel-Brisson *et al.* [6] recently proposed a property-based method which consider the reflectance of the material and propose an estimation of the shadow orientation. It increases the detection of shadow on objects of similar chromacity but is strictly related to a training sequence. Yang *et al.* [12] described a simple and expandable shadow detection method based on the combination of chromatic, texture and temporal information. But on the other side, Yang *et al.* do not considered particular problems as self cast shadow.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.89

3. Proposed Algorithm

The scheme of the proposed method is shown in Fig. 2. It is composed by a threshold estimator to detect changes, a stereo subtraction module [11], a periodical changes discard module, and a cast shadow detector.

3.1. Change Detection

We developed a method inspired by Su *et al.* [10] and described in details in [8] for the task of moving objects extraction. Moving objects are computed by dividing the input image in blocks and a threshold is estimated taking in consideration the brightness of each block. We refined the threshold by taking into account both the threshold average of neighbors' blocks and, where possible, tracking blocks.

The refine method can be resumed as follow:

$$\mathrm{Thr}_{(t+1,x,y)} = \frac{f(x,y,w,|A|)+\psi_{t,x,y}\cdot \mathrm{Thr}^{k}_{t,x,y}}{1+\psi_{t,x,y}} \qquad (1)$$

where Thr^{k} is the threshold calculated into a rectangular area whose size and position depends on the size of a detected object "k" on the scene in instant *t* and ψ assume value 1 if a object is detected, 0 otherwise. *f* a recursive function that smoothes the threshold value of the vicinity. The obtained threshold is adjusted with a salt and pepper detector function to reduce the noise which can be generated in empty areas. Then for each pixel a change is detected if its brightness is greater than the computed threshold.

3.2. Discarding Periodic Changes

Light changes or movement of leaves are generally unwished. To adjust the results obtained in 3.1, we have developed a method to discard periodic changes inspired by the concept of codebook described by Kim *et al.* in [3]. Its peculiarity is that it does not need any training phase (Fig. 3).

Periodic changes can be estimated considering the instances in an observation period. If color characteristics are retained continuously, a periodic event has probably occurred. On the basis of the structure of the codebook, defined as a collection of codewords, each simplified codeword is defined as:

- $\check{I}, \hat{I}$ the min and max brightness, respectively, of all pixels assigned to this codeword.
- An RGB vector $v = (\bar{R}, \bar{G}, \bar{B})$.
- f frequency with which the codeword has occurred.

The process of discarding periodic changes is composed of two phases: detection and updating. In detection, periodic phenomena are searched, and, in updating, the codebook is updated based on an extension of the rules in [3]. Since the computation time is not constant, the period of observation is normalized with the frames per second. Let M_t be the maximum observation time and τ the percentage of time when an event is expected; then, the maximum number of observation frames ϑ will be defined as:

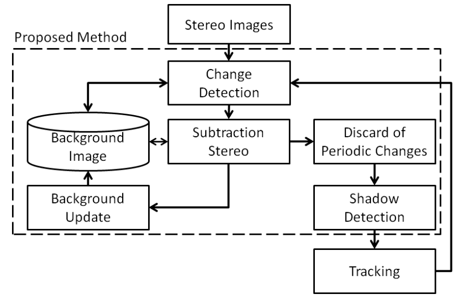

Fig. 2. The model of the proposed algorithm.

$$\vartheta = \frac{M_t \cdot FPS}{\tau} \qquad (2)$$

A pixel is detected if chromatic and intensity values are not registered in codebook (3) defined as follow:

$$\lambda'_{x,y} = \begin{cases} 0, \text{if } p_{x,y} \cong coldist(p_{x,y}, c_i) \le \sigma \wedge brightness(p_{x,y}, c_i) \\ 1, \text{otherwise.} \end{cases} \qquad (3)$$

where *coldist* is a function to estimate the chromatic difference and *brightness* the difference in intensity as shown in [3].

In the updating phase, the codebook is modified by considering the color properties and times of changes during the observation period. If the changes instances registered during the observation time are less than expected period of study, the codebook is initialized, otherwise register the new changes as follow:

$$C_{x,y} = \begin{cases} (C_{x,y} - c_i) \cup u(p_{x,y}, c_i), \text{if } \exists c_i \in C_{x,y}: p_{x,y} \cong c_i \\ C_{x,y} \cup a(p_{x,y}), \text{if } \nexists c_i \in C_{x,y}: p_{x,y} \cong c_i \\ \emptyset, \text{if } \sum_{i=i_0}^{i_0+\vartheta} \lambda_{i,x,y} \le \tau \end{cases} \qquad (4)$$

where $a(p_{x,y}, c_i)$ create a new codeword, $c = (\bar{R}, \bar{G}, \bar{B})$ and $\check{I}, \hat{I}$ are equal to pixel intensity and f is equal to 1, and $u(p_{x,y}, c_i)$ update the codebook $c_i = \left(\frac{f_i\bar{R}_i+R}{f_i+1}, \frac{f_i\bar{G}_i+G}{f_i+1}, \frac{f_i\bar{B}_i+B}{f_i+1}\right)$, reassign the intensity min and max value and f is increase by 1.

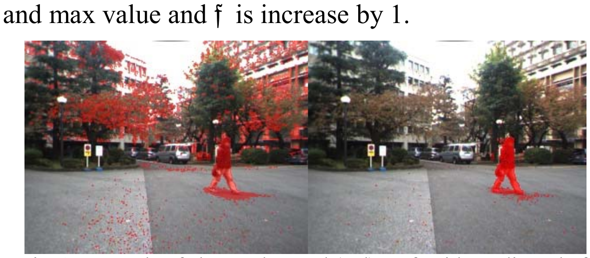

Fig. 3 Example of change detected (red). Left without discard of periodic changes, right discarding them.

3.3. Shadow Detection

Shadow detection is important for ghost regions and blob formation prevention. Based on Yang *et al.* [12] method we improved [8] and extended their method proposing a concept that combines stereo information with visual information in order to detect the shadow. The proposed algorithm differs with [5] where the stereo information is used to select the regions candidates to be shadow and subsequently analyzed with a chromatic algorithm.

Passive stereo vision in real-time application is generally inaccurate; however, as demonstrated in the literature and in [11], real-time stereo information can be used in a short range. If we consider the depth map obtained by a stereo vision system, a pixel will be a change if the variation of distance from background and foreground will be greater than 0, in ideal condition, or a threshold in a real environment.

Let "θ" be a shadow parameter that assumes a low value if a pixel is recognized like as shadow. Equation (5) combines the relationship between pixel of the foreground and background (Ψ), and within pixel (Λ) of background and foreground, and stereo distance (Φ) Eq. (6). α, β, γ and η are constant values that define the weight of textures, colors, distance and intensity I of a pixel of the last frame or intensity I' of background.

$$\theta_{(t+1,x,y)} = \begin{cases} \alpha\Psi_{x,y} + \beta\Lambda_{x,y} + \frac{\gamma}{d^2}\Phi_{x,y} + \left(1 - \alpha - \beta - \frac{\gamma}{d^2}\right)\theta_{t,x,y}, \text{ if } \frac{I_{x,y}}{\eta} < I'_{x,y} \\ \infty, \text{otherwise.} \end{cases} \quad (5)$$

where d is the distance calculated by the stereo system.

Stereo information is degraded with increasing the distance but, in the short range, gives high contribution to solve ambiguities, in particular when the background and foreground colors and textures are similar. It is considered that a small difference between background and foreground distances probably indicates that a pixel is a shadow. Stereo distance can be defined as

$$\Phi(x,y) = \begin{cases} \min\left(1, \frac{\left|d(x,y) - d'(x,y)\right|}{D}\right), \text{ if } \exists_{d,d'} \\ 0, \text{otherwise.} \end{cases} \quad (6)$$

where D is the maximal measurable distance and d' the background distance.

3.4. Background Update

In a long sequence video, elements on the scene can be removed, introduced or moved, and the light condition can change.

It is than reasonable to wish a good and stable model of background. We propose a method to update the background taking in consideration the number of changes in the observation period and the distance information obtained by the stereo vision system.

For each pixel, time information are computed as

$$T_{x,y} = \begin{cases} T, \text{ if } \prod_{i=i_0}^{i_0+\vartheta}(1 - \lambda_{i,x,y}) \cdot \sum_{i=i_0}^{i_0+\vartheta}\lambda_{i,x,y} > \vartheta \\ T_{x,y}, \text{otherwhise.} \end{cases} \quad (7)$$

$$T_{\vartheta,x,y} = \begin{cases} \left(1 - \frac{\sum_{i=i_0}^{i_0+\vartheta}\lambda_{i,x,y}}{\vartheta}\right) \cdot M_t, \text{ if } Z \geq \delta \\ \alpha, \text{otherwhise.} \end{cases} \quad (8)$$

Where T is the current elaboration time, λ assumes 1 if a change is detected (3.1), and ϑ is defined in (2). The background for each pixel will be defined as

$$B_{x,y} = \begin{cases} F_{x,y}, \text{ if } T - T_{x,y} \geq T_{\vartheta,x,y} \\ B_{x,y}, \text{otherwhise.} \end{cases} \quad (9)$$

3.5. Multilayer Background

The updating phase entails the risk of merging interesting objects with the background. We propose a multilayer background to support this detection (Fig 4).

Let $\mathcal{L}$ be the set of all the possible background layers, and the initial layer of the background is defined by $\mathcal{L}_0 \in \mathcal{L}$. *dist* is defined as the distance of the layer obtained by the stereo procedure. The pseudo code of the procedure is given below:

I. Create the $\mathcal{L}_0$ background.
II. For each incoming pixel p check if update the background.
IIIa. If $p \in \mathcal{L}$ and $\text{dist}_p \cong \text{dist}_{\mathcal{L}_{i-1}}$ restore $\mathcal{L}_p$ and delete the next layers.
IIIb. If $p \in \mathcal{L}$ but is not possible to calculate the distance, restore the $\mathcal{L}_p$.
IV. If the background is updated, register the information of color (equation 11) and add a layer $\mathcal{L}_i$.
IVa. If distance exist insert $\mathcal{L}_i$ at the correct position:
$$\text{dist}_{\mathcal{L}_{i-1}} \leq \text{dist}_{\mathcal{L}_i} \leq \text{dist}_{\mathcal{L}_{i+1}}$$
IVb. Otherwise than insert the layer at the level $\mathcal{L}_{i+1}$
V. Repeat the process from the Step II.

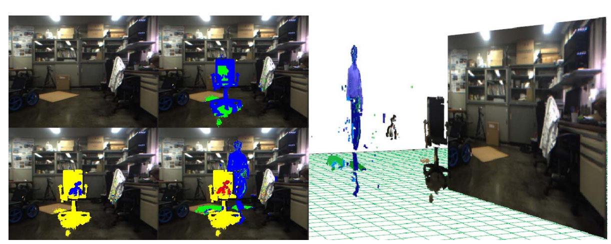

Fig. 4. Multilayer detection in indoor environment. In blue detected changes, green shadow, yellow and red memorized layers. On right image a 3D representation of the different layers.

4. Results

For the experiments, we acquired images with a Point Gray Research Bumblebee camera at 640 x 480 pixel resolutions and an Intel Core2 Quad CPU. In order to evaluate the real-time thresholding, and the detection rate and false alarm rate of shadows, it has been created a collection of ground truth labeled manually first. Frames in the videos are selected in approximately equal time interval, discarding the situation of no object showing up or just partial appeared. The performances of the algorithm are evaluated considering the changes detected in the images (detection rate, DR) and the total of changes corresponding to a real object or person (accuracy rate, AR). A low detection rate can be compensating by high accuracy rate. Four indoor and outdoor environments were considered at different day time.

For completeness, a PETS'2009 (Performance Evaluation of Tracking and Surveillance) sequence was used, which is challenging in term of multiple targets and change in lighting. The stereo information was unavailable, and the performance of the proposed method may have been unsatisfactory. We reached 84% of the DR with 31% of AR (Fig 1).

5. Conclusion

This work presents a stereo real-time object detection algorithm with shadow detection and removal of periodic changes suited for several contests. Main contributions are the use of visual stereo information to separate shadow from objects and a variant of codebook to remove periodic changes. In the future works we will extend the described concept to combine stereo and image information for segmentation purpose.

TABLE I - CHANGE DETECTION RATE

	Statistic [12]	Codebook [3]	Base Meth [10]	Proposed Method
DR	81.6%	91.9%	93.6%	91.1%
AR	9.1%.	19.1%	11.4%	30.8%
T (ns)	20	27	39	574

Detection Rate (DR), Accuracy Rate (AC) and time per pixel (T).

TABLE II
SHADOW DETECTION RATE

	Multiple Cues [12]	Unsupervised [2]	Proposed Method
DR	91.2%	90.3%	96.2%
FAR	5.7%	22%	4.9%
Time Pixel(ns)	39[19]	8100(our estimation)	41

Detection Rate (DR) and False Alarm Rate (FAR) of recent methods

6. References

[1] A.Bosch, et al., "Image classification using ROIs and multiple kernel learning", *IJCV*, 2008.

[2] C. Huang et al., "An unsupervised shadow detection algorithm", *ICS2008*, pp.190-195, 2008.

[3] K. Kim et al., "Real-time foreground background segmentation using codebook model", *Real-Time Imaging*, pp. 172-185, 2005.

[4] D.Lowe, "Distinctive image features from scale-invariant keypoints", *IJCV*, vol. 60, pp. 91-110, 2004.

[5] C.B. Madsen et al, "Shadow Detection in Dynamic Scenes Using Dense Stereo Information and an Outdoor Illumination Model", Dyn3D, 2009.

[6] N. Martel-Brisson and A. Zaccarin, "Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation", CVPR , pp. 1-8, 2008.

[7] T.B. Moeslund, et al., "A survey of advances in vision-based human motion capture and analysis", *CVIU*, vol. 104, Issues 2-3, pp.90-126, 2006.

[8] A.Moro et al., "Auto-adaptive threshold and shadow detection approaches for pedestrians detection", *AWSVCI*, pp.9-12, 2009.

[9] R.J.Radke, et al., "Image change detection algorithms: a systematic survey", *IEEE Trans. On Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.

[10] C. Su et al., "A Real Time Adaptive Thresholding for Video Change Detection, "*Proc. IEEE Int. Conf. in Image Processing 2006* ,pp. 157-160, 2006.

[11] K. Umeda, et al., "Subtraction Stereo. A stereo camera system that focuses on moving regions", *Proc. Of SPIE-IS&T Electronic Imaging*, VOL. 7239 Three Dimensional Imaging Metrology, 723908, 2009.

[12] M.-T. Yang et al., "Moving cast shadow detection by exploiting multiple cues", Image Processing, IET Vol 2, pp. 95-104, 2008.

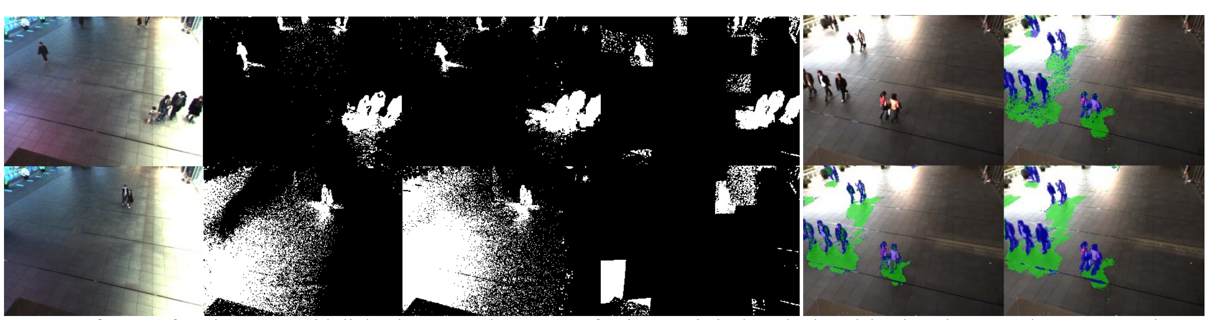

Fig. 5. Two frames of environment with light changes and presence of noise, Statistical method, codebook and proposed are compared.