〔RESEARCH〕

Normalized Separability Measures Based on Likelihood Criteria and Their Properties

Taiichi SAITO

As a multi-thresholding technique for gray images, the maximum likelihood method under the assumption of mixture of normal distributions and the thresholding method considering the quantization error have already been proposed. In this report, separability measures are defined from the likelihood criteria which are used as evaluation functions in these methods. The proposed separability measures are invariant under affine transformations of gray level scale and are normalized to be within values from 0 to 1. Binarization for a gray image whose distribution is uniform is considered. By investigating properties of the separability measures in this case, it is shown that consideration of the quantization error is effective and the defined separability measures are valid.

§ 1 Introduction

As a multi-thresholding technique for gray images, the methods[1] using the maximum likelihood criteria and the methods[2] considering the quantization error of an image have already been proposed. A purpose of these methods is to find the thresholds at which the likelihood criterion as an evaluation function becomes optimal. We usually only want to know the optimal thresholds. However, we sometimes want to use the value itself of the evaluation function. At this time, it is desirable that the evaluation functions are normalized so that the values of the functions may not change in the condition that there is no influence in the evaluation.

In section 2 of this report, separability measures are defined, which are invariant under affine transformations of gray level scale and are normalized to be within values from 0 to 1. In section 3, binarization for a gray image whose distribution is uniform is considered. By investigating properties of the separability measures in the case, it is shown that consideration of the quantization error is effective and the defined separability measures are valid.

§ 2 Likelihood criteria and definition of separability measures

In the case of multi-thresholding for gray images, the fundamental statistics and notation of the likelihood criteria under the assumption of mixture of normal distributions and normalized separability measures which evaluate the goodness of the separation are defined[3,4] here.

2.1 The fundamental statistics

N : The number of pixels of a gray image.

L : The number of gray levels of a gray image.

u_i : A gray level ($i = 0, 1, \ldots, L-1$).

δ : An interval between neighboring gray levels.

$$\delta = u_{i+1} - u_i = 1 \quad (\forall i). \tag{1}$$

$h(u_i)$: The frequency of a pixel whose gray level is u_i.

$$\sum_{i=0}^{L-1} h(u_i) = N. \tag{2}$$

$p(u_i)$: The relative frequency of a pixel

KEY WORDS: Separability Measure, Likelihood Criterion, Quantization Error, Multi-thresholding, Binarization

whose gray level is u_i.

$$p(u_i) = h(u_i)/N, \tag{3}$$

$$\sum_{i=0}^{L-1} p(u_i) = 1. \tag{4}$$

M : The number of classes into which gray levels of an image are divided ($2 \leq M \leq L$).

$\overline{k}$: A set of gray levels to use them as thresholds.

$$\overline{k} = (\ k_0, k_1, ..., k_m, ..., k_{M-2}\ ), \tag{5}$$

$$(\ m = 0,\ 1,\ ...,\ M-2,\ k_m < k_{m-1}\ ).$$

C_j : A set of gray levels to belong to the j class ($j = 0,\ 1,\ ...,\ M-1$).

$$C_j = \{\ u_i \mid k_{j-1} < u_i \leq k_j\ \}, \tag{6}$$

$$(\text{where, } k_{-1} = u_0 - \delta,\ k_{M-1} = u_{L-1}\ ).$$

ω_j : The occurrence probability of C_j.

$$\omega_j = \sum_{u_i \in C_j} p(u_i) = \sum_{u_i = k_{j-1}+\delta}^{k_j} p(u_i), \tag{7}$$

$$\sum_{j=0}^{M-1} \omega_j = 1. \tag{8}$$

μ_j : The mean of C_j.

$$\mu_j = \frac{1}{\omega_j} \sum_{u_i \in C_j} u_i p(u_i). \tag{9}$$

μ_T : The total mean.

$$\mu_T = \sum_{i=0}^{L-1} u_i p(u_i) = \sum_{j=0}^{M-1} \omega_j \mu_j. \tag{10}$$

σ_j^2 : The variance of C_j.

$$\sigma_j^2 = \frac{1}{\omega_j} \sum_{u_i \in C_j} (u_i - \mu_j)^2 p(u_i) = \frac{1}{\omega_j} \sum_{u_i \in C_j} u_i^2 p(u_i) - \mu_j^2. \tag{11}$$

σ_W^2 : The within-class variance.

$$\sigma_W^2 = \sigma_W^2(\overline{k}) = \sum_{j=0}^{M-1} \omega_j \sigma_j^2 = \sum_{i=0}^{L-1} u_i^2 p(u_i) - \sum_{j=0}^{M-1} \omega_j \mu_j^2. \tag{12}$$

σ_B^2 : The between-class variance.

$$\sigma_B^2 = \sigma_B^2(\overline{k}) = \sum_{j=0}^{M-1} \omega_j (\mu_j - \mu_T)^2 = \sum_{j=0}^{M-1} \omega_j \mu_j^2 - \mu_T^2. \tag{13}$$

σ_T^2 : The total variance.

$$\sigma_T^2 = \sum_{i=0}^{L-1} (u_i - \mu_T)^2 p(u_i) = \sum_{i=0}^{L-1} u_i^2 p(u_i) - \mu_T^2 = \sigma_W^2 + \sigma_B^2. \tag{14}$$

When the number of gray levels is small, it is required that the quantization error is considered, because the influence of error by quantizing gray values appears. About the above statistics, when the quantization error is considered, * is put on the right shoulder of the statistics. In the cases in which the quantization error is considered and is not considered, only the following equations are different.

$$\sigma_j^{*2} = \sigma_j^2 + \frac{\delta}{12}. \tag{15}$$

$$\sigma_W^{*2} = \sigma_W^2 + \frac{\delta}{12}. \tag{16}$$

$$\sigma_T^{*2} = \sigma_T^2 + \frac{\delta}{12}. \tag{17}$$

2.2 Likelihood criterion

By using the above notation, in the case that gray levels are divided into M classes under the assumption of mixture of M normal distributions, likelihood criteria[1,2] are shown next.

According to relations of statistics of M normal distributions, the likelihood criteria are divided into four cases in **Tab. 1**. The first case 'D' is the mixture model under the assumption of M normal distributions with different occurrence probabilities and with a common variance. The second case 'O' is the mixture model under the assumption of M normal distributions with a common occurrence probability and with a common variance. The third case 'K' is the mixture model under the assumption of M

Tab. 1 The classification of likelihood criteria

likelihood criteria	occurrence probabilities	variances	threshold selection method
$l_D^{(*)}$, $J_D^{(*)}$	different	equal	[1]
$l_O^{(*)}$, $J_O^{(*)}$	equal	equal	Otsu[5,6]
$l_K^{(*)}$, $J_K^{(*)}$	different	different	Kittler and Illingworth[7]
$l_Q^{(*)}$, $J_Q^{(*)}$	equal	different	Kurita[1]

normal distributions with different occurrence probabilities and with different variances. The fourth case 'Q' is the mixture model under the assumption of M normal distributions with a common occurrence probability and with different variances.

The likelihood criteria which do not consider the quantization error are the following[1].

$$l_D = \frac{L}{2}\sum_{j=0}^{M-1}\omega_j \ln\frac{1}{\sigma_j^2} - \frac{L}{2}(\ln 2\pi + 1). \quad (18)$$

$$l_O = \frac{L}{2}\ln\frac{1}{\sigma_W^2} - \frac{L}{2}(\ln 2\pi + 1). \quad (19)$$

$$l_K = \frac{L}{2}\sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_j^2} - \frac{L}{2}(\ln 2\pi + 1). \quad (20)$$

$$l_Q = \frac{L}{2}\sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_W^2} - \frac{L}{2}(\ln 2\pi + 1). \quad (21)$$

It has been shown[1] that l_O is equivalent to the discriminant and least squares criterion[5,6] by Otsu and that l_K is equivalent to the minimum error thresholding[7] by Kittler and Illingworth.

The likelihood criteria considering the quantization error are the following[2].

$$l_D^* = \frac{L}{2}\sum_{j=0}^{M-1}\omega_j \ln\frac{1}{\sigma_j^2 + \frac{\delta}{12}} - \frac{L}{2}(\ln 2\pi + 1). \quad (22)$$

$$l_O^* = \frac{L}{2}\ln\frac{1}{\sigma_W^2 + \frac{\delta}{12}} - \frac{L}{2}(\ln 2\pi + 1). \quad (23)$$

$$l_K^* = \frac{L}{2}\sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_j^2 + \frac{\delta}{12}} - \frac{L}{2}(\ln 2\pi + 1). \quad (24)$$

$$l_Q^* = \frac{L}{2}\sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_W^2 + \frac{\delta}{12}} - \frac{L}{2}(\ln 2\pi + 1). \quad (25)$$

Because the criterion l_O^* is only added a constant '$\delta/12$' to l_O, l_O and l_O^* give the same optimal thresholds[2].

The above criteria are written together with $l_X^{(*)}$, where, $X = D,\ O,\ K,\ Q$. $l_X^{(*)}$ means l_X or l_X^*. The following is also represented with the same notation.

2.3 Definition of separability measures

As a measure which evaluates the goodness of the separation, normalized separability measures based on the likelihood criteria are proposed in this section.

The thresholdings based on the maximum likelihood methods are not influenced by the increase and decrease of constant terms, because these only find maximum point of the above likelihood criteria. Therefore, the criteria are equivalent even if these are transformed as follows.

$$J_D = \sum_{j=0}^{M-1}\omega_j \ln\frac{1}{\sigma_j^2}. \quad (26)$$

$$J_O = \ln\frac{1}{\sigma_W^2}. \quad (27)$$

$$J_K = \sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_j^2} + \ln M^2. \quad (28)$$

$$J_Q = \sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_W^2} - \ln M^2. \quad (29)$$

$$J_D^* = \sum_{j=0}^{M-1}\omega_j \ln\frac{1}{\sigma_j^2 + \delta/12}. \quad (30)$$

$$J_O^* = \ln\frac{1}{\sigma_W^2 + \delta/12}. \quad (31)$$

$$J_K^* = \sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_j^2 + \delta/12} + \ln M^2. \quad (32)$$

$$J_Q^* = \sum_{j=0}^{M-1}\omega_j \ln\frac{\omega_j^2}{\sigma_W^2 + \delta/12} + \ln M^2. \quad (33)$$

In these criteria ($J_D^{(*)}$, $J_O^{(*)}$, $J_K^{(*)}$, $J_Q^{(*)}$), constant terms have been removed. In the criteria $J_K^{(*)}$ and $J_Q^{(*)}$, a constant '$\ln M^2$' has been added, in order to make it possible to compare the four (D, O, K, Q) criteria.

The following equations are defined toward these criteria.

$$\tau_X = \ln\frac{1}{\sigma_T^2} - J_X. \quad (34)$$

$$\tau_X^* = \ln\frac{1}{\sigma_T^{*2}} - J_X^*. \quad (35)$$

By dividing by the total variance ($\sigma_T^{(*)2}$), $\tau_X^{(*)}$ become the measures which are invariant under affine transformations of gray levels u_i. Using these equations, the following equations $\eta^{(*)}$ (eta) are defined.

$$\eta_X = 1 - \exp(\tau_X). \quad (36)$$

$$\eta_X^* = 1 - \exp(\tau_X^*). \quad (37)$$

Therefore, separability measures based on the likelihood criteria become the following.

$$\eta_D = 1 - \frac{\prod_{j=0}^{M-1}\left(\sigma_j^2\right)^{\omega_j}}{(\sigma_T^2)}. \quad (38)$$

$$\eta_O = 1 - \frac{\sigma_W^2}{\sigma_T^2} = \frac{\sigma_B^2}{\sigma_T^2}. \quad (39)$$

$$\eta_K = 1 - \frac{\prod_{j=0}^{M-1}\left(\sigma_j^2\right)^{\omega_j}}{M^2\left(\prod_{j=0}^{M-1}\omega_j^{2\omega_j}\right)\left(\sigma_T^2\right)}. \quad (40)$$

$$\eta_Q = 1 - \frac{\sigma_W^2}{M^2\left(\prod_{j=0}^{M-1}\omega_j^{2\omega_j}\right)\left(\sigma_T^2\right)}. \quad (41)$$

$$\eta_D^* = 1 - \frac{\prod_{j=0}^{M-1}\left(\sigma_j^2+\frac{\delta}{12}\right)^{\omega_j}}{\left(\sigma_T^2+\frac{\delta}{12}\right)}. \quad (42)$$

$$\eta_O^* = 1 - \frac{\sigma_W^2+\frac{\delta}{12}}{\sigma_T^2+\frac{\delta}{12}} = \frac{\sigma_B^2}{\sigma_T^2+\frac{\delta}{12}}. \quad (43)$$

$$\eta_K^* = 1 - \frac{\prod_{j=0}^{M-1}\left(\sigma_j^2+\frac{\delta}{12}\right)^{\omega_j}}{M^2\left(\prod_{j=0}^{M-1}\omega_j^{2\omega_j}\right)\left(\sigma_T^2+\frac{\delta}{12}\right)}. \quad (44)$$

$$\eta_Q^* = 1 - \frac{\sigma_W^2+\frac{\delta}{12}}{M^2\left(\prod_{j=0}^{M-1}\omega_j^{2\omega_j}\right)\left(\sigma_T^2+\frac{\delta}{12}\right)}. \quad (45)$$

(In the literatures[3,4], the equations $\eta_K^{(*)}$ and $\eta_Q^{(*)}$ do not contain the constant '$1/M^2$'.) $\eta_X^{(*)}$, $\tau_X^{(*)}$, $J_X^{(*)}$ indicate better separation as much as their values are large. On the likelihood criterion J_O, substituting equation (27) into (36), then equation (39) is given. This equation (39) is equivalent to the separability[5,6] defined in the thresholding based on the discriminant and least squares criteria by Otsu. This means that the proposed separability measures $\eta_X^{(*)}$ in this report are valid.

Next, it is shown that $0 \le \eta_X^{(*)} \le 1$, namely, $\eta_X^{(*)}$ are greater than or equal to 0 and are less than or equal to 1. Relations (46) are obvious.

$$0 \le \frac{\sigma_W^{(*)2}}{\sigma_T^{(*)2}} \le 1. \quad (46)$$

From the literature[8](p.908), relations (47) hold.

$$\prod_{j=0}^{M-1}\left(\sigma_j^{(*)2}\right)^{\omega_j} \le \sum_{j=0}^{M-1}\omega_j\sigma_j^{(*)2}. \quad (47)$$

From equations (11), (12), (15), (16), $\sum_{j=0}^{M-1}\omega_j\sigma_j^{(*)2} = \sigma_W^{(*)2}$. Therefore, the following relations hold.

$$\prod_{j=0}^{M-1}\left(\sigma_j^{(*)2}\right)^{\omega_j} \le \sigma_W^{(*)2} \ \left(\le \sigma_T^{(*)2} \right). \quad (48)$$

Because of relation (49)

$$\prod_{j=0}^{M-1}\omega_j^{2\omega_j} \ge \frac{1}{M^2}, \quad (49)$$

relation (50) holds.

$$0 < \frac{1}{M^2\left(\prod_{j=0}^{M-1}\omega_j^{2\omega_j}\right)} \le 1. \quad (50)$$

It is shown that the proposed separability measures $\eta_D^{(*)}$ (equations (38), (42)), $\eta_O^{(*)}$ (equations (39), (43)), $\eta_K^{(*)}$ (equations (40), (44)) and $\eta_Q^{(*)}$ (equations (41), (45)) are normalized (within $0 \le \eta_X^{(*)} \le 1$), from relations (48), from relations (46), from relations (48), (50) and from relations (46), (50), respectively. In equations (28), (29), (32), (33), if not adding the constant '$\ln M^2$', then the relations $0 \le \eta_K^{(*)}, \eta_Q^{(*)}$ can not hold.

When such a class that $\sigma_j^2 = 0$ exists, the value of η_D and η_K becomes 1 regardless of other conditions. Then, η_D, η_K are not reliable. Consideration of the quantization error makes it avoid such a unusual phenomenon.

§ 3 Properties of separability measures

Using the proposed separability measures, properties of the likelihood criteria are investigated in this section. In order to make it easy to understand, a case of binarization (M=2) is considered.

In the case binarizing under the assumption of a uniform distribution with L gray levels, when the occurrence probability of the class 0 is shown with $\omega = \omega_0$, the separability measures based on the likelihood criteria become the following.

$$\eta_D = 1 - \frac{\{L^2\omega^2 - \tilde{1}\}^{\omega}\{L^2(1-\omega)^2 - \tilde{1}\}^{(1-\omega)}}{L^2 - \tilde{1}}. \quad (51)$$

$$\eta_O = 1 - \frac{L^2 - \tilde{1} - 3L^2\omega(1-\omega)}{L^2 - \tilde{1}}. \quad (52)$$

$$\eta_K = 1 - \frac{\{L^2\omega^2 - \tilde{1}\}^{\omega}\{L^2(1-\omega)^2 - \tilde{1}\}^{(1-\omega)}}{4\omega^{2\omega}(1-\omega)^{2(1-\omega)}(L^2 - \tilde{1})}. \quad (53)$$

$$\eta_Q = 1 - \frac{L^2 - \tilde{1} - 3L^2\omega(1-\omega)}{4\omega^{2\omega}(1-\omega)^{2(1-\omega)}(L^2 - \tilde{1})}. \quad (54)$$

$$\eta_D^* = 1 - \omega^{2\omega}(1-\omega)^{2(1-\omega)}. \quad (55)$$

$$\eta_O^* = 3\omega(1-\omega). \quad (56)$$

$$\eta_K^* = \frac{3}{4}. \quad (57)$$

$$\eta_Q^* = 1 - \frac{1 - 3\omega(1-\omega)}{4\omega^{2\omega}(1-\omega)^{2(1-\omega)}}. \quad (58)$$

By replacing $\tilde{1}$ with 0, namely, by considering the quantization error, equations (51) – (54) become (55) – (58), respectively.

In the case which does not consider the quantization error, the values of the separability measures η_D,

η_O, η_K, η_Q are shown in **Fig. 1** (the number of gray levels $L = 256$) and **Fig. 2** (the number of gray levels $L = 8$). Equations (51) and (53) are not reliable, in the range $L^2\omega^2 - 1 \leq 0$, $L^2(1-\omega)^2 - 1 \leq 0$. When the number of gray levels is small, the unusual phenomenon appears at near ω =0 or 1 in Fig. 2. From the equations and Fig. 1, Fig. 2, η_Q does not appear the unusual phenomenon by the quantization error, η_O is not influenced by the quantization error.

In the case of considering the quantization error, the values of the separability measures η_D^*, η_O^*, η_K^*,

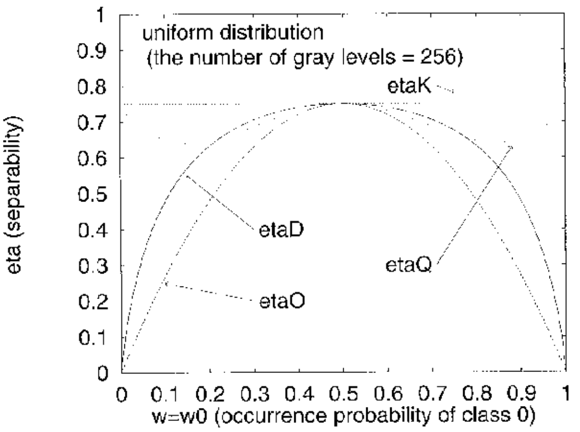

Fig. 1 The values of separability measures based on the likelihood criteria in the case of binarizing a uniform distribution. (the number of gray levels $L = 256$, $etaX = \eta_X$, $X = D, O, K, Q$)

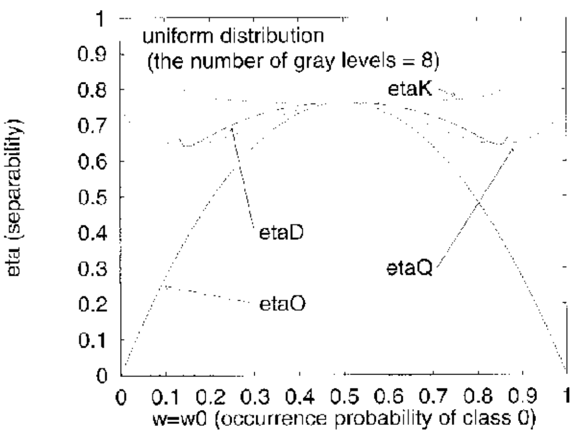

Fig. 2 The values of separability measures based on the likelihood criteria in the case of binarizing a uniform distribution. (the number of gray levels $L = 8$, $etaX = \eta_X$, $X - D, O, K, Q$)

η_Q^* are shown in **Fig. 3**. Because that the influence of the quantization error is small when the number of gray levels is large, the difference between the cases of Fig. 1 and Fig. 3 is very small. As it is clear from equations (55), (56), (57), (58), the separability measures are unrelated to the number of gray levels by considering the quantization error in the uniform distribution model. Especially, η_K^* (equation (57)) is unrelated to the occurrence probability ω too, namely, $\eta_K^* = 3/4$ is constant.

The method (l_K, J_K)[7] by Kittler and Illingworth was proposed, in order to avoid the influence by the bias of the occurrence probabilities. By considering the quantization error in the uniform distribution model, the fact clarifies the expectant effect that the criteirion η_K^* is constant regardless of the occurrence probability. (In the case of multi-thresholding into M classes, the criterion $\eta_K^* = 1 - 1/M^2$ is also constant, in the same way.)

From literature[9](p.44), $\lim_{\omega \to 0} \omega^\omega = 1$. In the case of binarizing the uniform distribution, when ω approaches 0 or 1, η_Q^* (equation (58)) approaches 3/4, η_D^* (equation (55)) approaches 0.

Although the four likelihood criteria have different conditions about occurrence probabilities and variances, in the case that a gray level histogram is symmetric such as a uniform distribution, the conditions about occurrence probabilities and variances be-

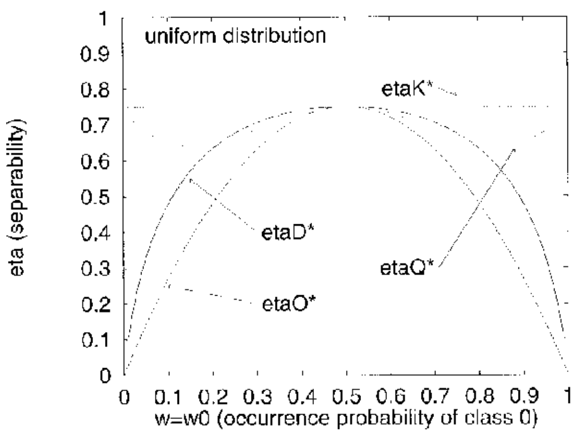

Fig. 3 The values of separability measures based on the likelihood criteria in the case of binarizing a uniform distribution. (in the case of considering the quantization error, $etaX = \eta_X$, $X = D, O, K, Q$)

come same at $\omega = 1/2$. As shown in equations (51) – (58) and Fig. 1, Fig. 2, Fig. 3, the values of the four (D, O, K, Q) defined separability measures become the same at $\omega = 1/2$. The above fact indicates that the defined separability measures are valid.

By the way, in the case of binarizing a mixture of two normal distributions $(N(u; 63, 16), N(u; 192, 16))$, the values of the separability measures η_X^* are shown in **Fig. 4**, where,

$$N(u; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(u-\mu)^2}{2\sigma^2}\right\}. \quad (59)$$

In this case, the values of the all separability measures become maximum at the midpoint between the two normal distributions. In the case of a mixture of ideal normal distributions (Fig. 4), the difference of the curves of each separability measure is very small regardless of the number of gray levels and consideration of the quantization error.

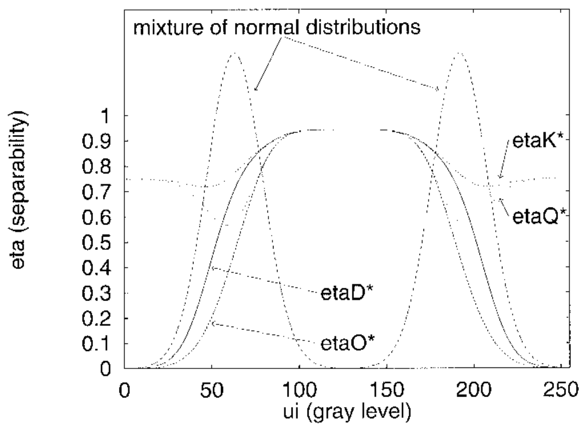

Fig. 4 The values of separability measures based on the likelihood criteria in the case of binarizing a mixture of two normal distributions. ($etaX^* = \eta_X^*$, $X = D, O, K, Q$)

§ 4 Conclution

In the case of thresholding under the assumption of mixture of normal distributions, the separability measures which are invariant under affine transformations of gray level scale and are normalized to be within values from 0 to 1 have been defined from likelihood criteria. Using these separability measures, the likelihood criteria have been investigated in the case of a uniform distribution model. It has been shown that consideration of the quantization error is effective and that the proposed separability measures are valid.

Acknowledgment

The author wishes to thank Dr. N.Otsu, Head of the Machine Understanding Division, and Dr. K.Sakaue, Leader of the Object Recognition Lab., for their hospitality and encouragement. Thanks are also due to the members of the Object Recognition Lab., for their helpful suggestions.

References

1) T.Kurita, N.Otsu and N.Abdelmalek: "Maximum likelihood thresholding based on population mixture models," Pattern Recognition, Vol.25, No.10 (1992-10) 1231–1240.

2) I.Sekita, T.Kurita, N.Otsu and N.Abdelmalek: "Thresholding Methods Considering the Quantization Error of an Image," Trans. IEICE, Vol.J78-D-II, No.12 (1995-12) 1806–1812 (in Japanese).

3) T.Saito: "Performance Comparison between Binarization Methods by Likelihood Criteria Using Hand-written Character Database ETL1," Bul. ETL, Vol.61, No.2 (1997-02) 97–109 (in Japanese).

4) T.Saito: "A Binarization Method by Likelihood Criteria with Gray Scale Modification," Trans. IEICE, Vol.J80-D-II, No.6 (1997-06) 1343–1351 (in Japanese).

5) N.Otsu: "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. on Syst., Man & Cybern., Vol.SMC-9, No.1 (1979-01) 62–66.

6) N.Otsu: "An Automatic Threshold Selection Method Based on Discriminant and Least Squares Criteria," Trans. IECE, Vol.63-D, No.4 (1980-04) 349–356 (in Japanese).

7) J.Kittler and J.Illingworth: "Minimum Error Thresholding," Pattern Recognition, Vol.19, No.1 (1986) 41–47.

8) The Mathematical Society of Japan: Mathematical Dictionary, Iwanami (1968) (in Japanese).

9) S.Moriguchi, K.Udagawa and S.Hitotsumatsu: Formulas of Mathematics I, Iwanami (1956) (in Japanese).

(Accepted August 25, 1999)

The Author

Taiichi SAITO
Object Recognition Lab., Machine Understanding Division
Image Processing, Analysis of Character Databases, Pattern Recognition.
E-mail= saito@etl.go.jp
URI= http://www.etl.go.jp/~saito/