'88 KACC 1988. 10. 21~22

SEGMENTATION AND ESTIMATION OF SURFACES FROM STATISTICAL PROBABILITY OF TEXTURE FEATURES

Mutsuhiro Terauchi, Mitsuo Nagamachi, Koji Ito and Toshio Tsuji

Department of Information Technology
Faculty of Engineering
Hiroshima University
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.

In general, an rigid body has several surfaces and many edges. But if it is not polyhedoron, it is necessary not only to describe the relation between surfaces, i.e. its line drawings but also to represent the surfaces' equations itself. In order to compute the surfaces' equation we use a probability of edge distribution.

At first it is extracted edges from a gray-level image as much as possible. These are not only the points that maximize the change of an image intensuty but candidates which can be seemed to be edges. Next, other character of a surface (color, coordinates and image intensity) are extracted. In our study, we call the all feature of a surface as "texture", for example color, intensity level, orientation of an edge, shape of a surface and so on. These features of a surface on a pixel of an image plane are mapped to a point of the feature space, and segmented to each groups by cluster analysis on this space. These groups are considered to represent object surface in an image plane. Finally, the states of object surface in 3-D space are computed from distributional probability of local and overall statistical features of a surface, and from shape of a surface.

1. 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.q. 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 and more meaningful feature in each image.

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

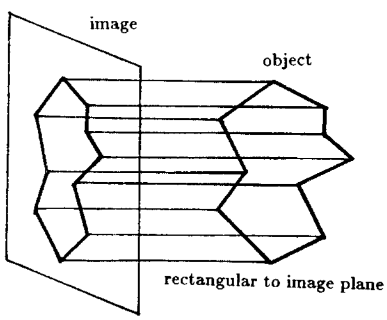
Now let us give attention to processing of an gray-level image. Generally when we extract edges from an image, we have used various operators like differential filters and so on, and reduced the amount of information to extract more essential features of an image. These are, for example, an extraction of boundary line which is edge segment at where image intensity changes largely. On general images, however, it is not possible to extract all the coutours 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 may be information to introduce these correction in an image essentially. In natural image surfaces almost have textures on itself [6][10]. When human see a scene, they must utilize a textures effectively to understand a scene. Then we analyze a texture on an image and collect data for this purpose. There are commonly no meaning in the most local feature of textures [8]. Therefore it must be analyzed by stochastic method, i.e. it needs some amount of data. After all we have to obtain much more information from an image itself.

2. Projection and Distortion

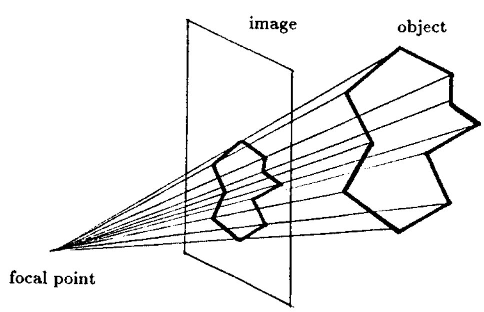
In monocular vision, all depth information is lost under projection of a 3-D scene onto a 2-D image. But human can infer 3-D spatial organization from a single view. A basic problem in computer vision is how to equip a capability like a human to machines.

There may mainly be two types of projections; a) Orthographic projection (Fig. 1.a) : This projection is often used to simplify the computation, and when a distance between an image plane and an object is sufficiently long in comparison with size of an object or when an object is located just on a frontal plane, the projection can be approximated to this orthographic projection. b) Perspective projection (Fig. 1.b) : In this projection, the lights from the object's surface converge to a point. Thus the optical system with a focal point like a human eye or a lense of camera distorts the object's image by the convergence.

Generally such a transformation is desirable for the computation of projection for application for common images. In our paper, we give an attention to this perspective projection (Fig. 1.b).



(a) Orthographic projection.



(b) Perspective projection.

Fig. 1 Projection from object to image.

2.1 Geometry of Perspective Projection

The viewer centered coordinate system we shall use has 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 space point projects onto the image along a line passing through the focal point. This is the perspective transform (Fig. 2). If the coordinates of the space point are x,y and z, the coordinates of its projection are xf/z, yf/z and f. Let us consider to set the Gaussian sphere centered at 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. This surface can be represented digitally as a 2-D grid with α, horizontal varying from $\pi/2$ to $3\pi/2$ and β, vertical varying from $-\pi/2$ to $\pi/2$. To summarize the orientation of any plane and the direction of any line will be represented as points on the Gaussian sphere. Eventually the line located in the scene, i.e. surface normal, is described by azimuth α and elevation β.

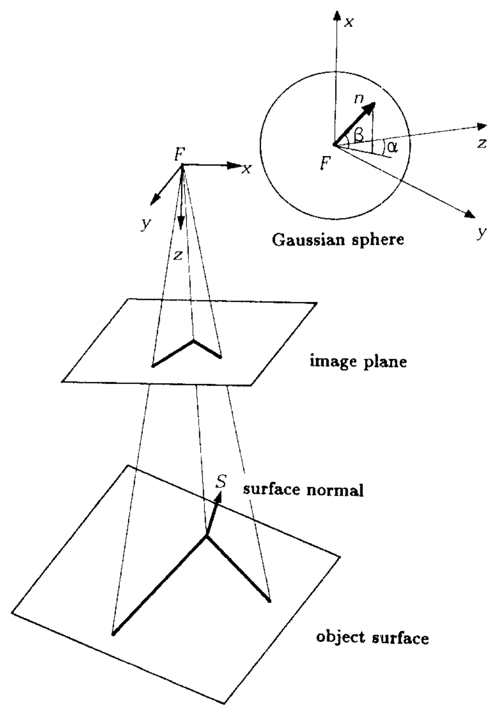
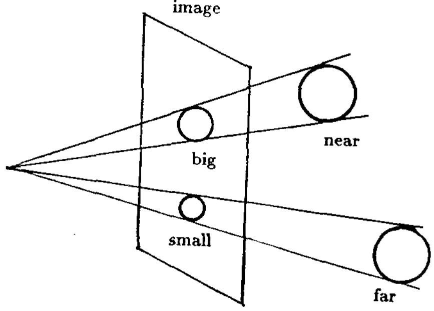


Fig. 2 Geometry of perspective transform.

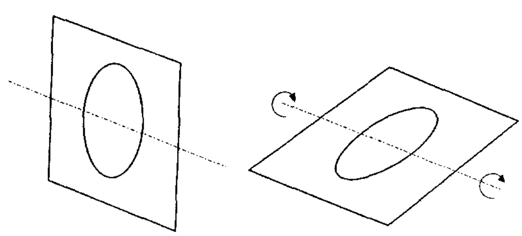
2.2 Distortion due to the Relation between Image Plane and Object's Surface

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 foreshotened or compressed in the direction of inclination (a tilted circle projects as an ellipse). See Fig. 3.a and Fig. 3.b.

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



(a) one example of "Rail road effect."



(b) Contracting distortion.

Fig. 3 Distortion of object shape.

3. Assumption for Objects and Environments

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

The assumption and the condition mentioned above are summarized in Table 1.

Table 1 Assumption and Condition on our study

texture	uniform distributed orientation
surface	planar or spheric
illumination	parallel lights
projection	perspective

4. Segmentation into Each Surface

It is needed that an image is segmented into each surfaces correctly. However it is very difficult to segment without error. But we have to cluster into each surface as well as possible. So it is valuable to gather many natures of the objects and their environments to use for segmentation.

4.1 Importance of Segmentation

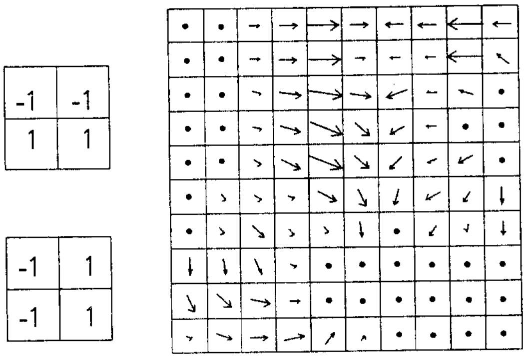
When we describe a feature of a surface, for example slant, shape, orientation and location of a surface, it is posibble to represent these features by using the statistical characters of an overall nature and local natures within a segmented area (a surface) on an image.

It is necessary for correct segmentation particularly when computing an overall feature of an surface or a local feature of an area near a boundary. Unless it were correct, then an overall feature of an surface would be misled and at neighborhood of boundary between surfaces distribution of local feature would be distorted because of misled population by segmentation error.

4.2 Feature Extraction for Segmentation

What kinds of features are nesecessary in particular for segmentation of surfaces? There are many factors of various 3D environment convolved in an image. According to such facts, it is desirable to extract as many features as possible. In the case of humans, they optimize the selection of noticeable features according to the situation. On the computer vision, it is also necessary such a "object oriented processing".

At first we extract the candidates of edge using operators shown in Fig. 4.a. These operators play a role of differentiation of an image to get direction and magnitude of intensity change at a pixel. These masks are the smallest square operators for a digital image. Therefore they can extract the change sensitive for small intensity change. These masks are not symmetry, but the error is negligible for large



(a) Masks. (b) Gradients.

Fig. 4 Extraction of gradient.

scale image. These parameter (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. For all pixel the orientation of gradient are computed (see Fig. 4.b). We make an assumption that there can be edges at right angles to the orientation, and assumed it as the candidate of edge component.

Next the color is considered as essntial feature for segmentation. This feature does not depend on non-local processsing. The color is represented by various measures. In our study we use R,G and B intensity of light for computation. The intensity of each color is devided by one intensity of colors for standardization, i.e. an arbitral 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. This mapping is illustrated in Fig. 5.

4.3 Clustering in Feature Space

The features obtained from the image are mapped into the feature coordinates space in terms of the features and their magnitude or size. An arbitrary pixel in an image is mapped to the coordinates space with n+3 orders, which contain n-features' axes, coordinates axes in an image (x and y axis) and depth axis (z axis). They are clusterred to several groups in the multi dimension space. To segment an image into surfaces, we used cluster analysis. In this method similarity (distance between data) or variance of data is utilized to share the whole pixels into some groups. However it does not guarantee whether a result of clustering is meaningful or not. Namely whether classification is valid or not is estimated on an interpretation of the result. The advantage of this method is that does not need any criteria for classification.

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

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

The magnitude of each feature is continuous and it is assumed that there are all data in M dimensional Euclidean space, then the distance between events I and J is obtained by

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

However each magnitude is not standardized on 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. And Euclidian distance in M dimensional space is used for classification as described above, and for

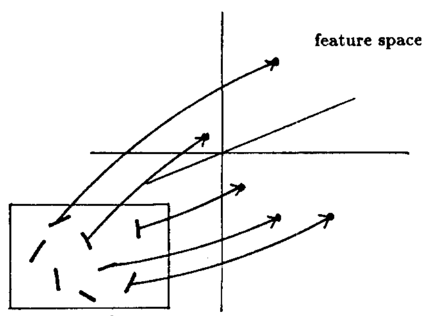


Fig. 5 Mapping to the feature space.

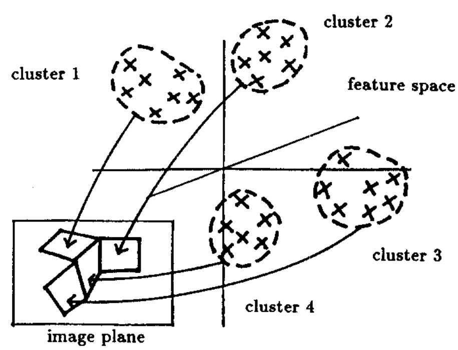


Fig. 6 Clustering and counter-mapping.

grouping process furthest neighbour method is used. This method enables us to segment into groups in the way that the size of each cluster is not different so much. Thus the groups clustered 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. 6. A distance between groups in partial space can be thought as similarity of surfaces on some mixed features.

5. Extraction and Description of Surface Features

The works that have been done in this area have reasonable results on the basis of their assumptions. In these works Gibson is the first researcher who tried to study how humans perceive surface orientation from texture. He made assumptions to perform the reconstruction as follws : 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, texture density is uniform [5]. But the texture density on the world plane is not uniform [1].

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. For this purpose many of previous researcher assumed the constraints for textures' features on the world plane that matches consistency with natural world scenes.

In the case of our study it is assumed that the texture edges are distributed uniformly on the plane in the scene as Witkin did. From this assumption if the object plane is parallel to the image plane, the distribution of texture edges is uniform shown in Fig. 7.a. But if it is located at some angle to the image plane, the distribution is distorted like an ellipse shown in Fig. 7.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 perspective transform the shape of distribution is not same as ellipse but little contracted ellipse to the direction of discending axis (for the vanishing point). So we can obtain the fact that whethere the plane is located upward or downward (which side of the plane is nearer than another side). To describe the plane in detail the distribution of a part of surface must be computed. Then the surfaces in the scene could be described by the orientation of overall and local surface.

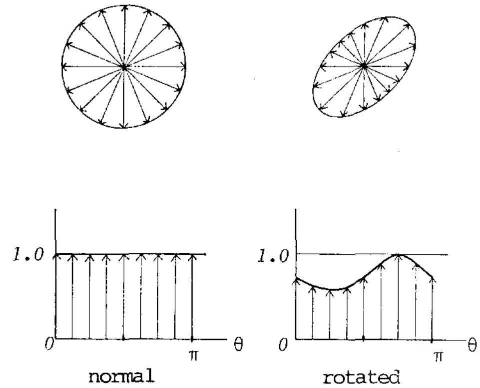


Fig. 7 Distribution of texture edges.

6. Discussion

We are trying to study intermediate level processing of vision. We have computed the orientation of the object surface on the basis of the constraints on the distribution of the texture edge of surface in the world scene. This constraint does not always match for natural scenes. When humans see the objects and try to know their structures, they obviously use not only the texture on surface but also its reflectance, i.e. a sense of material. Therefore it is necessary to research the global feature of gray-scale images further. It is also considered that humans use something like a priori information. In the area of image understanding, paticularly, it is necessary and of importance that general knowledge of human is analized, represented and implemented on a computer system.

The intermediated level vision that Marr had introduced is the processing to extract depth from images in order to obtain $2\frac{1}{2}$D sketch [11]. It is thought that the extraction of depth from a two dimensional image generally needs not only phisical laws but human 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]. There is a problem how to combine high level knowledge with 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 statistic processing establishing the hypothesis that the feature space would be considered as "hyper-column" in the visual cortex. (See Fig. 8)

It needs a large amount of computation for the extracting so much features from an image and for the statistic process. As the computation, however, is statistical processing of local information, it could be executed as parallel distributed processing (PDP). Considering that neural network is a device that can study the statistics as weights, it is considered that these process can be implemented on a network.

Finally we would have to extract, describe and utilize the knowledge of humans for these intermediate level of visual information processing.

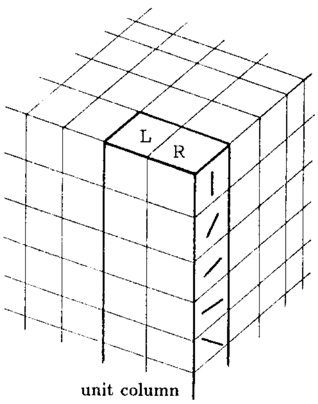


Fig. 8 Hyper - column.

7. Conclusion

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

Acknowledgment

The first author wishes to thank K.Matsushima and N.Enomoto of the University of Hiroshima 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", 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, "Recoverring Surface Shape and Orientation from Texture", Artificial Intelligence, Vol.17, pp.17/45 (1981)