Qualitative motion of human arm from a contour image sequence

M.Terauchi †, *Y.Yasutomi* ‡

† Faculty of Information Science, Hiroshima City University
Ozuka, Numata, Asaminami, Hiroshima, 73131 Japan
email: mucha@its.hiroshima-cu.ac.jp

‡ Faculty of Engineering, Hiroshima University
1-4-1 Kagamiyama, Higashi-Hiroshima, 724 Japan

Abstract

This paper presents an approach to recognize the human motion from a single viewed image sequence in order to realize a new man-computer interface. The human body is assumed to be a linked model of rigid segments which are approximated by ellipsoids with three different length of the diameters. The reconstruction of the human motion is mainly performed by interpretation of the 2D pattern deformation as the 3D motion. We show the analytical solution of the motion utilizing the framework of the visual servoing for the human motion reconstruction.

1 Introduction

In previous computer-work environments between human and computers the transmission of information has been mainly based on text representation on the computer terminals. Recently it has been evolving toward more interactive use of the visual and the sound effects. Yet the transmission is almost uni-directional, as ever, from computers to users, while the input device is restricted to pointing devices such as cursors, mouses and/or tablets. These input devices require contacts with human body or hands in exact manners. So a new input device for computers in the next generation is ought to possess a natural interface closing the gap between human and computers.

One of the most useful information is auditory one, which has been well studied and is about to be applied for real tasks especially in text inputs. Besides the above application, the analyzation of nonverbal information such as music or sound is too difficult to formulate or to be interpreted as people commonly do. In the exchange of information using auditory media, linguistic or numerical data recognition is well studied; for example, template structure representation of limited words or parsing of simple sentences. Hence a reliable data exchange must be based on text input/output using the computer keyboards and displays. Furthermore in text processing there remain a lot of difficult problems in analysis of pronounces, words' semantics in the contexts of sentences.

On the other hand, visual information is the most dominant in human senses. When it is too noisy or not able to understand speakers' language, the use of gestures or body movements is effective for understanding each other. It is called as a body language, and often used in real communication environments. Thus it is strongly desired for a computer-interface to make more humane by understanding visual patterns and body movements appearing in the human conversations.

In order to reconstruct and understand the human posture and motion from sequential visual images, it is necessary (1) to define the model of human shape, (2) to extract valid features from each image, (3) to reconstruct the 3D posture of the human body from them and (4) to give consistent interpretations of the motion using knowledge on the human motion and the link model. There are however a variety of problems in each step, for example, complicatedness of the shape of the human body in step (1), occlusions of segments in (2),(3) and hierarchical representation of motions with valid interpretation in (4).

2 Human motion analysis from an Image Sequence

2.1 Previous Study and Motivation of this Study

The research in the field of recognition of human motion was originally oriented to its application for measurement. The earlier study recognized the existence of many layers of problems from low level processing to high level one. Therefore it has become the standard to take an approach to divide the whole processes into some modules and study each in detail [1]. In general the process of motion recognition from an image sequence can be roughly divided into the following two stages,

1) detection of effective information from each image
2) linkage of information obtained from each of consecutive images.

In the first stage, it is necessary to obtain information on the spatial coordinate points on images which amount properly to feature points on a real human body. In order to extract the motion, temporal information incorporated by finding interrelationships among feature points on consecutive images as to be analyzed in the next stage.

It is, however, necessary to give a preliminary definition and a basic strategy before considering detail processes in the aforementioned stages. The following three matters must be deliberated from view points of human motion reconstruction,

i) modelling the human body,
ii) correspondence of feature points,
iii) analysis of occluding regions.

Modelling the human body provides us a methodology to represent links of a human body, for example the generalized cylinder (GC) or polyhedron is adopted for rigid parts, and a connected link system for the body. This also defines the features which should be detected on images. Occluding regions are recognized by using the human body model and feature points to segment overlapped patterns into body parts. Thus to divide the recognition process into certain stages is effective and some restrictions are imposed closely between stages.

In this section we explain the previous studies in this field. The studies in the early years were aimed for recognition of the moving light display (MLD). The MLD is a sequence of images on which the human body with lights on his/her main joints is projected. In the analysis of the MLD they used the minimum spanning tree as the link-model of the human body. The minimum spanning tree consists of links and joints like as a stick, and the change of its spatial orientations is interpreted as human body movement [2]. Recently the main problem in this field has been how to detect the joints and reconstruct its three dimensional location in ordinary environments. Among many approaches to the problems, Badler and O'Rourke adopted the spherical representation as the human body model and analyzed human motion image samples generated by a computer [3],[4]. Their system was designed to deliver a high-speed and reliable reconstruction, where a possible 3D motion inferred from the feature points movements on the image was used to limit the search space for finding feature-points on the next image in a given sequence. Etoh used the GC as the link-model of human arm and reconstructed the posture from a pair of stereo images [5]. On each image the curvature extremes were extracted in order to find vertices of the projected GC with high reliability. Akita proposed a method which uses the window code as a representative feature and matches the codes between the consecutive images [6].

The previous methods mentioned above reconstruct human motion by using the locations of feature points, and did not pay attention to the shape of the pattern projected onto images [7]. Accordingly we introduce the ellipsoid model in order to recognize the posture and the motion of a human body by evaluating the shape of the projected contours, and to provide the existing system this additional capability.

2.2 Outline of Reconstruction Process

The purpose of our research is first to reconstruct 3D posture and motion of the human body from a projected image sequence, secondly to give an interpretation of the reconstructed motion, and last to predict the future motion from a given image sequence.

The image sequence of human motion is given as a sequence of still images. After an image sequence is given, the projected contours are extracted from each image, then they are segmented into the parts of the human body which amount to rigid links. We consider each part as an individual object and reconstruct its 3D posture. Next we connect the body parts based on the connectivity and the metrics of the given human body model in order to generate the human-like posture. The human posture is reconstructed from each image, then motion can be derived from transitions between two neighboring posture images. In addition, we can infer the body motion directly from an image sequence by interpreting the deformation of a 2D projected contour as a 3D movement.

If the assumptions that the object is rigid and continuously smooth so that the projected contour is differentiable hold, then the deformation of the 2D pattern provides useful information for the 3D motion interpretation. Furthermore when there exists a discontinuous point, it is considered that the object is occluded by another object at this point.

2.3 Three Dimensional Object Model

When we recognize what objects exist in an image, it is necessary to have knowledge about the shape or the surface properties of the objects in advance. In the field of computer vision, the structure model representing a shape in 3D space is desired to be expressive with a small number of parameters. It is, however, impossible to satisfy both accuracy and economy requirements simultaneously. Therefore models of various shapes tend to fit themselves into the objects encountered frequently in our daily lives.

The models which are often used to represent three dimensional rigid objects in practical studies are
1) polyhedron model,
2) generalized cylinder model.

1)Polyhedron model

The polyhedron model consists of planar surfaces (or vertices and edges). The polyhedron model is able to represent various kinds of objects, especially artificial ones in the real world. The object represented by the polyhedron could be reconstructed from its projected image, if the exact model were given, and always has a possibility of reconstruction as a linear solution, because it consists of straight line segments [8]. There also has been an approach called Interpretation of Line Drawings in connection with studies in the field of Artificial Intelligence. In the approach, objects appearing in images are restricted in their existence according to their own constraints, which is called block world. The line in the drawing are labeled with signs and the object is recognized by using a dictionary which contains qualitative characteristics on edge-connectivities of a certain class of polyhedra [9]. Then the object is reconstructed from these labels without losing the consistency [10]. When the object contains one or more curved surfaces, the polyhedron model may approximate it by segmenting the curved surface into many planes, and the surface will be reconstructed by spline interpolations. Then data of this model become so massive that it is not adequate to adopt this model for curved objects.

2) Generalized Cylinder (GC)

The Generalized Cylinder has been regarded as a flexible model to represent real world objects and is often used for practical applications [11]. The reason for representational flexibility of the GC is that the GC is defined by a few functions with several free parameters so that many kinds of objects can be represented by the GC. The shape of a GC is defined by following functions, the spine function, the planar cross-section curve, the expansion function and the angle between the spine and the cross-section. The GC is divided into subclasses according to restrictions on the functions. The SHGC (Straight Homogeneous Generalized Cylinder) belongs to one of the subclasses of GC and it has been used in many applications and the method of reconstruction is studied well. The shape of SHGC is restricted as follows,

1) the spine is straight,

2) the planar cross-section is perpendicular to the spine,

3) expansion scale is given as a function with a single parameter along the spine.

Gross and Boult proposed a method that reconstruct the SHGC from a single intensity image using an assumption that the object has purely diffuse reflectances and that the coefficient of them is constant [12]. Horaud proposed another method for reconstruction of the SHGC [13]. He divided contours detected from a monocular image into two groups. One is the extremal contour and another is the discontinuity contour. The extremal contour occurs where a curved surface turns smoothly away from a viewer, and a discontinuity one occurs wherever a smooth surface terminates or intersects with another surface. As a pair of extremal contours are symmetric with respect to the spine, so the spine is reconstructed readily on the image. Furthermore the spatial tilt of the spine is obtained from the discontinuity contour, because it is one of the cross-sections and the expansion ratio of them can be calculated from the extremal contours. There is, however, few case that a cross-section can be observed.

Since a human arm is a curved object and roughly represented by a cylinder, it is of course desirable to adopt the GC from a viewpoint of expressive representation, setting aside the difficulty of its reconstruction; in reference [3] the whole human body was represented as a set of 310 spheres, while there was some studies on economical representations of the human body using ellipsoids [14]. This ellipsoid is considered a good approximation with less parameters. Therefore we introduce the human body model which consists of rigid ellipsoids (Fig. 1) with different diameters, where each link amounts to a real body part, e.g. head, arm, trunk, leg etc. Three diameters of the ellipsoid are given for all links in advance. The human body model contains also knowledge on the joint position, connectivity and geometrical relationship between links.

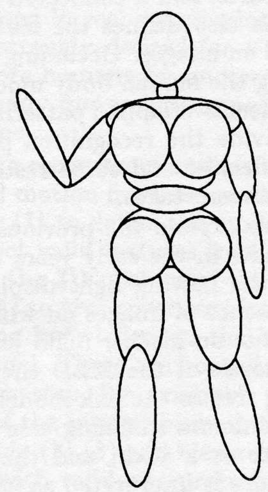

Fig.1 The human model and the elements (ellipsoids).

The advantages to introduce the ellipsoid model are that it only has a positive curvature everywhere on the surface and that the 3D motion of it is observable on its projection except for the translation along the viewing axis under the orthographic projection. However there remains a problem of deciding the interpretation of the ellipsoid postures which is not unique under the orthographic projection, due to the existence of real and mirrored interpretations [7].

2.4 Pre-processing of Images

There are some procedures which must be done before this approach. They are applied to real input images from the TV camera.

2.4.1 Feature extraction

In order to segment the overall contour into ellipse shaped parts, it is useful to utilize the features extracted on the contour shown in Fig.2. The representatives of them are as follows,

1) discontinuity point ,
2) curvature extreme .

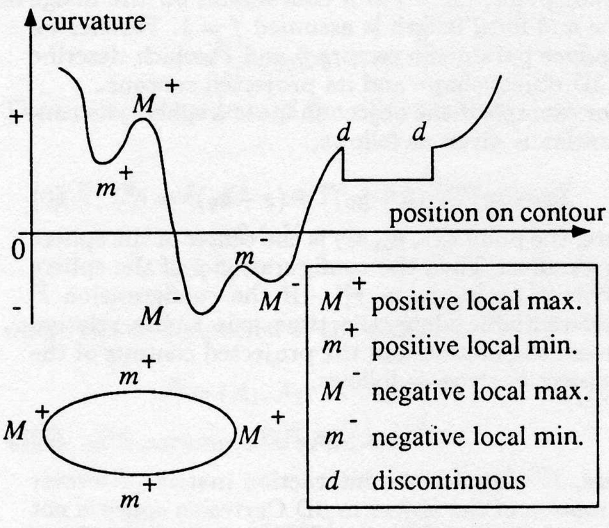

Fig.2 Contour extrema and discontinuity.

These are useful features for dividing the overall contour into links. Leyton proposed that a closed smooth planar curve is roughly represented by curvature extremes [15]. Fundamentally the projected contour of the human body is considered to be a closed smooth planar curve. So the contour of the ellipse can be roughly represented by curvature extreme points. In our human body model each link is approximated by an ellipsoid, which has only the positive curvature extreme, so that a negative curvature extreme or discontinuity point only appears around the joint points or on the intersection of occluding boundaries.

(1) Discontinuity point

When a human body is projected onto the 2D plane, there generally exist occluding regions. This is caused by overlapping of two or more object parts. These contour intersection points on the image become discontinuity points. The overlapped contour should be divided into parts (ellipses) to recognize the deformation (shape) of each ellipse. Then we introduce a method to generate subjective contours, which is usually applied in the psychophysical simulation of human visual phenomena [16], [17]. In the method, the terminal point, which is given as a discontinuous point here, is grown with the smooth continuation hypothesis.

(2) Curvature extreme

If a thin disk is rotated around its arbitrary diameter axis in the 3D space, the extreme curvature of the projected contour on the image changes in response to the rotation. From this point of view, it is considered that the object rotation in the 3D space can be computed from change of the extreme curvature of the projected contour.

2.4.2 Segmentation

As mentioned above, segmentation of a pattern into several parts which are arm, leg, trunk etc. in this approach is the most important in early vision process. In practical situations such a case is almost always observed, because the entity needs to occupy a certain 3D space for asserting its existence in the physical world. For the human body is regarded as a system of rotational links connected each other, this segmentation process is not avoidable.

In general, the segmentation can be performed by extracting the boundary of regions or growing small regions for example using the cluster analysis. In this paper the ellipsoids is the sole object and hence the discontinuity point and the negative curvature extreme would not appear unless the overlapping of parts do not occur. In other words the discontinuity point tells us overlapping of parts and their rough positions. Thus we can properly extract each link using the discontinuity points, symmetry of the projected link contours and zero-crossings of the curvature by Hough transform.

3 Posture Estimation of a Single Arm

In general a 3D scene is projected onto an image by perspective projection through the lens system of human eyes or an artificial camera system. But when the distance between the viewpoint and the object point is long enough compared to the object size, the projection can be approximated as the orthographic one. In this section, we formulate the relation between a 3D object and its projected pattern under the orthographic projection.

We assume that each element of the human body model is rigid and approximated by the ellipsoid,

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} + \frac{z^2}{C^2} = 1. \tag{1}$$

Then it is necessary to formulate the relationship between the posture of ellipsoids in the 3D space and the projected contour of the 2D pattern.

$$rot = F(k_x, k_y, k_z, \theta), \tag{2}$$

where the unit vector (k_x, k_y, k_z) represents the rotation axis and θ denotes the rotation angle.

When the ellipsoid is rotated according to the matrix equation Eq.(2), the above ellipsoid is transformed into that of Eq.(3). Here, our attention concentrates not on the position of the ellipsoid but on the posture so that we can assume the origin of the coordinate axes to be on the centroid of the ellipsoid,

$$f(x,y,z) = ax^2 + by^2 + cz^2 + dxy + eyz + fxz + g = 0. \quad (3)$$

This is resulted from an assumption that the translational motion of the ellipsoid can be easily eliminated so that the first order terms are eliminated in Eq.(3). In order to obtain the projected contour, the partial derivative of the ellipsoid surface equation (3) with respect to the z axis produces a plane equation which intersects with the ellipsoid,

$$\frac{\partial f(x,y,z)}{\partial z} = fx + ey + 2cz = 0. \quad (4)$$

The intersection of the plane and the ellipsoid has the elliptic shape in the three dimensional space. Substituting Eq.(4) into Eq.(3), we get the projected ellipse contour as follow,

$$(f^2 - 4ac)x^2 + (e^2 - 4bc)y^2 + 2(ef - 2cd)xy - 4cg = 0. \quad (5)$$

If an ellipse is extracted from the image, then the parameters of the ellipse, which are the coefficients of the above equation (5), are obtainable. Then we can calculate the ellipse contour in 2D image and the coefficients of the projected contour from these relationships. It follows in principle that reconstruction of the posture of the ellipsoid in the 3D space is possible from projected contour. It is, however, difficult to solve the simultaneous equations analytically, because each these equations is non-linear. In order to obtain analytical solutions we introduce a framework used in the visual servoing in the next section.

4 Motion estimation using Visual servoing

4.1 Visual servoing

Recently in the field of robot navigation the visual servoing is often introduced to achieve a manipulator to a given target position as an efficient visual feedback to the motion [18] [19]. The method is based on the differential relation between the motion of the camera and the projected image of the target object. If the object configuration is given, we can compute the shape of projected contour of the object. In such an application the problem is specified as to compute stepwise motion from the initial, goal and current object contours projected onto image plane. Finally if the goal image and the current image coincide, it is obvious that the camera (or the end-effector of the manipulator) achieved the goal position succesfully. In this paper we apply this method to motion estimation of ellipsoids whose configuration is already given.

4.2 Configuration of object and its projected contour

Here we define the shape of the rigid object $h(\bar{x}, \bar{p}) = 0$, and the projected contour on the image plane $g(\bar{X}, \bar{P}) = 0$. The constraint of point-wise correspondence between in the 3D space and on the image plane under the perspective projection is expressed as $\bar{X} = \bar{x}/z$, where (x, y, z) is a 3D coordinates of a space point, (X, Y) is a coordinates on the image plane and focal length is assumed $f = 1$. Further we introduce parameter vectors $\bar{p}$ and $\bar{P}$ which describe the 3D object shape and its projected contour.

For example if the object shape is a sphere, its configuration is given as follows,

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2, \quad (6)$$

where, the point (x_0, y_0, z_0) is the center of the sphere and r radius. Then the configuration $\bar{p}$ of the sphere is written as $[x_0, y_0, z_0, r]^t$. If the configuration $\bar{P}$ is differentiable along the time axis t, the relation between the motion and the projected contour of the 3D object is given as follows,

$$\dot{\bar{P}} = L^T(\bar{p}, \bar{P})T, \quad (7)$$

where, L^T denotes the interaction matrix. However the motion of the object in 3D Cartesian space is not observable, we can just see the change of the projected contour. As the derivative $\dot{\bar{P}}$ is observable, we can solve the motion T from the interaction matrix L^T. Next we derive the interaction matrix L^T. The object in 3D space is described by 3D coordinates $h(\bar{x}, \bar{p}) = 0$, we can however rewrite by using 2D coordinates and the depth z as follows,

$$h'(\bar{X}, z, \bar{P}) = 0 \quad (\bar{X} = \frac{1}{z}\bar{x}). \quad (8)$$

When $\partial h'/\partial z \neq 0$, the implicit function function theorem ensures the existence of a unique function μ around a solution $\bar{x}_0$ such that

$$z = \mu(\bar{X}, \bar{p}). \quad (9)$$

Then we can get the equation with 2D coordinates,

$$g(\bar{X}, \bar{P}) = 0. \quad (10)$$

This implies the projection onto the image plane $z = 1$ of the object. From the rigidity constraint $\dot{g} = 0$,

$$\frac{\partial g}{\partial \bar{P}}(\bar{X}, \bar{P})\dot{\bar{P}} + \frac{\partial g}{\partial \bar{X}}(\bar{X}, \bar{P})\dot{\bar{X}} = 0. \quad (11)$$

From above conditions we can obtain the interaction matrix,

$$\dot{\bar{P}} = L^T(\bar{p}, \bar{P})T.$$

The estimated motion is then solved by using the generalized inverse matrix of L^T,

$$T = L^T(\bar{p}, \bar{P})^{-1}\dot{\bar{P}}. \quad (12)$$

4.3 Motion Estimation of ellipsoid

An ellipsoid can be expressed as follows,

$$h(\bar{x},\bar{p}) = \frac{(x-x_0)^2}{k^2} + \frac{(y-y_0)^2}{m^2} + \frac{(z-z_0)^2}{n^2} - 1 = 0, \tag{13}$$

where the configuration of the ellipsoid can be written as

$$\bar{p} = [x_0, y_0, z_0, k, m, n]^t. \tag{14}$$

The projected contour of the ellipsoid is given as

$$\begin{aligned} g(\bar{X},\bar{P}) &= X^2 + A_1Y^2 + 2A_2XY + 2A_3X \\ &\quad +2A_4Y + A_5 \\ &= 0, \end{aligned} \tag{15}$$

where the configuration of the ellipse is

$$\bar{P} = (A_1, A_2, A_3, A_4, A_5)^t, \tag{16}$$

where each parameter is given as

$$\begin{aligned} A_0 &= k^2m^4n^4 - k^2m^2n^4{y_0}^2 - k^2m^4n^2{z_0}^2, \\ A_1 &= (k^4m^2n^4 - k^2m^2n^4{x_0}^2 - k^4m^2n^2{z_0}^2),/A_0 \\ A_2 &= (k^2m^2n^4x_0y_0)/A_0, \\ A_3 &= (k^2m^4n^2z_0x_0)/A_0, \\ A_4 &= (k^4m^2n^2y_0z_0)/A_0, \\ A_5 &= (k^4m^4n^2 - k^2m^4n^2{x_0}^2 - k^4m^2n^2{y_0}^2)/A_0. \end{aligned}$$

Rewriting the depth $1/z = aX + bY + c$,

$$\begin{aligned} a &= (m^2n^2x_0)/Q, \\ b &= (n^2k^2y_0)/Q, \\ c &= (k^2m^2z_0)/Q, \\ Q &= m^2n^2{x_0}^2 + n^2k^2{y_0}^2 + k^2m^2{z_0}^2 - k^2m^2n^2. \end{aligned} \tag{17}$$

Then the interaction matrix is givaen as

$$L^T = [{L^T}_{A_1}, {L^T}_{A_2}, {L^T}_{A_3}, {L^T}_{A_4}, {L^T}_{A_5}]^t, \tag{18}$$

where each component is given as follows,

$${L^T}_{A_1} = \begin{bmatrix} 2bA_2 - 2aA_1 & 2A_1(b-aA_1) & 2bA_4 - 2aA_1A_3 \\ 2A_4 & 2A_1A_3 & -2A_2(A_1+1) \end{bmatrix}$$

$${L^T}_{A_2} = \begin{bmatrix} b - aA_2 & bA_2 - a(2{A_2}^2 - A_1) & a(A_4 - 2A_2A_3) + bA_3 \\ A_3 & 2A_2A_3 - A_4 & A_1 - 2{A_2}^2 - 1 \end{bmatrix}$$

$${L^T}_{A_3} = \begin{bmatrix} c - aA_3 & a(A_4 - 2A_2A_3) + cA_2 & cA_3 - a({A_3}^2 - A_5) \\ -A_2 & 1 + 2{A_3}^2 - A_5 & A_4 - 2A_2A_3 \end{bmatrix}$$

$${L^T}_{A_4} = \begin{bmatrix} bA_3 + cA_2 - 2aA_4 & bA_4 + cA_1 - 2aA_2A_4 & bA_5 + cA_4 - 2aA_3A_4 \\ A_5 - A_1 & 2A_3A_4 + A_2 & -2A_2A_4 - A_3 \end{bmatrix}$$

$${L^T}_{A_5} = \begin{bmatrix} 2cA_3 - 2aA_5 & 2cA_4 - 2aA_2A_5 & 2cA_5 - 2aA_3A_5 \\ -2A_4 & 2A_3A_5 + 2A_3 & -2A_2A_5 \end{bmatrix}.$$

Therefore we can obtain the 3D motion from the projected contour as follows.

$$T = L^T(p,P)^{-1}\dot{P}.$$

5 Experimental Result

In order to confirm the performance of our method, we had some experiments with synthsized motion parameters. Input data is a set of coefficients of the configuration of the contour projected on the image plane. A set of three diameters of ellipsoid is given in advance. The inverse matrix ${L^T}^{-1}$ is computed as the generalized inverse, because the matrix is not regular. The geometry of the ellipsoid is given as,

$$(k, m, n) = (5, 7, 6), \quad (x_0, y_0, z_0) = (9, 11, 10),$$

and the focal length $f = 1$.

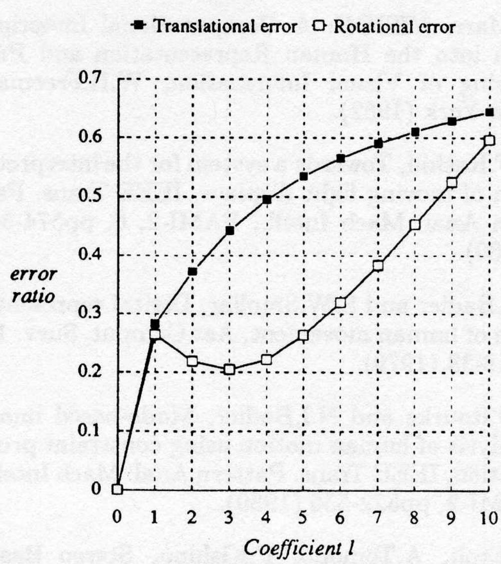

Fig.3 Estimation Error.

Then the true value of the motion

$$T = l[1, 2, 1, 0.0175, 0.0524, 0.0349]^t$$

is also set. The output of this system is the motion parameter vector with 6th order. The coefficients l is varied from 0 to 10, in order to show the degradation of the result with respect to the motion magnitude. The error is evaluated as followings, the translational and the rotational components respectively,

$$\frac{|\hat{v} - v|}{|v|}, \quad \frac{|\hat{\omega} - \omega|}{|\omega|}.$$

Table 1 and Fig.3 show that the estimation error becomes larger as the magnitude of motion increases. However the estimate for the motion is still sufficient, if the magnitude of it is small.

6 Concluding Remarks

In this paper we have considered the framework in which we interpret deformations of projected patterns in image sequence as 3D motion.

We formulate here the relationship between 3D posture of an ellipsoid and its projected ellipse contour. We reconstruct the single arm motion from the 2D contour deformation. The method must be expanded to the object which is consisted with two or more parts. And it is necessary to elucidate the limitation of this method. Furthermore we plan to predict the next motion from a given image sequence and to interpret what motion human is really doing.

References

[1] D.Marr, VISION: A Computational Investigation into the Human Representation and Processing of Visual Information, W.H.Freeman, New York (1982).

[2] R.F.Rashid, Towards a system for the interpretation of moving light displays, IEEE Trans. Pattern Anal. Mach Intell., PAMI-2, 6, pp574-581 (1980).

[3] N.I.Badler and S.W.Smoliar, Digital representation of human movement, Ass.Comput. Surv. 11, pp19-38 (1979).

[4] J.O'Rourke and N.I.Badler, Mode-based image analysis of human motion using constraint propagation, IEEE Trans. Pattern Anal. Mach Intell., PAMI-2, pp522-536 (1980).

[5] M.Etoh, A.Tomono, F.Kishino, Stereo Based Description by Generalized Cylinder Complexes from Occluding Contours, Trans.IEICE, Vol.J73-D-2, pp1402-1412 (1990).

[6] K.Akita, Image sequense analysis of real world human motion, Pattern Recognition, Vol.17,No.1, pp73-83 (1981).

[7] K.Shiraishi, M.Terauchi, K.Onaga, Recognition of Human Motion based on Interpretation of 2D Pattern Deformation, IAPR Workshop on Machine Vision Applications, Proc. of IAPR Workshop on MVA '90, pp.185-187 (1990).

[8] K.Sugihara, Machine Interpretation of Line Drawings, The MIT Press, Cambridge (1986).

[9] K.Kanatani, The Constraints on Images of Rectangular Polyhedera, IEEE Trans. Pattern Anal. Mach. Intell., PAMI- 8, No.4, pp.456-463 (1986).

[10] D.Waltz, Understanding Line Drawings of Scenes with Shadows, in P.H.Winston(eds.), The Psychology of Computer Vision, McGraw-Hill, New York (1975).

[11] T.O.Binford, Visual Perception by Computer, Proc. of IEEE Conference on Systems and Control, Miami (1971).

[12] A.D.Gross, T.E.Boult, An Algorithm to Recover Generalized Cylinders from a Single Intensity View, IEEE, pp.790-795 (1990).

[13] R.Horaud, M.Brady, On the Geometric Interpretation of Image Contour, Artificial Intelligence, 37 pp.33-353 (1988).

[14] C.M.Ginsberg, D.Maxwell, Graphical Marionette, Motion: Representation and Perception, N.I.Badler (eds.), Elsevier (1986).

[15] M.Leyton, A Process-Grammar for Shape, Artificial Intelligence, Vol.34, pp213-247 (1988).

[16] T.Kitahashi, J.Yin, T.Kobayashi, Three Aspects of Decomposition of a Silhouette for pattern recognition, IPSJ Technical Report, Vol.CV64, pp49-56 (1990).

[17] H.Yasuda, N.Onishi, N.Sugie, Generation Algorithm for the Subjective Contours, Proc. on the 3rd Biophysical Engineering Symposium, pp.329-332 (1989).

[18] B.Espiau, F.Chaumete, P.Rives, A new approach to visual servoing in Robotis, IEEE Trans. Robotics and Automation, RA-8, pp313-326 (1992).

[19] K.Kinoshita, K.Deguchi, Determination of Camera Pose and Intrinsic Parameters by Visual Servoing, IPSJ Technical Report, Vol.CV87, pp33-40 (1994).

[20] S.Ullman, Interpretation of Visual Motion, MIT Press, Cambridge (1979).

Table 1 Experimental results (motion estimation).

l	True Values						Estimated values					
	Translational components			Rotational components			Translational components			Rotational components		
	x axis	*y axis*	*z axis*	*x axis*	*y axis*	*z axis*	*x axis*	*y axis*	*z axis*	*x axis*	*y axis*	*z axis*
0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	2	1	0.0175	0.0524	0.0349	0.8673	1.3323	1.1293	0.0059	0.0407	0.0404
2	2	4	2	0.0349	0.1047	0.0698	1.4669	2.2538	1.9146	0.0205	0.0821	0.0794
3	3	6	3	0.0524	0.1571	0.1047	1.9583	2.9688	2.4412	0.0433	0.1207	0.1187
4	4	8	4	0.0698	0.2094	0.1396	2.4257	3.5852	2.7532	0.0741	0.1551	0.1592
5	5	10	5	0.0873	0.2618	0.1745	2.9181	4.1654	2.8756	0.1125	0.1841	0.2013
6	6	12	6	0.1047	0.3142	0.2094	3.4654	4.7476	2.8234	0.1583	0.2067	0.2451
7	7	14	7	0.1222	0.3665	0.2443	4.0868	5.3563	2.6069	0.2111	0.2901	0.2906
8	8	16	8	0.1396	0.4189	0.2793	4.7933	6.0067	2.2344	0.2704	0.2302	0.3375
9	9	18	9	0.1571	0.4712	0.3142	5.5902	6.7073	1.7141	0.3357	0.2299	0.3854
10	10	20	10	0.1745	0.5236	0.3491	6.4778	7.4617	1.0553	0.4059	0.2212	0.4339