Recognizing Functionality

Z. Duric

Computer Vision Laboratory
Center For Automation Research
University of Maryland
College Park, MD 20742-3275

J. Fayman, E. Rivlin

Department of Computer Science
Israel Institute of Technology – Technion
Haifa, Israel

Abstract

In this paper, we look into a new area of function recognition: determining the function of an object from its motion. Given a sequence of images of a known object performing some function, we attempt to determine what that function is. We show that the motion of an object, when combined with information about the object and its normal uses, provides us with strong constraints on possible functions that the object might be performing.

1 Introduction

In the field of robotics, researchers have long pursued the goal of enabling a robot to act autonomously in its environment. For robots, as for humans, recognizing the functions of objects is a prerequisite to autonomous interaction with them. Functionality can be defined as the usability of an object for a particular purpose [1].

Recent research has focused on the problem of recognizing object functionality [1]. The goal of this research has been to determine functional capabilities of an object based on characteristics such as shape, physics and causation [2]. Little attention has been given to the problem of determining the functionality of an object from its motion. We believe that motion provides a strong indication of function. In particular, velocity, acceleration, and force of impact resulting from motion strongly constrain possible function. As in other approaches to functional recognition, the object (and in our case, its motion) should not be evaluated in isolation, but in context. The context includes the nature of the agent and the frame of reference it uses.

In this paper, we address the following problem: given a model of an object, how can we use the motion of the object, while it is being used to perform a task, to determine its function? Our method of answering this question takes into consideration the angular relationships between three vectors obtained from image sequence analysis. These vectors are compared with angular relationships that arise in known motion-to-function mappings. If the analyzed motion is consistent with one of the known motion-to-function mappings, we identify object functionality.

In Section 2 we review related work. In Section 3 we cover some preliminaries related to the problem. Section 4 considers the problem of determining the functionality of a known object by analyzing an image sequence showing that object performing the function. The motion estimation machinery needed for this task is developed in Section 5. In Section 6 we present experimental results demonstrating that motion analysis can indeed be used in determining functionality. We conclude with section 7.

2 Related Work

Our research is concerned with the problem of determining the function of an object by analyzing its motion. Motion and functionality have appeared in the literature in several contexts. Early work on functional recognition can be found in [3, 4, 5]. More recently, Stark and Bowyer [6, 7, 2, 8] used these ideas to solve some of the problems presented by more traditional model-based methods of object recognition.

By analyzing the trajectories followed by points on an object, Gould and Shah [9] attempt to identify the object. This is accomplished by recognizing "significant" events in the trajectory such as changes in direction, speed and acceleration.

Motion analysis for recognition of activities was described by Polana and Nelson [10]. They use Fourier analysis to detect and localize periodic activities such as walking or flying in a sequence of images. This work is similar in nature to our work in that both use motion as a basis for identifying activities. However, Polana and Nelson are concerned only with detecting the activities, without concern for the source of the motion.

3 Preliminaries

3.1 Rigid Body Motion

To facilitate the derivation of the motion equations of a rigid body $\mathcal{B}$ we use two rectangular coordinate frames, one $(Oxyz)$ fixed in space, the other $(Cx_1y_1z_1)$ fixed in the body and moving with it. The coordinates X_1, Y_1, Z_1 of any point P of the body with respect to the moving frame are constant with respect to time t, while the coordinates X, Y, Z of the same point P with respect to the fixed frame are functions of t. It is assumed that these functions are differentiable with respect to t. The position of the moving frame at any

0-8186-7190-4/95 $4.00 © 1995 IEEE

instant is given by the position $\vec{d}_c = (X_c \; Y_c \; Z_c)^T$ of the origin C, and by the nine direction cosines of the axes of the moving frame with respect to the fixed frame. Let $\vec{\imath}$, $\vec{\jmath}$, and $\vec{k}$ be the unit vectors in the directions of the Ox, Oy, and Oz axes, respectively; and let $\vec{\imath}_1$, $\vec{\jmath}_1$, and $\vec{k}_1$ be the unit vectors in the directions of the Cx_1, Cy_1, and Cz_1 axes, respectively. For a given position $\vec{p}$ of P in $Cx_1y_1z_1$ we have the position $\vec{r}_p$ of P in $Oxyz$

$$\vec{r}_p \equiv \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \vec{\imath}\cdot\vec{\imath}_1 & \vec{\imath}\cdot\vec{\jmath}_1 & \vec{\imath}\cdot\vec{k}_1 \\ \vec{\jmath}\cdot\vec{\imath}_1 & \vec{\jmath}\cdot\vec{\jmath}_1 & \vec{\jmath}\cdot\vec{k}_1 \\ \vec{k}\cdot\vec{\imath}_1 & \vec{k}\cdot\vec{\jmath}_1 & \vec{k}\cdot\vec{k}_1 \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} + \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \equiv R\vec{p} + \vec{d}_c$$

where R is the matrix of the direction cosines (the frames are taken as right-handed so that $\det R = 1$). The velocity of $\vec{r}_p$ is then given by

$$\dot{\vec{r}}_p = \vec{\omega} \times (\vec{r}_p - \vec{d}_c) + \vec{T}$$

where $\vec{\omega} = (A \; B \; C)^T$ is the rotational velocity of the moving frame; $\dot{\vec{d}}_c = (\dot{X}_c \; \dot{Y}_c \; \dot{Z}_c)^T \equiv (U \; V \; W)^T \equiv \vec{T}$ is the translational velocity of the point C. This can be written as

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{pmatrix} = \begin{pmatrix} 0 & -C & B \\ C & 0 & -A \\ -B & A & 0 \end{pmatrix} \begin{pmatrix} X - X_c \\ Y - Y_c \\ Z - Z_c \end{pmatrix} + \begin{pmatrix} U \\ V \\ W \end{pmatrix}. \quad (2)$$

Let the rotational velocity in the moving frame be $\vec{\omega}_1 = (A_1 \quad B_1 \quad C_1)^T$; we can write $\vec{\omega} = R\vec{\omega}_1$ and $\vec{\omega}_1 = R^T\vec{\omega}$.

3.2 The Motion and Optical Flow Fields

The instantaneous velocity of the image point (x, y) under weak perspective projection can be obtained by taking derivatives of (3) with respect to time and using (2):

$$x = \frac{X}{Z_c}f, \quad y = \frac{Y}{Z_c}f. \quad (3)$$

$$\begin{aligned} \dot{x} &= f\frac{\dot{X}Z_c - X\dot{Z}_c}{Z_c^2} = f\frac{[-C(Y - Y_c) + B(Z - Z_c) + U]Z_c - XW}{Z_c^2} \\ &= \frac{Uf - xW}{Z_c} - C(y - y_c) + f\,B\left(\frac{Z}{Z_c} - 1\right), \quad (4) \\ \dot{y} &= f\frac{\dot{Y}Z_c - Y\dot{Z}_c}{Z_c^2} = f\frac{[C(X - X_c) - A(Z - Z_c) + V]Z_c - YW}{Z_c^2} \\ &= \frac{Vf - yW}{Z_c} + C(x - x_c) - f\,A\left(\frac{Z}{Z_c} - 1\right) \quad (5) \end{aligned}$$

where $(x_c, y_c) = (fX_c/Z_c, fY_c/Z_c)$ is the image of the point C. Let $\vec{\imath}$ and $\vec{\jmath}$ be the unit vectors in the x and y directions, respectively; $\dot{\vec{r}} = \dot{x}\vec{\imath} + \dot{y}\vec{\jmath}$ is the projected motion field at the point $\vec{r} = x\vec{\imath} + y\vec{\jmath}$.

If we choose a unit direction vector $\vec{n}_r$ in the image point $\vec{r}$ and call it the normal direction, then the *normal motion field* at $\vec{r}$ is $\dot{\vec{r}}_n = (\dot{\vec{r}} \cdot \vec{n}_r)\vec{n}_r$. $\vec{n}_r$ can be chosen in various ways; the usual choice (as we shall now see) is the direction of the image intensity gradient.

Let $I(x, y, t)$ be the image intensity function. The time derivative of I can be written as

$$\frac{dI}{dt} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = (I_x\vec{\imath} + I_y\vec{\jmath})\cdot(\dot{x}\vec{\imath} + \dot{y}\vec{\jmath}) + I_t =$$

$$\nabla I \cdot \dot{\vec{r}} + I_t$$

where ∇I is the image gradient and the subscripts denote partial derivatives.

If we assume $dI/dt = 0$, i.e. that the image intensity does not vary with time [13], then we have $\nabla I \cdot \vec{u} + I_t = 0$. The vector field $\vec{u}$ in this expression is called the *optical flow*. If we choose the normal direction $\vec{n}_r$ to be the image gradient direction, i.e. $\vec{n}_r \equiv \nabla I/\|\nabla I\|$, we then have

$$\vec{u}_n = (\vec{u} \cdot \vec{n}_r)\vec{n}_r = \frac{-I_t \nabla I}{\|\nabla I\|^2} \quad (6)$$

where $\vec{u}_n$ is called the *normal flow*.

It was shown in [14] that the magnitude of the difference between $\vec{u}_n$ and the normal motion field $\dot{\vec{r}}_n$ is inversely proportional to the magnitude of the image gradient. Hence $\dot{\vec{r}}_n \approx \vec{u}_n$ when $\|\nabla I\|$ is large. Equation (6) thus provides an approximate relationship between the 3-D motion and the image derivatives. We will use this approximation later in this paper.

4 Function from Motion

Following [11, 12] we regard objects as composed of primitive parts. On the most coarse level we consider four types of primitive parts: sticks, strips, plates, and blobs, which differ in the values of their relative dimensions. Basic or primitive motions (which can be rotational or translational) are motions relative to the main axes of a primitive object.

In this work, we are interested in the mapping $f : M \mapsto F$ from motion to function. Given a moving object as seen by an observer we would like to infer the function being performed by the acting agent. We are interested in the object's motion over time in the object's coordinate system and its relation to the object it acts on (the actee). Both of these measurements are necessary for the mapping. The object's motion over time in the object coordinate system gives us the relationship between the main axis of the object and its direction of motion. Given an object, these relationships help to determine the intended function. For example, we would expect the motion of a knife that a person is using to "cut" to be parallel to the main axis of the knife, whereas if the person is "chopping" with the knife we would expect motion perpendicular to the main axis.

When determining function from motion, attention must be paid to the intended recipient. The relation to the actee is essential for establishing the mapping and creating a frame of reference. Once this frame is established, motion of a knife in one direction could signify murder while motion in the opposite direction could signify suicide. Humans usually employ reference frames in which one axis represents the gravity vector, but this is not necessary. We can slice bread on a wall as well as on a table; what matters is the motion of the knife relative to the actee.

In the next section, we develop the motion estimation machinery needed for this class of examples and we formalize our procedure for obtaining $f : M \mapsto F$. In the following section, we present experimental results.

5 Motion of Sticks and Strips

Consider a moving object $\mathcal{B}$. There is an *ellipsoid of inertia* associated with $\mathcal{B}$. The center of the ellipsoid is at the center of mass C of $\mathcal{B}$; the axes of the ellipsoid are called the *principal axes*. We associate the coordinate system $Cx_1y_1z_1$ with the ellipsoid and choose the axes of $Cx_1y_1z_1$ to be parallel to the principal axes. Let $\vec{\imath}_1$ be the unit vector in the direction of the longest axis l_c (this axis corresponds to the smallest principal moment of inertia); let $\vec{k}_1$ be the unit vector in the direction of the shortest principal axis (this axis corresponds to the largest moment of inertia); and let $\vec{\jmath}_1$ be the unit vector in the direction of the remaining principal axis with the direction chosen so that the vectors $(\vec{\imath}_1, \vec{\jmath}_1, \vec{k}_1)$ form a right-handed coordinate system.

In this paper we consider only planar and approximately straight strips and sticks. For a planar strip the axis of the maximal moment of inertia is orthogonal to the plane of the strip; if the strip is approximately straight, the axis of the minimal moment of inertia is approximately parallel to the medial axis l_c of the strip. In the case of a straight stick we have the center of mass C at the middle of its medial axis l_c: in this case l_c corresponds to the longest principal axis of the ellipsoid of inertia; the other two principal axes are orthogonal to l_c and can be chosen arbitrarily. We assume that there is no rotational velocity around l_c.

We choose the center of mass C of a stick or a strip $\mathcal{B}$ as the origin of the object coordinate system $Cx_1y_1z_1$; the coordinates of C expressed in the fixed frame are (X_c, Y_c, Z_c). We choose the unit vector $\vec{\imath}_1$ along l_c with the orientation chosen to be in the direction of the acting part of the tool. Let Π_{l_c} be the plane orthogonal to the plane $Z = Z_c$ in which the line l_c lies (we can obtain Π_{l_c} by sliding the line parallel to the $\vec{k}$ along l_c). We chose $\vec{k}_1$ to lie in the plane Π_{l_c} with the orientation of $\vec{k}_1$ chosen so that $\vec{k} \cdot \vec{k}_1 \geq 0$; the unit vector $\vec{\jmath}_1$ is then normal to the Π_{l_c} plane. We assume that strips are orthogonal to the Π_{l_c} plane.

The orthographic image of l_c in the plane $Z = Z_c$ is the line l'_c which is the intersection of the planes $Z = Z_c$ and Π_{l_c}; let the unit vector in the direction of l'_c be $\vec{\imath}_c$ and let it be oriented so that $\vec{\imath}_c \cdot \vec{\imath}_1 \geq 0$; and let the angle between l_c and l'_c be φ. The rotation $R_{Y_c}(\varphi)$ through the angle φ around the normal $\vec{\jmath}_1$ of Π_{l_c} transforms $\vec{\imath}_1$ into $\vec{\imath}_c$ and $\vec{k}_1$ into $\vec{k}$. The rotation $R_Z(\alpha)$ through the angle α (this is the angle between $\vec{\imath}_c$ and $\vec{\imath}$) around the Oz axis transforms $\vec{\imath}_c$ into $\vec{\imath}$. The rotation matrix $R = R_Z(\alpha)R_{Y_c}(\varphi)$ in (2) is then given by

$$R = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix}.$$

By our assumption about the rotational velocity and the choice of the object coordinate system we have $\vec{\omega}_1 = B_1\vec{\jmath}_1 + C_1\vec{k}_1$. The expression for the rotational velocity in the fixed frame is given by

$$\vec{\omega} = \begin{pmatrix} A \\ B \\ C \end{pmatrix} = R(B_1\vec{\jmath}_1 + C_1\vec{k}_1) = R_Z(\alpha) \begin{pmatrix} C_1 \sin\varphi \\ B_1 \\ C_1 \cos\varphi \end{pmatrix}. \tag{7}$$

Similarly, since the translational velocity of the object is $\vec{T}_1 = (U_1 \quad V_1 \quad W_1)^T$ and $\vec{T} = R_Z(\alpha)R_{Y_1}(\varphi)\vec{T}_1$ we have

$$\vec{T} = \begin{pmatrix} U \\ V \\ W \end{pmatrix} = R_Z(\alpha) \begin{pmatrix} U_1 \cos\varphi + W_1 \sin\varphi \\ V_1 \\ -U_1 \sin\varphi + W_1 \cos\varphi \end{pmatrix}. \tag{8}$$

We now consider the term $(Z - Z_c)/Z_c$ for the points on the object $\mathcal{B}$. The equations we derive are valid for points in the plane in which l_c lies and is orthogonal to Π_{l_c}; the unit vector $\vec{k}_1$ is normal to this plane. The expression for $\vec{k}_1$ in the fixed frame is $R\vec{k}_1 = (\cos\alpha\sin\varphi \quad -\sin\alpha\sin\varphi \quad \cos\varphi)^T$. The equation of the plane orthogonal to $R\vec{k}_1$ and in which the point (X_c, Y_c, Z_c) lies is given by

$$(X - X_c)\cos\alpha\sin\varphi - (Y - Y_c)\sin\alpha\sin\varphi + (Z - Z_c)\cos\varphi = 0.$$

Multiplying by $f(Z_c\cos\varphi)^{-1}$ and using (3) we obtain

$$f\frac{Z - Z_c}{Z_c} = -(x - x_c)\cos\alpha\tan\varphi + (y - y_c)\sin\alpha\tan\varphi. \tag{9}$$

This is an exact formula for thin planar strips; in the case of sticks this formula is exact for an occluding contour.

From (4–5), (7), and (9) we obtain the equations of projected motion for points on $\mathcal{B}$ under weak perspective:

$$\dot{x} = \frac{Uf - xW}{Z_c} - C(y - y_c) \\ -B\tan\varphi\,[(x - x_c)\cos\alpha - (y - y_c)\sin\alpha]. \tag{10}$$

$$\dot{y} = \frac{Vf - yW}{Z_c} + C(x - x_c) \\ +A\tan\varphi\,[(x - x_c)\cos\alpha - (y - y_c)\sin\alpha]. \tag{11}$$

Equations (10-11) relate the projected motion field, α, and (x_c, y_c) to the scaled translational velocity $Z_c^{-1}\vec{T} = Z_c^{-1}(U\ \ V\ \ W)^T$ and the three parameters of rotation and slant $(A\tan\varphi, B\tan\varphi, C)$.

Now, from (8) we have

$$Z_c^{-1}\begin{pmatrix} U_1\cos\varphi + W_1\sin\varphi \\ V_1 \\ -U_1\sin\varphi + W_1\cos\varphi \end{pmatrix} = \tag{12}$$

$$R_Z^T(\alpha)\begin{pmatrix} U/Z_c \\ V/Z_c \\ W/Z_c \end{pmatrix} \equiv \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

and by rearrangement we obtain

$$\frac{V_1}{Z_c} = c_2, \quad \begin{pmatrix} U_1/Z_c \\ W_1/Z_c \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}\begin{pmatrix} c_1 \\ c_3 \end{pmatrix} \tag{13}$$

When $Z_c^{-1}\vec{T} = Z_c^{-1}(U\ \ V\ \ W)^T$ and α are known c_1, c_2, and c_3 are computable. From (13) we then can compute V_1/Z_c. However, φ and $(U_1/Z_c, W_1/Z_c)$ cannot be estimated without some additional assumptions. If we assume a fronto-parallel surface, i.e., $\varphi = 0$, we obtain $U_1/Z_c = c_1$ and $W_1/Z_c = c_3$; similarly, a bound on φ (e.g. $\varphi < \pi/6$) gives bounds on the motion parameters too. Finally, if we assume that $W_1 = 0$ we obtain $\varphi = -\arctan(c_3/c_1)$ and then from (13) we have U_1/Z_c; similarly, if we assume that $U_1 = 0$ we obtain $\varphi = \arctan(c_1/c_3)$ and then from (13) we have W_1/Z_c.

From (7) we have

$$R_Z(\alpha)\begin{pmatrix} C_1\tan\varphi\sin\varphi \\ B_1\tan\varphi \\ C_1\cos\varphi \end{pmatrix} = \tag{14}$$

$$I_\varphi R_Z(\alpha)\begin{pmatrix} C_1\sin\varphi \\ B_1 \\ C_1\cos\varphi \end{pmatrix} = \begin{pmatrix} A\tan\varphi \\ B\tan\varphi \\ C \end{pmatrix}$$

where I_φ is a diagonal matrix with the first two diagonal elements $\tan\varphi$ and the third diagonal element 1; we have used the fact that I_φ and $R_Z(\alpha)$ are commutative matrices. From (14) we have

$$\begin{pmatrix} C_1\tan\varphi\sin\varphi \\ B_1\tan\varphi \\ C_1\cos\varphi \end{pmatrix} = R_Z^T(\alpha)\begin{pmatrix} A\tan\varphi \\ B\tan\varphi \\ C \end{pmatrix} = \begin{pmatrix} c_4 \\ c_5 \\ c_6 \end{pmatrix}. \tag{15}$$

When $(A\tan\varphi\ \ B\tan\varphi\ \ C)^T$ and α are known c_4, c_5, and c_6 are computable. Since $\varphi \in [0, \pi/2)$ both c_4 and c_6 should have the same sign, otherwise we can assume that $\varphi = 0$. If $\operatorname{sgn} c_4 = \operatorname{sgn} c_6$ we have

$$\begin{aligned} C_1 &= \operatorname{sgn}(c_6)\sqrt{c_4 c_6 + c_6^2}, \\ \varphi &= \pm\arctan\sqrt{c_4/c_6}, \\ B_1 &= \pm\frac{c_5}{\sqrt{c_4/c_6}}. \end{aligned} \tag{16}$$

where sgn is the sign function. Note that if $\operatorname{sgn} c_5 = 1$ the signs of B_1 and φ must be equal, otherwise they must be different.

If translation is non-zero we can combine (13) and (16) to estimate φ and the motion parameters. When $U_1 \equiv 0$ we have $\tan\varphi = c_1/c_3$; when $W_1 \equiv 0$ we have $\tan\varphi = -c_3/c_1$. In both of these cases B_1 can be determined directly, as well as φ and C_1. If $\varphi \approx 0$ (or $\varphi < 0.1$) then $C_1 \approx c_6$, and $|c_5| = |B_1|\tan\varphi \approx |B_1|\varphi < 0.1|B_1|$ and thus $|B_1| > 10|c_5|$. If (16) is used to compute φ we have two solutions for the $(U_1/Z_c, W_1/Z_c)$ pair.

6 Experiments

In this section we illustrate how our methods can be applied to real images. Due to space limitations, only two experiments can be shown.

In our experiments we assumed a table-top scenario, with a stationary observer on one side of the table. Based on this assumption we used a coordinate system that was fixed to the center of the image, with the X axis horizontal and pointing toward the right side of the image, the Y axis pointing upward, and the Z axis chosen to yield a right-handed coordinate frame (pointing toward the scene). All measurements were made relative to this coordinate system.

Estimation of the medial axis of the object was done by taking the median of all edge orientations at those points for which the normal flow was computed. We estimated (x_c, y_c)—the image position of C (the reference point and the center of mass of the object)—as the average of the coordinates of all edge points for which the normal flow was computed.

In the following subsections we describe our method of motion estimation for sticks and strips. This motion estimation is then used to discriminate between two different functionalities of a knife (chopping and stabbing).

6.1 Motion Estimation from Normal Flow

In what follows we show how the different motion parameters defined in Section 5 can be estimated based on normal flow data computed from an image sequence.

Let $g_1(x,y) = (x - x_c)\cos\alpha - (y - y_c)\sin\alpha$. We can then define the vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{d}}$:

$$\vec{\mathbf{a}} = \begin{pmatrix} f\,n_x \\ f\,n_y \\ -x\,n_x - y\,n_y \\ n_y\,g_1(x,y) \\ -n_x\,g_1(x,y) \\ -n_x(y - y_c) + n_y(x - x_c) \end{pmatrix}, \quad \vec{\mathbf{d}} = \begin{pmatrix} U/Z_c \\ V/Z_c \\ W/Z_c \\ A\tan\varphi \\ B\tan\varphi \\ C \end{pmatrix}$$

For a given $\vec{n}_r = n_x\vec{\imath} + n_y\vec{\jmath}$ we then have from (10-11)

$$\dot{x}\,n_x + \dot{y}\,n_y = \vec{\mathbf{a}}\cdot\vec{\mathbf{d}}. \tag{17}$$

If we use the spatial image gradient as the normal direction $\vec{n}_r \equiv \nabla I/\|\nabla I\| = n_x\vec{\imath} + n_y\vec{\jmath}$ and $\dot{\vec{r}}_n \approx \vec{u}_n$ we can obtain an approximate equation by replacing the left hand side of (17) by the normal flow $-I_t/\|\nabla I\|$. In this way we obtain one approximate equation in the six unknown elements of $\vec{d}$. For each point (x_i, y_i), $i = 1, \ldots, N$ of the image at which $\|\nabla I(x_i, y_i, t)\|$ is large we can write one equation. If we have more than six points we have an over-determined system of equations $A\vec{d} = \vec{b}$; the rows of the $N \times 6$ matrix A are the vectors $\vec{a}_i$. and the elements of the N-vector $\vec{b}$ are $-(\partial I(x_i, y_i, t)/\partial t)/\|\nabla I(x_i, y_i, t)\|$.

We seek the solution for which $\|\vec{b} - A\vec{d}\|$ is minimal. This solution is the same as the solution of the system

$$A^T A\vec{d} = A^T\vec{b} \equiv \vec{e}.$$

We solve the system $A^T A\vec{d} = \vec{e}$ using the Cholesky decomposition. Since the matrix $A^T A$ is a positive definite 6×6 matrix there exists a lower triangular matrix L such that $L L^T = A^T A$. We solve two triangular systems $L\vec{f} = \vec{d}$ and $L^T\vec{d} = \vec{f}$ to obtain the parameter vector $\vec{d}$.

In the case when $\varphi \approx 0$ (fronto-parallel case) and the rotation B_1 around the Cy_1 axis is small the equations (10-11) become

$$\begin{aligned} \dot{x} &\approx \frac{Uf - xW}{Z_c} - C(y - y_c), \\ \dot{y} &\approx \frac{Vf - yW}{Z_c} + C(x - x_c). \end{aligned}$$

In this case we need to estimate only four parameters U/Z_c, V/Z_c, W/Z_c, and C in the parameter vector $\vec{d}$; thus, $A^T A$ is a 4×4 matrix. We then have from(14-16) that $C_1 \approx C$ and from (12-13) we have

$$\begin{pmatrix} U_1/Z_c \\ V_1/Z_c \end{pmatrix} \approx \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} U/Z_c \\ V/Z_c \end{pmatrix},$$

$$\frac{W_1}{Z_c} = \frac{W}{Z_c}. \tag{18}$$

For the following experiments we use these approximations to compute the object motion from images.

Let β be the angle between the vector $(U_1 \ \ V_1 \ \ 0)^T$ (the projection of $\vec{T}_1$ onto the plane $Z = Z_c$) and $\vec{\imath}_c$ (the unit vector along the projection of the medial axis l_c onto the plane $Z = Z_c$). We have

$$\beta = \arctan\frac{V_1}{U_1}. \tag{19}$$

Let θ be the total rotation angle as a function of time. For a fronto-parallel surface the total rotation angle is approximately equal to the change in α and we have

$$\theta = \int_0^t C_1 dt \approx \alpha - \alpha_0. \tag{20}$$

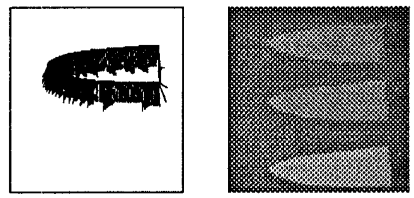
We use the triples (α, β, θ) to recognize the functionalities of simple objects.

6.2 Action recognition for a class of manipulation tasks: Cutting

Our example shows how our techniques can be used to differentiate between two examples of simple functions performed by knifes: chopping and stabbing.

6.2.1 Chopping

Chopping is defined as the cutting motion of a knife in which α (the angle between the projection of l_c onto the plane $Z = Z_c$ and the Ox axis) is close to either 0 or π, β is close to $\pi/2$ ($\alpha \approx \pi$) or $-\pi/2$ (when $\alpha \approx 0$), and θ is small and approximately constant.



(a) (b)

Figure 1: (a) Flow vectors for Chopping. (b) Chopping motion.

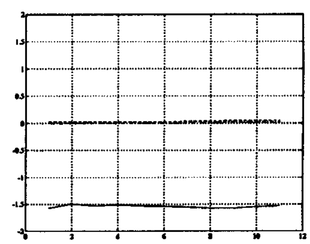


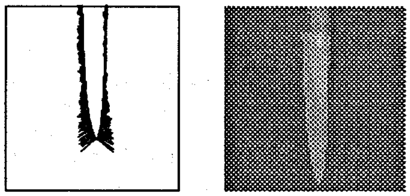
Figure 2: Angles α, β, and θ for Chopping. α is given by a dashed line, β is given by a solid line, and θ is given by a dash-dot line.

Figure 1(a) shows the flow vectors taken from the 6th sample and (b), a composite image of the knife taken from the 1st, 6th and 11th samples of the chopping experiment. Figure 2 shows a plot of the triple (α, β, θ) with respect to time (frame numbers). We can see that the values of α are very close to 0, as was expected, β is close to $-\pi/2$ and θ is around 0.

6.2.2 Stabbing

Stabbing is defined as the cutting motion of a knife in which α (the angle between the projection of l_c onto the plane $Z = Z_c$ and the Ox axis) is close to either $-\pi/2$ or $\pi/2$, β is approximately 0, and θ is small and approximately constant. The difference between jabbing and stabbing is in α.

Figure 3(a), shows the flow vectors taken from the 6th sample and (b), a composite image of the knife taken from the 1st, 6th and 11th samples of the stabbing experiment. Figure 4 shows a plot of the triple (α, β, θ) with respect to time (frame numbers). We can see that the values of α are very close to $-\pi/2$, as was expected, β is close to 0 and θ is around 0.



(a) (b)

Figure 3: (a) Flow vectors for Stabbing. (b) Stabbing motion.

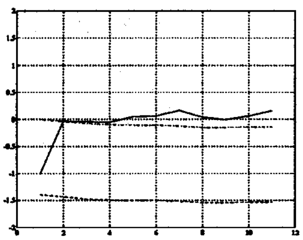


Figure 4: Angles α, β, and θ for Stabbing. α is given by a dashed line, β is given by a solid line, and θ is given by a dash-dot line.

7 Conclusions

Perceiving function from motion provides an understanding of the way an object is being used by an agent. To accomplish this we combined information on the shape of the object, its motion, and its relation to the actee (the object it is acting on). Assuming a decomposition of the object into primitive parts, we analyzed a part's motion relative to its principal axes. Primitive motions (translation and rotation relative to the principal axes of the object) were dominating factors in the analysis. We used a frame of reference relative to the actee. Once such a frame is established, it can have major implications for the functionality of an action; for example, motion of a knife in one direction can signify murder while motion in the opposite direction can signify suicide.

References

[1] L. Bogoni and R. Bajcsy. Active investigation of functionality. In *Workshop: Role of Functionality in Object Recognition, CVPR,*, Seattle, WA, June 1994.

[2] L. Stark and K. Bowyer. Indexing function-based categories for generic recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 795–797, Champaign, IL, June 1992.

[3] P. Freeman and A. Newell. A model for functional reasoning in design. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 621–640, August 1971.

[4] F. Solina and R. Bajcsy. Shape and function. In *Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision*, Volume 726, pages 284–291, 1983.

[5] P.H. Winston, T.O. Binford, B. Katz, and M. Lowry. Learning physical descriptions from functional descriptions, examples, and precedents. In *Proceedings of the National Conference on Artificial Intelligence*, pages 433–439, 1983.

[6] L. Stark and K. Bowyer. Achieving generalized object recognition through reasoning about association of function to structure. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10), pages 1097–1104, October 1991.

[7] L. Stark and K. Bowyer. Generic recognition through qualitative reasoning about 3-d shape and object function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 251–256, Maui, HI, 1991.

[8] L. Stark, A. Hoover, D. Goldgof, and K. Bowyer. Function based recognition from incomplete knowledge of shape. In *Proceedings of the IEEE Workshop on Qualitative Vision*, pages 11–22, New York, NY, 1993.

[9] K. Gould and M. Shah. The trajectory primal sketch: A multi-scale scheme for representing motion characteristics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 79–85, June 1989.

[10] R. Polana and R. Nelson. Detecting activities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2–7, New York, NY, June 1993.

[11] E. Rivlin, S.J. Dickinson, and A. Rosenfeld. Recognition by functional parts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 267–274, Seattle, WA, June 1994.

[12] E. Rivlin, A. Rosenfeld, and D. Perlis. Recognition of object functionality in goal-directed robotics. In *Proceedings of the AAAI Workshop on Reasoning about Function*, 1993.

[13] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:189–203, 1981.

[14] A. Verri and T. Poggio. Against quantitative optical flow. In *Proceedings of the International Conference on Computer Vision*, pages 171–180, London, England, June 1987.