Segmentation and Object Detection with Gabor Filters and Cumulative Histograms

Tadayoshi SHIOYAMA, Haiyuan WU and Shigetomo MITANI
Department of Mechanical and System Engineering
Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
E-mail: shioyama@ipc.kit.ac.jp

Abstract

This paper proposes an algorithm for segmentation and extracting an object region by using Gabor filters. Gabor filters are exploited to extract spatial frequency in some orientations, and not only the outputs of Gabor filters but also color informations are used to construct the features at each image pixel. The criterion is devised so as to consider the similarity, the region size and the region shape factors in order to efficiently merge the features. In general, a complex object may be segmented into multiple regions. However for purpose of detecting such complex object, we represent the object region by the normalized cumulative histogram of features. From experimental results, it is found that the proposed algorithm is able to efficiently detect the object regions such as cars in images of usual traffic scenes.

1 INTRODUCTION

The detection of objects in images with complex background is an important subject in the computer vision and the scene analysis, but yet unresolved problem. This problem has been studied by many researchers[1][2][3][4]. As indicated by Jain et al.[1], the problem of object detection is considered as an equivalent problem of segmenting images on the basis of the discrimation between, and the spatial continuity of, local image features. The efficiency of segmentation methods depends on the types of features used and the criteria used for merging the features.

As a useful set of features, image textures are used for segmenting images, but they possess spatial characteristics at both local and global scales. Hence a multi-channel filtering approach has been focused[5], and a bank of Gabor filters has been used by Jain et al.[1]. In this paper, we use Gabor filters in order to extract spatial characteristics such as spatial frequencies in some orientations at each image pixel. We construct the features at image pixel from not only the outputs of Gabor filters but also color informations. The criterion is devised to efficiently merge the extracted features. In general, a complex object may be segmentred into multiple regions. However, we propose the method for detecting such complex object region by representing the object region with histograms[6]of features. The differences of our work from that by Jain et al. are the criterion for merging features and the method for detecting object region. The difference of our work from of that of [6] is that we use the normalized cumulative histogram while in [6] histogram is used. In order to evaluate the proposed method for segmentation and detection of object, experimental results are shown for images of usual traffic scenes.

2 GABOR TRANSFORMATION

Let f(x,y) be the intensity at the coordinate (x,y) in a gray scale image. Denoting by θ the counterclockwise angle from the positive x-axis, the Gabor transformation z(x,y) of f(x,y) in the direction of $\theta = 0^o$ is given by

$$z(x,y) \quad = \quad \int_{-\infty}^{\infty} f(x-t,y-s)exp\{-\frac{1}{2}(\frac{t^2}{\sigma_x^2}+\frac{s^2}{\sigma_y^2})\}cos(2\pi u_0 t)dtds, \qquad (1)$$

where u_0 represents the radial frequency and σ_x and σ_y the space constants of the filter. The Fourier transformation of equation (1) is as follows:

$$Z(u,v) = F(u,v)G(u,v), \tag{2}$$

$$G(u,v) = A(exp\{-\frac{1}{2}[\frac{(u-u_0)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}]\}+exp\{-\frac{1}{2}[\frac{(u+u_0)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}]\}), \tag{3}$$

$$Z(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z(x,y)exp\{-2\pi i(ux+vy)\}dxdy, \tag{4}$$

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)exp\{-2\pi i(ux+vy)\}dxdy, \tag{5}$$

where $A \equiv 2\pi\sigma_x\sigma_y$, $\sigma_u \equiv 1/2\pi\sigma_x$, $\sigma_v \equiv 1/2\pi\sigma_y$ and $i \equiv \sqrt{-1}$. The Gabor filter is a band pass filter near the spatial frequency point $(\pm u_0, 0)$ as shown in equation (3). Gabor transformation with arbitrary orientation θ is obtained from the filter by rotation of its x-y coordinates. The output of the filter represents the energy at spatial frequency u_0 in an orientation θ in the original image.

3 FEATURE VECTORS

Let $z(x,y;u_0,\theta)$ be the output of a selected Gabor filter with parameters u_0 and θ. In the sequel, we use a nonlinear transformation $P(x,y;u_0,\theta)$ to which $z(x,y;u_0,\theta)$ is subject;

$$P(x,y;u_0,\theta) = [1+tanh\{(\mid z(x,y;u_0,\theta)\mid^2 -\alpha)/\beta\}]/2. \tag{6}$$

The function $P(x,y;u_0,\theta)$ takes the value in the range (0,1) and represents the intensity at the frequency u_0 in the orientation θ. For notational convenience, we denote by $P(u_0,\theta)$ the function $P(x,y;u_0,\theta)$. Parameters u_0 and θ are set as $u_0 = Low(=1/32)$, $u_0 = High(=1/8)$ and $\theta = (0^o, 45^o, 90^o, 135^o)$. These parameters are empirically determined from experimental results of segmentation.

Figure 1 shows $\{P(u_0,\theta), u_0 = Low, High, \theta = 0^o, 45^o, 90^o$ and $135^o\}$ as an intensity image for an original real image (256×256 pixel) where

$$\begin{aligned} Low &: u_0 = 1/32,\ \sigma_u = 1/64,\ \sigma_v = 1/128,\ \alpha = 50000,\ \beta = 75000, \\ High &: u_0 = 1/8,\ \sigma_u = 1/16,\ \sigma_v = 1/32,\ \alpha = 200, \beta = 300. \end{aligned}$$

The feature vector $\vec{P}(x,y)$ at a coordinate (x,y) in an image is constructed from $P(x,y;u_0,\theta)$ as follows:

$$\vec{P}(x,y) = \{P(x,y;u_0,\theta), u_0 = Low, High, \theta = 0^o, 45^o, 90^o, 135^o\}.$$

4 SEGMENTATION

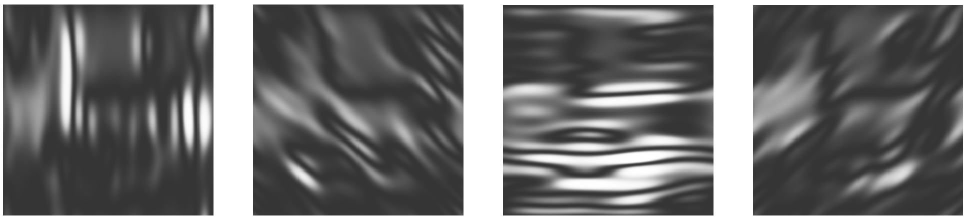
For segmentation, we use both the feature vector $\vec{P}(x,y)$ and the color vector $\vec{C}(x,y)$. The color vector $\vec{C}(x,y)$ is defined as the vector whose component is given by the polar coordinates in the H-S plane in the HLS(Hue-Lightness-Saturation) color model. The segmentation is carried out by merging similar regions neighboring to a considered region.

First, the initial regions are generated in the following steps:
(1)Assign a label to an unlabelled pixel found by scanning as illustrated in Fig.2 where a label 2 is assigned.
(2)Compute the distanceD^2 between the last labelled pixel and its 4-neighboring pixel, and assign the same label to the pixel where distance is less than a threshold T_1;

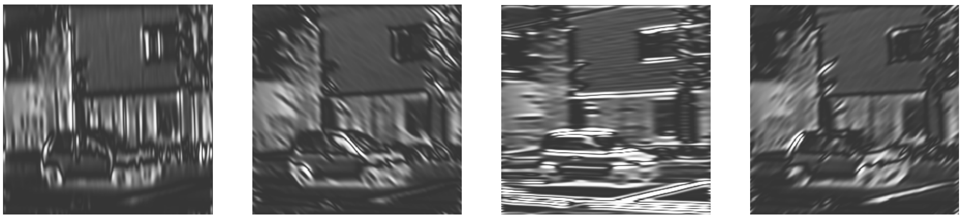
$$D^2 \equiv \parallel \vec{P}(a,b) - \vec{P}(c,d) \parallel^2 + \eta \parallel \vec{C}(a,b) - \vec{C}(c,d) \parallel^2, \tag{7}$$

where (a,b) and (c,d) are the coordinates of pixels in an original image, and η denotes a weighting coefficient.
(3)Iterate the same process as the step 2 for the pixel which is assigned the label in the step 2. When there is no pixel which is to be assigned the same label, go to the step 4.
(4)If all pixels are already labelled, then stop. Otherwise, return to step 1.

(a) Original image (256×256pixel).



(b) P(Low,0^o). (d) P(Low,45^o). (f) P(Low,90^o). (h) P(Low,135^o).



(c) P(High,0^o). (e) P(High,45^o). (g) P(High,90^o). (i) P(High,135^o).

Figure 1: The example of the intensities $P(u_0, \theta)$.

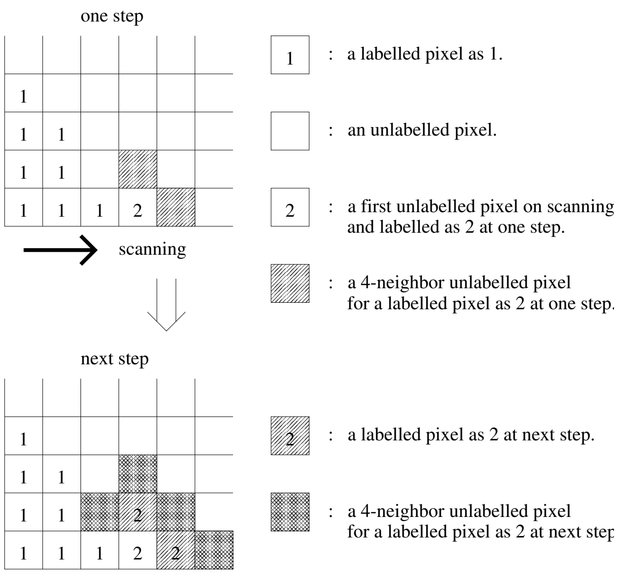


Figure 2: The method generating initial regions.

For each region, we compute the average of the feature vector, and also compute the center of gravity and the second central moments such as covariances from the coordinates of pixels belonging to the region. Utilizing these informations, we efficiently merge regions by the following algorithm:
(1)For each region i, find a candidate region for merging which satisfies the following condition:

$$min_j \ E_{ij} \quad \equiv \quad U_{ij}^2\{R_{ij} + \kappa(M_{ij}^2 + M_{ji}^2)\} \ \ and \ \ U_{ij}^2 < T_2, \tag{8}$$

where κ and T_2 are constants, and

$$U_{ij}^2 \quad \equiv \quad \| \vec{P}_i - \vec{P}_j \|^2 + \eta \| \vec{C}_i - \vec{C}_j \|^2, \tag{9}$$

$\vec{P}_i$, $\vec{C}_i$: the averages of $\vec{P}$ and $\vec{C}$ in region i,

$$R_{ij} \quad \equiv \quad n_i n_j/(n_i + n_j), \tag{10}$$

n_i:the number of pixels in region i,

$$M_{ij}^2 \quad \equiv \quad \frac{1}{(1-\rho_i^2)V_{x_i}V_{y_i}}\{V_{y_i}(x_i - x_j)^2 - 2\rho_i\sqrt{V_{x_i}V_{y_i}}(x_i - x_j)(y_i - y_j) + V_{x_i}(y_i - y_j)^2\}, \tag{11}$$

$$\begin{aligned} (x_i, y_i) \quad &: \quad the\ center\ of\ gravity\ of\ region\ i, \\ V_{x_i}, V_{y_i} \quad &: \quad variances\ in\ x-axis\ and\ y-axis\ in\ region\ i, \\ \rho_i \quad &: \quad correlation\ coefficient\ in\ region\ i, \\ M_{ij}^2 \quad &: \quad Maharanobis\ distance\ from\ region\ i\ to\ region\ j. \end{aligned}$$

(2)Merge two regions which are mutually candidates each other. Compute the average of feature vector and the second central moments for the new region after merging. In Fig.3, regions 1 and 2 are mutually candidates each other and the new region is made from the two regions.
(3)Iterate the procedure of steps 1 and 2 until there is no region which satisfies the condition (8).

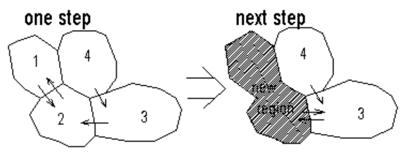


Figure 3: The merging method.

The criterion E_{ij} includes the factor of similarity U_{ij}^2, the size factor R_{ij} and the Maharanobis distance M_{ij}^2 which depends on the shape of the region. In the algorithm, the region with small size (i.e. small R_{ij}) is superiorly merged , and the merging process is dependent on the shape of region by using Maharanobis distances.

The results of segmentation by the above algorithm are shown in Fig.4 for the case in Fig.1. Figure 4(b) and (c) show the segmentations at the initial stage and the final stage, respectively as the average intensity of each region. Figure 4(d) shows the result (c) with region boundaries.

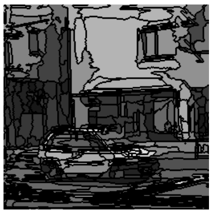
5 CUMULATIVE HISTOGRAM

As shown in Fig.4, in general an object is segmented into multiple regions. Hence, we use the cumulative histogram for representing an object in order to detect the obeject. Given an image which includes an object as a model, we segment the image and compute the feature vectors for each region. We set a rectangular window so as to fit the object, compute the normalized cumulative histogram of feature for the regions belonging to the window. As a feature, the components of $\vec{P}_j$(the average feature vector of

(a) Original image.

(b) At the initial stage.

(c) At the final stage.



(d) With boundary.

Figure 4: The results of segmentation for the case in Fig.1: $T_1 = 0.02,\ T_2 = 0.2,\ \eta = 4,\ \kappa = 1000$, where 13186 regions at the initial stage decreased to 1021 at the final stage.

region j) are used.

Denote by $P_j(u_0, \theta)$ the average intensity at frequency u_0 in orientation θ in the region j. For each parameter (u_0, θ), we obtain the cumulative histogram and normalize it deviding by the sum of pixel number of all regions in the window. The normalized cumulative histogram $H(\gamma; u_0, \theta)$ is given by

$$H(\gamma; u_0, \theta) = \frac{1}{N} \sum_{j \in Q(\gamma)} n_j, \tag{12}$$

$$Q(\gamma) = \{j \in W : P_j(u_0, \theta) < \gamma\}, \tag{13}$$

$$\gamma \in \{\xi \ell : \ell = 1, 2, ..., \ell_{max}\},\ \xi \equiv 1/\ell_{max}, \tag{14}$$

$$N \equiv \sum_{j \in W} n_j, \tag{15}$$

$$W \equiv \{j : (x_j + \sqrt{V_{x_j}} < x_w + \frac{W_x}{2})\ and\ (x_j - \sqrt{V_{x_j}} > x_w - \frac{W_x}{2})$$

$$and\ (y_j + \sqrt{V_{y_j}} < y_j + \frac{W_y}{2})\ and\ (y_j - \sqrt{V_{y_j}} > y_j - \frac{W_y}{2})\}, \tag{16}$$

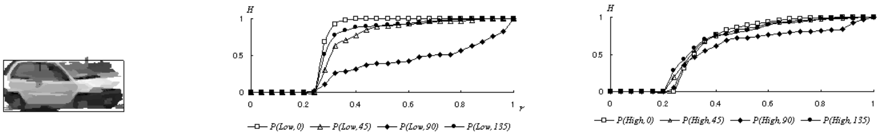
where W_x and W_y are the sizes of the window in x-axis and y-axis, and W is the set of regions in the windows. (x_w, y_w) is the center of window. Figure 5(a) shows the regions in the window obtained from equation (16) for the result in Fig.4. Figure 5(b) and (c) show the normalized cumulative histograms for u_0=Low and u_0=High, respectively.

From these 8 normalized cumulative histograms, we construct the normalized cumulative histogram model for the model object.

Denote by $H_M(\gamma; u_0, \theta)$ the normalized cumulative histogram for the model and by $H_T(\gamma; u_0, \theta)$ that for a test window. Then, the distance between the model and the test window is given by

$$S^2 = \sum_{\gamma, u_0, \theta} \{H_M(\gamma; u_0, \theta) - H_T(\gamma; u_0, \theta)\}^2. \tag{17}$$

We define the shape parameter μ as $\mu \equiv W_y/W_x$.



(a)Model region (μ=0.44). (b) Low side (ξ=0.04). (c) High side (ξ=0.04).

Figure 5: The normalized cumulative histogram for the model region.

6 OBJECT DETECTION

In this sectin, we extract an object region in an arbitrary image by searching the rectangular window which has the most similar noramalized cumulative histogram to that for the model object. We assume that the shape parameter of rectangular window for a test image (called a test window) is the same μ as that for the model. On sweeping the test window, we set the test window so that the center of the test window coincides with the center of gravity in each region and the initial size of the test window and the increment of the size are dependent on variances of the region. Here, we do not set the window at a small region with V_{x_i} and V_{y_i} less than 1.

The initial sizes $W_x^i(0)$ and $W_y^i(0)$ in x-axis and y-axis directions, respectively at region i are given by

$$\begin{aligned} W_x^i(0) &= max(2\sqrt{V_{x_i}}, L),\ W_y^i(0) = \mu W_x^i(0) \quad for\ V_{x_i} \geq V_{y_i}, \\ W_x^i(0) &= W_y^i(0)/\mu,\ W_y^i(0) = max(2\sqrt{V_{y_i}}, L) \quad for V_{x_i} < V_{y_i}, \end{aligned} \tag{18}$$

where L is a constant. The symbol max(a,b) denotes the maximaum value among a and b. Setting a window such that the center of the window coincides with the center of gravity in region i, we enlarge the size of window in the following manner: for the k-th step

$$W_x^i(k) = W_x^i(0) + kW_x^i(0)/d, \quad W_y^i(k) = W_y^i(0) + kW_y^i(0)/d, \quad k = 1, 2, ..., k_{max}. \tag{19}$$

As the test window we adopt the window defined as W in equation (16) substituting $W_x^i(k)$ and $W_y^i(k)$ for W_x and W_y. Computing the distance S^2 in equation (17) between the test window for the k-th increment and that for the model, we stop the increment at the k_{max}-th step given by

$$k_{max} = sup\{k \geq 3 : \sum_{\ell=2}^{k-1} max(S(\ell)^2 - S(\ell-1)^2, 0) = 0\}, \tag{20}$$

and use $W_x^i(k_{max}-1)$ and $W_y^i(k_{max}-1)$ as the sizes W_x and W_y for a test window. The equation (21) gives the number k_{max} of increment step such that $S(k_{max}-1)^2$ attains to the minimum distance among $S(\ell)^2, \ell = 1, 2, ...$ Figure 6 shows $S(k)^2$ for the case where the window is set so that the center of window coincides with the center of model region in Fig.5, the initial size is 1/20 of the model region and d=5. Since $S(k_{max}-2)^2$, $S(k_{max}-1)^2$ and $S(k_{max})^2$ take similar values, we use as the distance between the test window and that for model, the mean value S_m^2 defined as

$$S_m^2 \equiv \{S(k_{max}-2)^2 + S(k_{max}-1)^2 + S(k_{max})^2\}/3, \tag{21}$$

in order to prevent from the numerical instability due to the discrete computation.

7 EXPERIMENT IN OBJECT DETECTION

Figure 7 shows the experimental results for five test images which include various sizes of an object model. The object model is the same as in Fig.5. In Fig.7, the distance S_m^2 on detection is also shown. As shown in Fig.7(c), (d) and (g), (h), even for small sizes of objects, the objects are detected. From Fig.7 it is found that the proposed method can detect the objects for real traffic scenes.

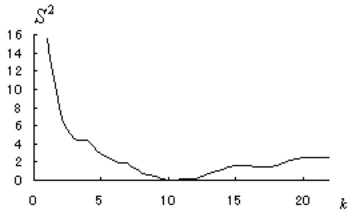


Figure 6: The distance from the model as enlarging window size.

8 CONCLUSIONS

We have proposed an algorithm for segmentation and detecting an object by using Gabor filters. In order to efficiently extract features both local and global scales simultaneously, we exploit Gabor filters. For purpose of precise segmentation, we construct the feature vectors from not only the outputs of Gabor filters but also the color informations at each image pixel. In order to efficiently merge the features, the criterion is devised so as to consider the factor of similarity, the size factor of region and the shape factor of region. Furthermore, to overcome the case where a complex object is in general segmented into multiple regions, we adopt the normalized cumulative histogram of features. From experimental results, the proposed method is found to detect objects for usual traffic scenes.

References

[1] A.K.Jain, N.K.Ratha and S.Lakshmanan, Object detection using Gabor filters, Pattern Recognition **30**, 295-309 (1997).

[2] N.R.Pal and S.K.Pal, A review on image segmentation techniques, Pattern Recognition **29**, 1277-1294 (1993).

[3] C.S.Lu P.C.Chung and C.F.Chen, Unsupervised texture segmentation via wavelet transform, Pattern Recognition **30**, 729-742 (1997).

[4] J.Peng and B.Bhanu, Closed-loop object recognition using reinforcement learning, IEEE Trans. on PAMI **20**, 139-154 (1998).

[5] A.K.Jain and F.Farrokhnia, Unsupervised texture segmentation using Gabor filters, Pattern Recognition **24**, 1167-1186 (1991).

[6] V.V.Vinod and Hiroshi Murase, Focused color intersection with efficient searching for object extraction, Pattern Recognition **30**, 1787-1797 (1997).


(a) Test image A. (b) Extracted region from A: S_m^2=0.08.


(c) Test image B. (d) Extracted region from B: S_m^2=0.67.


(e) Test image C. (f) Extracted region from C: S_m^2=0.23.


(g) Test image D. (h) Extracted region from D: S_m^2=0.81.


(i) Test image E. (j) Extracted region from E: S_m^2=0.43.

Figure 7: The results of object detection: d=5, L=10.