ACCV2002: The 5th Asian Conference on Computer Vision, 23--25 January 2002, Melbourne, Australia

Graph Representation of Images in Scale–Space with Application to Face Detection

Hidenori Maruta, Tatsuo Kozakaya, Yasuharu Koike, Makoto Sato
Precision and Intelligence Laboratory
Tokyo Institute of Technology
4259, Nagatsuta-machi, Midori-ku, Yokohama, Japan
hmaruta@hi.pi.titech.ac.jp

Abstract

In the image recognition problem, it is very important how we represent the image. Considering this, we propose a new representational method of images based on the stability in scale-space. In our method, the image is segmented and represented as a hierarchical region graph in scale-space. The object is represented as feature graph, which is subgraph of region graph. In detail, the region graph is defined on the image with the relation of each segment hierarchically. And the feature graph is determined based on the "life-time" of the graph of the object in scale-space. This "life-time" means how long feature graph lives when the scale parameter is increased. We apply our method to the face detection problem, which is foundmental and difficult problem in face recognition. We determine the feature graph of the frontal human face statistical point of view. We also build the face detection system using this feature graph to show how our method works efficiently.

1. Introduction

In the field of image recognition, it is very important how we represent the image. If we want to build a image recognition system, image processing is depend on how the image is represented. Generally, an image is represented as just a multi–dimensional array of pixel value. If we can develop the image representation method from this array with considering the structure of the object, we can develop a suitable object recognition method. However, especially on object detection problem, almost all researches are based on multivariate statistical or neural network approach. In this approach, the image is simply represented as a high–dimensional vector, and compared with template images which are sometimes very large. These vector-based methods don't consider the structure of objects. So these methods are poor in changes of the object such as relative size in the image, individual variation and so on. To solve these problem, vector-based methods must prepare many templates. But this is expensive in the sense of computation and the disadvantage in computation effects the system directly.

In this paper, we propose a new representational method of an image which is considering the structure of the object. In the proposed method, an image is segmented and represented as a graph structure called *region graph* in scale–space. And the target object in the image is also segmented into some segments and represented as a subgraph of the region graph. To represent the object, we define the *feature graph* of the object by selecting possible subgraphs spanned on the objects, based on the stability in scale–space. We also apply this feature graph representation of the object to face detection problem, which has difficulties with mentioned above.

2 Image segmentation and scale–space

In this section, we consider the image segmentation and scale–space representation of images. Both are well–known methods in image recognition and so many researches are developed.

2.1 Image segmentation

There exists a large number of image segmentation technique. In this paper, we use image segmentation methods, sometimes categorized into region growing, based on *digital flows* defined as follow.

Definition 1 (Digital flow) *[10] A* **digital flow** *between two connected pixels P_i and P_j of an image is defined to be a vector from the higher intensity pixel to the lower intensity pixel with a magnitude of $|P_i - P_j|$.*

Using the digital flow, we can construct a image segmentation method. For constructing this, we define some terms.

Definition 2 (Effective inflow) *[10] A* **effective inflow** *of a pixel is defined in as the steepest digital flow entering to the pixel.*

Definition 3 (Effective outflow) *[10] A* **effective outflow** *of a pixel is defined as the steepest digital flow leaving to the pixel.*

From effective inflows of all pixels, we obtain steepest uphill paths. As we see pixels lying in uphill paths to the same peak are grouped to be the same segment, we obtain the segmented image. We call this segmentation method *hill segmentation*. In parallel using effective outflow, we can obtain *dale segmentation*. Figure 1 shows an example of these segmentation methods. We notice that these segmentation methods have a same property that each segment of a segmented image has one and only one extrema. So we can identify all segments by the extrema of each segment.

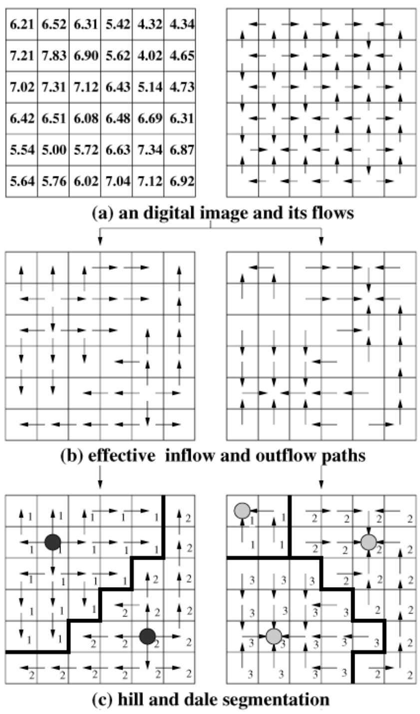

Figure 1. an example of image segmentation

2.2 Scale–Space representation

A scale–space representation([4], [6], [8], [11], [12])is known as the technique to obtain a multiresolution hierarchical representation of image. The scale–space representation of an image is constructed by convolving the original image $f_0(x, y)$ with Gaussian kernels $G(x, y; \sigma)$:

$$\begin{aligned} f(x, y, \tau) &= f_0(x, y) * G(x, y; \sigma) \\ &= \int f_0(\xi, \eta) G(x - \xi, y - \eta; \sigma) d\xi d\eta \quad (1) \end{aligned}$$

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \ \sigma > 0 \quad (2)$$

where σ is called *scale parameter*. We can obtain multi-scale representation of an image varying the scale parameter. This convolved image is also called *blurred image*. It is known that the scale–space representation has many well–behavedness, however, the property of the non-creation of extrema does not hold in 2 or more higher dimensions. As this property, we must consider *how to use* the extrema in scale–space when we want use it.

3 Graph representation of the image in scale–space

As we apply the image segmentation method in scale–space, we obtain a sequence of blurred, segmented images. To utilize these images as a sequence, we consider the representation method of images in scale–space.

3.1 Graph representation of the image

To represent the image, we first define a graph structure (called *region graph*) on a segmented image as follows.

Definition 4 (Region graph) *A* **region graph** *of image is defined as a graph spanned on the segmented image whose nodes are extrema of each segment and connections of nodes(edges) are determined with connections of segments.*

We can obtain the region graph both hill segmentation and dale segmentation. Generally, we can define the region graph on *watersheds-based* segmented images. In scale–space, an image is represented as a sequence of region graph structures with each scale. An object of our interest in a image is also segmented into some segments. So we can represent the object as a subgraph of the region graph spanned on the object. This subgraph representation of the object is unstable with varying the scale parameter because of the property of the scale–space representation. However, if the subgraph is stable in some range of the scale parameter, we obtain the stable subgraph representation of the object. Now we define the stability of the subgraph structure as a *life time*.

Definition 5 (Life time of a subgraph) *A* **life time of a subgraph** *which spanned on the object is defined as the range of the scale parameter such that the subgraph holds its structure continuously.*

Assume that subgraph–structure appears at $\sigma = \sigma_a$ and disappears at $\sigma = \sigma_d$, the lifetime is $\mid \sigma_d - \sigma_a \mid$. We notice that there exists a case that once vanishing structure appears again with varying scale–parameter since the property of non–creation of the extrema. In this paper, we just call life time of a subgraph, *life time*. Based on the life time, we select the suitable subgraph representation of the object that has the longest life time in scale–space.

Definition 6 (Feature graph) *Assume that the subgraph G has the longest life time in subgraphs spanned on the object and nodes of G are given by pixels' coordinates* $P_i = (x_i, y_i)(i = 1, 2, \cdots, n)$*. The* **feature graph** G_f *of the objects is defined the subgraph as follows:*

- G_f *has same connections of nodes as G*
- *each node* $P'_i(i = 1, 2, \cdots, n)$ *of* G_f *is obtained by transforming of* G_f *with the next formula:*

$$P'_i = \frac{P_i - P_G}{\sqrt{\frac{1}{n}\sum_{k=1}^{n} ||P_k - P_G||^2}} \tag{3}$$

$$where\ P_G = (\sum_{k=1}^{n} x_k, \sum_{k=1}^{n} y_k)$$

We notice that, in the definition of the feature graph, we can define the transformation of the feature graph as *Affine transformation*, i.e. $P'_i = \Lambda P_i + \vec{a}$ (where $\det \Lambda > 0$ and $\vec{a}$ is a parallel transport) generally. However, in this paper, we consider the simplest case of transformation and call this transformation *normalization* of the feature graph.

The feature graph representation of the object has following properties.

- As the graph representation, the object has the same structure even if the size in the image is different.
- Considering in scale–space, objects have the same structure in its life time even if there exist individual variations.

The feature graph is expected to provide the representation method that considers the structure of the object. However, there remains a problem that the feature graph of the target object exists or not. This problem is depend on the target object. So we concretely apply the proposed method to the face detection problem in the next section, and show how the feature graph works in this problem efficiently .

4 Application to the face detection

In this section, we apply the proposed representation method to the human–face detection. The face detection([1], [2], [3], [7]) is one of the important problem in the face recognition. It is a difficult problem because there exist size and individual variations of the face. Since the feature graph representation of the image has suitable properties for such difficulties. So we apply the proposed method to this problem and show the feature graph representation works efficiently.

We first consider the feature graph structure of the face statistical point of view and determine the feature graph of the face. Next, we apply the feature graph determined above to detect faces in a test image which includes the plural number of human. Notice that we only treat frontal-faces with non-occlusion because we want to see only how the feature graph representation works.

4.1 Feature graph of the face

To determine the feature graph of the human–face, we investigate subgraphs spanned on the face of sample images with two segmentation methods, *Hill* and *Dale*. Figure 2 shows how the region graph changes in scale–space.

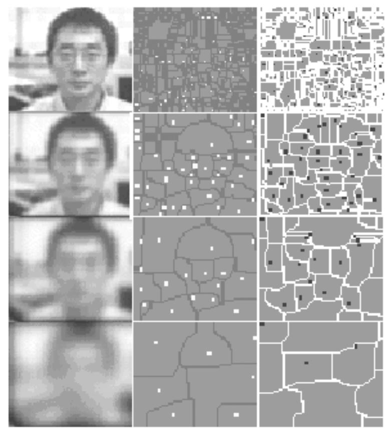

Figure 2. How the region graph changes in scale–space. From right side, bullered images, results of Hill segmentation and results of Dale segmentation. σ=0.0, 1.0, 2.0, 4.0 from up to bottom.

We prepare 19 sample images (Sample set 1, containing 19 persons' facial images) that are 80×60 and include one frontal–face(Figure 3 shows examples of Sample set 1). Varying the scale–parameter σ from 0.0 to 6.0 by 0.1 step, we obtain the region graph representation of each sample image in scale–space. From observations of subgraphs spanned on the face of all sample images, we can select 2 typical subgraphs with Hill segmentation and 3 typical subgraphs with Dale segmentation manually.

Figure 4 is typical subgraphs with Hill segmentation, we call this two subgraphs H_1 with 4–nodes and H_2 with 6–

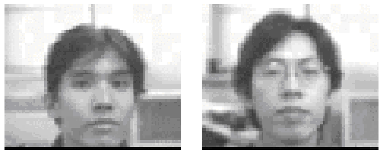

Figure 3. Examples of Sample set 1.

nodes. Figure 5 is typical subgraphs with Dale segmentation, we call this three subgraphs D_1 with 6–nodes, D_2 with 4–nodes and D_3 with 3–nodes. We notice that H_1 is included in H_2 as a subgraph, i.e. $H_1 \subset H_2$, this including property hold between D_1 and D_2 and D_3, i.e. $D_3 \subset D_2 \subset D_1$.

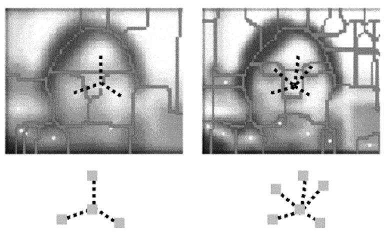

Figure 4. Typical subgraphs with Hill segmentaion, H_1 (left side) and H_2 (right side).

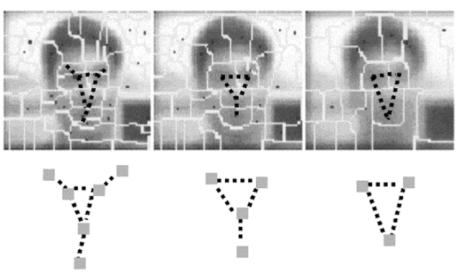

Figure 5. Typical subgraphs with Dale segmentaion, D_1(left side), D_2(middle) and D_3(right side).

To select a suitable subgraph for the feature graph, we investigate the mean value of the *life time* of each subgraph. The result is shown in Table 1 and Table 2. From these result, the subgraph H_1 has suitable properties for the feature graph that (1)H_1 can be observed all samples in Sample set 1, (2)H_1 has the longest mean value of the life time. So we determine that the feature graph of the human–face has the same graph–structure as H_1 and Hill segmentation is the most suitable segmentation method in this case. We also obtain that the common range of the scale parameter σ that H_1 exist in all samples is from $\sigma = 1.0$ to $\sigma = 2.0$. As this result we set the scale parameter as constant, $\sigma = 1.7$ without noticing.

	number of appeared images	mean of life time
H_1	9 (of 19)	0.644
H_2	19 (of 19)	1.372

Table 1. Result of Hill Segmentation. (Number of appeared images represents how many samples include that subgraph.)

	number of appeared images	mean of life time
D_1	11 (of 19)	0.436
D_2	14 (of 19)	0.418
D_3	14 (of 19)	0.700

Table 2. Result of Dale Segmentation. (Number of appeared images represents how many samples include that subgraph.)

Next we consider how to detect the feature graph(= human–face) in the region graph of the image. To detect the feature graph from all possible subgraphs in the region graph, we define the similarity of the selected subgraph to the feature graph at a statistical point of view. We prepare 80 sample images (Sample set 2, containing 5 persons with 16 images per one person) that are 80×60 and include one frontal–face(Figure 6 shows examples of Sample set 2) as the Sample set 1 is. We set nodes $P_i = (x_i, y_i)(i = 1, \cdots, 4)$ of H_1 as shown in Figure 7.

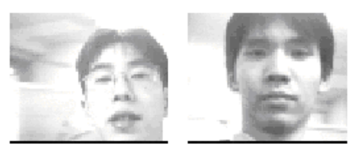

Figure 6. Exmaples of Sample set 2.

For this H_1, the normalization formula defined in **Definition 4** becomes as follows.

$$P_i' = \frac{P_i - P_G}{\sqrt{\frac{1}{4}\sum_{k=1}^{4} ||P_k - P_G||^2}} \qquad (i = 1, \cdots, 4) \tag{4}$$

As we seem the featugraph (*normalized*) H_1 as a 8–dimensional vector $\mathbf{x} = (x_1, y_1, \cdots, x_4, y_4)$, we obtain a

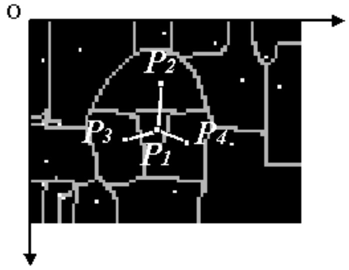

Figure 7. Nodes of H_1

8–dimensional statistical distribution of $\mathbf{x}$ from Sample set 2 (we set P_1 as the origin). Assume that $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrices of the distribution, we obtain

$$\begin{aligned}\boldsymbol{\mu} \;=\; &(0.0290\ 0.2428\ -0.0785\ -1.0504\\ &-1.0420\ 0.4546\ 1.0915\ 0.3530)^t\end{aligned}$$

and

$$\boldsymbol{\Sigma}=$$

$$\begin{bmatrix}
0.02183 & 0.00252 & -0.02113 & 0.00633 & -0.00386 & 0.00104 & 0.00316 & -0.00989 \\
0.00252 & 0.01161 & -0.00411 & 0.00148 & 0.00084 & -0.00535 & 0.00075 & -0.00774 \\
-0.02113 & -0.00411 & 0.06578 & -0.09365 & -0.01843 & -0.00274 & -0.02622 & 0.01621 \\
0.00633 & 0.00148 & -0.00936 & 0.03120 & -0.01848 & -0.01420 & 0.02151 & -0.01848 \\
-0.00386 & 0.00084 & -0.01843 & -0.01848 & 0.02512 & 0.01176 & -0.00283 & 0.00587 \\
0.00104 & -0.00535 & -0.00274 & -0.01420 & -0.01176 & 0.01856 & -0.01006 & 0.00099 \\
0.00316 & 0.00075 & -0.02622 & 0.02151 & -0.00283 & -0.01006 & 0.02589 & -0.01219 \\
-0.00989 & -0.00774 & 0.01621 & -0.01848 & 0.00587 & 0.00099 & -0.01219 & 0.02523
\end{bmatrix}$$

Using $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the *similarity* of subgraphs is defined as *Mahalanobis distance* of this distribution.

Definition 7 (Similarity) *When a $\boldsymbol{G}$ has the same structure as $\boldsymbol{H_1}$, the* **similarity** *of $\boldsymbol{G}$ is defined as follows.*

$$S(G) = (x - \mu)^t \Sigma^\dagger (x - \mu) \tag{5}$$

where x is the vector representation of G and $\Sigma^\dagger$ is Moore–Penrose generalized inverse of Σ.

From the similarity S, we can detect the feature graph from any possible subgraphs of the region graph.

4.2 Experiment

We apply the proposed method to face detection. The test image includs some frontal–faces. The detection of the face is carried out by ordering the similarity of all possible obtained subgraphs in the image.

In this experiment, we prepare 2 test images , image A (Figure 8) and image B (Figure 9). Image A is 160×120 and includes 4 frontal–faces. On the other hand, image B is 320×240 and includes 6 frontal–faces. Because of the relative difference of size between test images and sample images of Sample set 2, we set $\sigma = 1.0$ in the case of image A, and $\sigma = 1.5$ in the case of B. The experimental results are shown in Figure 10 (image A) and Figure 11 (image B) as numbers with circle. The number is order of the subgraph $\boldsymbol{H_1}$ detected on the test image.

From these results, the proposed method works efficiently except one face with both images. One exceptional face has the low similarity because it has just a few inclination. It is thought that the statistical distribution of the feature graph is so shape that one exceptional face become far from this distribution.

Figure 8. Image A. (160×120)

Figure 9. Image B. (320×240)

Figure 10. Detected result of image A.

Figure 11. Detected result of image B.

5 Conclusion and future research

In this paper, we propose a new representation method of images in scale–space. We apply this method to face detection problem and show it works efficiently. There remain some important problems about this method that include (1)how to determine adaptive scale parameter automatically (2)how to determine the feature graph automatically (3)what is the suitable segmentation method. There are some approaches to these problem ([2], [5], [8]), but we must analyze *good* properties (e.g. invariability , monotonicity) of the image in scale–space at first. If we define the *feature* of the image and the object considering these properties, we can obtain the suitable image representation method that has *good* properties for image processing.

References

[1] R. Brunelli and T. Poggio: ”Face recognition: Feature versus Template” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.15, no.10, pp1042–1052, 1993.

[2] I. Craw, N. Costen, T. Kato and S. Akamatsu: ”How Should We Represent Faces for Automatic Recognition?” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.21, no.8, pp725–736, 1999.

[3] M. Kirby and L. Sirovich: ”Application of the Karhunen-Loѐve Procedure for the Characterization of Human faces” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.12, no.1, pp103–108, 1990.

[4] J. J. Koenderink: ”The structure of Images”, *Biological Cybernetics*, Vol.50,pp363–370, 1984.

[5] T. Lindeberg: ”Effective Scale: A Natural Unit for Measuring Scale–Space Lifetime” , *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.15, no.10, pp1068–1074, 1993.

[6] T. Lindeberg: *Scale-Space Theory in Computer Vision*, Kluwer Academic Publisher, 1994.

[7] H.A. Rowley, S. Baluja and T. Kanade: ”Neural Network-Based Face Detection” , *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, no.1, pp23–38, Jan., 1998.

[8] J. Sporring, M. Nielsen, L. Florack and P. Johansen(Eds): *Gaussian Scale–Space Theory*, Kluwer Academic Publisher, 1997.

[9] S. Chinveeraphan, R. Takamarsu and M. Sato: ”Ridge–Valey Lines of Images in Scale–Space”, *Forma*, Vol.10, pp119–131, 1995.

[10] S. Chinveeraphan: *A Hierarchical Description of Gray-scale Images based on Image Dipoles*, Ph.D thesis!$Tokyo Institute of Technology, 1996!%

[11] A. P. Witkin: ”SCALE-SPACE FILTERING” , *Proceedings of 8th International Joint Conference on Artificial Intelligence*, pp1019–1022, Aug. 1983.

[12] A. L. Yuille, T. A. Poggio: ”Scaling Theorem for Zero Crossings”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.8, no.1, pp15–25, Jan., 1986.