MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
8-19

Pedestrian Detection and Identification using Two Cameras

Naoya Ohta
Department of Computer Science
Gunma University
1-5-1 Tenjin-cho, Kiryu 376-8515
Gunma Japan
ohta@cs.gunma-u.ac.jp

Anthony R. Dick
School of Computer Science
University of Adelaide
Adelaide, SA 5005
Australia
ard@cs.adelaide.edu.au

Abstract

This paper describes a method for pedestrian detection, identification and tracking using image information. The method makes use of two cameras with a shared field of view and is robust to changes in illumination and shadows. After a brief calibration process, in which the scene is divided coarsely into planar pieces (which are later optimised), the process requires no interaction and automatically compensates for pairs of cameras with very different optical properties. Individual pedestrians are identified by the novel application of a process similar to the Wiener filter, which we call the regularised inverse filter. *Experiments on outdoor scenes demonstrate that the method is robust to changes of illumination and shadows, successfully tracking over 9 out of every 10 pedestrians in challenging conditions.*

1 Introduction

This paper describes an algorithm for detecting, identifying and tracking pedestrians as they move through the shared field of view of a pair of cameras. The algorithm avoids some of the problems of other detection algorithms: it ignores shadows and is robust to severe changes in lighting due to its use of two cameras.

The use of a pair of cameras to eliminate shadows was previously described in [1]. In this paper we improve the robustness of the method by allowing the cameras to have unknown and different internal settings (e.g. different apertures) so long as they have the same gamma value. We also extend the method to work on piecewise planar backgrounds, rather than requiring that the background be a single plane.

The extraction of individual pedestrians from detected foreground pixels is achieved by the application of a function similar to the Wiener filter which we call a *regularised inverse filter*. This is based on a model of what a pedestrian looks like rather than simply trying to cluster or segment pixels morphologically. This technique can therefore provide more accurate and robust results, especially when pedestrians overlap in the cameras' field of view.

2 Camera calibration

Before operation this algorithm requires some basic calibration information from the user. This involves outlining quadrilateral planar pieces of the scene in each camera view, as shown in Figure 1(b). This gives a set of approximate correspondences between the two views. To overcome inaccuracies in the manual specification of corresponding points, their placement is optimised using Powell's method [2]. The same optimisation procedure accounts for cameras with different sensitivity to light and other optical settings, as follows. It is assumed that the cameras have the same gamma value relating pixel intensity I to incident light energy E, and that I is also affected linearly by each camera's specific settings such as gain and contrast:

$$I_1 = a_1 E_1^\gamma + b_1 \qquad (1)$$

$$I_2 = a_2 E_2^\gamma + b_2 \qquad (2)$$

In the above equations, suffixes 1 and 2 designate images 1 and 2, respectively. $\gamma = 0.45$ is a standard value for many cameras. In this case, assuming a Lambertian surface, because E_1 and E_2 are in a linear relation, corresponding pixel intensities are linearly related:

$$I_1(x, y) = \alpha I_2(\hat{x}, \hat{y}) + \beta \qquad (3)$$

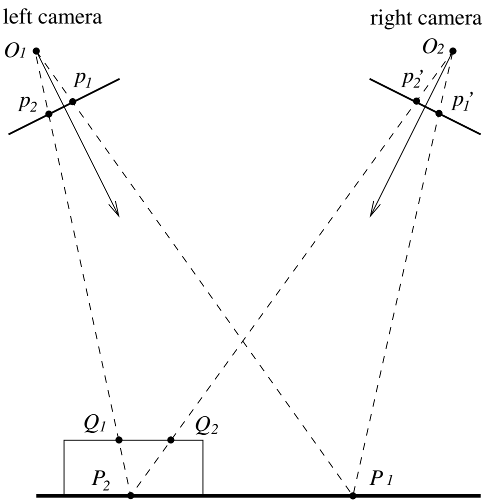
where (x, y) and $(\hat{x}, \hat{y})$ are corresponding image points in images 1 and 2. The scale and offset parameters α and β are estimated simultaneously with the image coordinates of the planar region boundaries, using Powell's method. The quantity to be minimized is:

$$D^2 = \sum (I_1(x, y) - \alpha I_2(\hat{x}, \hat{y}) - \beta)^2 \qquad (4)$$

and the variables to be optimized are α, β and corner positions of the quadrilateral patches in image 2. This optimisation occurs once, before the system begins detecting and tracking pedestrians.

3 Pedestrian detection

The four corners of each quadrilateral specified in the calibration stage define a homography $\mathbf{H}$ which maps any point lying on the plane of the quad from one image to another. Conversely, any point not lying on that plane will not be correctly mapped between the images by $\mathbf{H}$ (see Figure 1(a)). Thus to detect pixels not lying on the background plane we compare each pixel within each quad defined in the left image with the pixel it maps to in the right image. Those pixels whose difference (after being adjusted by the α and β calibration parameters) is above a threshold are considered occupied by an object not in the background plane, and are flagged (Figure 1(d)).



(a)

(b) (c) (d)

Figure 1: Object detection. (a) Visible points P_1 on the background match in both cameras, whereas hidden ones (P_2) do not. (b) User marked plane boundaries. (c) Person walks across detection area. (d) Pixels detected as not lying in the background (white).

This method of detection is more robust than standard techniques such as background subtraction [4, 5] and frame differencing. Because shadows and other changes in lighting affect both cameras equally, our detector is unaffected by them. Unlike the other techniques mentioned, this detector is not based on motion. It can therefore detect objects which have become immobile.

4 Pedestrian identification

To identify an individual pedestrian among the detected foreground (high difference) pixels, we use a template describing what a "typical" pedestrian looks like. This template is obtained by manually highlighting several pedestrians in the difference image, and averaging them. The difference image for each frame is then correlated with this template, using the following filter represented in Fourier (frequency) domain:

$$C(u,v) = \frac{T^*(u,v)}{|T(u,v)|^2 + \eta} F(u,v) \tag{5}$$

where $T(u,v)$, $F(u,v)$ and $C(u,v)$ are Fourier spectrums of the template, the image and the correlation signal, respectively, and a star on the shoulder denotes complex conjugate. Like the Wiener filter [3] this separates out point sources in the source signal (the image). The amount of regularisation applied during filtering is determined by a single parameter η. This parameter can affect the location and number of peaks in the correlation and therefore needs to be chosen empirically for a given camera setup. In the case shown in Figure 2, $\eta = 0.1$ results in too many peaks, $\eta = 1.0$ gives too smooth a correlation surface, but $\eta = 0.5$ gives the correct result.

This technique is not as sophisticated as that of Viola et al. [6], but nor does it require a large volume of training data. It has the advantage over standard morphological methods of clustering and segmentation of taking into account the appearance of a pedestrian. By taking peaks in correlation with a pedestrian template, it is more effective than a generic morphological operator at separating out pedestrian-like shapes.

5 Pedestrian tracking

An estimate of the current position and velocity of each identified pedestrian is updated every frame. Tracking involves a predict-update cycle, with the predicted position at time t calculated using the position and velocity observed at time $t-1$:

$$\hat{x}[t] = x[t-1] + v[t-1] \tag{6}$$

The pedestrian closest to the predicted position at time t is chosen as the corresponding object. The velocity is then

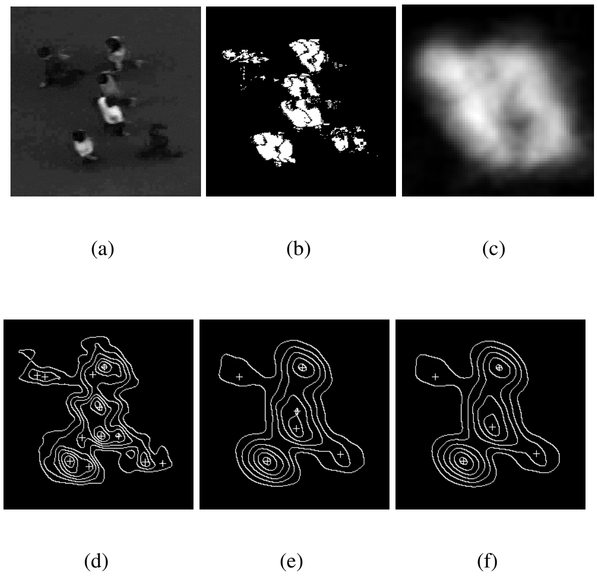


Figure 2: Pedestrian identification. (a) Original image. (b) Difference image. (c) Pedestrian template. (d)-(f) Correlation contours, with $\eta = 0.1, 0.5, 1.0$.

updated as follows:

$$v[t] = cv[t-1] + (1-c)(x[t] - x[t-1]) \qquad (7)$$

where c is a constant, empirically set to 0.2. The radius of the area searched around the predicted position of each pedestrian is determined by their velocity in the previous frame.

6 Results

This system has been used to successfully track pedestrians in challenging conditions. Figure 3 shows a couple of examples, involving strong shadowing and a number of pedestrians walking close to each other. Table 1 shows the proportion of pedestrians that were successfully tracked during our experiments on an outdoor scene. The tracks were gathered during two periods of 30 and 60 minutes, at different times of day, which included significant changes in lighting. Pedestrians would occasionally stop and remain motionless for a period of time, a case which often defeats motion based detection algorithms.

Tracked pedestrians were divided into 3 categories: those who did not come near other pedestrians (A), those who came near to another pedestrian while moving in the same direction (B), and those who came near another pedestrian while moving in the opposite direction (C). The most problematic case is B, in which 2 pedestrians have a similar position and velocity. This is mainly due to our basic tracking algorithm (Section 5), which relies on proximity to a single predicted position at each frame. We anticipate that a more sophisticated tracker, particularly a multi-hypothesis tracker, will cope better with these situations.

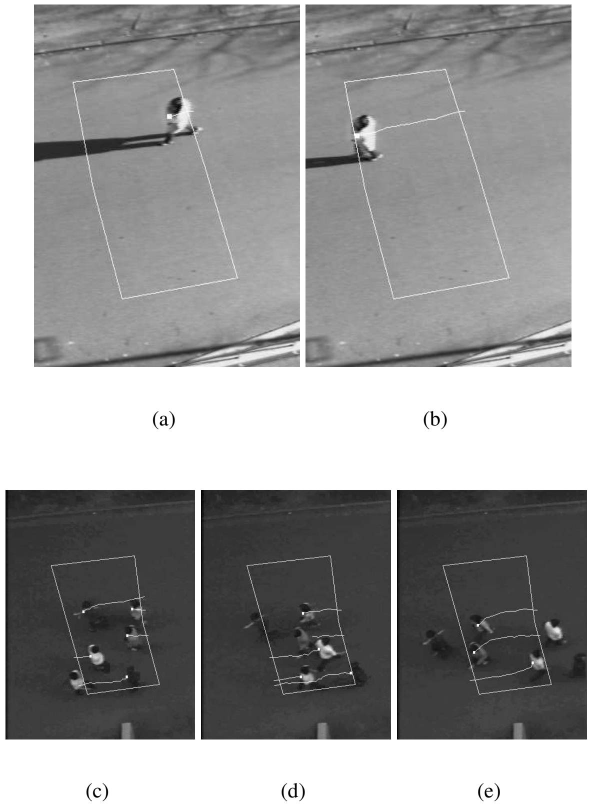


Figure 3: Tracked pedestrians. (a)(b) Pedestrian with strong shadow, which is ignored. (c)(d)(e) Tracking several pedestrians moving in close proximity.

7 Conclusion

In this paper we have presented a system for tracking pedestrians in outdoor environments using two cameras. The system is robust to shadows and changes in illumination, and the cameras can have different optical parameters which are compensated for by a straightforward prior calibration technique. Pedestrians are identified using a novel application of a Wiener-like filter to isolate point sources.

Acknowledgement

The authors acknowledge the implementation work of Mr. Masanobu Kiyama and Mr. Kenichi Yoshikawa. Anthony

	A	B	C	Total
Observed	132	41	6	179
Tracked	125	36	6	167
Success rate	95%	88%	100%	93%

Table 1: Successfully tracked pedestrians. Results are divided into 3 categories: A - pedestrian does not come close to any other pedestrian; B - pedestrian comes close to another pedestrian who is moving in the same direction; C - pedestrian comes close to another pedestrian who is moving in the opposite direction.

Dick acknowledges the support of the CRC for Sensor Signal and Information Processing (CSSIP).

References

[1] K. Onoguchi. Shadow elimination method for moving object detection. In *Proc. International Conference on Pattern Recognition*, pages 583–587, 1998.

[2] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.

[3] A. Rosenfeld and A.C. Kak. *Digital Picture Processing*. Academic Press, 1982.

[4] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

[5] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. IEEE International Conference on Computer Vision*, pages 255–261, 1999.

[6] P. Viola, M.J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *Proc. IEEE International Conference on Computer Vision*, pages 734–741, 2003.