Contour Extraction by Mixture Density Description Obtained from Region Clustering

Minoru ETOH[1], *Yoshiaki SHIRAI*[2] *and Minoru ASADA*[2]

[1] Central Research Laboratories, Matsushita Electric Ind., Moriguchi, Osaka 570, Japan
[2] Mech. Eng. for Computer-Controlled Machinery, Osaka University, Suita, Osaka 565, Japan

Abstract. This paper describes a contour extraction scheme which refines a roughly estimated initial contour to outline a precise object boundary. In our approach, mixture density descriptions, which are parametric descriptions of decomposed sub-regions, are obtained from region clustering. Using these descriptions, likelihoods that a pixel belongs to the object and its background are evaluated. Unlike other active contour extraction schemes, region- and edge-based estimation schemes are integrated into an energy minimization process using log-likelihood functions based on the mixture density descriptions. Owing to the integration, the active contour locates itself precisely to the object boundary for complex background images. Moreover, C^1 discontinuity of the contour is realized as changes of the object sub-regions' boundaries. The experiments show these advantages.

1 Introduction

The objective of this work is to extract an object contour from a given initial rough estimate. Contour extraction is a basic operation in computer vision, and has many applications such as cut/paste image processing of authoring systems, medical image processing, aerial image analyses and so on.

There have been several works which takes an energy minimization approach to the contour extraction(e.g.[1][2]). An active contour model (ACM) proposed by Kass [3] has demonstrated an interactivity with a higher visual process for shape corrections. It results in smooth and closed contour through energy minimization. The ACM, however, has the following problems:

Control : It is common for typical ACMs to look for maxima in intensity gradient magnitude. In complex images, however, neighboring and stronger edges may trap the contour into a false, unexpected boundary. Moreover, if an initial contour is placed too far from an object boundary, or if there is not sufficient gradient magnitude, the ACM will shrink into a convex closed curve even if the object is concave. In order to avoid these cases, a spatially smoothed edge representation[3], a distance-to-edge map[4][5], successive lengthening of an active contour[6] and an internal pressure force[5] have been introduced. Unfortunately, even if those techniques were applied, the edge-based active contours might be trapped by unexpected edges.

Scaling : Many optimization methods including ACMs suffer from a scaling problem[7]. Even if we can set parameters of an optimization method through experiments for a group of images, the validity of the parameters is not assured for images with different contrasts.

Discontinuity : The original ACM requires the object contour to be C^1 continuous. This requirement implies the approximation errors for C^1 discontinuous boundaries which often appear due to occlusion. A popular approach to the discontinuity control is that

the discontinuities are just set at the high curvature points[4]. Generally, however, it is difficult to interpret the high curvature points as corners or occluding points automatically without knowledge about the object.

Taking into account the interactivity for correction, we adopt an ACM for the object contour extraction. In light of the above problems, we will focus not on the ACM itself but on an underlying image structure which guides an active contour precisely to the object boundary against the obstacles. In this paper we propose a *mixture density description* as the underlying image structure. *Mixture density* has been noted to give mathematical basis for clustering in terms of distribution characteristics[8][9][10]. It refers to a probability density for patterns which are drawn from a population composed of different classes. We introduce this notion to a low level image representation. The features of our approach are:

- Log-likelihoods that a pixel belongs to an object(inside) and its background(outside) regions enforce the ACM to converge to their equilibrium. On the other hand, the log-likelihood gradients enforce the ACM to locate the contour on the precise object boundary. They are integrated into an energy minimization process. These log-likelihood functions are derived from the mixture density descriptions.
- Parameter setting for the above enforcing strengths is robust for a variety of images with various intensity. Because, the strengths are normalized by statistical distributions described in the mixture density description.
- The object boundary is composed of sub-regions' boundaries. The C^1 discontinuity of the ACM can be represented as changes of sub-regions' boundaries.

First, we present a basic idea of the mixture density description, its definition with assumptions and a region clustering. Second, we describe the active contour model based on the mixture density descriptions. The experimental results are presented thereafter.

2 Mixture Density Description by Region Clustering

2.1 Basic Idea

Our ACM seeks for a boundary between an object and its background regions according to their mixture density descriptions. The mixture density descriptions describe the positions and measurements of the sub-regions in the both sides of the object boundary. In our approach, the mixture density description can be obtained from a region clustering in which pixel positions are considered to be a part of features. Owing to the combination of the positions and the measurements, the both side regions can be decomposed into locally distributed sub-regions. Similarly, Izumi et al.[11] and Crisman et al.[12] decomposed a region into sub-regions and integrated the sub-regions' boundaries into an object boundary. We do not take such boundary straightforwardly because they may be not precise and jagged for our purpose.

For a pixel to be examined, by selecting a significant sub-region, which is nearest to the pixel with respect to the position and the measurement, we can evaluate *position-sensitive* likelihoods of inside and outside regions. Fig. 1 illustrates an example of mixture density descriptions. In Fig.1, suppose that region descriptions were not for decomposed sub-regions. The boundary between inside "black" and outside "blue" might not be correctly obtained, because the both side regions include blue components and the likelihoods of the both side regions would not indicate significant difference. On the other hand,

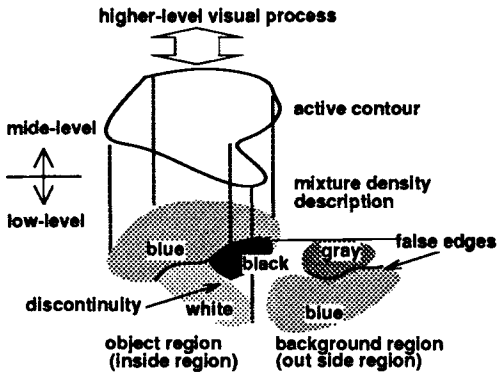


Fig. 1. Example of mixture density description.

using the mixture density description, the likelihoods of the both regions can indicate significant difference knowing the sub-regions' positions. Moreover, the false edges can be canceled by the equal likelihoods in the both sides of the false edges.

2.2 Definitions and Assumptions

We introduce the probability density function for mixture densities [8]. In the mixture densities, it is known that the patterns are drawn from a population of c classes. The underlying probability density function for class ω_i is given as a conditional density $p(\mathbf{x}|\omega_i, \boldsymbol{\theta}_i)$ where $\boldsymbol{\theta}_i$ is a vector of unknown parameters. If a priori probability of class ω_i is known as $p(\omega_i), \sum_{i=1}^{c} p(\omega_i) = 1$, a density function for the mixture densities can be written as:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^{c} p(\mathbf{x}|\omega_i, \boldsymbol{\theta}_i) p(\omega_i) \ , \tag{1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_c)$. The conditional densities $p(\mathbf{x}|\omega_i, \boldsymbol{\theta}_i)$ are called *component densities*, the priori probabilities $p(\omega_i)$ are called *mixing parameters*.

We assume that a region of interest $\mathbf{R}$ consists of c classes such that $\mathbf{R} = \{\omega_1, \omega_2, ..., \omega_c\}$. In order to evaluate the likelihood that a pixel belongs to the region $\mathbf{R}$, we take a model that overlapping of the component densities are negligible and the most significant (highest) component density can dominate the pixel's mixture densities. Thus, the mixing parameters of the region $\mathbf{R}$ should be conditional with respect to position $\mathbf{p}$(e.g. row,column values) and measurements $\mathbf{x}$(e.g. RGB values) of the pixel.

The mixing parameters are given by:

$$p(\omega_i|\mathbf{R}, \mathbf{x}, \mathbf{p}) \approx \begin{cases} 1 \text{ if } p(\mathbf{x}, \mathbf{p}|\omega_i, \boldsymbol{\theta}_i) > p(\mathbf{x}, \mathbf{p}|\omega_j, \boldsymbol{\theta}_j), \forall j, j \neq i, 1 \leq i, j \leq c \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

Thus, we can rewrite the mixture density function of the region $\mathbf{R}$ for $(\mathbf{x}, \mathbf{p})$ as:

$$p(\mathbf{x}, \mathbf{p}|\mathbf{R}, \boldsymbol{\theta}) = \sum_{i=1}^{c} p(\mathbf{x}, \mathbf{p}|\omega_i, \boldsymbol{\theta}_i) p(\omega_i|\mathbf{R}, \mathbf{x}, \mathbf{p}) \ . \tag{3}$$

Owing to the approximation of (2), a log-likelihood function of a region $\mathbf{R}$ with a mixture density description $\boldsymbol{\theta}$ can be given by:

$$\ln p(\mathbf{x}, \mathbf{p}|\mathbf{R}, \boldsymbol{\theta}) = \max_i \ln p(\mathbf{x}, \mathbf{p}|\omega_i, \boldsymbol{\theta}_i) \ . \tag{4}$$

For convenience of notation, we introduce the notation $\mathbf{y}$ to represent the joint vector of $\mathbf{x}$ and $\mathbf{p}$: $\mathbf{y} = \begin{pmatrix} \mathbf{x} \\ \mathbf{p} \end{pmatrix}$. We assume that the component densities take general multivariate normal densities. According to this assumption, the component densities can be written as:

$$p(\mathbf{y}|\omega_i, \boldsymbol{\theta}_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp[-\frac{1}{2}\chi^2(\mathbf{y}; \mathbf{u}_i, \Sigma)] \ , \tag{5}$$

and

$$\chi^2(\mathbf{y}; \mathbf{u}_i, \Sigma_i) = (\mathbf{y} - \mathbf{u}_i)^t \Sigma_i^{-1} (\mathbf{y} - \mathbf{u}_i) \ , \tag{6}$$

where $\boldsymbol{\theta}_i = (\mathbf{u}_i, \Sigma_i)$, d, $\mathbf{u}_i$ and Σ_i are the dimension of $\mathbf{y}$, the mean vector and the covariance matrix of $\mathbf{y}$ belonging to the class ω_i, $(\cdot)^t$ denotes the transpose of a matrix, and $\chi^2(\cdot)$ is called a Mahalanobis distance function. For the multivariate normal densities, the mixture density description is a set of means and covariance matrices for c classes:

$$\textbf{Mixture Density Description:} \boldsymbol{\theta} = ((\mathbf{u}_1, \Sigma_1), (\mathbf{u}_2, \Sigma_2), ..., (\mathbf{u}_c, \Sigma_c)) \ . \tag{7}$$

The log-likelihood function for the multivariate normal densities is given by:

$$\textbf{Log-Likelihood Function:} l(\mathbf{y}|\mathbf{R}, \boldsymbol{\theta}) = \max_i -\frac{1}{2}(d \ln(2\pi) + \ln|\Sigma_i| + \chi^2(\mathbf{y}; \mathbf{u}_i, \Sigma_i)) \ . \tag{8}$$

Henceforth, we use the log-likelihood function and the mixture density description of (7) and (8).

2.3 Region Clustering

A mixture density description defined by (7) is obtained from decomposed sub-regions. If the number of classes and parameters are all unknown, as noted in [8], there is no analytical singular decomposition scheme. Our algorithm is similar to the Supervising ISODATA[3]. of Carman et al.[14] in which Akaike's information criterion(AIC)[15] is adopted to reconcile the description error with the description compactness, and to evaluate goodness of the decomposition. The difference is that we assume the *general* multivariate normal distribution for the component densities and use a distance metric based on (5), while Carman et. al. assume a multivariate normal distribution with diagonal covariance matrices and used Euclidian distance metric.

By eliminating the constant terms from the negated logarithm of (5), our distance metric between sample $\mathbf{y}$ and class ω_i is given by:

$$d(\mathbf{y}; \mathbf{u}_i, \Sigma_i) = \ln|\Sigma_i| + \chi^2(\mathbf{y}; \mathbf{u}_i, \Sigma_i) \ . \tag{9}$$

For general multivariate normal distributions (with no constraints on covariance matrices), AIC is defined as:

$$AIC(\mathbf{R}, \boldsymbol{\theta}) = -2\sum_{\mathbf{y}} l(\mathbf{y}|\mathbf{R}, \boldsymbol{\theta}) + 2||\boldsymbol{\theta}|| = \sum_{i=1}^{c} \{n_i(d(1+\ln(2\pi)) + \ln(|\Sigma_i|)) + (2d + d(d+1))\} \ , \tag{10}$$

[3] For review of clustering algorithms see [13].

where n_i and $||\theta||$ represent the number of samples $\mathbf{y}$ in the ith class and the degree of free parameters to be estimated, respectively (note that $E(\chi^2(\cdot)) = d$). In (10), the first term is the description error while the second term is a compactness measure which increases with the number of classes.

The algorithm starts with a small number of clusters, and iterates split-and-merge processes until the AIC has minima or other stopping rules, such as the number of iterations, the number of class samples and intra-class distributions, are satisfied.

3 Active Contour Model based on the Mixture Density Descriptions

We assume that θ^{in}: a mixture density description of an object region $\mathbf{R}^{\text{in}}$ and θ^{out}: a mixture density description of the object background $\mathbf{R}^{\text{out}}$ have been given by the region clustering. According to a maximum likelihood estimate (MLE) which maximize the sum of the two log-likelihoods of $\mathbf{R}^{\text{in}}$ and $\mathbf{R}^{\text{out}}$, an active contour can be localized to a balanced position as illustrated in Fig.2. In addition, we will provide an edge-based estimate for precise boundary estimation. Because there are vague features or outliers caused by shading, low-pass filtering and other contaminatiors. Assuming that the log-likelihoods indicate step changes at the boundary, the boundary position can be estimated as the maxima of the log-likelihood gradients. In our ACM, the both estimates are realized by a region force and an edge force, and they are integrated into an energy minimization process[4].

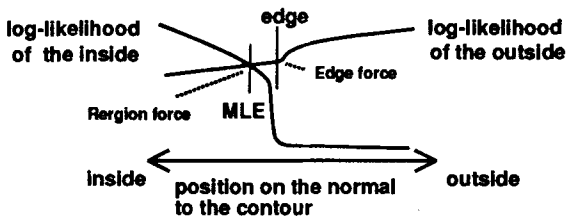


Fig. 2. Boundary location.

3.1 Energy Minimization Process

We employ a dynamic programming(DP) algorithm for the energy minimizing of the ACM. It provides hard constraints to obtain more desirable behavior. Equation (11) is a standard DP form as is of Amini[5] except the external forces.

$$S_i(\mathbf{v}_{i+1}, \mathbf{v}_i) = \begin{cases} \min_{\mathbf{v}_{i-1}} \{S_{i-1}(\mathbf{v}_i, \mathbf{v}_{i-1}) + E_{stretch}(\mathbf{v}_i, \mathbf{v}_{i-1}) + E_{bend}(\mathbf{v}_{i+1}, \mathbf{v}_i, \mathbf{v}_{i-1}) \\ + E_{region}(\mathbf{v}_i) + E_{edge}(\mathbf{v}_i)\} \end{cases}. \tag{11}$$

[4] For more details see[16]

[5] For review of an ACM implemented by a DP see[17].

In (11), we assume that N control points $\mathbf{v}_i$ on a active contour are spaced at equal intervals. The energy minimization process proceeds by iterating the DP for the N control points until the number of position changes of the control points converges to a small number or the number of iteration exceeds a predefined limit. At each control point $\mathbf{v}_i$, we define two coordinates $\mathbf{t}_i$ and $\mathbf{n}_i$, which are an approximated tangent and an approximated normal to the contour, respectively. In the current implementation, each control point $\mathbf{v}_i$ is allowed to move to itself and its two neighbors as $\{\mathbf{v}_i, \mathbf{v}_i+\xi\mathbf{n}_i, \mathbf{v}_i-\xi\mathbf{n}_i\}$ at each iteration of the DP. For convenience of notation, we express them as {middle, outward, inward }. $E_{stretch}(\cdot)$ and $E_{bend}(\cdot)$ in (11) are the first and the second order continuity constraints, and they are called *internal energy*. In the ACM notation, $E_{region}(\cdot)$ and $E_{edge}(\cdot)$ in (11) are called *external forces*(image forces).

In the following subsections, we will describe the external forces.

3.2 External Forces

Two external forces of our ACM are briefly illustrated in Fig. 3. A region force has the capability to *globally* guide an active contour to a true boundary. After guided nearly to the true boundary, an edge force has the capability to *locally* enforce an active contour to outline the precise boundary.

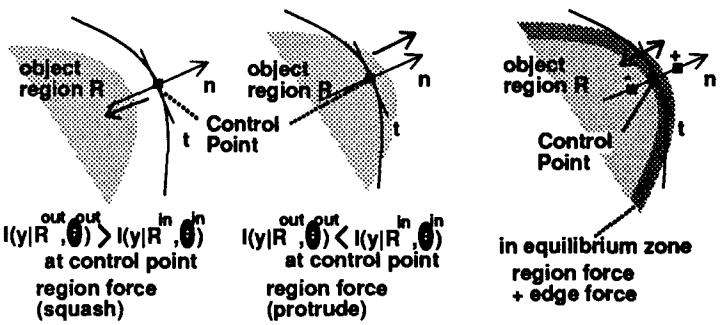


Fig. 3. Boundary seeking model: a region force and an edge force.

The region force is given by:

$$-E_{region}(\mathbf{v}_i) = \begin{cases} \rho_{out}(\mathbf{v}_i) & \text{if } \mathbf{v}_i \text{ moves inward} \\ \rho_{in}(\mathbf{v}_i) & \text{if } \mathbf{v}_i \text{ moves outward} \\ 0 & \text{otherwise (if } \mathbf{v}_i \text{ stays in the middle),} \end{cases} \tag{12}$$

and

$$\rho_{out}(\mathbf{v}_i) = -\rho_{in}(\mathbf{v}_i) = l(\mathbf{y}|\mathbf{R}^{\text{out}}, \theta^{\text{out}})_{\mathbf{v}_i} - l(\mathbf{y}|\mathbf{R}^{\text{in}}, \theta^{\text{in}})_{\mathbf{v}_i}\ , \tag{13}$$

where $l(\cdot)_{\mathbf{v}}$ denotes the log-likelihood taking feature vectors at or nearly at the position $\mathbf{v}$.

In order to introduce the edge force, two auxiliary points $\mathbf{v}_i^+$ and $\mathbf{v}_i^-$ are provided for the control point $\mathbf{v}_i$, where $\mathbf{v}_i^+ = \mathbf{v}_i + \eta\mathbf{n}_i, \mathbf{v}_i^- = \mathbf{v}_i - \eta\mathbf{n}_i$. The inside parameter $\theta_{\mathbf{v}_i}^-$ is selected from θ^{in} so that $\theta_{\mathbf{v}_i}^-$ is the parameter of the highest component density at $\mathbf{v}_i^-$.

The outside parameter $\boldsymbol{\theta}^{+}_{\mathbf{v}_i}$ is selected at $\mathbf{v}_i^{+}$ from $\boldsymbol{\theta}^{\text{out}}$ in the same way. The edge force is given by:

$$-E_{edge}(\mathbf{v}_i) = \begin{cases} \frac{\partial W(t,u;\sigma_t,\sigma_n)*\chi^2(\mathbf{y};\boldsymbol{\theta}^{-}_{\mathbf{v}_i})}{\partial u} + \frac{\partial W(t,u;\sigma_t,\sigma_n)*\chi^2(\mathbf{y};\boldsymbol{\theta}^{+}_{\mathbf{v}_i})}{\partial(-u)} & \text{if } \mathbf{v}_i \text{ is in equilibrium} \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where W, t and u are a window function and variables on $\mathbf{t}_i$, $\mathbf{n}_i$ coordinates of $\mathbf{v}_i$, respectively. In the current implementation, W has been implemented as a Gaussian window(σ_t and σ_n are scale factors). In (14), differential of log-likelihoods is calculated from Mahalanobis distances. Note that the edge force can be applied on condition that $\mathbf{v}_i^{+}$ and $\mathbf{v}_i^{-}$ are placed across the true boundary. We call this condition *equilibrium* condition which can be easily determined by likelihoods of $\mathbf{v}_i^{+}$ and $\mathbf{v}_i^{-}$.

3.3 Discontinuity Control

In our ACM, the second order continuity constraint will be ignored at point where the $\mathbf{v}_i$ is in equilibrium and the inside parameter $\boldsymbol{\theta}^{-}_{\mathbf{v}_i}$ is different from its neighbors $\boldsymbol{\theta}^{-}_{\mathbf{v}_{i-1}}$ or $\boldsymbol{\theta}^{-}_{\mathbf{v}_{i+1}}$.

4 Experiments

Throughout the experiments, the conditions are 1.input image size:512 × 480 pixel RGB , 2. feature vectors: five dimensional vectors (r,g,b,row,column).

Fig. 4 shows an input image with an initial contour drawn by a mouse. The initial contour is used as a "band" which specifies the inside and outside regions. According to the band description, the region clustering result is modified by splitting classes crossing the band. Given an initial contour in Fig.4, we have obtained 24 classes for the inside and 61 classes for the outside through the region clustering. Mixture density descriptions are obtained from the inside and outside classes. Using these mixture density descriptions, the ACM performs the energy minimizing to outline the object boundary. Fig. 5 shows that the discontinuous parts are precisely outlined with the discontinuity control.

Fig. 6 (a) shows a typical example of a trapped edge-based active contour. In contrast with (a), (b) shows a fair result against the stronger background edges.

We can apply our ACM to an object tracking by iterating the following steps:1) extract a contour using the ACM, 2) refine mixture density descriptions using the extracted contour, 3) apply the region clustering to a successive picture taking the mixture density description into the initial class data. In Fig.6 (c), the active contour tracked the boundary according to the descriptions newly obtained from the previous result of (b).

5 Conclusion

We have proposed the precise contour extraction scheme using the mixture density descriptions. In this scheme, region- and edge-based contour extraction processes are integrated. The ACM is guided against the complex backgrounds by the region force and edge force based on the log-likelihood functions. Owing to the statistical measurement, our model is robust for parameter setting. Throughout the experiments including other pictures, the smoothing parameter has not been changed. In addition, the mixture density descriptions have enabled to represent the C^1 discontinuity. Its efficiency is also demonstrated in the experiment.

Regarding the assumptions for the mixture density description, we have assumed that the component densities take general multivariate normal densities. To be exact, the position vector $\mathbf{p}$ is not in accordance with normal distribution. So far, however, the assumption has not exhibited any crucial problems.

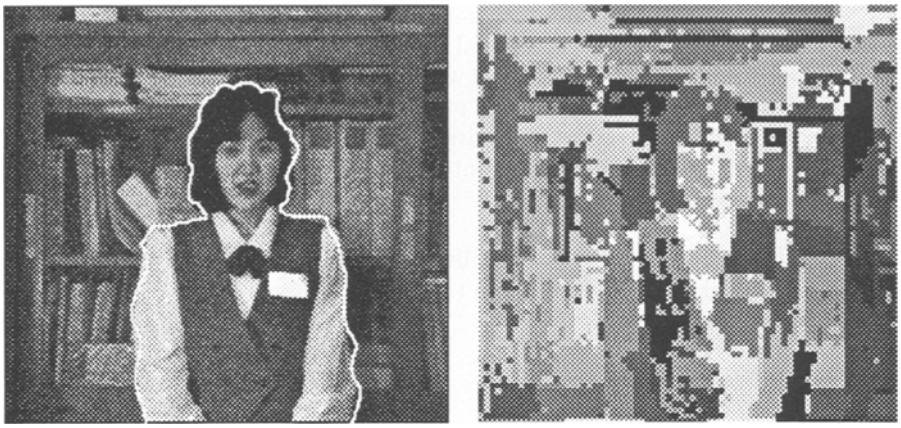
Further work is needed in getting initial contours in more general manner. Issues for feature research include the initial contour estimation and extending our scheme to describe a picture sequence.

Acknowledgments: The authors would like to thank Yoshihiro Fujiwara and Takahiro Yamada for their giving the chance to do this research.

References

1. Blake, A., Zisserman, A.: *Visual Reconstruction.* The MIT Press (1987)
2. Mumford, D., Shah, J.: Boundary Detection by Minimizing Functionals. Proc. CVPR'85 (1985) 22–26
3. Kass, M., Witikin, A., Terzopoulos, D.: SNAKES: Active Contour Models. Proc. 1st ICCV (1987)259–268
4. Menet, S., Saint-Marc, P., Mendioni, G.: B-snakes: Implementation and Application to Stereo. Proc. DARPA Image Understanding Workshop '90 (1990) 720–726
5. Cohen, L., Cohen, I.: A Finite Element Method Applied to New Active Contour Models and 3D Reconstruction from Cross Sections. Proc. 3rd ICCV (1990) 587–591
6. Berger, M., Mohr, R.: Towards Autonomy in Active Contour Models. Proc. 11th ICPR (1990) 847–851
7. Dennis, J., Schnabel, R.: *Numerical Methods for Unconstrained Optimization and Linear Equations.* Prentice-Hall (1988)
8. Duda, R., Hart, P.: *Pattern Classification and Scene Analysis.* John Wiley and Sons (1973)
9. Scolve, S.: Application of the Conditional Population-Mixture Model to Image Segmentation. IEEE Trans. on Patt. Anal. & Mach. Intell. **5** (1983) 429–433
10. Yarman-Vural, F.: Noise,Histogram and Cluster Validity for Gaussian-Mixtured Data. Pattern Recognition **20** (1987) 385–501
11. Izumi, N., Morikawa, H., Harashima, H.: Combining Color and Spatial Information for Segmentation. IEICE 1991 Spring Nat. Convention Record, Part 7 (1991) 392 (in Japanese)
12. Crisman, J., Thorpe, C.: UNSCARF, A Color Vision System for the Detection of Unstructured Roads. Proc. Proc. Int. Conf. on Robotics & Auto. (1991) 2496–2501
13. Jain, A., Dubes, R.: *Algorithms for Clustering Data.* Prentice-Hall (1988)
14. Carman, C., Merickel,M: Supervising ISODATA with an Information Theoretic Stopping Rule. Pattern Recognition **23** (1990) 185-197
15. Akaike, H.: A New Look at Statistical Model Identification. IEEE Trans. on Automat. Contr. **19** (1974) 716–723
16. Etoh, M., Shirai, Y., Asada, M.: Active Contour Extraction by Mixture Density Description Obtained from Region Clustering, SIGPRU Tech. Rep. 91-81. IEICE of Japan (1991)
17. Amini, A., Weymouth, T., Jain, R.: Using Dynamic Programming for Solving Variational Problems in Vision. IEEE Trans. on Patt. Anal. & Mach. Intell. **12** (1990) 855–867

This article was processed using the LaTeX macro package with ECCV92 style

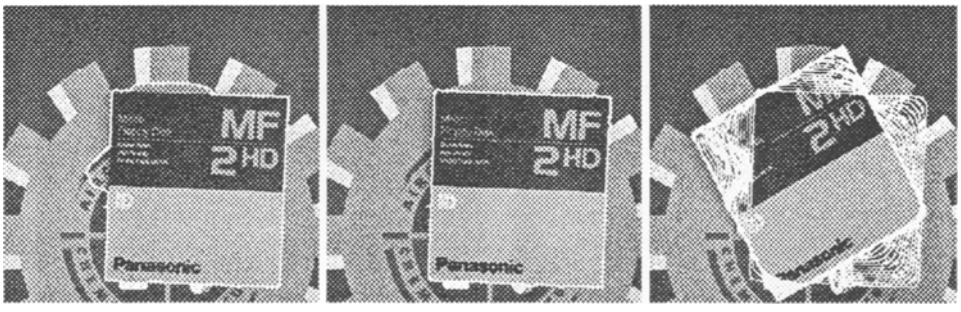


(a) an initial Contour (b) a clustering result

Fig. 4. An example image.

(a) without the discontinuity control (b) with the discontinuity control

Fig. 5. Extracted contours of Fig.4.



(a) an edge-based AC (b) the proposed AC (c) loci to another picture

Fig. 6. Comparison with a conventional edge-based ACM and an application to tracking.