SEGMENTATION AND ESTIMATION OF SURFACE FROM STATISTICAL PROBABILITY OF OBJECT FEATURES

Mutsuhiro Terauchi†, Mitsuo Nagamachi†, Koji Ito† and Toshio Tsuji†
†Faculty of Engineering
Hiroshima University

Shitami, Saijo,
Higashi-Hiroshima 724, Japan

ABSTRACT

This paper presents an approach to segment an image into areas of surfaces, and to compute the surface properties from a gray-scale image in order to describe the surfaces for reconstruction of the 3-D shape of the objects.

At first edges are extracted from a gray-level image as many as possible. Next, the other characters of a surface (color, coordinates and image intensity) are extracted. These features of a surface on a pixel of an image plane are mapped to a point of the feature space, and segmented to several groups by cluster analysis on this space. Finally, the states of object surface in 3-D space are computed from distributional probability of local and overall statistical features of a surface.

INTRODUCTION

In the works of computer vision, there have been many approaches to recover the structure of objects in the physical world from one or a few image. For example, in "Interpretation of Line-drawings" 3-D shape of objects is inferred from the line drawing which consists of the extracted edges. Here, a basic difficulty is that the problem is highly underconstrained, i.e. any given set of edges could be the projection of an infinite number of different scenes. The way to converge the interpretation of these scenes is to constrain the variance of the structure of 3-D objects (e.g. to the block world, where the objects' planes are located orthogonally) and the environment around the objects [7][9]. But there is open problem that natural image cannot be interpreted at all. And these algorithms still rely upon man-made heuristic rules [14].

On the other hand, in "Stereo vision" two images are used to reconstruct 3-D shape like two eyes in human visual system. In these study, there is a problem that what feature in one image matches to what feature in another image. It is called "stereo matching problem". To solve this problem for implementation, various methods are proposed. However the key point to make a successful processing is to extract more essential feature in each image.

Eventually it is important to analyze one image throughly and to extract micro and macro-structure of objects meaningfully.

Now let us give attention to processing of an gray-scale image. Generally when we extract edges from an image, we have used various operators like differential filters, and reduced the amount of information to extract more essential features of an image. These are, for example, an extraction of boundary line where image intensity changes largely. On general images, however, it is not possible to extract all the contours perfectly. There have been various approaches in order to correct the unperfect boundary lines. For instance, Rosenfeld utilized relaxation algorithm to combine the edge components [13]. Tsuji et al. computed the image segmentation by using degree of centrality, i.e. the measure of centrality of pixel location in an area [12]. But there must be any other essential information for 3D reconstruction. In natural image almost the surfaces have textures on them [6][10]. When we see a scene, we must utilize textures effectively to understand a scene. Then we analyze textures on an image and collect data for this purpose. Each element which composes the texture has no meaning [8]. Therefore it is significant to analyze it by stochastic method, although it needs some amount of data. After all we have to obtain much more information from an image itself.

PROJECTION AND DISTORTION

In monocular vision, all depth information is lost under projection of a 3-D scene onto a 2-D image. But human can infer 3-D spatial organization from a single view. A basic problem in computer vision is how to equip the capability to machines.

Generally, there may be two types of projections; a) Orthographic projection : When the distance between an image plane and an object is sufficiently long in comparison with size of the object or when an object is located just on a frontal plane, the projection can be approximated to this orthographic projection. The orthographic projection is often used to simplify the computation. b) Perspective projection : In this projection, the lights from the object's surface converge to a point. Thus the optical system with a focal point like human eye or lens of camera distorts the object's image by the convergence.

Such a transformation is desirable for the computation of projection for application to natural scene. In our paper, we pay an attention to this perspective projection.

Geometry of Perspective Projection : Let us the viewer centered coordinate system have its origin at the focal point F. The image plane is parallel to the xy-plane at distance f (the focal length) from the origin along the z-axis. A point in the space is projected onto the image along a line passing through the focal point. This is the perspective transform (Fig. 1). If the coordinates of the point are x,y and z, the coordinates of its projection are given as xf/z, yf/z and f. Let us consider to set the Gaussian sphere centered at the origin (the focal point), where this sphere is represented as unit vectors whose direction are distributed uniformly. A point on this sphere has two angles as coordinates, the azimuth α and the elevation β. The azimuth is the angle measured from the z-axis in the yz-plane. The elevation is the angle measured from the yz-plane toward the x-axis. The relationship between the Cartesian coordinates and the spherical coordinates (azimuth and elevation) of a point on the Gaussian sphere is :

$$x = \sin\beta$$

$$y = \sin\alpha\cos\beta$$

$$z = \cos\alpha\cos\beta$$

In practice only half of the sphere (the one side oriented toward the viewer) is important. Eventually the orientation of the point in the 3D space, is described by azimuth α and elevation β.

Distortion due to the Transformation between Image Plane and Object's Surface : The appearance of surface edges in the image is subjected to two simple geometric distortions: (1) As a surface recedes from viewer, the surface edges appear smaller (the railroad track effect) ; and (2) as a surface is inclined off from the frontal plane, the surface edges appear foreshortened or compressed in the direction of inclination (a tilted circle projects as an ellipse).

Thus, any method for recovering surface orientation from texture must be concerned with some concrete description of the image texture that is sensitive to these two types of distortion.

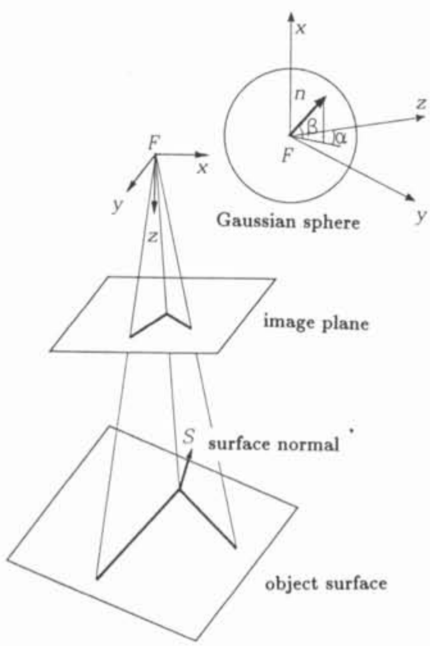


Fig. 1 Geometry of perspective transform.

ASSUMPTION FOR OBJECTS

To simplify the overall process we assumed the constraints on the objects. First for the objects, the surface of object in a scene is planar or curved smoothly (Here we only use spheric surface). Therefore a point where the feature of texture edge changes quite largely is considered to a discontinuous point, i.e. boundary of surfaces. And for the texture on the surface, the edge elements of the texture are assumed to constitute uniform distribution of their orientation in a unit area on the surface. Next we assume that the object in the scene is illuminated by parallel lights from upper side like Sun-light. By this assumption it is thought to be applicable to natural scenes without so much modification.

SEGMENTATION INTO EACH SURFACE

It is required that an image is segmented into each surfaces correctly. However it is very difficult to segment without error. But we have to cluster the image into each surface as well as possible. So it is necessary to gather many features of the objects and environments to be used for segmentation.

Importance of Segmentation : When we describe a surface, for example, slant, shape, orientation and location of the surface, it is possible to represent these properties by using the statistical characters of overall feature and local feature within a segmented area (a surface) on an image.

It is necessary for correct segmentation particularly to compute the overall feature of the surface or local features of the area near a boundary. Unless it were correct, at neighborhood of the boundary between the surfaces the distribution of local features would be distorted because of population misled by segmentation error.

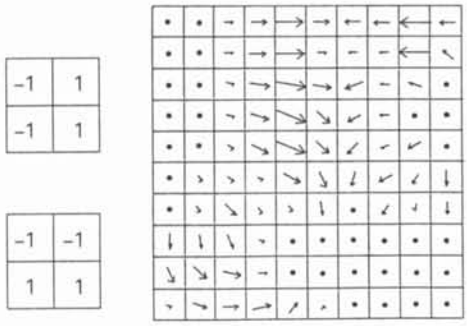
Feature Extraction for Segmentation : What kinds of features are required in particular for segmentation into surfaces? In surfaces there are many convolved factors of various 3D environment. Therefore it is desirable to extract as many features as possible. We can select the optimum features depending on the objects and environment. On the computer vision, it is also necessary to gain such a "object oriented processing".

At first we extract the candidates of edge using the operators shown in Fig. 2.a. These operators play a role of differentiation of an image to get direction and magnitude of intensity change at a pixel. They are the smallest square operators for a digital image. They are not symmetry, but the error is negligible for large scale image. The parameters (the orientation and the magnitude of the gradient) are given as follows :

$$\theta = \tan^{-1}\frac{G_v(i,j)}{G_h(i,j)}$$

$$|\vec{G}| = \sqrt{G_v(i,j)^2+G_h(i,j)^2}$$

where $G_v(i,j)$ and $G_h(i,j)$ are the vertical and horizontal components of the gradient obtained from the mask applied at pixel i,j. The orientation of gradient are



(a) Masks. (b) Gradients.

Fig. 2 Extraction of gradient.

computed for every pixel (see Fig. 2.b). We make an assumption that edges intersect vector of intensity change at right angle, and assumed it as the candidate of edge component.

Next the color is considered as essential feature for segmentation. This feature does not depend on non-local processing. The color is represented by various measures. In the presented paper we use R,G and B intensity of light for computation. The intensity of each color is divided by the one of them for normalization, i.e. an arbitrary color is mapped to the plane that has two dimension axes R/B and G/B. Also the intensity of the image is taken account. The feature space would be established by these parameters. It follows that each pixel on an image is mapped into the feature space according to its characters.

Clustering in Feature Space : The features obtained from the image are mapped into the feature space according to the features (axis). An arbitrary pixel in the image is mapped into the coordinates space with $n+2$ dimensions, which contain n-features' axes and coordinates axes in an image (x and y axis). They are clustered to several groups in the multi-dimension space. Cluster analysis is used to segment the image into surfaces, in which similarity (distance between data) or variance of data is utilized to share the whole pixels into some groups. However it does not guarantee whether the result of clustering is significant or not. Namely, whether the classification is valid or not depends on an interpretation of the result. It is the advantage of this method that needs no criteria for classification.

Let us the two data X'_i and X'_j out of the whole image pixel have m features represented as matrices as follows :

$$X'_i = (X_{i1}, X_{i2}, \cdots, X_{im})$$
$$X'_j = (X_{j1}, X_{j2}, \cdots, X_{jm}) .$$

It is assumed that each feature is continuous and all the data lies in M dimensional Euclidean space. Then the distance between events I and J is given by

$$d_{ij}^2 = \sum_{k=1}^{m} (X_{ik} - X_{jk})^2 .$$

However each magnitude is not normalized by the same criterion. So we have to weight the magnitude of each feature in advance. Then we adopted heuristic values of weights common for several scenes. Euclidean distance in M dimensional space is used for the classification and furthest neighbor method is used for grouping process. It enables us to segment into groups in the way that the size of each cluster is not different so much. Then the clustered groups in the feature space are counter-mapped to the image plane to segment the image into sub-area which correspond to surfaces in the scene. The process is shown in Fig.3. A distance between groups in partial space can be thought as similarity of each surface on some mixed features.

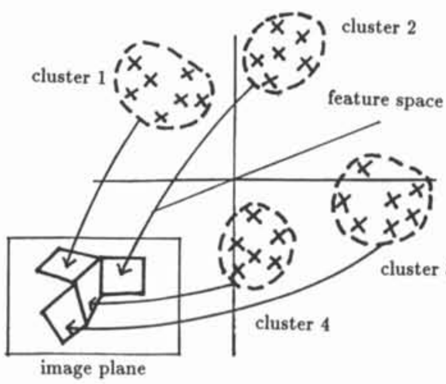


Fig. 3 Clustering and counter-mapping.

DESCRIPTION OF SURFACE

The works that have ever been done in this area had reasonable results on the basis of their assumptions. In these works Gibson is the first researcher who tried to study how surface orientation from texture is perceived. He made assumptions to perform the reconstruction as follows : The individual elements that constitute the texture are uniformly distributed on the world plane, i.e. each unit area on the world plane contains approximately the same number of texture elements; in other words, the density of the texture edge is uniform [5]. But it is not uniform in the real world[1].

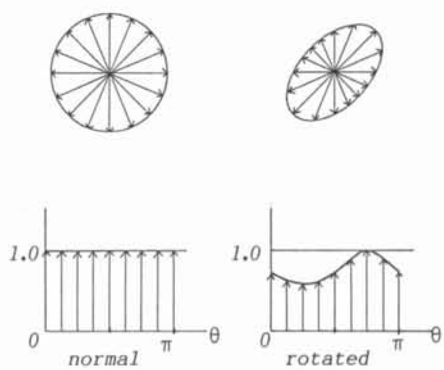


Fig. 4 Distribution of texture edges.

Witkin presented a statistical approach without assuming spatial homogeneity. His assumption is directional isotropy that texture edges are distributed uniformly over all orientations. He derived the estimators of the slant and tilt angle based on an orthographic model [15].

Thus we have to separate the distortion due to orientation of a surface (contraction distortion) from original textures' features.

In the presented paper, it is assumed that the texture edges are distributed uniformly on the plane in the scene as Witkin did. From this assumption it is known that if the object plane is parallel to the image plane, the distribution of texture edges is uniform as shown in Fig. 4.a. But if it is inclined at some angle to the image plane, the distribution is distorted like an ellipse as shown in Fig. 4.b, where the proportion of long axis to short axis of the ellipse represents tilt angle, and the orientation of short axis of ellipse does the direction of descending. However in the perspective transformation the shape of distribution is not same as ellipse but little contracted in the direction of descending axis (for the vanishing point) because of "rail road effect". So we can find that whether the plane is located upward or downward (which side of the plane is nearer than another side). The distribution of a part of surface must be computed to describe the plane in detail. Then the surfaces in the scene could be described by the orientations of overall and local surface.

DISCUSSION

We are trying to find the intermediate level processing of vision. We have computed the orientation of the object surface on the basis of the constraints that the texture edge of surface is distributed uniformly in the world scene. The constraints do not always match for natural scenes. When we see the objects and try to know their structures, we obviously use not only the texture on surface but also its reflectance, i.e. a sense of material. Therefore it is necessary to find the global feature of gray-scale images furthermore. It is also considered that we use something like a priori information.

The intermediated level of vision that Marr had introduced is considered as a process to extract depth from images in order to obtain 2½ sketch [11]. The extraction of depth from a two dimensional image generally needs not only physical laws but some latent knowledge. We can not obtain the macro-structures in an image (also in a scene) only by using the bottom-up processing (Of course we can obtain them, after we have described the production rule for each object like an expert system) [4]. The problem is how to combine the high level knowledge with the bottom-up data efficiently in top-down process. The concept of "schema" proposed by Arbib is also interesting for their application [2].

It is popular to utilize cluster analysis in the works of pattern recognition. We computed statistical processing from the hypothesis that the feature space would be considered as "hyper-column" in the visual cortex.

A large amount of computation for the extracting various features from an image and for the statistic process. Since the computation, however, is statistical processing of local information, it could be executed as parallel distributed processing (PDP).

CONCLUSION

The orientation of objects' surfaces is obtained by using statistical property of texture edges. The segmentation into the surfaces is based on clustering in the feature space where features of pixel in an image are mapped. In perspective projection it has been found that the surface vertical to the descending plane is upward or downward. However the constraint on the uniformity of distribution of the texture edge are too restricted to apply to the general natural scenes. It is desired that more proper constraint is found and utilized for the reconstruction of the 3-D scene.

ACKNOWLEDGMENT

The authors wish to thank K.Matsushima and N.Enomoto at Hiroshima University for the help to draw pretty fine figures.

REFERENCES

[1] J.Aloimonos, "Shape from Texture", Biol.Cybern., vol.58 pp.345/360 (1988)

[2] M.A.Arbib, "Schemas and Perception", Pattern Recognition by Humans and Machines, Academic Press, New York, pp.121/157 (1986)

[3] J.B.Burns, A.R.Hanson and E.M.Riseman, "Extracting Straight Lines", IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-8, pp.425/455 (1986)

[4] M.A.Fischler and O.Firschein, "Intelligence", Addison Wesley, Massachusetts, Part3, pp.248/280 (1987).

[5] J.J.Gibson, "The perception of visual surfaces", Amer. J. Psychology 63, pp.367/384 (1950)

[6] D.C.He, L.Wang and J.Guibert, "Texture Discrimination Based on an Optimal Utilization of Texture Feature", Pattern Recognition, Vol.21, No.2, pp.141/146 (1988)

[7] D.A.Huffman, "Impossible objects as nonsense sentences," in Machine Intelligence, Vol.6, B.Meltzer and D.michie, Eds. Edinburgh Univ. Press, Edinburgh, pp.295/323 (1971)

[8] B.Julesz, "Textons, the elements of texture perception, and their interactions", Nature, Vol.290, No.12, pp.91/97 (1981)

[9] K.Kanatani, "The constraints on Images of Rectangular Polyhedra", IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-8, No.4, pp.456/463 (1986)

[10] D.G.Lowe and T.O.Binford, "Recovery of Three Dimensional Structure from Image Curves", IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-7, pp.320/326 (1985)

[11] D.Marr, "VISION", W.H.Freeman, San Francisco, (1982)

[12] W.G.Oh and S.Tsuji, "Image Segmentation Based on Distribution of a Measure of Centrality", Trans. IEICE Vol.J71-D, No.3, pp.561/566 (1988)

[13] A.Rosenfeld, "Iterative methods in image analysis", Pattern Recognition, Vol.10, No.3, pp.181/187 (1978)

[14] M.Terauchi, M.Nagamachi, K.Ito and T.Tsuji, "3-D Reconstruction of an Object from a Projected Image by Using Prior Knowledge", Proc. IEEE Workshop Intell. Robot. Systems, (1988) (to be appeared)

[15] A.P.Witkin, "Recovering Surface Shape and Orientation from Texture", Artificial Intelligence, Vol.17, pp.17/45 (1981)