Paper:

Human Limb Extraction Based on Motion Estimation Using Optical Flow and Image Registration

Toru Tamaki

Graduate Schoold of Science and Technology, Niigata University
8050 Ikarashi2nocho, Niigata City, Niigata Pref. 950-2181, Japan
E-mail: tamaki@ie.niigata-u.ac.jp

[Received August 8, 2003; accepted December 1, 2003]

We propose a method for extracting human limb regions by the 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 cost and avoids the initial state problem of optimization. According to estimated parameters, the limb region is extracted by a Bayesian classifier to obtain accurate region boundaries. Experimental results on real images are shown.

Keywords: extraction, motion estimation, optical flow, EM algorithm, gesture recognition

1. Introduction

It is important to extract human regions from a movie as a part of a human activity recognition system, such as 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] because it is a crucial cue to know where a subject is and what he/she acts, especially when recognizing gestures in which arm movements mainly determine meaning.

In the last decade, many human activity recognition studies often used parameterized volumetric human body models to reconstruct actual human posture [2, 3]. However, a common problem is that these methods require that a background is known or at least is uniform in color to make subtraction easy; otherwise there must be no moving object except the subject. The assumptions about background are hurdles to developing methods so that a recognition system adapts to various real environments.

To overcome these problems, we have proposed a method [4] for finding and recognizing human arms in a general scene. The method extracts regions of rotating human limbs represented by a stick model and estimates their motion parameters. The extraction method proposed is an *indirect* method; that is, with optical flow of two successive images calculated in advance, it segments an image into several moving regions and estimates motion parameters of each region. It can extract arm regions from optical flow of a real image sequence contaminated by much noise. It is essentially impossible, however, to compute optical flow where the motion correspondence cannot be found, especially at the edge of motion, and the indirect method would fail to extract the exact arm region boundary.

In contrast, a method of motion segmentation by comparing intensities of two successive frames directly [5, 6] has been proposed. This *direct* method can deal with the motion discontinuity at motion edges and estimate more precise motion parameters because it does not use optical flow that causes failure with the indirect method. A 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 certain initial parameters that 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. First, limb regions are extracted by the indirect method using optical flow and motion parameters are estimated. Then, accurate boundaries of regions are obtained by the direct method based on the extracted region and estimates produced by the indirect method; the results of the indirect method are used as initial values for the direct method. This combination is expected to decrease computational cost and improve extraction and estimation results 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. Then an extraction with MAP estimation is discussed in section 4. Finally, we provide experimental results of real images in section 5, and discuss the performance of the proposed method in section 6.

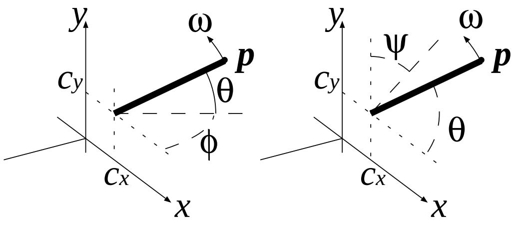

Fig. 1. A stick model moving on a plane rotated by ϕ about the y axis (left) and by ψ about the x axis (right).

2. Indirect Method Using Optical Flow

2.1. Motion Model

In this section, we describe the indirect method of extracting limb regions using optical flow. To achieve rough extraction and estimation, we use a rather simplified motion model because optical flows in real images usually have so much noise that parameters of a precise motion model can not be recovered. Therefore, by using the simple model, we divide an image into regions with coherent movement and determine which region moves or rotates.

Here a limb is assumed to rotate on a plane that is not parallel to the image plane. As shown in **Fig.1**, there are two cases. In general, a rotation in three dimensions is represented as a revolution about an axis, however, both cases can be used as an approximated 3D motion model. Then the limb is projected onto the image plane by orthographic projection. In both cases, a motion of point $\mathbf{p}_j = (x_j, y_j)$ on a rotating limb and its velocity $\dot{\mathbf{p}}_j = (u_j, v_j)$ in the image plane are modeled as follows [4],

$$\dot{\mathbf{p}}_j = A_j \mathbf{q}, \quad \ldots \ldots \ldots \quad (1)$$

where

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

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

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

Similarly, ϕ and ψ are retrieved from $\mathbf{q}$. But it is important that we discriminate rotating regions by ω; a rotating arm region has a large angular velocity, while in a moving region but without rotation ω becomes quite small.

2.2. Motion Clustering

Although motion of a point on a limb is modeled as explained above, optical flow computed from real images involves inevitable noise. Therefore, we find groups of flows corresponding to each moving region by clustering optical flow.

We assume that the distribution of $\dot{\mathbf{p}}_j$ is subject to a two-dimensional Gaussian,

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

where $\Sigma = \left(\begin{smallmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{smallmatrix}\right)$ is a covariance matrix that assumes that the errors for u and v are mutually independent because α, β and γ, δ are estimated separately.

Then, posterior probability of the optical flow is modeled as a mixture density comprised of M densities of clusters R_i,

$$p(\dot{\mathbf{p}}_j | \mathbf{p}_j) = \sum_i^M \xi_i \, p(\dot{\mathbf{p}}_j | \mathbf{p}_j, \mathbf{q}_i, \Sigma_i), \quad \ldots \ldots \quad (5)$$

where ξ_i are weights for each density.

The problem becomes an estimation of $\mathbf{q}_i$ for each cluster R_i and a segmentation of points $\mathbf{p}$ based on the weights of densities [4]. Assuming that each moving object has its own motion parameter $\mathbf{q}$ and that Eq.(4) is the distribution of optical flow within the object region, the following algorithm uses the EM algorithm[7] to perform estimation and segmentation:

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 the optical flow based on the direction of velocity to obtain initial clusters R_i $(i = 1, \ldots, M)$ (see section 5).
   Then, set initial values of weights w_{ij} as probabilities that point $\mathbf{p}_j$ belongs to cluster R_i,
   $$w_{ij} = \begin{cases} 1 & (\mathbf{p}_j \in R_i) \\ 0 & (\mathbf{p}_j \notin R_i) \end{cases} \quad \ldots \ldots \ldots \quad (6)$$
2. Normalize weights w_{ij};
   $$w'_{ij} = \frac{w_{ij}}{\xi_i}, \quad \ldots \ldots \ldots \quad (7)$$
   where $\xi_i = \frac{1}{N}\sum_j^N w_{ij}$.
3. Find parameters $\mathbf{q}_i = (\alpha_i, \beta_i, \gamma_i, \delta_i)$ of each cluster R_i by solving the following overdetermined system of equations:
   $$\begin{pmatrix} \sqrt{w'_{i1}}\,\dot{\mathbf{p}}_1 \\ \sqrt{w'_{i2}}\,\dot{\mathbf{p}}_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \sqrt{w'_{i1}}A_1 \\ \sqrt{w'_{i2}}A_2 \\ \vdots \end{pmatrix} \mathbf{q}_i, \quad \ldots \quad (8)$$
   with QR decomposition [9].
   The solution $\mathbf{q}_i$ is the weighted mean of the motion parameters for cluster R_i.