Extracting Human Limb Region using Optical Flow and Nonlinear Optimization

Toru Tamaki [1] Tsuyoshi Yamamura [3] Noboru Ohnishi [24]

[1] *Department of Information Engineering, Faculty of Engineering, Niigata University, Niigata 950–2181 Japan*
[2] *Center for Information Media Studies, Nagoya University, Nagoya 464–8603 Japan*
[3] *Faculty of Information Science and Technology, Aichi Prefectural University, Aichi 480–1198 Japan*
[4] *Bio-Mimetic Control Research Center, RIKEN, Nagoya 463–0003 Japan*

Abstract

We propose a method for extracting human limb regions by combination of optical flow based motion segmentation and nonlinear optimization based image registration. First, rotating limb regions with rough boundaries are extracted and motion parameters are estimated for an approximated model. Then the extracted region and estimated parameters are used as initial values for nonlinear optimization that minimizes residuals of two successive frames and estimates motion parameters. Combining the two steps reduces computational costs and avoids the initial state problem of optimization. According to estimated parameters, the limb region is extracted by Bayesian classifier to obtain accurate region boundaries. Experimental results on real images will be shown.

1. Introduction

It is important to extract human regions from a movie as a part of a human activity recognition system, including gesture recognition for human interfaces and motion reconstruction in virtual reality. For such applications, detecting and extracting human arms in a scene plays a key role [1] showing where a subject is and how he/she acts, especially for recognition of gestures, which are mainly determined by the arm movements.

Many human activity recognition studies have been developed; and they often use parameterized human body models to reconstruct actual human posture [2, 3]. However, these methods require that a background is known or at least is of uniform color to make subtraction easy, otherwise there must be no moving object except the subject. The assumption about background is one hurdle to developing methods so that a recognition system adapts to a changing real environment.

To overcome the problem, we have proposed a method [4] to extract regions of rotating human limbs represented by a stick model and estimate their motion parameters. The extraction method we proposed is an indirect method; i.e., from optical flow of two successive images calculated in advance, segmenting an image into motion regions and estimating the motion parameters of each region. This method can extract arm regions from optical flow of a real image sequence contaminated by much noise. However, it is impossible to compute optical flow where the motion correspondence can not be found, especially at the edge of motion, and the indirect method would fail to extract the exact arm region boundary.

On the other hand, a method of motion segmentation by comparing intensities of two successive frames directly [6] has been proposed. This direct method can deal with motion discontinuity at the motion edge and estimate accurate motion parameters because it doesn't use optical flow, which causes failure of the indirect method. The problem of the direct method is its high computational cost because it uses nonlinear optimization to minimize intensity residuals of two frames all over the image with initial parameters which may sometimes deviate greatly from true values.

In this paper, we propose a method to extract regions of rotating human limbs with an accurate boundary by combined use of indirect and direct methods. At first, limb regions are extracted by the indirect method using optical flow and estimated its motion parameters. Then, the accurate boundary of the region is obtained by the direct method using the extracted region and estimates of the indirect method as initial values. This combination is expected to decrease computational costs and improve the extraction result of the indirect method. We describe the indirect method of extraction and estimation based on optical flow in section **2**, and the direct method using nonlinear optimization in section **3**. Final extraction with MAP estimation is discussed in section **4**. Finally, we provide experimental results of real images in section **5**.

2. Indirect method using optical flow

In this section, we describe the indirect method of extracting limb region using optical flow. The limb is assumed to rotate on a plane that is not parallel to the image plane (Fig.1). There are two cases in which the plane slants; and

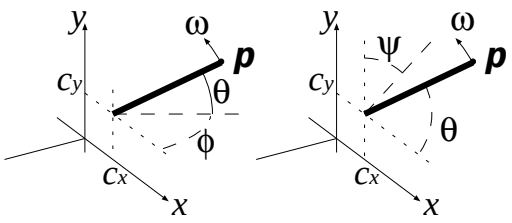


Fig. 1. A stick model moving on a plane rotated by ϕ about the y axis (left) and by ψ about the x axis (right).

both can be used as an approximated 3D motion model. In both cases, motion of a point $\mathbf{p}_j = (x_j, y_j)$ on a rotating limb and its velocity $\dot{\mathbf{p}}_j = (u_j, v_j)$ are modeled as follows [4].

$$\dot{\mathbf{p}}_j = A_j \mathbf{q} \tag{1}$$

where

$$A_j = \begin{pmatrix} y_j & 1 & 0 & 0 \\ 0 & 0 & x_j & 1 \end{pmatrix}, \quad \mathbf{q} = (\alpha, \beta, \gamma, \delta)^T \tag{2}$$

Here, motion parameters (angular velocity ω and rotation center (c_x, c_y) of the rotating limb) are calculated from $\mathbf{q}$ as follows.

$$c_x = -\delta/\gamma, \quad c_y = -\beta/\alpha, \quad \omega = -\text{sign}(\alpha)\sqrt{-\alpha\gamma} \tag{3}$$

Also, ϕ and ψ are retrieved[4] from $\mathbf{q}$.

Motion of a point on a limb is modeled as above, but optical flow computed from real images involves inevitable noise. We assume that distribution of $\dot{\mathbf{p}}_j$ is subject to a two-dimensional Gaussian,

$$P(\dot{\mathbf{p}}_j |\ \mathbf{p}_j, \mathbf{q}, \Sigma) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left\{ \frac{-1}{2} \Big(\dot{\mathbf{p}}_j - A_j\mathbf{q}\Big)^T \Sigma^{-1} \Big(\dot{\mathbf{p}}_j - A_j\mathbf{q}\Big) \right\} \tag{4}$$

where $\Sigma = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$ is a covariance matrix which assumes that the errors for u and v are mutually independent from each other because α, β and γ, δ are estimated separately.

Next, we show below the algorithm to segment the limb region and estimate its parameters [4]. This is an application of the EM algorithm[10] which assumes that each moving object has its own motion parameter $\mathbf{q}$ and that optical flow distribution within the object region is Eq.(4).

1. Compute optical flow $\dot{\mathbf{p}}_j = (u_j, v_j)^T$ at each point $\mathbf{p}_j = (x_j, y_j)^T$ $(j = 1, \ldots, N)$.

   Perform initial segmentation of optical flow based on direction of velocity to obtain initial clusters R_i $(i = 1, \ldots, M)$ (see section 5).

   Then, set weight w_{ij} as a probability that a point $\mathbf{p}_j$ belongs to a cluster R_i. Set initial values of w_{ij} as follows.

   $$w_{ij} = \begin{cases} 1 & (\mathbf{p}_j \in R_i) \\ 0 & (\mathbf{p}_j \notin R_i) \end{cases} \tag{5}$$

2. Normalize weights w_{ij}.

   $$w'_{ij} = \frac{w_{ij}}{\xi_i}, \qquad \xi_i = \frac{1}{N}\sum_j w_{ij} \tag{6}$$

3. Find parameters $\mathbf{q}_i = (\alpha_i, \beta_i, \gamma_i, \delta_i)$ of each cluster R_i solving the following system of equations by QR decomposition[11].

   $$\begin{pmatrix} \sqrt{w'_{i1}}\,\dot{\mathbf{p}}_1 \\ \sqrt{w'_{i2}}\,\dot{\mathbf{p}}_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \sqrt{w'_{i1}}A_1 \\ \sqrt{w'_{i2}}A_2 \\ \vdots \end{pmatrix} \mathbf{q}_i \tag{7}$$

4. Compute the weighted variances $\sigma^2_{x_i}$ and $\sigma^2_{y_i}$ for each cluster R_i.

   $$\sigma^2_{x_i} = \frac{1}{N}\sum_j w'_{ij}(u_j - \alpha_i y - \beta_i)^2 \tag{8}$$

   $$\sigma^2_{y_i} = \frac{1}{N}\sum_j w'_{ij}(v_j - \gamma_i y - \delta_i)^2 \tag{9}$$

5. Update weights w_{ij} with the following equation.

   $$w_{ij} = \frac{\xi_i P(\dot{\mathbf{p}}_j |\ \mathbf{p}_j, \mathbf{q}_i, \sigma^2_{x_i}, \sigma^2_{y_i})}{\sum_k \xi_k P(\dot{\mathbf{p}}_j |\ \mathbf{p}_j, \mathbf{q}_k, \sigma^2_{x_k}, \sigma^2_{y_k})} \tag{10}$$

6. For all clusters, if the difference between estimated parameter values of the current iteration and that of the previous iteration is larger than a threshold, return to step 2. Otherwise, proceed.

7. Perform segmentation by making each point $\mathbf{p}_j$ belong to cluster R_i with the largest weight w_{ij}.

   $$\mathbf{p}_j \in R_{i^*} \quad \text{where} \quad i^* = \underset{1 \le i \le M}{\text{argmax}}\, w_{ij} \tag{11}$$

   Then, extracting a region with the largest angular velocity (calculated by Eq.(3)) as the rotating limb region R_Ω.

   $$\Omega = \underset{1 \le i \le M}{\text{argmax}}\, \omega_i \tag{12}$$

For simplicity, we assume that there is only one limb in a scene. However, the method described in the following sections can be applied to each moving region separately because one can determine the number of movements in a scene [4], place of the background (no motion), and extract each region. Actually, section 5 shows results of a scene with two moving arms.

3. Direct method using nonlinear optimization

As mentioned previously, the indirect method cannot be applicable to motion edge discontinuity (shaded line area of Fig.2) because optical flow cannot be calculated there. However, we have extracted the rough region of rotating limb R_Ω and estimated motion parameters of approximated motion model $\mathbf{q}_\Omega$ by the indirect method. These estimates help us to improve the direct method in terms of computational cost and initial value problems.

The direct method models motion using eight parameters[5, 6], while the method in the previous section uses only four. Let I_t and I_{t+1} be images at times t and $t+1$. A point $\mathbf{p}_j$ in I_t corresponds to $\mathbf{p}_j + \mathbf{u}(\mathbf{p}_j;\boldsymbol{\theta})$ in I_{t+1}, where $\mathbf{u}$ is a motion vector and $\boldsymbol{\theta}$ represents motion parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_8)^T$.

Estimation is done by minimizing the square of residuals of intensities r_i of two points as

$$\min_{\boldsymbol{\theta}} \sum_{\mathbf{p}_j \in R_\Omega} r_j^2 \tag{13}$$

where

$$r_j = I_t(\mathbf{p}_j) - I_{t+1}(\mathbf{p}_j + \mathbf{u}(\mathbf{p}_j;\boldsymbol{\theta})) \tag{14}$$

$$\mathbf{u}(\mathbf{p};\boldsymbol{\theta}) = \begin{pmatrix} x & y & 0 & 0 & 1 & 0 & x^2 & xy \\ 0 & 0 & x & y & 0 & 1 & xy & y^2 \end{pmatrix} \boldsymbol{\theta} \equiv M\boldsymbol{\theta} \tag{15}$$

We use Gauss-Newton method [7] to estimate $\boldsymbol{\theta}$ minimizing the cost function (Eq.(13)). By iteration of optimization, estimates are modified as $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \delta\boldsymbol{\theta}$, and the modification $\delta\boldsymbol{\theta}$ is obtained by solving the following systems of equations [6].

$$\sum_{l}^{8} \sum_{\mathbf{p}_j \in R_\Omega} \frac{\partial r_j}{\partial \theta_k} \frac{\partial r_j}{\partial \theta_l} \delta\theta_l = - \sum_{\mathbf{p}_j \in R_\Omega} r_j \frac{\partial r_j}{\partial \theta_k} \tag{16}$$

for $k = 1, \ldots, 8$, where

$$\frac{\partial r}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{u}}{\partial \boldsymbol{\theta}} \frac{\partial r}{\partial \mathbf{u}} = -M^T \nabla I_{t+1}(\mathbf{p} + \mathbf{u}(\mathbf{p};\boldsymbol{\theta})) \tag{17}$$

The estimation procedure repeats to solve the system of equations and update estimates. This requires appropriate initial values. According to Eqs.(2) and (15), estimates $\mathbf{q}_\Omega$ obtained by the indirect method correspond to $\boldsymbol{\theta}$ as $\alpha = \theta_2$, $\beta = \theta_5$, $\delta = \theta_3$, and $\gamma = \theta_6$. So $\mathbf{q}_\Omega$ is used as the initial value of the four of $\boldsymbol{\theta}$, and others of $\boldsymbol{\theta}$ initialized to 0.

After the iteration converges, final estimates for $\boldsymbol{\theta}$ (we write as $\hat{\boldsymbol{\theta}}$) are obtained.

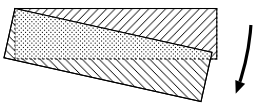


Fig. 2. An example of the motion edge

4. Extraction by Bayesian classifier

Since estimation in the previous section is performed only in the region R_Ω (extracted by the indirect method), we need to determine the human limb region according to final estimates $\hat{\boldsymbol{\theta}}$.

To extract the limb region, we use a Bayesian classifier which maximizes posterior probability assuming that residual r_j at each pixel is subject to Gaussian distribution. So, conditional probability is defined as

$$P(r_j|\boldsymbol{\theta}_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{{r_j}^2}{2\sigma_i^2}\right) \tag{18}$$

where, $i = 1$ means the human limb region with $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}$, and $i = 0$ is the background with parameters of zero $\boldsymbol{\theta}_0 = \mathbf{0}$ (no movement). Therefore, σ_1^2 is calculated within R_Ω and σ_0^2 is calculated outside of R_Ω (or in the background region).

Let $|R_\Omega|$ be the number of pixels in the region R_Ω, and N be the number of all pixels in the image. We define prior probabilities of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$ by the ratio of areas, that is,

$$P(\hat{\boldsymbol{\theta}}) = \frac{|R_\Omega|}{N} \tag{19}$$

$$P(\mathbf{0}) = 1 - P(\hat{\boldsymbol{\theta}}) \tag{20}$$

Then computing and comparing posterior probabilities for each pixel $\mathbf{p}_i$ (without regarding denominators $P(r_j)$) as follows;

$$P(\hat{\boldsymbol{\theta}})P(r_j|\hat{\boldsymbol{\theta}}) > P(\mathbf{0})P(r_j|\mathbf{0}) \tag{21}$$

If the above inequality holds, then the pixel belongs to the limb region; otherwise, the pixel is the background.

5. Experimental results

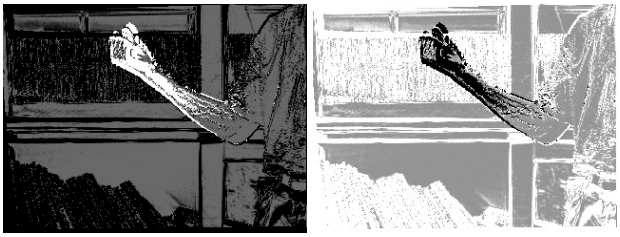
The proposed method has been implemented on PC using C++. Computation of optical flow (algorithm step 1.) was performed by code released by [8, 9], and initial clusters (step 1.) were made by a simple histogram clustering which divides directions of velocity vector into 24 sections and finds peaks in the direction histogram as the center of the clusters.

Figure 3 shows the experimental result on a real image sequence of bending arm toward the shoulder fixing the elbow position. Figure 3(a) is the first frame and Fig.3(b) shows superimposed optical flow of motion between the first and second frame. Fig.3(c) is the indirect method result showing that, because of inaccurate optical flow at the motion edge, the indirect method cannot extract the top of the arm where the motion is large. We can also see that the region boundary is not identical with that of the actual arm.

Figure 3(d) is the result of extraction by the direct method which uses the result (estimated parameters and extracted region) of Fig.3(c) as the initial value. Compared with

(a) original image (b) optical flow

(c) result of indirect method (d) result of direct method



(e) posterior probability $P(\hat{\boldsymbol{\theta}})P(r_j|\hat{\boldsymbol{\theta}})$ (f) posterior probability $P(\mathbf{0})P(r_j|\mathbf{0})$

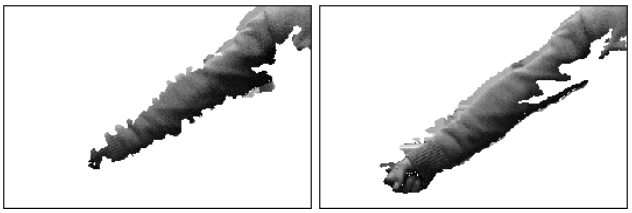
Fig. 3. Experimental result.

Fig.3(c), we can see that the lower arm region is extracted and boundary is close to the actual contour in Fig.3(d). However, the region around the elbow is not extracted because the posterior probability of the area where change of intensity is flat is small in either class. It is therefore ambiguous to which class the point should belong. Figure 3(e) and (f) illustrate posterior probabilities in which high probability is painted in white and low is in black. In Fig.3(e), the arm area is white except around the elbow and the background with uniform intensity is same gray level in both (e) and (f).

Another experiment is shown in Fig.4. The arm moves downward, and Fig.4(b) shows that the extracted region by the indirect method becomes narrow at the top of the arm because of the motion edge problem. On the other hand, extraction of the direct method in Fig.4(c) is better in terms of accuracy of the boundary of extracted region. Ambiguity at the flat intensity (near the shoulder) is also occurs in this case.

As mentioned at the end of section 2, the proposed method can deal with multiple arm motions. Figure 5 shows

(a) Original image



(b) Result of indirect method (c) Result of direct method

Fig. 4. Another experiment

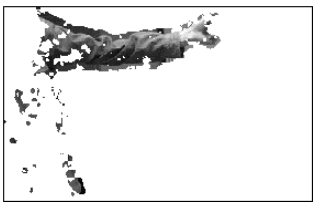
that the two arms move simultaneously; the left arm moves downward an the right arm moves upward. We can see extraction improvement for the left arm in Figs.5(b) and (c), but right arm extraction is not improved so much (Figs.5(d) and (e)). The reason is that there are many areas with flat intensity.

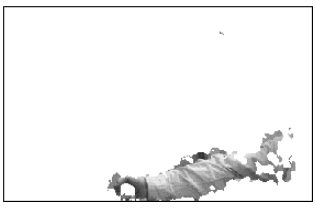
Finally, note that small regions were removed and holes buried in Figs.3(c) and (d) and Figs.4(b) and (c).

6. Conclusions

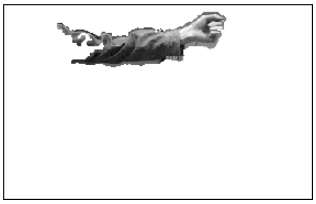
We have proposed a method to extract human limb regions by the direct method which uses two frames directly with an initial value that is the result of the indirect method based on optical flow. At first, rotating limb regions which have rough boundaries are extracted and motion parameters of the approximated model are estimated. Then the extracted region and estimated parameters are used as initial values of nonlinear optimization that minimizes residuals of two successive frames, and estimates motion parameters. According to estimated parameters, the limb region is extracted by Bayesian classifier to obtain accurate boundary of the region. Experimental results on real images show that the result of the direct method is better than that of the indirect method from the viewpoint of dealing with the motion edge. However, there is still difficulty in determining whether an area is in motion or not when the area has flat intensity. We will try to handle this problem by applying a shape model of an arm or using results of several frames.

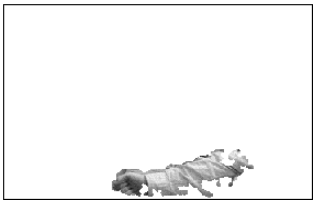
(a) Original image





(b) Extraction of arm 1 (c) Extraction of arm 2
Result of indirect method





(d) Extraction of arm 1 (e) Extraction of arm 2
Result of direct method

Fig. 5. Experiment where there are two moving arms.

7. References

[1] L. Goncalves, E. D. Bernardo, E. Ursella, and P. Perona, "Monocular tracking of the human arm in 3D," in *Proc. of ICCV'95*, 1995, pp. 764–770.

[2] M. Yamamoto, A. Sato, S. Kawada, T. Kondo, and Y. Osaki, "Incremental tracking of human actions from multiple views," in *Proc. of CVPR'98*, 1998, pp. 2–7.

[3] D. M. Gavrila and L. S. Davis, "3-D model-based tracking of humans in action: a multi-view approach," in *Proc. of ICPR'96*, 1996, pp. 73–80.

[4] Toru Tamaki, Tsuyoshi Yamamura, and Noboru Ohnishi, "Extraction of human limb regions and parameter estimation based on curl of optical flow," in *Proc. of ACCV2000*, 2000, vol.2, pp. 1008–1013.

[5] R. Szeliski, "Video mosaics for virtual environment," *IEEE Computer Graphics and Applications*, vol.16, no.3, pp. 22–30, 1996.

[6] H. S. Sawhney and S. Ayer, "Compact representations of videos through dominant and multiple motion estimation," *IEEE Trans. on PAMI*, vol.18, no.8, pp. 814–830, 1996.

[7] G. A. Seber and C. J. Wild, *Nonlinear Regression*, Wiley, 1989.

[8] N. Ohta, "`http://www.ail.cs.gunma-u.ac.jp/Labo/Program/Flow-R00.tar.gz`, 1996.

[9] N. Ohta, "Image movement detection with reliability indices," *IEICE Transactions*, vol.E74, no.10, pp. 3379–3388, 1991.

[10] A.P.Dempster, N.M.Laird, D.B.Rubin : "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Statist. Soc. B*, Vol.39, pp.1–38, 1997.

[11] W.H.Press, S.A.Teukolsky, W.T.Vetterling, B.P. Flannery : *Numerical recipes in C*, Cambridge University Press, 1992.