Person following with a mobile robot using a modified optical flow

Ankur Handa[1], Jayanthi Sivaswamy[1], Madhava Krishna[1], Sartaj Singh[1]

Robotics Research Center, IIIT Hyderabad, Hyderabad, 500032, India
[1] *E-mail: mkrishna@iiit.ac.in*

Paulo Menezes[2]

Department of Electrical Engineering, Laboratory, University of Coimbra, Pine, Morocco - 3030, Coimbra, Portugal
[2] *E-mail: pm@deec.uc.pt*

This paper deals with the tracking and following of a person with a camera mounted mobile robot. A modified energy based optical flow approach is used for motion segmentation from a pair of images. Further a spatial relative veolcity based filering is used to extract prominently moving objects. Depth and color information are also used to robustly identify and follow a person.

Keywords: Tracking, Person following, Energy minimization, non-parametric density estimation.

1. Introduction

Tracking a moving object has numerous applications in survellience, security and monitering [7–9]. It is relatively simple to extract moving objects from a static background but when the background itself is changing, it becomes more challenging to segment moving object motion. In this paper we deal with the problem of following a person from a camera mounted mobile robot. We propose a modified energy based optical flow technique which robustly computes the smooth flow vector field. Next, we employ a spatial relative velocity based filering to extract the regions which have abrupt change in their relative velocity (compared with their neighbourhood) and intensity profile around them. Further depth and color information are used to accurately identify and follow a person. The proposed method has been extensively tested on our robot, called SPAWN, in different environments. The tested environments include moderate changes in ambient light, pres-

ence of many stationary objects.

2. Methodology

Optical flow based techniques have been widely used to extract motion information [1,2]. However, these techniques are quite susceptible to noise since they also depend on intensity gradient. To overcome these problmes we formulate here the flow field determination in an energy minimization framework which takes into account the correlation of an intensity patch in two successive frames and the direction of flow vectors of its neighboring patches by introducing a new energy term E_{dir}. We explain this in a detail as follows. From an image pair I_t and I_{t-1}, consider a patch of size pWxpH at location (x, y) in an image I_t and define an energy function as

$$E_{corr}(i,j) = \sum_{y=\frac{-pH}{2}}^{y=\frac{pH}{2}} \sum_{x=\frac{-pW}{2}}^{x=\frac{pW}{2}} (I_t(x,y) - I_{t-1}(x+i, y+j))^2 \,,$$

This represents the correlation of an intensity patch in I_t with intensity patches in I_{t-1} at locations (i, j) within a window W centered around (x, y). Next we define *direction energy* E_{dir} as the penalty imposed for a direction when the patch is at (i, j) given that the direction of the neighboring patches is known. We denote $d^t_{n_k}$ as the direction of a neighboring patch for which the direction has already been determined and the direction associated with the patch when it is at (i, j) is denoted by d^{t-1}_p.

$$E_{dir}(i,j) = \sum_{k=1}^{m} \alpha_k (d^{t-1}_p - d^t_{n_k})^2 \,,$$

$$d^{t-1}_p = \tan^{-1}\left(\frac{j}{i}\right)$$

The net energy function E_{net} at each (i, j) is defined as

$$E_{net}(i,j) = E_{corr}(i,j) + E_{dir}(i,j)$$

The final direction d^t_p and net spatial displacement $(\hat{i},\hat{j})$ is defined as the one which minimizes this energy in that window W

$$(\hat{i},\hat{j}) = \arg\min_{i,j}(E_{net}(i,j)W(i,j))$$

$$d^t_p = \tan^{-1}\left(\frac{\hat{j}}{\hat{i}}\right)$$

Here, the window W of size wHxwW is defined as

$$W(i,j) = \begin{cases} 1 \text{ if } \frac{-wW}{2} \leq i \leq \frac{wW}{2} \\ \quad and \\ \quad \frac{-wH}{2} \leq j \leq \frac{wH}{2} \\ 0 \; otherwise \end{cases}$$

and α_k is a smoothening constant. Since the background pixels inherit the motion from the camera, their motion will be locally similar. On the other hand the pixels belonging to the moving objects will have motion incoherent with the background. Thus the boundaries of moving objects should correspond to the discontinuity in their relative displacements in local neighborhood. Once the flow vectors are determined, we employ spatial relative veolcity based filtering to robustly extract the boundaries of moving objects. This spatial relative velocity based filtering includes labelling $\mathbf{L}$ of patches according to the following criteria.

$$\mathbf{L} = \begin{cases} 0 : \text{if}((\delta_x + \delta_y < th_1) \vee ((\delta_x + \delta_y > th_1) \wedge (\sigma_i < th_2))) \\ 1 : otherwise \end{cases}$$

where δ_x and δ_y are the sum of relative displacements in x and y directions respectively of a patch in its neighborhood and σ_i is the standard deviation in intensity around that patch. A given patch is labelled 0 if sum of relative velocities in x and y directions is below a certain threshold th_1. Other patches which surpass the threshold are again processed to check for false alarms. If the intensity profile around these patches is smooth ($\sigma_i < th_2$), it is very unlikely that they belong to a moving object boundary. Hence they are labelled as 0. Patches finally labelled 1 are the ones which are likely to have prominent motion.

3. Fusing color and depth information

In order to accurately classify each prominently moving patch as to belonging to a person, we incorporate color and depth information. Patches are first clustered on the basis of their depth values and then classified on the basis of their color information. The color model of a person is non-parametrically estimated offline. We model the color density of the upper part of persons body using non-parametric kernel density estimation. Given a sample data for color values $D^c = \{c_i\}$ where $i = 1..N$ and c_i is a k-dimensional vector, kernel density estimation is used to estimate the probability that a given color sample C is from the distribution given by

D^c as

$$P(C) = \frac{1}{N}\sum_{i=1}^{N} K(C - c_i)$$

Choosing a zero mean and $\sum$ bandwidth Gaussian function as a kernel estimator function K, we assume independence between the different k channels. Then for each kernel, the bandwidth is

$$\sum = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & . & . & . \\ 0 & . & . & \sigma_k^2 \end{pmatrix}.$$

Hence, the density can be written as

$$P(C) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{(2\pi)^{\frac{k}{2}}|\sum|^{\frac{1}{2}}} e^{-\frac{1}{2}(C-c_i)^T \sum^{-1}(C-c_i)}$$

The bandwidths were estimated from image regions of the upper part of the person. The bandwidth for the Gaussian function was estimated as $\sigma \approx 1.06\widehat{\sigma}n^{-1/5}$ where $\widehat{\sigma}$ is the standard deviation and n is the sample size. To speed up the computation of the probabilities, the values of the Gaussian kernel, given the color value difference and kernel function bandwidth, were precalculated and stored in a Look Up Table (LUT). Thus, the values could be fetched in $O(1)$, avoiding excessive floating point computations. Also, the color values for the models were stored as $< r^j, g^j, b^j, n^j >$ where $< r^j, g^j, b^j >$ is the sample color data for person and n^j denotes the number of times the j^{th} color tuple has occured in the sample data. Hence, the likelihood of the pixel to a person was computed efficiently as

$$P(\mathbf{p}|D^c) = \frac{1}{N}\sum_{j} n^j K_{\sigma_r}(r - r^j) K_{\sigma_g}(g - g^j) K_{\sigma_b}(b - b^j)$$

where

$$\mathbf{p} = < r, g, b >$$

If this likelihood is more than a particular threshold p_{th}, the pixel is classified as belonging to a person. This is done for every pixel in the patch and using a majority rule. The centriod of a person is computed from the patches classified as belonging to a person. Color model is then periodically updated by the new intensity values obtained from the cluster after identification is done.

The robot velocities are controlled by the disparity and the angle of the centroid of person in image plane. The translational veolcity v_{tx} of robot is proportional to the disparity of the centroid and the rotational velocity v_{rl} of robot is proportional to the angle of centroid in the image plane.

$$v_{tx} = c_1 d_c^t$$
$$v_{rl} = c_2 \theta_c^t$$

where d_c^t is disparity of the centroid of person and θ_c^t is the angle the centroid makes in the image plane at time t.

$$\theta_c^t = \tan^{-1}\left(\frac{x_c^t - c_x}{f}\right)$$

c_x is the center of the image in x direction. Proximity of the vector (x_c^t, y_c^t, d_c^t) to its previous position is used as a consitency check, where (x_c^t, y_c^t) is the position of the centriod of a person in the image at time t. Figure 1 shows the robot following a person.

Fig. 1. Robot following a person

4. Results

The proposed method was implemented in C++ on a linux platform (FC7) with AMD Athlon 64-bit processor. The image resolution was kept at 320x240. The algorithm was extensively tested on our lab robot, SPAWN in indoor environments under different conditions such as similar background color, varying lightening conditions and presence of many static objects. Figure 2 and Figure 3 show that the robot is able to keep track of the person. The media files (format: .avi) showing the robot following a person can be obtained from the following website: `http://students.iiit.ac.in/~ankurhanda/robot.html`.

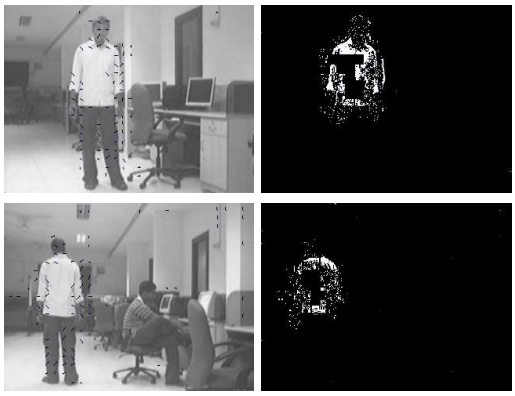

Fig. 2. Left images showing the motion segmentation after spatial relative velocity based filtering while right images showing the results of segmentation of person after fusing depth and color information.

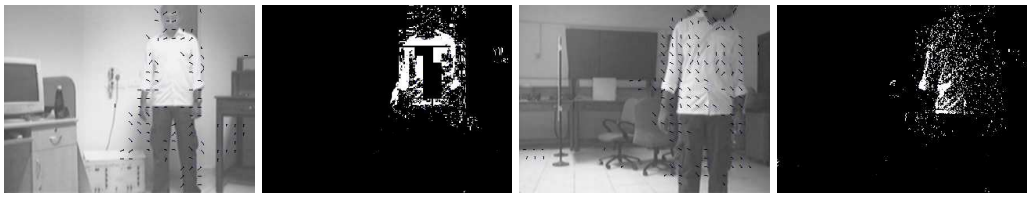

Fig. 3. Tracking of a person under different testing environment,left: same background color and right: poor lightening conditions

References

1. M. Piaggio, P. Fornaro, A. Piombo, L. Sanna and R. Zaccaria. An optical flow based person following behaviour. In *Proceedings of the IEEE ISIC/CIRNISAS Joint Conference*, 1998.
2. C. Schlegel, J. Illmann, H. Jaberg, M. Schuster and R. Worz. Vision based person tracking with a mobile robot. In *The British Machine Vision Conference*, 1998.
3. Z.Zivkovic. Improved adaptive Gausian mixture model for background subtraction. *International Conference Pattern Recognition*, Vol.2, pages: 28-31 , 2004.
4. Wren C., A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-Time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, pages:780-785, 1997.
5. Y. Raja, S. McKenna, S. Gong. Object Tracking Using Adaptive Colour Mixture Models,*Proc. ACCV 98*, Vol. I, pp 615-62266.

6. B.K. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*,Vol.17,pages:185-203,1981.
7. A. Behrad, A. Shahrokni, S. A. Motamedi and K. Madani.A Robust Vision-based Moving Target Detection and Tracking System. In *Proceedings of Image and Vision Computing conference (IVCNZ2001)*, University of Otago, Dunedin, New Zealand, November 26-28, 2001
8. B. Jung and Gaurav S. Sukhatme. Detecting Moving Objects using a Single Camera on a Mobile Robot in an Outdoor Environment. In *International Conference on Intelligent Autonomous Systems*, pp. 980-987, Amsterdam, The Netherlands, Mar 2004.
9. H. Kwon, Y. Yoon, J. B. Park and A. C. Kak. Person tracking with a mobile robot using two uncalibrated independently moving cameras. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2005.