Proceedings of the 1998 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
Victoria, B.C., Canada • October 1998

Fast Object Recognition Using Salient Line Groups

Dong Jung Kang* and In So Kweon**
*Signal Processing Lab., Samsung Advanced Institute of Technology, Suwon, KOREA 440-600
**Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology,
207-43, Cheongryangri-dong, Dongdaemoon-gu, Seoul, KOREA, e-mail: {djkang,iskweon}@cais.kaist.ac.kr

Abstract: *This paper presents an effective recognition method based on perceptual organization of low level features detected in an image. The method uses a dynamic programming (DP) based formulation to represent various line groups such as convex, concave, and more complex patterns consisting of convex and concave shapes. The essential features of perceptual organization such as endpoint proximity, collinearity, parallelism, and connectivity of lines, are incorporated into the DP based formulation as energy terms. As endpoint proximity, we detect two line junctions from image lines. We then search for junction groups by using collinearity constraint between the junctions. A DP-based search algorithm is used to detect a junction chain similar to the model chain, based on a local comparison. The proposed system is able to find line groups from images with broken lines and strong background clutters. We demonstrate the feasibility of our DP-based matching method based on perceptual organization using real images.*

Indexing Keywords: Feature matching, Dynamic programming, Perceptual grouping, Connected line chain

1. Introduction

This paper describes an algorithm that robustly locates collections of salient line segments in an image. In computer vision and related applications, we often wish to find objects based on stored models from an image containing objects of interest [1,2,3,4]. To achieve this, a model-based object recognition system first extracts sets of features from the scene and the model, and then it looks for matches between members of the respective sets. The hypothesized matches are then verified and possibly extended to be useful in various applications. Verification can be accomplished by hypothesizing enough matches to constrain the geometrical transformation from a 3-D model to a 2-D image under perspective projection.

Guiding object recognition by matching perceptual groupings of features was suggested by Lowe [2]. In the SCERPO, his approach is to match a few significant groupings made of certain arrangements of lines found in images. Lowe has successfully incorporated grouping into an object recognition system. First, he groups together lines thought particularly likely to come from the same object. Then, the SCERPO looks for groups of lines that have some property invariant with the camera viewpoint. For the purpose, he proposes three major line groups - proximity, collinearity, and parallelism.

Recent results in the field of object recognition including Jacobs [5], Grimson and Huttenlocher [6] demonstrate the necessity of some type of grouping, or feature selection, to make the combinatorics of object recognition into manageable level. Grouping, as the non-accidental image features, overcomes the unfavorable combinatorics of recognition by removing the need to search the space of all matches between image and model features. Grimson [7] has shown that the combinatorics of the recognition process in cluttered environments using constrained search reduces from an exponential to a low order polynomial if we use an intermediate grouping process. Only those image features considered likely to come from a single object must be included together in hypothetical matches. And these groups need only be matched with compatible groups of model features. For example, in case of constrained tree search, grouping may tell us which parts of the search tree to explore first, or allow us to prune sections of the tree in advance.

As done in several previous works [5,8,9,10], the use of perceptual groups is not new now. However, the usage has mostly involved just simple, small groups of features such as parallels [8], only convex lines [5], ellipses [9], and rectangles [10]. This is partly because of the rarity of fast frameworks to form large feature groups. The role of large feature groups, as the non-accidental features, is very important [11].

This paper is basically related to Lowe's work in using perceptual organization for object recognition. However, the SCERPO to grouping has a limitation: forming only small groups of lines limit the search that we may reduce. In this work, however, we use a DP based formulation for perceptual organization to represent more complex structures that cannot be grouped in the SCERPO. Among Lowe's organization groups, the *proximity* consisting of two or more image lines is an important clue to start object recognition. When projected to the image plane, most manmade objects have a polyhedral plane in which two or several sides give line junctions. First, we introduce a *quality measure* to detect meaningful line junctions denoting the proximity. The quality measure must be carefully defined not to skip salient junctions in the input image. Then extracted salient junctions are combined to form more complex and important local line groups. The combination between junctions is guided by the *collinear condition* that is also Lowe's perceptual group. In sum, our intended contributions are as follows:

(1) Define a robust and stable geometric representation that is based on the perceptual organizations (i.e., the representation as a primitive search node includ-

es two or more perceptual grouping elements); and

(2) Introduce a consistent search framework combining the primitive geometric representations, based on the dynamic programming (DP) formulation.

The DP-based search method has an advantage in greatly reducing the time complexity for a candidate search, based on the local similarity. Silhouette or boundary matching problems that satisfy the locality constraint can be solved by DP-based methods using local comparison of the shapes. In these approaches, both the model and matched scene have a connected chain form of lines [12], ordered pixels [13], or chained points [14]. However, there also exist many vision problems, in which the ordering or local neighborhood can not be easily defined. For example, meaningful groupings of lines in Fig. 3(c) are very difficult due to real complex scenes. In this paper, we do not assume any fixed connections of lines or junctions. We keep any connection possibilities for two arbitrary lines or junction groups in DP-based search. That is, the given problem is a local comparison between a pre-defined model chain and all possible scene chains in an energy-minimization framework.

2. Line Junction

Two-line junction is defined as any pair of line segments which intersect, and whose intersection point either lies on one of the line segments, or does not lie on either of the line segments. An additional requirement is that the acute angle between the two lines must lie in a range $\theta_{\min}$ to $\theta_{\max}$. In order to avoid ambiguity with parallel or collinear pair [2], $\theta_{\min}$ is chosen to be a pre-defined threshold. Various junction types are well defined by *Etemadi et al.* [8].

Fig. 1 shows the schematic diagram of a typical junction. Now a perfect junction is defined as one in which the intersection point $\mathbf{P}$ lies precisely at the end points of the line segments. Note that there are now two virtual lines that share the end point $\mathbf{P}$. The points P_1 and P_4 locating opposite side of P_2 and P_3 denote the remaining end points of the virtual lines, respectively.

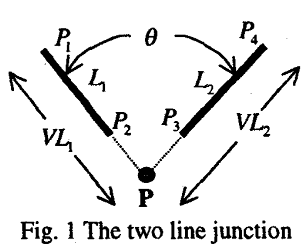


Fig. 1 The two line junction

Then, the junction quality factor is

$$Q_J = \frac{|L_1| - \sigma_1^{\parallel} - \sigma_2^{\perp}}{|VL_1|} \cdot \frac{|L_2| - \sigma_2^{\parallel} - \sigma_1^{\perp}}{|VL_2|}, \quad (1)$$

where $|VL_i|$ ($i = 1,2$) are the lengths of the virtual lines, as shown in Fig. 1. The standard deviations $\sigma_i^{\parallel}$ and $\sigma_i^{\perp}$, incorporating the uncertainties in the line extraction process for the position of the end points of the line segments along and perpendicular to its direction respectively, may be replaced by constants without affecting the basic grouping algorithms [8]. In this paper, two variance factors $\sigma_i^{\parallel}$ and $\sigma_i^{\perp}$ are ignored. The defined relation penalizes pairings in which either line is far away from the junction point. The quality factor also retains the symmetry property.

3. Energy Model for a Line Chain

The relational representation, made from each contextual relation of the model and scene features, provides a reliable means to compute the correspondence information in the matching problem. Suppose that the model consists of M nodes of two line junctions. Then, a linked node chain, given by the sequential connection of the nodes, can be constructed.

If the selected features are sequentially linked, then it is possible to calculate a potential energy from the enumerated feature configuration. That is, assume that any two junctions of the model correspond to two features f_I, f_{I+1} of the scene. If the relational configuration of each node depends only on the connected *neighboring* nodes, then the energy potential obtained from the M nodes can be represented as:

$$\begin{aligned} E_{total}(f_1, f_2, \ldots, f_M) \\ = E_1(f_1, f_2) + E_2(f_2, f_3) + \ldots + E_{M-1}(f_{M-1}, f_M) \end{aligned} \quad (2)$$

where

$$\begin{aligned} E_I(f_I, f_{I+1}) &= \alpha_1 \cdot \mathrm{F}_1(\theta(f_I), \Theta(I)) \\ &\quad + \alpha_2 \cdot \mathrm{F}_2(r(f_I, f_{I+1}), R(I, I+1)) \\ &= \alpha \cdot \mathrm{U}_1(\theta, \Theta) + \beta \cdot \mathrm{U}_2(r, R) + \gamma \cdot \mathrm{U}_3(r, R). \end{aligned} \quad (3)$$

Here F_1 denotes a function representing the unary relation and F_2 is for the binary relations defined between two junctions. The functions describing the unary and binary relations are detailed as:

$$\mathrm{U}_1 = \begin{cases} \infty, & \Theta_1(I) - \Theta_{c_1} > \theta_1(f_I) > \Theta_1(I) + \Theta_{c_1} \\ -\dfrac{1}{\sqrt{2\pi}\sigma_{c1}} e^{-\frac{(\theta_1(f_I) - \Theta_1(I))^2}{2\sigma_{c1}^2}}, & \text{else} \end{cases} \quad (4a)$$

$$\mathrm{U}_2 = \begin{cases} 0, & \mathrm{Rot}(f_I, f_{I+1}) \equiv \mathrm{Rot}(I, I+I) \\ \infty, & \text{else} \end{cases} \quad (4b)$$

$$U_3 = \begin{cases} \infty, \Theta_2(I,I+1) - \Theta_{c_2} > \theta_2(f_I, f_{I+1}) > \Theta_2(I,I+1) + \Theta_{c_2} \\ -\dfrac{1}{\sqrt{2\pi}\sigma_{c2}} e^{-\frac{(\theta_2(f_I, f_{I+1}) - \Theta_2(I,I+1))^2}{2\sigma_{c2}^2}}, \quad \text{else} \end{cases} \tag{4c}$$

Fig. 2 shows the unary and binary relations made from any two connected junction nodes. In (4b) and Fig. 2(c), Rot(*I,I+1*) denotes a circulating direction given by connected chain of two junctions *I* and *I+1*. If we change the constraint (4b) as $\mathrm{Rot}(f_I, f_{I+1}) \equiv \mathrm{Rot}(I+1, I)$, a complex pattern with c.c.w and c.w directions for the two constituting junctions is allowed.

In (4a) and (4c), Θ_{c_1}, σ_{c1}^2, and Θ_{c_2}, σ_{c2}^2 present permitted angle ranges and their variances for the unary and binary relations, respectively. In (4c), θ_2 is defined by outer two lines 1 and 3 in Fig. 2(b), (c).

Hence, each junction has the *unary* angle relation by two lines in a single junction. In (4a), $\theta(f_I)$ is corresponding junction angle in a *scene* for a reference or model property $\Theta(I)$. The binary relations for the scene (r) and model (R) in the last two terms of (3) are defined as a topological constraint of (4b) and another angle relation θ_2 between two junctions. For example, the following topological descriptions can represent the binary relations, as shown in Fig. 2(c) and (4b), (4c).

(1) Two lines 1 and 3 can be approximately parallel (*parallelism*) and the representation is covered by (4c);
(2) Scene lines corresponding to two lines 2 and 2' must be a collinear pair [2] or the same line for each other. That is, two junctions are combined by the *collinear constraint*; and
(3) Line ordering for two junctions J_1, J_2 should be maintained such as *counter-clockwise* (c.c.w).

The relation defined by two connected junctions (i.e., in the aspect of the binary relation as the second order that does not seriously increase the time complexity of the DP-based search) includes all three perceptual organizations that Lowe used in SCERPO. These local relations can be selectively imposed according to the type of the given problems. For example, a convex line triplet [15] is simply defined, by removing the constraint (1) and letting line 2 and line 2' of constraint (2) be equal to each other. We do not use a relation depending on line length, because lines in a noisy scene are easily broken. The weighting coefficients α, β, and γ of the energy potential are experimentally given.

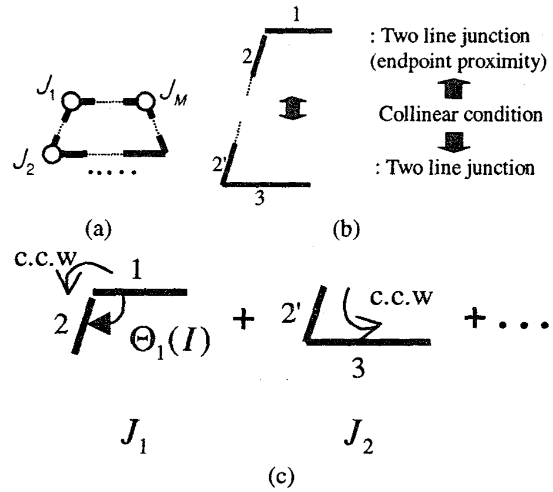


Fig. 2 Unary and binary relations made from any two connected junction nodes on a linked chain; (a) A model; (b) Combination of two junctions by a collinear constraint; (c) A node chain of junctions from model (a).

4. Energy Minimization

Dynamic programming (DP) [16,17] is an optimization technique good for problems where not all variables are inter-related simultaneously. Suppose that the global energy can be decomposed into the following form:

$$E(f_1, \ldots, f_M) = E_1(f_1, f_2) + E_2(f_2, f_3) + \ldots + E_{M-1}(f_{M-1}, f_M) \tag{5}$$

in which M is the number of the model nodes such as junctions, and f_I is a scene label that can be assigned to the model node I.

Fig. 3(d) shows a schematic DP diagram to find a given model of Fig. 3(a) in the scene lines of Fig. 3(c). Fig. 3(c) shows a typical case in which we can not define an ordering for the scene lines due to the cluttered background. Therefore, it is difficult to extract a meaningful object boundary that corresponds to the given model. In this case, the DP-based search structure is formulated as the column in Fig. 3(d), in which all detected scene features are simultaneously included in each column. Each junction of the model can get a junction of the scene. The potential matches are defined as the energy accumulation form of (6). From binary relations of junctions (i.e., 'arrow' in Fig. 3(d)) defined between neighboring two columns, the local comparison-based method using the recursive energy-accumulation table of (6) can give a fast matching solution.

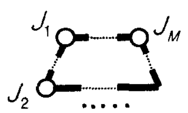
The DP algorithm generates a sequence that can be written in the recursive form. For $I = 1, \ldots, M-1$,

$$D_I(f_{I+1}) = \min_{f_I}[D_{I-1}(f_I) + E_I(f_I, f_{I+1})], \tag{6}$$

with $D_0(f_1) = 0$. The minimal energy solution is obtained by

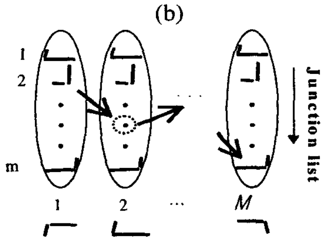
$$\min_{f} E(f_1,\dots,f_M) = \min_{f_M} D_{M-1}(f_M) \cdot \quad (7)$$

If each f_l takes on m discrete values, then to compute $D_{l-1}(f_l)$ for each f_l value, one must evaluate the summation $[D_{l-1}(f_{l-1}) + E_{l-1}(f_{l-1}, f_l)]$ for the m different f_{l-1} values. Therefore, the overall minimization involves $(M-1)m^2$ evaluations of the summations. This is an enormous reduction from the exhaustive search m^M, for the total evaluation of $E(f_1,\dots,f_M)$. Here m is the number of junctions satisfying a threshold for the junction quality in the scene.



(a)

(b)



(c) (d)

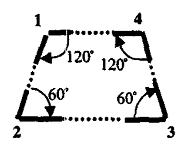
Fig. 3. The DP algorithm searches a scene node corresponding to each model node. A model feature can be matched to at least one node, among scene nodes, $1,\dots,m$ of a column: (a) A model; (b) Original image for the rear view of a vehicle; (c) Detected scene lines for the image; (d) Junction chain-based search. m is the number of detected junctions from (c) and M is the number of pre-defined model junctions.

5. Experiments

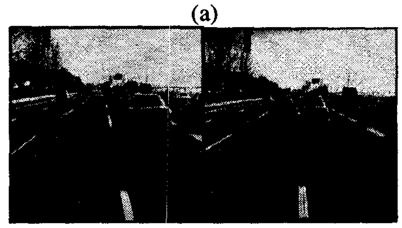
In this section, we apply the proposed method to find a group of optimal line junctions. Test scenes include highly distorted line patterns by a wide variation of viewpoints. First, input images are processed to detect only the strongest step edges. Edges are further filtered by discarding shorter edge segments. Junctions are then inferred between the remaining line segments. When junctions are extracted from the extracted lines, relative relations between the junctions are searched using the criteria of Section 3, with the collinear constraint that link two junctions.

To reduce the repeated computation for relations between line segment, possible all previously calculated relations such as inter-angles, collinear properties, and junction qualities between two lines are saved. The weighting coefficients in (4) are all set 1.

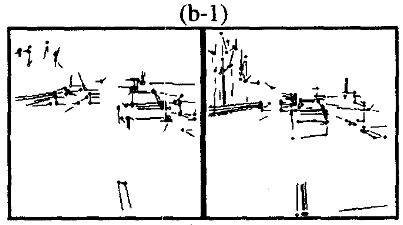
As an example of 2-D line matching under weak perspective projection, the rear-view of a vehicle is used. Description of the model lines are given by a user and shown in Fig. 4(a). Fig. 4(b-1) shows the first and the last image among a sequence of 30 images. Fig. 4(b-2) shows the detected junctions in the two frames. A threshold for the quality measure Q_J is set 0.5. Fig. 4(b-3) shows optimal matching lines having smallest accumulation energy of (7). In spite of distorted shapes from the original model shape, a reasonable matching result is obtained. Unary and binary properties of (4) are both used. Fig. 4(c) shows a few optimal matching results. In each row of Fig. 4(c), the model shape is well matched as minimum DP-energy of (7), in spite of highly distorted shapes. Matching was successful for 25 frames out of thirty frames - the success ratio 83%. Failing cases are resulted from line extraction errors in low-level vision processes, in which junctions can not be defined by missing lines. Fig. 5 shows an experimental result for matching 2-D polyhedra. Recognition of 2-D occluded objects is guided by detection of salient lines such as the convex patterns. Fig. 5(a) and (b) present an interesting salient pattern of convex four-line type and three models to be recognized, respectively. All models have at least one convex pattern. With previous automatic detection for the model images, input scene of Fig. 5(c) is subjected to the DP-based search with a low-level process detecting line segments and the junctions. The result is presented as thick lines in center figure of Fig. 5(c) with total 7 hypotheses. In each model, the convex pattern is detected 5, 2, and 4 times, respectively. Therefore, the total hypotheses are generated 11x7 times to match the three models. Recognition result is shown with model overlap in Fig. 5(c).



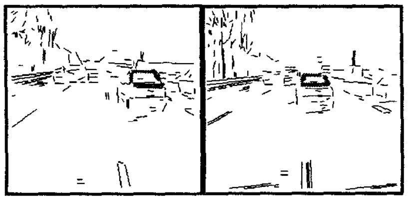
(a)



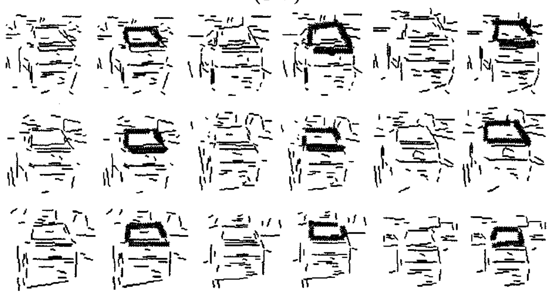
(b-1)



(b-2)

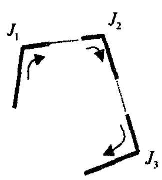


(b-3)

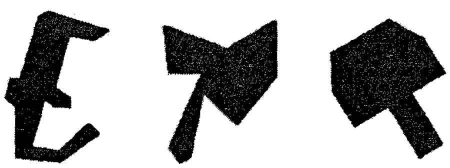


(c)

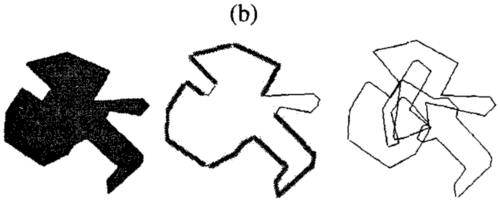
Fig. 4 An example of approximated 2-D matching under weak perspective projection: A rear-window of a vehicle on the highway is used. (a) Model; (b-1) The first and last images to be tested; (b-2) Junction detection for Q_J =0.5; (b-3) Optimal model matching; (c) A few optimal matching results between the first and last images.



(a)



(b)



(c)

Fig. 5 A 2-D occluded object recognition guided by detection of salient lines such as the convex patterns. (a) A convex salient pattern; (b) Three models; (c) Salient pattern detection as the hypotheses and object recognition.

6. Conclusions

A fast and reliable matching and grouping method by the dynamic programming is proposed to extract collections of the salient line segments. We consider the classical dynamic programming as an optimization technique for the geometric matching and grouping problems. The importance of grouping to object recognition is emphasized. By grouping together line features that are likely to have been produced by a single object, this paper shows that significant speedups in a recognition system can be achieved, compared to performing a random search.

A main contribution in this paper is a new DP-based formulation for matching and grouping of line patterns by introducing a robust and stable geometric representation that is based on perceptual organizations. We detect the junctions as the endpoint proximity for the grouping of line segments. Then, we search again a junction group, in which each junction is combined by the collinear constraint between them. These local primitives, by including Lowe's perceptual organizations and acting as the search node, are consistently linked as a line-chain form in the DP-based search structure.

False matching shapes usually denote highly increased energy values, while a shape similar to the reference pattern shows a small accumulation in energy. Accumulated energy may be used as a threshold to decide the success or failure of the matching or grouping.

Through experiments using images from cluttered scenes including outdoor environments, we show that the method can be applied to the optimal matching, grouping of line segments, and 2-D/3-D recognition problems, with a simple and consistent shape description sequentially represented.

References

1. W.E.L Grimson and T.Lozano-Perez, Localizing overlapping parts by searching the interpretation tree, *IEEE Trans. Pattern Analysis and Machine Intelligence* **9**, 469-482 (1987).
2. D.G. Lowe, Three-Dimensional Object Recognition from Single Two-Dimensional Images, *Artificial Intelligence* **31**, 355-395 (1987).
3. R.A.Hummel and S.W.Zucker, On the foundation of relaxation labeling processes, *IEEE Trans. Pattern Analysis and Machine Intelligence* **5**(3), 267-286 (1983).
4. N.Ayache, O.D.Faugeras, HYPER: A New Approach for the Recognition and Positioning of Two - Dimensional Objects, *IEEE Trans. Pattern Analysis and Machine Intelligence* **8**(1), 44-54 (1986).
5. D.W.Jacobs, Robust and efficient detection of salient convex groups, *IEEE Trans. Pattern Analysis and Machine Intelligence* **18**, 23-37 (1996).
6. W.E.L.Grimson and D.Huttenlocher, On the Sensitivity of the Hough Transform for Object Recognition, *Proc. of the Second International Conference on Computer Vision*, 700-706 (1998).
7. W.E.L.Grimson, The combinatorics of heuristic search termination for object recognition in cluttered environments, *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**, 920-935 (1991).

8. A.Etemadi, J.P.Schmidt, G.Matas, J.Illingworth and J.Kittler, Low-level grouping of straight line segments, *in Proc. of 1991, British Machine Vision Conference*, 118-126 (1991).
9. G.Roth and M.D.Levine, Geometric primitive extraction using a genetic algorithm, *IEEE Trans. Pattern Analysis and Machine Intelligence* **16**, 901-905 (1994).
10. R.Mohan and R.Nevatia, Using perceptual organization to extract 3-D structures, *IEEE Trans. Pattern Analysis and Machine Intelligence* **11**, 1121-1139 (1989).
11. D.G.Lowe, Three – dimensional object recognition from single two – dimensional images, *Artificial Intelligence* **31**, 355-395 (1987).
12. Y.Ohta and T.Kanade, Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **7**(2), 139-154 (1985).
13. I.J.Cox, S.L.Higorani, S.B.Rao, A Maximum Likelihood Stereo Algorithm, *Computer Vision and Image Understanding* **63**(3), 542-567 (1996).
14. H.Bunke and U. Buhler, Applications of Approximate string Matching to 2D Shape Recognition, *Pattern Recognition* **26**, 1797-1812 (1993).
15. J.Henikoff and L.G.Shapiro, Representative Patterns for Model-based Matching, *Pattern Recognition* **26**, 1087-1098 (1993).
16. M.Fischler and R.Elschlager, The representation and matching of pictorial structures, *IEEE Trans. on Computer* **C-22**, 67-92 (1973).
17. A.A.Amini, T.E.Weymouth, and R.C.Jain, Using dynamic programming for solving variational problems in vision, *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**(9), 855-867 (1990).