An Improved 2-D Moment Algorithm for Pattern Classification

Myoung-Young Yoon*

요 약 화상 데이터의 특성을 표현하는데 적합한 깁스분포를 바탕으로 특징벡터를 추출하여 패턴을 분류하는 새로운 알고리즘을 제안하였다. 특징벡터는 화상의 크기, 위치, 회전에 대해서 불변이며 접영에 대해서도 덜 민감한 특징을 갖는 2차원 모멘트들의 원소로 만들어진다. 알고리즘은 공간정보를 갖는 2차원 모멘트를 이용하여 특징벡터를 추출하는 과정과 거리함수를 이용하여 패턴을 분류하는 과정으로 구축하였다. 특징벡터는 깁스분포의 묘수를 추정하여 2차원 조건부 모멘트를 추출하여 구성한다. 패턴 분류 과정은 추출된 특징벡터로부터 제안된 판별거리함수를 계산하여 여러 원형 패턴 가운데 최소거리를 산출한 미지의 패턴을 원형패턴으로 분류한다. 제안된 방법의 성능을 검증하기 위하여 대문자와 소문자 52자로 구성된 훈련 데이터를 만들어 SUN ULTRA 10 워크스테이션에서 실험을 한 결과 98%이상의 분류성능이 있음을 밝혔다.

Abstract We propose a new algorithm for pattern classification by extracting feature vectors based on Gibbs distributions which are well suited for representing the characteristic of an images. The extracted feature vectors are comprised of 2-D moments which are invariant under translation, rotation, and scale of the image, less sensitive to noise. This implementation contains two parts: feature extraction and pattern classification. First of all, we extract feature vector which consists of an improved 2-D moments on the basis of estimated Gibbs distribution. Next, in the classification phase, the minimization of the discrimination cost function for a specific pattern determines the corresponding template pattern. In order to evaluate the performance of the proposed scheme, classification experiments with training document sets of characters have been carried out on SUN ULTRA 10 Workstation. Experiment results reveal that the proposed scheme has high classification rate over 98%.

1. Introduction

An essential issue in the field of pattern analysis is the classification of objects and characters regardless of their positions, sizes, and orientations. In the recent computer vision literature there has been increasing interest in use of statistical techniques for classifying and processing image data. Moments and function of moments have been extensively employed as the invariant global features of an image in pattern recognition, image classification, target identification, and scene analysis[1]. The goal of a typical computer vision system is to analyze images of a given scene and classify the content of the scene. "Good" features are those satisfying the following requirements: (i)small interclass invariance-slightly different shapes with similar general characteristics should have numerically close values; (ai)large interclass separation- features from different classes should be quite different numerically[2].

These features may be divided into five groups-Visual features(edges, texture and shape), Transform coefficient features(Fourier descriptors), Algebraic features(based on matrix composition of the image), Statistical features (moment invariant), Differential invariant (used especially for curved objects). Since statistical features are invariant under translation, rotation, size of the patterns, the moments are very useful features for pattern classification. Statistical image analysis concerns the measurement of quantitative information from an image to produce a probabilistic description. In [3, 4], their proposed

* Dip. of Computer Science, Churchgoing College

moments that provide features for classification of patterns have been used for a number of image classifying applications. Their proposed moments are calculated by using the intensity at each point. However, their performance for pattern classification is poor since the moments did not included spatial information which is the characteristic of the most images. The spatial information represents statistical dependence(or spatial continuity) of the pixel value at a lattice point on the those of its neighbors. Previously, many researchers considered image processing using spatial information[5, 6. 7, 8]. Gibbs random field. However, their works are concerned with only both restoration and segmentation.

In this paper, we propose a new algorithm for pattern classification using an improved two dimensional (2-D) moments on the basis of Gibbs distributions. Gibbs distribution which are well suited for representing spatial continuity that is the characteristic of the most images[6, 9]

2. GRFs for Pattern Classification

We review the basic definition and the properties of Gibbs random fields(GRFs). And we also present a particular class of Gibbs distribution that is used in the image model of this paper.

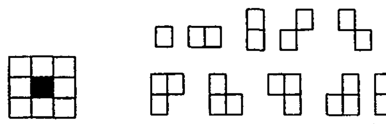
2.1 Gibbs random fields on finite lattices

We focus our attention on discrete 2-D random fields defined over a finite $N_1 \times N_2$ rectangular lattice of points(pixels) defined as $L = \{(x, y) : 1 \le x \le N_1,\ 1 \le y \le N_2\}$. Suppose $M = \{m_{xy}\}$ represents a image, where m_{xy} measures the grey-level of the pixel in the $x-th$ row and $y-th$ column. Let η be neighborhood system defined over the finite L. A random field $M = \{M_{xy}\}$ on L has a Gibbs distribution or equivalently is a Gibbs random field with respect to η if and only if its joint distribution is of the form [9, 10]

$$P(M=m) = \frac{1}{Z}\exp\{-E(m)\}$$
$$Z = \sum_m \exp\{E(m)\} \qquad (1)$$
$$E(m) = \sum_{c \in C} V_c(m)$$

where Z is a normalizing constant, called the partition function; $E(m)$ is energy function; c is a clique, a set of sites(including single sites) such that any two elements in the set are neighbors of each other; C is the set of all cliques of a lattice-neighborhood pair (L, η); and $V_c(m)$ is the potential associated with clique C, arbitrary except for the fact that it depends only on the restriction of m to C. Let η^r be the rth order neighborhood system. Clique types for the first-order and second-order neighborhoods systems are depicted in Figure 1. The source of the revived interest in Gibbs distribution(GD), especially in the context of image modeling and processing, is an important result known as the Hammersley-Clifford theorem. Besag [10] derives an expression for the joint probability $P(M=m)$ in terms of the conditional probabilities(local characteristics) $P(M_{xy} | \eta_{xy})$.

(a) η^1-neighborhood system (b) Cliques in η^1



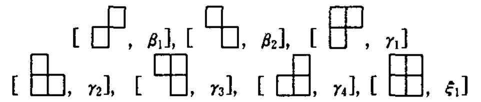
(a) η^2-neighborhood system (b) Cliques in η^2

<Figure 1> Neighborhood systems η^1 and η^2, and their associated clique types.

Equivalently, $P(M_{xy} = m_{xy} | \eta_{xy}) \propto \exp\{-E(m_{xy})\}$ where $E(m_{xy})$ is the energy function for pixel site (x, y). By choosing $V_c(m)$ properly, a wide variety of distributions both for discrete and continuous random fields can be formulated as GD. The GD characterization in some applications provides a more workable spatial model [11].

2.2 Gibbs distribution for pattern classification

In this subsection, we present a particular class of (GD), which is used to estimate the parameters of Gibbs distributed image. We assume that the random field M consists of binary-valued discrete random variables $\{M_{xy}\}$ taking values in $\Omega = \{\omega_1, \omega_2\}$. To define GD it suffices to specify the neighborhood system η, the associated cliques and the clique potentials $V_c(m)$'s. Here, it is assumed that the random field is homogeneous, that is the clique potentials depend only on the clique type and the pixel values in clique, but not on the position of the clique in L. The distribution is specified in terms of the second order neighborhood system η^2. Figure 2 shows the parameters associated with clique types, except for the single pixel clique.



<Figure 2> The parameters associated with clique types.

The clique potentials associated with η^2 are defined as follows.

$$V_c(m_{xy}) = \begin{cases} -\zeta & \text{if all } m_{xy}\text{'s in } c \text{ are equal} \\ \zeta & \text{otherwise} \end{cases} \quad (2)$$

where ζ is the parameter specified for the clique type c. For the single pixel cliques, the clique potential is defined as

$$V_c(m_{xy}) = \alpha_k \text{ for } m_{xy} = \omega_k. \quad (3)$$

The parameters α_k control the percentage of pixels in each site, that is the marginal distribution of the single random variables M_{xy}'s, while the other parameters control the size and direction of clustering.

3. Estimation of Parameter in a GD

In this section, we describe a parameter estimation method of Gibbs distributed image since calculation of conditional 2-D moment requires the estimated parameters of Gibbs distribution. The most commonly used parameter estimation method to date is the so-called "coding method," first presented by Besag[10]. It requires the solution of a set of nonlinear equations. Therefore, it is cumbersome and difficult to use reliably. In view of the practical difficulties involved in using the coding method[12], we describe an alternative parameter estimation scheme for finite range space GRF, which consists of histograming and a standard, linear, least squares estimation as its components. We present the formulation in terms of a second order neighborhood system η^2, although its extension to any order is possible.

Suppose M is a GD of the class described in Section 2.2, with a discrete range space of $\Omega = \{\omega_1, \omega_2\}$. A realization m of this random field is available to be used in estimating the parameters of the distribution. Consider a site (x, y) and its neighborhood η_{xy}. For convenience of notation, let S represent m_{xy} and λ' represent the vector of the neighboring values of m_{xy}, that is,

$$\lambda' = [u_1, u_2, u_3, u_4, v_1, v_2, v_3, v_4]^T \quad (4)$$

where the location of u_i's and v_i's with respect to S are shown in Figure 3.

v_1	u_2	v_2
u_1		u_3
v_4	u_4	v_3

<Figure 3> m_{xy} and η_{xy}.

We define indicator functions

$$I(\delta_1, \delta_2, \ldots, \delta_k) = \begin{cases} -1 & \text{if } \delta_1 = \delta_2 = \cdots = \delta_k \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

and

$$J_m(s) = \begin{cases} -1 & s = \omega_m \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

We can express the potential functions of the GD in terms of these quantities. Let $V(s, \lambda', \theta)$ be the sum of the potential functions of all the cliques that contain (x, y), the site of s. That is $V(s, \lambda', \theta) = \sum_{c:s \in C} V_c(m)$

where θ is the parameter vector

$$\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \xi_1). \quad (7)$$

Using the clique potentials for this class of GD we can write $V(s, \lambda', \theta)$ as $V(s, \lambda', \theta) = \rho^T(s, \lambda')\theta$ where

$$\begin{aligned} \rho(s, \lambda) = \; & [J_1(s), J_2(s), (I(s, v_2) + I(s, v_4)), (I(s, v_1) + \\ & I(s, v_3)), (I(s, u_2, v_2) + I(s, u_4, u_3) + I(s, u_1, v_4)), \\ & (I(s, u_4, u_3) + I(s, u_2, u_3) + I(s, u_1, v_1)), \\ & (I(s, u_2, v_1) + I(s, u_1, u_4) + I(s, u_3, v_3)), \\ & (I(s, u_1, u_2) + I(s, u_4, v_4) + I(s, u_3, v_2)), \\ & (I(s, u_1, v_1, u_2) + I(s, u_2, v_3, u_3) + I(s, u_3, v_3, u_4) \\ & + I(s, u_4, v_4, u_1))]^T. \end{aligned} \quad (8)$$

Now suppose $P(s, \lambda')$ is the joint distribution of the random variables on the 3×3 window centered at (x, y) and $P(\lambda')$ is the joint distribution of the random variables on η_{xy} only. Then the conditional distribution $P(s|\lambda')$ is given by the ratio of $P(s,\lambda')$ to $P(\lambda')$. It follows from the GRF-MRF equivalence and the resulting local characteristic that

$$P(s \mid \lambda') = \frac{P(s,\lambda')}{P(\lambda')} = \frac{e^{-V(s,\lambda',\theta)}}{Z(\lambda',\theta)} \tag{9}$$

where $Z(\lambda',\theta)$ is the appropriate normalizing constant. Hence

$$\frac{e^{-V(s,\lambda',\theta)}}{P(s,\lambda')} = \frac{Z(\lambda',\theta)}{P(\lambda')} \tag{10}$$

is obtained. Note that the right-hand side of (10) is independent of s. Considering the left-hand side of (10) for any two distinct values of s, e.g., $s=j$ and $s=k$, we have

$$(\rho(k,\lambda') - \rho(j,\lambda'))^T\theta = \ln\frac{P(j,\lambda')}{P(k,\lambda')} \tag{11}$$

where $\rho^T(k,\lambda')\theta = V(k,\lambda',\theta)$. Consideration of all possible triplets (j,k,λ'), $j<k$, generates from equation (11) a large set of linear equations, which may be solved for θ by least squares procedures. The question that remains to be answered, now, is how to determine or estimate $P(s,\lambda')$ for all (s,λ') combinations using a single or a few realizations. We propose to estimate $P(s,\lambda')$ using histogram techniques.

4. Improved 2-D Moments and Classification

The geometric moments proposed by many researchers[4, 13] have not included spatial information which is the characteristic of most images. However, we propose conditional 2-D moments which include spatial information by using the estimated conditional Gibbs distribution.

4.1 Improved 2-D moment based on GD

The basic and classical moment, A regular 2-D moment of order $(k+l)$ is defined by [2, 13]

$$\pi_{kl} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^l f(x,y), \tag{12}$$

where $f(x,y)$ is the intensity at a point (x,y) in the image and k, $l=0,1,2,\cdots$. Since this two dimensional integration can be viewed as if the image irradiance function $f(x,y)$ is projected to onto the moment kernel $\{x^k y^l\}$, the regular moment will be referred as geometric moment (GM). We propose an improved 2-D moments by using the estimated Gibbs distribution, instead of $f(x,y)$. Let $\hat{\theta}$ be the estimated parameter vector of Gibbs distributed image described in Section 3. The parameter vector $\hat{\theta}$ measures the strength of interaction between pixels. Also, the clique potentials $\{V_c(m_{xy})\}$ specify the local characteristics $P(M_{xy}=m_{xy} \mid \eta_{xy})$. By the MRF property we see that $P(M_{xy}=m_{xy} \mid \eta_{xy}) = P(M_{xy}=m_{xy} \mid L\backslash(m_{xy}))$ where $L\backslash(m_{xy})$ is denotes the set $\{m_{kl}:(k,l) \neq (i,j)\}$. In the general 2-D form and for binary-valued images, the corresponding conditional 2-D moments is given by the following steps.

● Step 1) Calculate the centroids $\bar{x}$, $\bar{y}$ of the considered shape as follows. Let $I(\cdot)$ be the indicator function.

$$\bar{x} = \sum_{x=1}^{N_1}\sum_{y=1}^{N_2} x\hat{P}(M_{xy}=m_{xy} \mid \eta_{xy}) \tag{13}$$

$$\bar{y} = \sum_{y=1}^{N_2}\sum_{x=1}^{N_1} y\hat{P}(M_{xy}=m_{xy} \mid \eta_{xy}) \tag{14}$$

● Step 2) Calculate σ_x and σ_y are the standard deviation of the image with respect to the coordinates x and y, given by

$$\sigma_x = \sqrt{\sum_{x=1}^{N_1}\sum_{y=1}^{N_2}(x-\bar{x})^2\hat{P}(M_{xy}=m_{xy} \mid \eta_{xy})} \tag{15}$$

$$\sigma_y = \sqrt{\sum_{y=1}^{N_2}\sum_{x=1}^{N_1}(y-\bar{y})^2\hat{P}(M_{xy}=m_{xy} \mid \eta_{xy})}. \tag{16}$$

● Step 3) Calculate the 2-D conditional moments from (15) and (16) for $k=0,1,2,\ldots$, and $l=0,1,2,\ldots$. And then, we store these moments to a feature vector.

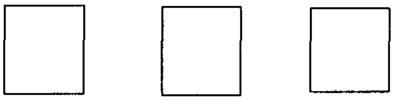
$$\pi_{kl} = \sum_{x=1}^{N_1}\sum_{y=1}^{N_2}\left(\frac{x-\bar{x}}{\sigma_x}\right)^k\left(\frac{y-\bar{y}}{\sigma_y}\right)^l\hat{P}(M_{xy}=m_{xy} \mid \eta_{xy}), \tag{17}$$

The above moments are invariant under translation and magnification of the image, but not under rotation. Thus, In order to use them as classification features we have to normalize the moments with respect to rotation by multiplying the coordinates of the image by $e^{-j\phi}$, where ϕ is the rotation change of the object.

<Table 1> The values of the proposed moments

Moments	π_{3}	π_{4}	π_{5}	π_{6}	π_{7}	π_{8}	π_{0}	π_{0}	π_{0}	π_{0}	π_{0}	π_{0}
Template	1.0	2.5	1.2	4.11	3.9	7.0	1.11	2.9	1.0	3.0	1.5	3.0
Shape 1	1.0	2.6	1.31	4.1	3.9	7.2	1.0	2.8	1.2	3.1	3.0	3.1
Shape 2	1.01	2.4	1.01	2.0	1.0	5.7	-1.1	3.2	-1.9	4.2	-2.9	7.9

Table 1 shows the 2-D normalized moments of template letter C, as well as two similar letters(distorted C and O) of Figure 4.



(a) template (b) shape 1 (c) shape 2

<Figure 4> The template and the shapes to be tested.

4.2 Classification

In order to classify patterns, we define a the discrimination cost function (DCF) $F(i, v)$ which is defined by

$$F(i, v) = \sum_{j=1}^{d} [T_{vj} - U_{ij}]^2 \qquad (18)$$

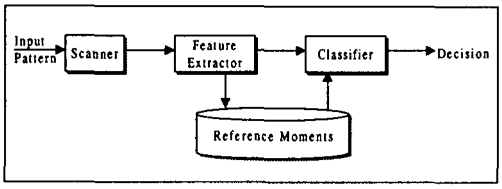
where T_{vj} denotes the $j-th$ feature of the $v-th$ template, U_{ij} denotes the $j-th$ feature of the $i-th$ shape under consideration and d is the dimension of the feature vectors. The minimization of the index $F(i, v)$, $v = 1, 2, \ldots$ for a specific shape i determines the corresponding template v. The proposed DCF is a kinds of Euclidean distance between an arbitrary pattern vector U_{ij} and the $v-th$ prototype vector T_{vj}. Since the proposed DCF only require some simple analytic algebraic calculations, It is characterized by low computation cost. The ideal discrimination of a shape corresponding exactly to a template, without any noise and computational error, the index $F(i, v)$ should be zero. However, in practice, the discrimination is clear if $F(i, v)$ is sufficiently smaller in comparison with the other templates, as well as small enough itself. Table 2 shows the discrimination functions of the letters of Figure 4. In Table 2 it is seen that $F(i, v)$ is sufficiently smaller for the distorted C.

<Table 2> The DCF of the shapes of Fig. 4.

F(template "C", shape 1 "C") =	2.17
F(template "C", shape 2 "O") =	80.53

5. Experimental Results

In order to illustrate the performance of the proposed moment for pattern classification, we carried out the following experiments was carried out. The training document consists of 10 lines of 52 characters each. Two documents were created for testing the performance of the proposed classification method on the basis of the extracted feature vector. Each document consists of 24 lines 52 characters each. Figure 5 shows the overall block diagram of the proposed method for classification of patterns, where it is shown that a document to be processed is at first scanned. Then the classification feature vectors are extracted by formulae (13) through (17). These features are sent to a classifier, which is described by formula (18), for a decision in order to identify the input character.



<Figure 5> Overview of the proposed method for classification of patterns.

The gross structural features of the shape can be better characterized by the proposed moments derived from the silhouette. In our experiments we use only silhouette moments since these moments are less sensitive to noise. The used feature vector F_v for the templates is considered to be $F_v = [\pi_{03}, \pi_{04}, \pi_{05}, \pi_{06}, \pi_{07}, \pi_{08}, \pi_{30}, \pi_{40}, \pi_{50}, \pi_{60}, \pi_{70}, \pi_{80}]^T$.

A classification simulation was run six times. The first simulation used a library set of 52 feature vectors derived from the first line of characters of the training document. The second simulation used two library sets derived from the first two lines of the training document. The third, fourth, fifth and sixth simulations used four, six, eight and ten library sets, respectively. The classification rates resulting from these simulations are presented in Table 3. As Table 3 reveals, we can achieve better than 98% increase in classification rates when we use eight or ten library sets. Since the proposed 2-D the improved moments have properties of the affine or geometric moments, as well as spatial information which describe dependance between pixels, our proposed method was superior to other methods using the affine moments and the geometric moments, respectively.

<Table 3> The classification rates

No of library sets	Flusser's method using the affime moments (%)	Tsirikolias's method using the geometric moments (%)	Proposed method using the 2-D moments (%)
1	73	72	80
2	82.5	81	86
4	84	85	88
6	90.5	89	95
8	93	91	96
10	95	94.5	98.5

6. Concluding Remarks

In this paper we propose a new algorithm for pattern classification using an improved two dimensional (2-D) moments based on GD. Experiment results reveal that the proposed scheme has high classification rate over 98%. The proposed method appears to be efficient with respect to the existing ones, since it shares the following advantages. (i)The discrimination process is invariant under translation, scaling and rotation of the considered shape. (ii)Fast processing, since calculations of the moment are simple. (ii)Each shape is uniquely described (iv)It is possible to consider the pattern itself rather that its contour.

The success of pattern classification depends on how good the used clique parameter $\hat{\theta}$ fits characteristic of the image. Upon completion of the pattern classification, we will focus our efforts on further development of the clique functions.

REFERENCES

[1] R.J. Prokop and A.P. Reeves, "A Survey of Moment-based Techniques for Unconcluded Object Representation and Recognition," Graphical Models and Image Processing, Vol. 54, No. 5, 1992, pp. 438-460.

[2] J. Flusser and T. Suk, "Classification Objects and Patterns by Affine Moment Invariants," Image Proc.: Theory and Applications, 1993, pp. 345-348.

[3] B. C. Li and J. Shen, "Fast computation of moment invariants," Pattern Recognition Vol. 24, 1991, pp. 807-813.

[4] H. Arezki, H. T. Huynh, and G. A. Lampropoulos, "A modified moment Inavariants Technique for Classification of Noisy Objects," Conference on Electrical and Computer Engineering, Canada, 1991, pp. 25-27.

[5] S. X. Liao and M. Pawlak, "On Image Analysis by Moments," IEEE Transactions on PAMI, Vol.18 No.13, 1996, pp. 254-266.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. 6, 1984, pp. 721-741.

[7] O. Shunichiro, "Texture Segmentation Method by Using Two-Dimensional AR Model and Kullback Information," Pattern Recognition, Vol. 26, 1993, pp. 237-244.

[8] A. M. Tekalp and G. Pavlovic, "Multichannel Image Modeling and Kalman Filtering for Multispectral Image Restoration," IEEE Trans. on ASSP, Vol. 19, 1990, pp. 221-232.

[9] H. Derin and H. Elliott, "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," IEEE Trans. on PAMI, Vol. 9, No. 1, 1987, pp. 39-55.

[10] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," Journal of Royal Statistics Society B, Vol. 36, 1974, pp. 192-236.

[11] C. H. Wu and P. C. Doerschuk, "Cluster Expansions for the Deterministic Computation of Bayesian Estimators Based on Markov Random Fields," IEEE Transactions on PAMI, Vol. 17, No. 3, 1995, pp. 275-293.

[12] A. J. Gray, J. K. Kay, and D. M. Titterington, "On the Estimation of Noisy Binary Markov Random Fields," Pattern Recognition, Vol. 25, No. 7, 1992, pp. 749-768.

[13] W. Chung and E. M. Tzanakou, "A Simulation for Different Moment sets," Proceedings in SPIE, Vol. 2181: Document Recognition, San Jose, California, 1994, pp. 378-389.