# Person Tracking by Integrating Optical Flow and Uniform Brightness Regions*

Tsuyoshi Yamane, Yoshiaki Shirai and Jun Miura

Department of Computer-Controlled Mechanical Systems

Osaka University, 2-1, Yamadaoka, Suita, Osaka 565, JAPAN

{yamane,shirai,jun}@cv.mech.eng.osaka-u.ac.jp

## Abstract

*This paper describes a method to track a person by integrating two cues: optical flow and uniform brightness regions, where optical flow cannot be obtained. This method works even if tracking with either optical flow or uniform brightness regions may fail. The proposed method is implemented on a realtime image processor with multiple DSPs and successfully tracked a target person using a real image sequences.*

## 1 Introduction

Visual object tracking is useful for various applications such as visual surveillance and gesture recognition.

Several techniques for visual object tracking have been proposed. Some of them are based on subtraction between frames[1]. This method cannot be applied to the case in which the camera moves because the background changes. A correlation based method[2] works even when camera moves, but object tracking is difficult if the appearance of the object changes. An optical flow based method[3] can be applied to such a case. Another methods use regions without texture [6][7] or color regions[8] as features. But these methods use only one cue: either optical flow or region.

cannot discriminate a target from other objects, If a target and other objects have similar in terms of the used cue, they are hard to discriminate.

Several methods use multiple cues for reliable tracking. Okada at el.[4] used optical flow and depth for object tracking. However this method doesn't work well if the target person has low-contrast regions (*uniform brightness region*), where both optical flow and depth cannot be obtained . Nordlund at el.[5] uses uniform brightness regions in addtion to flow and depth, but such a region is used only when flow and depth are available.

---

This paper proposes a method to track a person by integrating optical flow and uniform brightness regions when a target has uniform brightness regions. This method works even if tracking with either optical flow or uniform brightness regions may fail. The proposed method is implemented on a realtime image processor and successfully tracked a target person using a real image sequence.

## 2 Outline of Person Tracking

The outline of our method is shown in Figure 1. We assume that a target person moves almost parallel to the image plane. When the person moves, the flow vectors of the person become nearly uniform. The region of the uniform flow vectors of the person is defined as *target motion region*. But the flow vectors are not obtained in the low-contrast region (*uniform brightness region*). The uniform brightness region of the person is defined as *target brightness region*. The target person is tracked by updating a rectangular window which circumscribes both target regions.

Initially, a moving person is searched for while the camera is stationary. When a region with large optical flow is detected, it is circumscribed by a rectangle (*motion window*). Uniform brightness regions are extracted in the search area generated by extending the motion window toward both the upper and the lower directions. The extracted regions are circumscribed by one rectangle ( *brightness window*) as shown in Figure 2.

In tracking, optical flow and uniform brightness regions are calculated at each frame and mean optical flow vector is calculated in the motion window. Then the motion and the brightness windows are predicted from the mean optical flow vector. The target motion region is determined around the predicted position and the new motion window is set to circumscribe it. The target brightness regions are determined around the predicted position and the new brightness window

is set to circumscribe all of them.

We integrate optical flow and unifrom brightness regions in order to track the target person reliably. Each block of the diagram is described in the following sections.
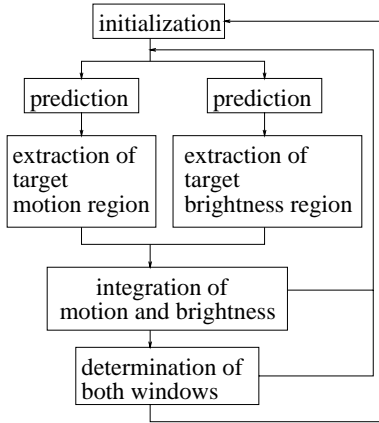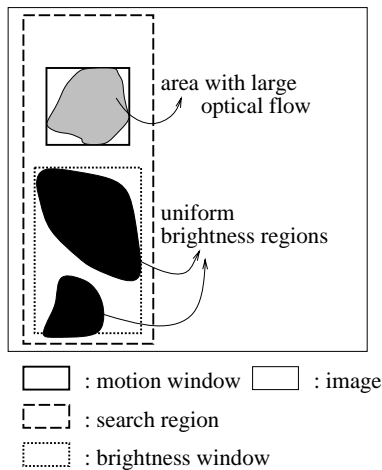


Figure 1: Block diagram of person tracking



Figure 2: Setting the initial windows

## 3 Determination of Target Motion Region

Optical flow is extracted by the generalized gradient method based on spatio-temporal filtering (see Appendix). Figure 3(a) shows an example of optical flow.

A target motion region is determined in the predicted motion window. If the difference of the flow vector of a pixel and the mean flow vector is within a certain error range, the pixel belongs to the target motion region. This range is determined as

$$(\tilde{u}-u)^2+(\tilde{v}-v)^2 < A(\tilde{u}^2+\tilde{v}^2)+B(C_u^2+C_v^2)+C, \quad (1)$$

where $(u,v)$ denotes a flow vector, $(\tilde{u},\tilde{v})$ denotes the mean flow vector in the predicted area and $(C_u,C_v)$ denotes the flow vector generated by the motion of the camera. A and B denote thresholds (A = 0.5 and B = 0.01 were used in our experiments).

But the method using only optical flow cannot discriminate the target from other objects if they have similar flow. Figure 3(b) shows that two persons are included in the target motion region (the white rectangle is the motion window).
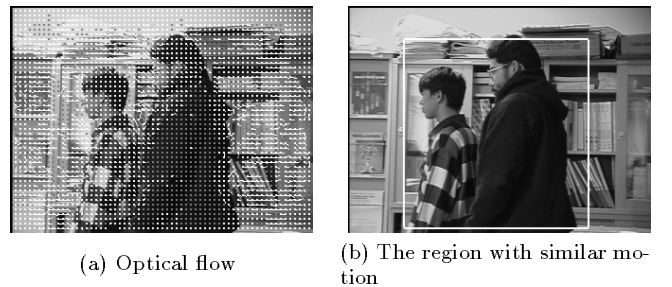


(a) Optical flow

(b) The region with similar motion

Figure 3: Failure in determination of the target motion region

## 4 Determination of Target Brightness Regions

A uniform brightness region is defined as the region where optical flow cannot be obtained (see Appendix). Figure 4 shows an example of uniform brightness regions extraction. Small regions are removed because they are hard to track due to noise. Some of uniform brightness regions may be the background regions in Figure 5(a) . A region whose centroid is outside the motion window is regarded as the background and removed. The rest regions are circumscribed by the *brightness window* as shown in Figure 5(b) .

### 4.1 Making Correspondence

For each uniform brightness region being tracked, target brightness regions are determined in the predicted brightness window as shown in Figure 6. Uniform brightness regions in the predicted brightness window are selected as candidates for the target brightness regions. Next, the position of each target brightness region obtained at the previous frame is predicted from the mean flow vector for a candidate
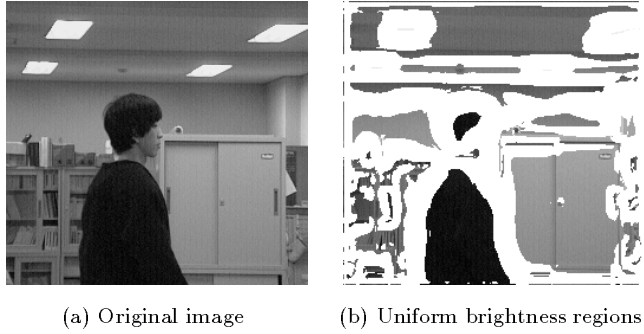
(a) Original image     (b) Uniform brightness regions

Figure 4: Extraction of uniform brightness region



(a) Background regions included     (b) Background regions removed

Figure 5: Removing background regions



: predicted target brightness region

: predicted brightness window

Figure 6: Determination of the target brightness region



(a) Original image     (b) The region with similar brightness
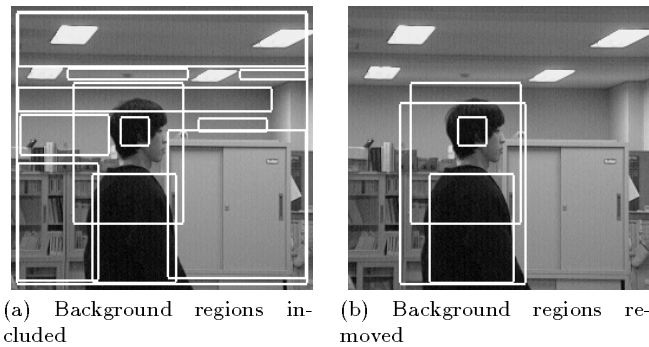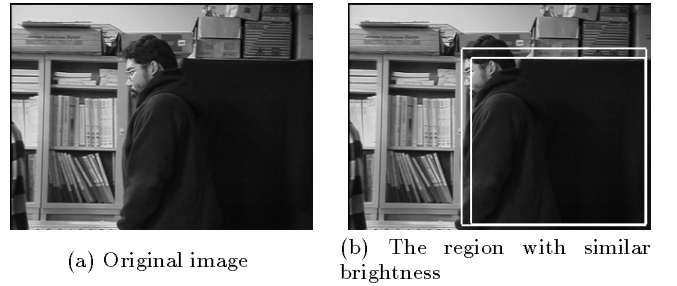
Figure 7: Failure in determination of the target brightness region

which overlaps with the target brightness region at the predicted position. If the brightness of the candidate is similar to that of the target brightness region at the predicted position, the candidate is regarded as the target brightness region of this frame. The similarity measure is defined as:

$$|m_p - m_c| < \sigma_p, \qquad (2)$$

where $m$ and $\sigma$ denote the average and the standard deviation of the brightness of a uniform brightness region. Subscriptions $p$ and $c$ denote the previous and the current frame, respectively.

But the method using only uniform brightness regions cannot discriminate the target from other objects if they have similar brightness. Figure 7(b) shows that a person and a curtain are included in the target brightness region (the outer rectangle is the brightness window).

## 4.2 Segmentation

Uniform brightness regions may split or merge. When the size of a region suddenly changes, it is con-
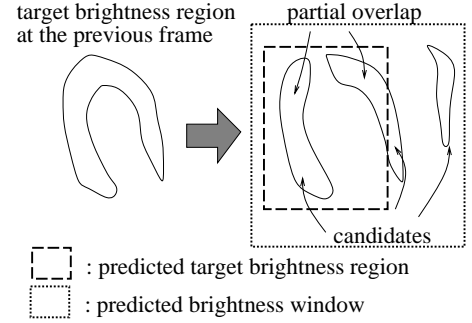
sidered that a split or a merge occurs. When a region splits, its standard deviation before the split is large and Equation (2) is satisfied. When the regions merge, however, their standard deviation before the merge may not be large and Equation (2) may not be satisfied.

For the candidates which were not matched with the target brightness region in Equation (2), we check correspondence backward, namely, check if the candidate can be matched with the target brightness region at the predicted position, using the following equations:

$$|m_p - m_c| < \sigma_c, \qquad (3)$$

where subscriptions are the same in Equation(2). If this Equation is satisfied, the candidate is regarded as the target brightness region.

Figure 8 shows the result of split and merge of region. Small rectangles approximates each target brightness region at each frame. The rectangle that circumscribes small ones is the brightness window. The largest window is the motion window. A region determined at the 19th frame was split into two regions: the region on the arm and that on the shirt by

the side of the arm in Figure 8(b) . The region before split were matched with the two regions in Equation (2). In Figure 8(c) the two regions merged into one region. The region was matched with the two regions in Equation (3).
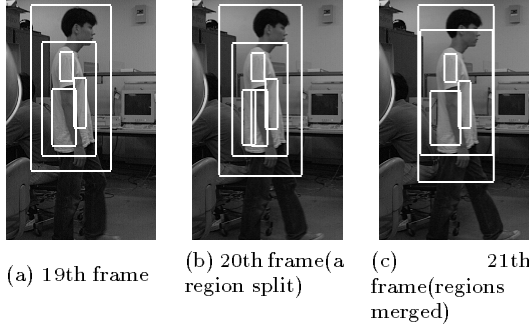


(a) 19th frame    (b) 20th frame(a region split)    (c) 21th frame(regions merged)

Figure 8: split and merge of region

# 5 Integration of Motion and Brightness

The brightness window and the motion window have almost the same horizontal position and the same width. Possible situations of the windows are classified into the following cases.

- Both windows are obtained correctly.

- The width of a window is much larger than that of the other. The larger one is considered to be less reliable and its width is reduced to that of the smaller one (see 5.1 and 5.2).

- The target motion region and/or target brightness regions are lost. A search area is determined and is searched for the lost region again (see 5.3).

## 5.1 Modification of Brightness Window by Motion

Some of them may connect to the background regions with similar brightness as shown in Figure 7(b) . In such a case the width of the motion window becomes large. If the brightness window does not satisfy the following equation, the brightness region is less reliable and its part out of the motion window is cut as the background:

$$O_b + 2W_f \leq \alpha(O_m + 2W_f), \qquad (4)$$

where $O$ denotes the width of the window, the subscriptions $b$ and $m$ denote brightness and motion, respectively. $W_f$ denotes the filter size when optical flow

is extracted (see Appendix), and $\alpha$ denotes a threshold ($\alpha = 1.5$ in our experiments). Figure 9(a) shows an expamle of modification by motion. It is shown that the target person is tracked acculately.

## 5.2 Modification of Motion Window by Brightness

When the target overlaps with other objects and they have simialr flow, the width of the motion window becomes large. If the motion window does not satisfy the following equation, the motion window is less reliable and the part out of the brightness window is cut as other objects with similar flow:

$$O_m + 2W_f \leq \alpha(O_b + 2W_f), \qquad (5)$$

where subscriptions are the same in Equation (4).

Figure 9(b) shows an expamle of modification of brightness. It is shown that the target person is tracked accurately.



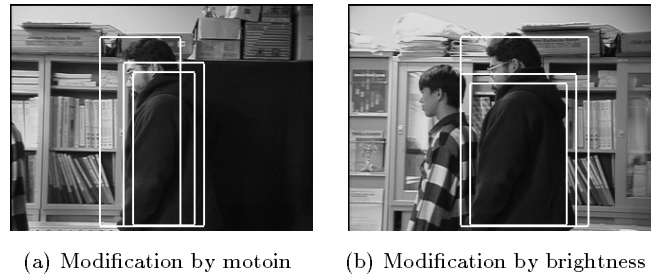(a) Modification by motoin    (b) Modification by brightness

Figure 9: Tracking by modification

Some of the uniform brightness regions of other objects may be included in the motion window. In this case, it is difficult to modify the motion window correctly due to the spurious brightness regions. We use a measure of reliability to solve this problem. A region which has been tracked over a certain number of frames is regarded as a reliable one. We set a counter to each region as the reliablity measure. It is incremented as the region is tracked. A region outside reliable regions is removed as that of other objects.

Figure 10 shows an example. The regions of the foreground person have been tracked over certain frames and they are regarded as reliable ones in Figure 10(a) ,(b). After overlapping, the spurious regions of the background person appear in Figure 10(c) . Based on the reliable regions, they are removed in Figure 10(d) .
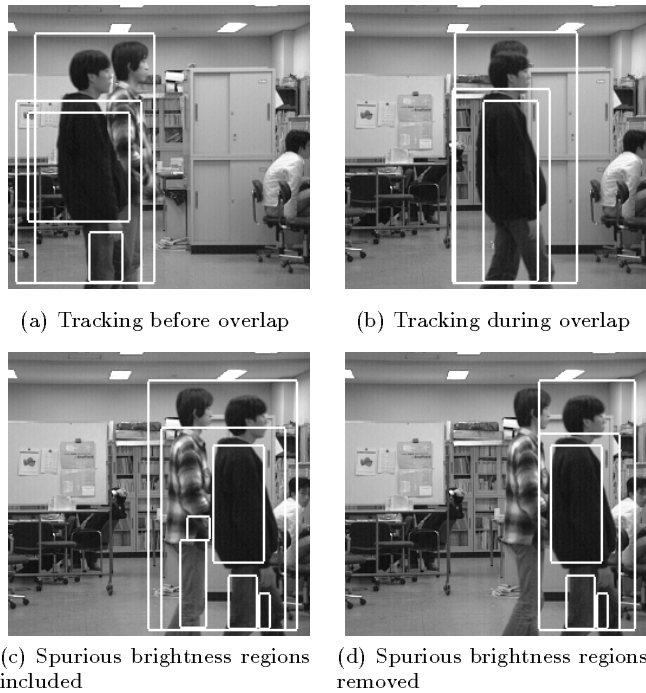
(a) Tracking before overlap



(b) Tracking during overlap



(c) Spurious brightness regions included



(d) Spurious brightness regions removed

Figure 10: Modification by reliable regions

## 5.3 Determination of Search Area for Target Region

Figure 11 shows the association of the target motion region and target brightness regions.

In case that either the target motion region or the target brightness region is lost, the search area is generated by extending the window of the other toward the upper and lower directions. If both of them are lost, they are searched for the target motion region in the same way as the initial search described in Section 2.
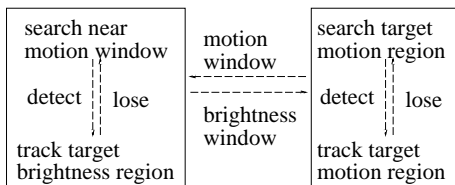


Figure 11: Association of the motion and the brightness windows

# 6 Implementation of Tracking on The Realtime Image Processor

The hardware configuration of our system is illustrated in Figure 12. The image processor inputs images at frame rate 15Hz and outputs the result of tracking out to the monitor. The resolution of a image is 160 x 120 pixels.

The image processor is composed of several DSP(Digital Signal Processor) boards. Each DSP board has one DSP chip(TMS320C40 40MHz). Each DSP board can send data to other DSP boards through the data bus. Each DSP chip can exchange data with another DSP chip through the communication port. DSP boards can be connected flexibly.

The proposed algorithm is implemented on the image processor as shown in Figure 13. Optical flow is extracted in DSP 1-9,13,14. Uniform brightness regions are extracted in DSP 10. The position, the average and the standard deviation of each region are calculated in DSP 11. DSP 12 performs tracking using uniform brightness regions. DSP 15 performs tracking using optical flow, puts the result out to the monitor and sends data of camera control.
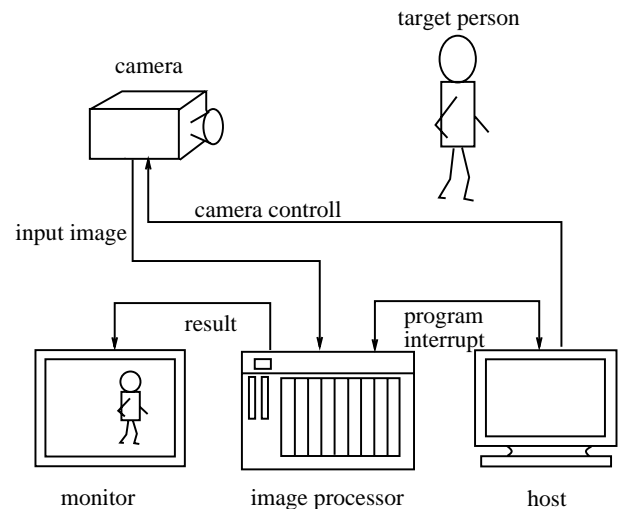


Figure 12: Overview

# 7 Conclusion

We proposed a method to track a person by integrating optical flow and uniform brightness regions. This method worked even if tracking with either optical flow or brightness may fail. This method was implemented on the DSP-based realtime image processor.
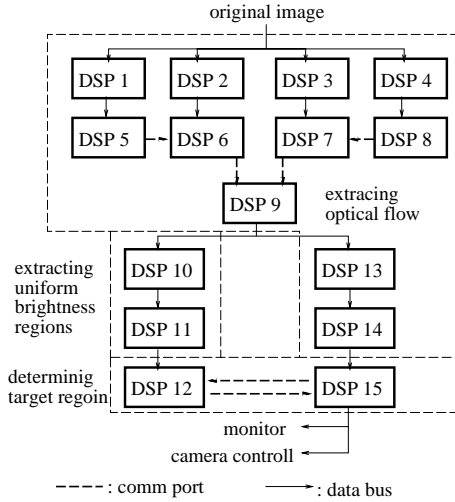
Figure 13: Realtime image processor

The current method cannot track the target person if it is completely occluded by other objects. We can handle such a case by using the the brightness information which was obtained before occulusion. Another complicated case occurs, when other objects have similar motion and simlar brightness to those of the target. To cope with this complicated case we are planning to use three kinds of information together: optical flow, uniform brightness regions and depth.

## Appendix: Extracting Optical Flows and Contrast

Optical flow extraction is based on the generalized gradient method[9]. By applying four orientation-selective spatial Gaussian filters to the original image, we obtain the following constraint equations

$$f_{ix}u + f_{iy}v + f_{it} = 0, \quad i = 1 \sim 4 \qquad (6)$$

where $f_i$ denotes the $i$th filtered image and subscriptions x,y,t denote partial differentiation. We define the contrast as

$$c_i = \sqrt{f_{ix}^2 + f_{iy}^2} \qquad (7)$$

If $c_i$ is small enough, the optical flow cannot be obtained. We can obtain the optical flow vector as the following equation which minimizes the sum of weighted squared distance from the solution to four constraint equations (6)[10],

$$\binom{u}{v} = \begin{bmatrix} \sum w_i \frac{f_{ix}^2}{c_i^2} & \sum w_i \frac{f_{ix}f_{iy}}{c_i^2} \\ \sum w_i \frac{f_{ix}f_{iy}}{c_i^2} & \sum w_i \frac{f_{iy}^2}{c_i^2} \end{bmatrix} \left( \sum w_i \frac{f_{ix}f_{it}}{c_i^2} \atop \sum w_i \frac{f_{iy}f_{it}}{c_i^2} \right) \cdot (8)$$

We define the weight which increases as the contrast is large:

$$w_i = c_i^2 / \sum_j c_j^2. \qquad (9)$$

## References

[1] M. Yachida, M. Asada and S. Tsuji, "Automatic analysis of moving image", IEEE Trans. Pattern Anal. Mach. Intell, Vol. PAMI-3, No. 1, pp. 12-20, 1981.

[2] H. Inoue, T. Tachikawa and M. Inaba, "Robot vision system with a correlation chip for real-time tracking, optical flow and depth map generation", Proc. IEEE Int. Conf. on Robotics and Automation, pp. 1621-1626, 1992.

[3] S. Yamamoto, Y. Mae, Y. Shirai and J. Miura, "Real-time multiple object tracking based on optical flows", Pro. Robotics and Automation, Vol. 3, pp. 2328-2333, 1995.

[4] R. Okada, Y. Shirai and J. Miura, "Object tracking based on optical flow and depth", Proc. IEEE/SICE/RSJ Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems, pp. 565-571, 1996.

[5] P. Nordlund and J.-O. Eklundh, "Figure-ground segmentation as a step towards deriving object properties", Proc. 3rd Int. Workshop on Visual Form, May, 1996.

[6] B. Bascle and R. Deriche "Region tracking through image sequences", Proc. 5th Int. Conf. on Computer Vision, pp. 302-307, 1995.

[7] J. Badenas, M. Bober and F. Pla, "Motion and intensity-based segmentation and its application to traffice monitoring", Proc. Int. Conf. Image Analysis and Processing, Vol. 1, pp. 502-509, 1997.

[8] K. Sato, M, Inaba and H, Inoue, "Realtime Human Detection, Tracking and Interpretation of Action based on Parallel Processing", ROBOMEC' 97, pp. 591-592, 1997.

[9] M.V. Srinivasan. "Generalized gradient schemes foe the measurement of two-dimensional image motion", Biological Cybernetics, Vol. 63, pp. 421-431, 1990.

[10] H.J. Chen, Y. Shirai and M. Asada, "Detecting multiple rigid image motions from an optical flow field obtained with multi-scale filters", IEICE Trans. Inf. & Syst. Vol. E76-D, No. 10, pp. 1253-1262, 1993.