DBpia

論文 96-21-7-5

Customised Feature set Selection for Automatic Signature Verification

Younglae J. Bae*, Dong-Uk Cho**, Ji-Yeong Kim*** *Regular Members*

서명자동검정을 위한 개인별 특징 세트 선택

正會員 배영래*, 조동욱**, 김지영***

ABSTRACT

This paper covers feature extraction for automatic handwritten signature verification. Several major feature selection techniques are investigated from a practical perspective to realise an optimal signature verification system, and customised feature set selection based on set-to-set distance measurement is presented. The experimental results have proved the proposed methods to be efficient, offering considerably improved verification performance compared to conventional methods. Also, they dramatically reduce the processing complexity in the verification system.

요약

본 논문에서는 서명자동검증을 위한 특징 선택 방법에 대해 논한다. 기존에 사용되어 온 주요한 특징선택방법들이 최적한 서명자동검증시스템의 구현을 위해서 실용적인 측면에서 조사되었는데 특히 최근 관심을 가져온 개인별 서명 특징 추출 방법에 초점이 맞추어졌다. 월등한 검증 성취력을 갖고 있으나 과도한 계산 복잡도로 실용화되지 못하고 있는 기존의 경험적 탐색 방법이 본 논문에서는 집단과 집단 사이의 거리에 기초한 통계적 평가 방법으로 개선되었다. 제안되는 방법의 실험결과 기존의 방법에 비해 검증율을 현저히 향상시키는 유효한 방법임이 증명되었는데, 처리 복잡도도 대단히 감소시켰다.

Ⅰ. Introduction

Signatures have for centuries provided a primary means of legal authorization and personal authentication. Unlike other traditional identification tools, such as the fingerprint, signatures have become widely accepted and gained respect throughout history.

The recognition of handwritten signatures, however, still remains a difficult problem, as, unlike other biometric characteristics such as the fingerprint or retinal pattern, a signature is a personal representation which reflects the writer's own intellectual, psychological

---

* 시스템공학연구소(SERI)
** 서원대학교 정보통신공학과
Dept. of Information & Communication Engineering, Seowon Univ.
*** 서원대학교 전자계산학과
Dept. of Computer Science & Engineering, Seowon Univ.
論文番號:96101-0323
接受日字:1996年 3月 23日

www.dbpia.co.kr

and physical characteristics, and, hence, the characteristics may vary depending on his mental state, health and writing environment at any instance [4, 7]. This causes the main barrier in realising their successful recognition.

One of the major problems encountered, as in other fields of pattern recognition, is to find the optimal feature set which can best represent the characteristics of the pattern data. Particularly in signature tecognition, many efforts have been devoted to the problem of *feature extraction*, extraction of a feature vector that contains as much as possible of the significant information of a signature pattern. However, a relatively small number of studies have been carried out for *feature selection*, reduction of the dimensionality of the extracted feature vector while maintaining all the discriminating power of the original feature set by excluding or combining the redundant or irrelevant features in the pattern data. [1, 4, 5, 6, 10, 11, 12, 15, 16, 18, 21, 22].

Feature selection could play a more important role than has been anticipated to data in the context that reduction of the dimensionality not only has a direct bearing on the computational demands but also contributes to the improvement in error rate performance or tolerance towards errors [4, 6, 22].

Most of the work in feature selection during the past years has been carried out by heuristic or syntactic approaches, which are basically known to be lacking in objective selection standards, i.e., theoretical grounds of evaluation criteria for feature selection, and to have a huge computational complexity in processing time.

In this paper, a statistical approach using customised feature sets with an acceptable computational latency is proposed. Unlike the previous statistical approaches, this approach has derived the set-to-set distance measurement in terms of variances [29], which does not require the pre-assumption of equal covariances.

It not only provides a feasible methodology to objectively evaluate each feature's discriminating ability, but also maintains an affordable computational complexity as it replaces the exhaustive heuristic searching processes with a one-time-ranking process for each individual, in which features are ordered according to their discriminating power in terms of the correlation coefficients to the major feature sets.

Ⅱ. Signature Data Base

The signature data base for this study has been built from public trials with the KAPPA[1)] signature verification system [8], covering 8,500 signature samples: From these samples, 64 individuals who had donated 10 signatures in 4 sessions and had no serious degree of noise in their samples have been randomly selected, which constitutes 2,560 samples.

In building this signature data base, the Summagraphics Plus graphics tablet with wired pen, which was interfaced to a standard PC host, was chosen. The Summagraphics Plus tablet is capable of sampling the x-y coordinates of a special pen at the rate of 100 Hz and with a resolution of 100 points per inch, and additionally detects paper contact.

The generated information $z(t)$ from this device could be easily transformed to a position function, $f(x(t), y(t))$, using Equation (1).

$$I[x(t),\ y(t)] = z(t)\ \forall t \tag{1}$$

Where $I[.]$ is the two-dimensional static image, $x(t)$, $y(t)$ and $z(t)$ are the Cartesian coordinates in the time domain t.

Ⅲ. Feature Extraction

[1)] KAPPA is a signature verification system based on image analysis algorithms developed at the University of Kent, England, and available under licence from the British Technology Group.

www.dbpia.co.kr

To define an optimal set of signature features, in the first place, all the primitive features of signatures which contain significant information should be extracted. For this, the following parametric features have been extracted.

Total Duration (TimT): This is the total duration of a signature signal, which is only applied to on-line systems. In this study, it was annotated as the variable name "TimT".

Width (WidT, WidF, WidL): The width of the signature image. WidT is for the total image, WidF is for the first segment and WidL for the last segment.

Height (HigT, HigF, HigL): The height of the signature image. HigT for the total, HigF for the first segment and HigL for the last segment.

Total Length (LenT): This is the total geometrical length of the signature.

Convex Length (CVLT, CVLX1, CVLX2, CVLX3): The sum of lengths of convex strokes. CVLT is for the total, CVLX1 for the 1st section, CVLX2 for the 2nd and CVLX3 for the 3rd.

Concave Length (CCLT, CCLX1, CCLX2, CCLX3). This is for the case of concave strokes.

Pen Down Time (PDTT, PDTF, PDTL): The duration while the pen tip is in contact with the writing surface. PDTT is for the total, PDTF is for the first segment and PDTL is for the last segment.

Pen Down Length (PDLT, PDLF, PDLL): This is the case for the geometrical length.

Height to Width (HtoW): The ratio of the height to the width of the signature.

Width to Height of First Segment (WtoHF): The ratio of the width to the height of the first segment.

Width to Height of Last Segment (WtoHL): The ratio of the width to the height of the last segment.

Longest Stroke to Width (LStoW): The ratio of the longest stroke to the width.

Shortest Stroke to Width (SStoW): The ratio of the shortest stroke to the width.

Longest Stroke to Preceding Stroke (LStoP): The ratio of the longest stroke to the preceding.

Left to Begin (LtoB): The distance from the left most position of a signature image to the beginning position of the signature writing.

Number of Zero Crossings (NoCY1, NoCY2, NoCY3): The number of zero crossing points in the signature, which is similar to the features "maxima and minima" reported in some studies. NoCY1 is for the 1st vertical section, NoCY2 is for the 2nd and NoCY3 for the 3rd.

Sum of Velocity (SMVX1, SMVX2, SMVX3): The sum of velocity. SMVX1 is for the 1st horizontal section, SMVX2 for the 2nd, SMVX3 for the 3rd.

Number of Segments (NoST, NoSX1, NoSX2, NoSX3): The number of segments for the total, the 1st horizontal section, the 2nd section and the 3rd section.

Number of Junction Points (NoJT, NoJX1, NoJX2, NoJX3): This is the case for jurction points in the signature image.

Number of Points (NoPX1, NoPX2, NoPX3): The number of pixel points for the horizontal sections. These are same as the duration in on-line systems.

Histogram Width to Height Ratio (HWH1, HWH2): The histogram built on the vectical projection of a signature can be used as a feature. HWH1 is the ratio of the width of the left half to the height and HWH2 is for the right half.

Vertical Density (VDY1, VDY2, VDY3): These are the integral values of individual sections of the vertical histogram. VDY1 is for the 1st section and so on.

Horizontal Density (HDX1, HDX2, HDX3): These are for the horizontal histogram.

Fourier Transform Coefficients (FFT1, FFT2, FFT3, FFT4): The sequence of x, y coordinate pairs was applied as real and imaginary parts of a Fourier transform function and the coefficients in the frequency domain are used as features.

Ⅳ. Feature Evaluation

4-1. Heuristic Approach [4]

www.dbpia.co.kr

Given an initial feature vector

$$y=(y_1, y_2, \ldots, y_D) \tag{2}$$

where y_i is the i^{th} feature of the measurement vector y of dimensionality D.

To find the optimum set of features

$$x=(x_1, x_2, \ldots, x_d) \tag{3}$$

where x_i is the i^{th} feature of the measurement vector x of dimensionality d,

$$R(X)=max\ R(\Upsilon)\forall\Upsilon \tag{4}$$

where $R(\cdot)$ is the criterion and Υ the set of $\left[\begin{matrix} D \\ d \end{matrix}\right]$ features from Y

will be

$$\left|\begin{matrix} \begin{pmatrix} D \\ d \end{pmatrix} \\ D \end{matrix}\right| = \left|\begin{matrix} \dfrac{D!}{(D-d)!\,d!} \\ D \end{matrix}\right|$$

This number of calculations is too huge to be realised in most practical situations:

In the work by Brittan [4], the number for the case of $D=20$ and $d=5$, i.e., the case that five features are to be selected from twenty features, was calculated as 168 billion combinations for five features, which is computationally prohibitive. Hence, the need for an alternative search scheme has given rise to several "sub-optimal" searching methods as follows: (Nevertheless these compromises cannot guarantee finding the "optimal" set of features.)

• Sequential Forward Selection (SFS)

This method starts from an initially empty feature vector. Features can be sequentially added to the feature vector through proving their relative merits in relation to the rest of candidate features in conjunction with the existing feature vector:

$$K_n=\max\nolimits_{n=1}^{n=d} R(K_{n-1}+\kappa_i) \tag{6}$$

where $\kappa_i \in K_{n-1}$ and $1 \leq n \leq D$ is the size of the selected feature vector.

• Sequential Backward Selection (SBS)

This is the top-down approach where features are sequentially eliminated from the initially full installed feature vector. A feature must be excluded from the feature vector if its absence has resulted in a better feature vector than both the existing feature vector and the results of the rest of the remaining features:

$$K_n=\max\nolimits_{n=1}^{n=d} R(K_{n+1}-\kappa_i) \tag{7}$$

where $\kappa_i \in K_{n+1}$ and $1 \leq n \leq D$ is the size of the selected feature vector.

• SFS-SBS Combined Method

The SFS-SBS combined method is to overcome these problems by selectively combining SFS and SBS, such as employing both m features to add and n features to delete to the active feature vector.

The ratio of n to m will depend upon the required selection approach:

If $n>m$, then the feature vector will increase in size, implying a bottom-up approach.

If $n<m$, then the feature vector will decrease in size, implying a top-down approach.

And the magnitude of n and m will reflect the degree of tolerance towards redundance and the processing requirements.

4-2. Statistical approach

A few studies [15, 17, 20] have applied statistical approaches to comparison processes to provide an objective criterion tool. However, despite all their analytical merits, they have been, apart from the fact that

www.dbpia.co.kr

they stil have not resolved the exhaustive searching complexity, generally regarded to have two irrevocable defects relevant to their use for signature verification:

1. Contrary to the general assumption in those studies that the set of features is distributed as a multivariate Gaussian density and that the covariance matrices are equal, some features have been found not to satisfy that condition.
2. The lack of a valuable statistical model of true signatures versus forgeries distribution makes the use of statistical feature selection methods impossible [19].

Ⅴ. Customised Feature Selection

Over the years, many well known statistical approaches have been used in relation to the problem of feature evaluation: *Principal Component Analysis, Hotelling Transform, Karhunen-Loève Expansion, Entropy Minimisation, Eigenvalue Analysis, Factor Analysis, Canonical Analysis, Divergence Maximisation* and *Discriminant Analysis*, etc. [2, 6, 9, 17, 21]. Unlike the specific studies particularly devoted to signature verification [15, 17, 28], in which comparison methods using simple distance measurements, such as *Euclidean distance, Quadratic Discriminant Model, Bayes criterion, Fisher criterion, F-ratio analysis* or *Mahalanobis distance* were applied, these approaches have derived the set-to-set distance measurement in terms of variances [21] and have not required the pre-assumption of equal covariances. Their basic concept is to analyse the divergence between patterns and to find the most effective features to discriminate them. Although they have originated from different motivations and may have used different methodologies, all of them are closely related in the context that they are based on the Eigenvalue problem.

Among these approaches, as signature verification in this study is equivalent to the practical problem of discriminating between authentic specimens and forgeries, the feature evaluation problem surely belongs to the case of Divergence Maximisation or the two classes-case of Discriminant Analysis.

In this paper, the feature evaluation and feature selection processes were mainly performed using two-class-Discriminant Analysis for customised feature set selection.

- Eigenvalue Analysis [2, 13]

A pattern which has N attributes (features) can be arranged in the form of a pattern vector:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \qquad (8)$$

where x_i is the i^{th} feature of the measurement vector x.

If we assume a set which has M patterns of this kind, this set can be defined as a matrix:

$$[X] = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1N} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{M1} & \cdot & \cdot & \cdot & \cdot & x_{MN} \end{bmatrix} \qquad (9)$$

where $x_{i,}$ is the i^{th} feature of the measurement vector x which belongs to the i^{th} pattern.

And the covariance maxtrix between features can be defined as

$$C_x = \begin{bmatrix} c_{11} & c_{12} & \cdot & \cdot & \cdot & c_{1N} \\ c_{21} & c_{22} & \cdot & \cdot & \cdot & c_{2N} \\ c_{31} & c_{32} & \cdot & \cdot & \cdot & c_{3N} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{M1} & \cdot & \cdot & \cdot & \cdot & c_{MN} \end{bmatrix} \qquad (10)$$

www.dbpia.co.kr

where C_{ij} is the covariance between the i^{th} and the j^{th} features of the measurement vector x, and C_{ii} is the variance of the ith feature of x:

$$= E\{(x - m_x)(x - m_x)'\}$$

$$\cong \frac{1}{M} \sum_{i=1}^{M} (x - m_x)(x - m_x)' \tag{11}$$

where $m_x = E\{X\} \cong \frac{1}{M} \sum_{i=1}^{I} X$

Then the correlation matrix between features is

$$R_x = \begin{bmatrix} R_{11} & R_{12} & \cdot & \cdot & \cdot & R_{1N} \\ R_{21} & R_{22} & \cdot & \cdot & \cdot & R_{2N} \\ \cdot & \cdot & & & & \cdot \\ \cdot & R_{ij} & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ R_{N1} & R_{N2} & \cdot & \cdot & \cdot & R_{NN} \end{bmatrix} \tag{12}$$

where $R_{ij} = C_{ij}/\sigma_i \sigma_j$
where $\sigma_i = \sqrt{C_{ii}}$, $\sigma_j = \sqrt{C_{jj}}$

Consider the vector equation

$$R_x = \lambda x \tag{13}$$

where λ is a number.

A value of λ for which Equation (13) has such a solution as $\lambda \neq 0$ is called an eigenvalue of the matrix R and there can exist N eigenvalues.

Let these values be λ_k, k = 1, 2, ..., N, and the corresponding eigenvectors e_k, $k = 1, 2, \ldots, N$.

If we define a matrix A

$$A = \begin{bmatrix} e_1' \\ e_2' \\ \cdot \\ \cdot \\ \cdot \\ e_N' \end{bmatrix}, \quad A^{-1} = (e_1, e_2 \ldots e_N) \tag{14}$$

such that $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_k| > \ldots > |\lambda_N|$,

then the largest of the absolute values among the eigenvalues of R, $|\lambda_1|$, is called the spectral radius of R and the corresponding eigen vector, e_1, is called the first principal component (axis).

This principal axis represents the maximum divergent axis and is considered to maximally discriminate the patterns with respect to any other axes.

Each feature has a contribution level to this principal axis as well as to the following major axes, which is represented as a correlation coefficient to this axis.

Assume a transformed variable, y_1, from the first eigen vector, e_1, such that

$$y_1 = e_1' x. \tag{15}$$

Then the correlation coefficient for the j^{th} component of x, x_j, to the first eigen vector is

$$R_{x_j} = \frac{C_{x_j y_1}}{\sigma_{x_j} \sigma_{y_1}} \tag{16}$$

Actually these correlation coefficients in this study are similar to the coefficients of the eigenvectors in the matrix A in Equation (14) as the eigenvectors are calculated from the correlation matrix between features, R in Equation (12).

The correlation coefficient represents how much a feature contributes to the feature set and, hence, can be regarded as a measurement tool which can evaluate each feature.

Several numerical methods have been developed for solving this eigenvalue problem and a popular iterative approximation method, *Jacobian Iteration*, was implemented in this study.

- Divergence Maximisation [2, 13, 21]

Divergence is a measure of similarity between pairs of arbitrary distributions and a measure of the difficulty in discriminating between these two distributions. It can be used to evaluate the discriminating

www.dbpia.co.kr

power of features and to determine feature ranking.

Let the probability of occurrence of pattern x of Equation (8) in the multi-dimensional feature space be represented as:

$$\varphi(x)=\frac{1}{(2\pi)^{\frac{1}{2}N}\,|C|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-m_x)^0C^{-1}(x-m_x)\right] \tag{17}$$

where C is the covariance matrix and mx the mean vector of this pattern respectively,

Then the probabilities of x to belong to given pattern classes G_i and G_j are

$$P(x|G_i)=\frac{1}{(2\pi)^{\frac{1}{2}N}\,|C_i|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-m_i)^0C_i^{-1}(x-m_i)\right] \tag{18}$$

where C_i is the covariance matrix and m_i the mean vector of class G_i, respectively, and

$$P(x|G_j)=\frac{1}{(2\pi)^{\frac{1}{2}N}\,|C_j|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-m_j)^0C_j^{-1}(x-m_j)\right] \tag{19}$$

where C_j is the covariance matrix and m_j the mean vector of class G_j, respectively.

The discriminating information for class G_i versus class G_j and for class G_j versus class G_i may be measured by the logarithm of the likelihood ratio

$$U_{ij}=\ln\frac{P(x|G_i)}{P(x|G_j)} \tag{20}$$

and

$$U_{ji}=\ln\frac{P(x|G_j)}{P(x|G_i)} \tag{21}$$

respectively.

Then the average discriminating information for class G_i is given by

$$I(i,j)=\int_x P(x|G_i)\ln\frac{P(x|G_i)}{P(x|G_j)}\,dx \tag{22}$$

and

$$I(j,i)=\int_x P(x|G_j)\ln\frac{P(x|G_j)}{P(x|G_i)}\,dx \tag{23}$$

respectively, also.

The divergence for these two classes is defined as

$$\begin{aligned}J_{ij}&=I(i,j)+I(j,i)\\&=\int_x [P(x|G_i)]\ln\frac{P(x|G_i)}{P(x|G_j)}\,dx\\&=\frac{1}{2}\,\mathrm{tr}[(C_i-C_j)(C_j^{-1}-C_i^{-1})]\\&+\frac{1}{2}\,\mathrm{tr}[(C_i^{-1}+C_j^{-1})(m_i-m_j)(m_i-m_j)']\end{aligned} \tag{24}$$

where tr is the trace of a matrix.

In many practical applications, Equation (24) is apprizimated using *pooled variance*, P, as

$$J_{ij}=(m_i-m_j)'\,P^{-1}(m_i-m_j) \tag{25}$$

and P is defined as

$$P=\frac{1}{M_i+M_j-2}\,((M_i-1)C_i+(M_j-1)C_j) \tag{26}$$

where M_i and M_j are the number of members in class G_i and G_j, respectively.

- Feature ordering using Divergence

Let A be a transformation matrix of dimension $T\times N$, such that $T<N$, from Equation (14).

Then the pattern vector x of N dimension can be reduced to a T dimensional vector, y

www.dbpia.co.kr

$$y = Ax. \tag{27}$$

The mean vectors and the covariances of this new vector y for the class G_i and G_j will be

$$m_i^* = Am_i, \quad m_j^* = Am_j \tag{28}$$

and

$$C_i^* = AC_iA', \quad C_j^* = AC_jA' \tag{29}$$

respectively.

Under these conditions, the divergence is given by

$$J_{ij} = \frac{1}{2}\ \mathrm{tr}[(C_j^*)^{-1}C_i^* + (C_i^*)^{-1}C_j^*] - T + \frac{1}{2}\ \mathrm{tr}\{[(C_i^*)^{-1} + (C_j^*)^{-1}]\,\delta^*(\delta^*)'\} \tag{30}$$

where

$$\delta^* = A\delta = A(m_i - m_j) = m_i^* - m_j^* \tag{31}$$

Since the trace of a matrix is equal to the sum of its eigenvalues,

$$J_{ij}^* = \frac{1}{2}\sum_{k=1}^{T}(\lambda_k + \lambda_k^{-1}) - T + \frac{1}{2\lambda_{T+1}} + \frac{1}{2}\lambda_{T+2} \tag{32}$$

where λ_k are the eigenvalues of $(C_j^*)^{-1}(C_i^*)$,
λ_{T+1} are the eigenvalues of $(C_i^*)^{-1}\delta^*(\delta^*)'$,
and λ_{T+2} are the eigenvalues of $(C_j^*)^{-1}\delta^*(\delta^*)'$.

Then a necessary condition for the divergence J_{ij}^* to be an extremum is that the matrix A satisfy the relation

$$G = \sum_{k=1}^{r}(1 - \lambda_k^{-2})(C_jA' - \lambda_k C_jA')e_k e_k' + (\delta\delta'A' - \lambda_{T+1}C_iA')e_{T+1}e_{T+1}' + (\delta\delta'A' - \lambda_{T+2}C_jA')e_{T+2}e_{T+2}' = 0 \tag{33}$$

where λ_k and e_k are the eigenvalues and eigenvectors of $(AC_jA')^{-1}(AC_iA')$,
λ_{T+1}, e_{T+1} and λ_{T+2}, e_{T+2} are the eigenvalues and eigenvectors of $(C_i^*)^{-1}(A\delta\delta A')$ and $(C_j^*)^{-1}(A\delta\delta A')$, respectively,
and $\delta = m_i - m_j$.

In general cases, where $C_i^{-1}C_j$ and $m_i \neq m_j$, this solution is solved by a *steepest-ascent approach* in which we lead a set of difference equations to a convergence by iteration.

$$A(s+1) = A(s) + \theta G(s) \tag{34}$$

where s is an iteration index and θ is some suitable convergence factor.

The solution matrix A of Equation (34) has M eigenvectors (*canonical discriminant functions*) that are rank-ordered. The first function is the maximum divergent axis and is considered to maximally discriminate the two classes and so on.

Again, for practical applications, these eigenvectors can be derived by solving Equation (26). In this study, the eigenvectors are calculated from this approximation using *Jacobian Iteration*, along with using the well-known statistical package SPSS as a proofing tool.

Each feature has the *pooled-within-groups correlation* to these vectors.

$$R_{ij} = \frac{1}{M_i + M_j - 2}\ ((M_i - 1)R_i + (M_j - 1)R) \tag{35}$$

where M_i and M_j are the number of members in class G_i and G_j, and R_i and R_j are the correlation coefficients of class G_i and G_j to the function, respectively.

The correlation coefficients R_i and R_j can be calculated by the similar methods as Equation (15) and Equation (16).

• Computational Time Complexity

www.dbpia.co.kr

Given a feature vector of dimensionality D, assuming that the maximum dimensionality of the optimum feature vector is d and the time to calculate one criterion function is equivalent for all combinations of features, then computational time complexity for feature selection can be measured in terms of the number of calculations for the criterion function, which will be

1. Optimal heuristic search

$$\binom{D}{1}+\binom{D}{2}+\cdots+\binom{D}{d} \tag{36}$$

2. Suboptimal heuristic searches

$$D+(D-1)+(D-2)+\cdots+(d+1) \tag{37}$$

3. The proposed statistical method

$$d \tag{38}$$

VI. Experimentation

A total of 2,560 handwritten signatures from 64 writers during 4 sessions was subjected to feature selection, and 56 parametric features, which include 47 amplitude features, 5 histogram features and 4 transform coefficient features, were initially extracted.

Features were then evaluated using Discriminant Analysis for both the case of the universal feature set and the case of each customised features set.

Table 1 is the result for universal feature set, in

Table 1. Pooled-within-groups correlations

Functions / Features	Funct 1	Funct2	Funct3	Funct4	Funct5
PDLF	[illegible]	0.19883	0.10136	0.24388	0.03531
PDLT	.48055*	0.21993	[illegible]	[illegible]	[illegible]
CCLX2	.48876*	0.20213	[illegible]	0.11958	0.10119
WIDT	.57204*	0.01359	[illegible]	0.14624	[illegible]
WIDL	.50627*	0.13479	[illegible]	[illegible]	[illegible]
WIDF	.56570*	0.02501	[illegible]	[illegible]	[illegible]
CCLX3	.41974*	0.16891	[illegible]	0.18117	[illegible]
TIMT	0.25688	.45566*	[illegible]	[illegible]	[illegible]
PDTL	[illegible]	.35391*	[illegible]	[illegible]	[illegible]
PDLL	0.06457	0.10993	[illegible]	[illegible]	[illegible]
TENT	[illegible]	[illegible]	[illegible]	0.20251	[illegible]
HDX2	[illegible]	0.23198	[illegible]	0.20167	[illegible]
SMVX3	[illegible]	0.02213	[illegible]	[illegible]	[illegible]
VDY3	0.28413	0.06644	.38797*	[illegible]	[illegible]
HDX3	0.26351	0.20369	[illegible]	0.23536	[illegible]
HIGF	[illegible]	[illegible]	[illegible]	.43379*	[illegible]
HIGT	0.27411	0.06024	[illegible]	.39226*	[illegible]
FFT2	0.29012	0.09449	0.29016	.36763*	0.15017
HTOW	-0.18036	0.07326	0.14517	.35860*	0.11385
PDTT	0.15183	0.28366	0.30049	0.21596	.51916*
FFT1	0.01747	0.01076	[illegible]	[illegible]	[illegible]
NOSX3	[illegible]	0.31922	[illegible]	[illegible]	[illegible]
NOSX1	0.26398	0.29792	0.11676	0.34053	0.18897
SMVX1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
FFT3	0.04389	0.00006	[illegible]	[illegible]	[illegible]
PDTF	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SMVX2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
LTOB	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
VDY2	[illegible]	0.13461	0.01793	0.05738	[illegible]
WTOHF	[illegible]	[illegible]	0.07087	0.12976	[illegible]
HDX1	0.13343	[illegible]	0.20390	0.01212	[illegible]
CCLX1	0.12577	0.11051	[illegible]	0.01515	[illegible]
[illegible]	[illegible]	[illegible]	0.0181	0.06278	0.16168
[illegible]	[illegible]	[illegible]	0.25055	[illegible]	[illegible]
FFT4	[illegible]	[illegible]	[illegible]	0.22628	[illegible]
HIGL	0.13079	[illegible]	[illegible]	0.09499	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.07757	[illegible]
[illegible]	[illegible]	[illegible]	0.07227	[illegible]	[illegible]
[illegible]	0.22236	[illegible]	0.01002	0.06878	[illegible]
[illegible]	0.04508	[illegible]	0.1506	0.0495	[illegible]
NOPX1	[illegible]	[illegible]	0.01581	0.03563	[illegible]
VDY1	[illegible]	[illegible]	0.02166	0.06124	[illegible]
NOHX3	[illegible]	[illegible]	[illegible]	0.11331	[illegible]
HWTH	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
LSTOP	[illegible]	[illegible]	0.01398	0.00441	[illegible]
WTOHL	[illegible]	[illegible]	0.003	0.03654	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.00813	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00291
[illegible]	[illegible]	[illegible]	[illegible]	0.00497	[illegible]
[illegible]	[illegible]	[illegible]	0.01802	[illegible]	[illegible]
HWTH2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
NOPX2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.13402	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.10508	[illegible]

www.dbpia.co.kr

which *pooled-within-groups correlations* between discriminating variables *canonical discriminant functions* are listed:

In the first row of the table, the first five major functions (canonical discriminant functions) have been enlisted. Funct1 is the function for the first principal axis, Funct2 is the next, and so on.

In the first column, features are ordered according to their ranks in terms of correlation coefficients within major functionss2. The corresponding next five columns are the coefficients in the five major functions.

These correlation coefficients can be considered as the indicator to the contribution level of each feature as a member of the universal feature set to the Maximum Likelihood Classifier [10]. Hence, in this context, the features, from the best one, are sequentially selected for the universal feature set:

They are sequentially fed into the Maximum Likelihood Classifier until the classifier satisfies the pre-defined desired conditions in terms of error rates and the number of features.

The same procedure has been performed for the customised feature set for each individual class.

Figure 2 is the MLM(Maximum Likelihood Method) verification result for the universal feature sets.

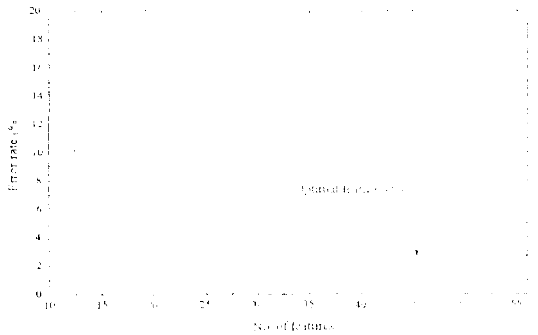

Fig. 2 MLM using the universal feature sets

The figure shows that reducing the number of features does not always degrade the MLM performance. Rather, the best performance has been recorded by using 46 features, not using all 56 features.

For customised feature sets, each writer has his own optimum feature set and, accordingly, the customised feature sets differ in number between individual writers. However, the average number of optimum feature sets was significantly reduced with an impressive error rate performance improvement. Table 2 shows the results of MLM for both types of feature sets.

In the table, using customised feature sets has shown a significant performance improvement, an average error rate at 0.7 percent with an average number of features per signer of 30, from the result using the whole 56 features of an error rate at 2.5 percent.

Ⅶ. Conclusions

This paper has investigated various feature selection methods from a practical perspective to realize an optimal signature verification system, and a statistical approach using the *divergence* measurement concept for customised feature set selection has been proposed. The experimental results show a significant performance improvement in terms of error rate. Also they showed a dramatic reduction in the computational time complexity as the proposed method replaces the exhaustive heuristic searching processes with a one-time-ranking process for each individual.

Practically, it often becomes a problem to collect a sufficient number of samples to build a statistical model, which is equally applied to the heuristic approach while the result related to this problem is not

Table 2. MLM results

ID	Use whole features	Feature optimisation (universal set)	Feature customisation (average)
No. of features	56	46	30
Equal error rate(%)	2.50	2.385	0.709

www.dbpia.co.kr

easily exposed in such a black-box type process. Although future work should seek to solve this issue, the proposed method is thought to have provided a feasible methodology to objectively evaluate the discriminating ability of each feature and an optimal personal feature set for each individual with an acceptable computational latency in practical situations.

We are thanksful Ho-Chul Jeon & Seung-Hyub Yoo for their fine word processing job.

References

1. M. Ammar, Y. Yoshida and T. Fukumura, "Structural Description and Classification of Signature Images," Pattern Recognition, Vol.23, No.7, pp. 697-710, 1990.
2. Y. J. Bae, D. U. Cho and B. U. Choi, "MHSC & Post-Processing for the Automatic Image Segmentation," Trans. on KITE, Vol.24, No.1, 1987.
3. Y. J. Bae and M. C. Fairhurst, "Parallelism for Dynamic Time Warping for Automatic Signature Verification," ICDAR '95: Int'l Conference on Document Analysis and Recognition, Montreal, Canada, Aug. 1995.
4. P. Brittan, "High Performance Parallel Image Analysis Architectures applied to Handwritten Signature Verification," Ph. D. Thesis, University of Kent at Canterbury, 1992.
5. H. Cardot and M. Revenu, "A Static Signature Verification System Based on a Cooperative Neural Networks Architecture," Int'l Journal of Pattern Recognition and Artificial Intelligence, Vol.8, No. 3, pp.679-692, 1994.
6. P. A. Devijver and J. Kittler, "Pattern Recognition : A Statistical Approach," Prentice-Hall Publication, ISBN 0-13-654236-0.
7. I. W. Evett, "A Study of the Verification in the Dimension of Genuine Signatures," Forensic Science Society, pp.206-215, 1985.
8. M. C. Fairhurst et al., "KAPPA-Automatic Signature Verification-Signature Verification Public Trials and Public Survery on Biometric," British Technology Group Ltd., 1994.
9. R. C. Gonzalez and R. E. Woods, "Digital image processing," Addison-Wesley publishing Co., ISBN 0-201-11026-1.
10. D. D. Kerrick and A. C. Bovik, "Microprocessor-Based Recognition of Handprinted Character from a Tablet Input," Pattern Recognition, Vol. 21, No.5, pp.525-537, 1988.
11. C. F. Lam and D. Kamins, "Signature Recognition through Spectral Analysis," Pattern Recognition, 22, 1, pp.39-44, 1989.
12. F. Leclerc and R. Plamondon, "Automatic signature verifiaction: The state of the art-1989-1993," Int'l Journal of Pattern Recognition and Artificial Intelligence, Vol.8, No.3, pp.643-660, 1994.
13. S. Mori, K. Yamamoto and M. Yasuda, "Research on Machine Recognition of Handprinted Characters," IEEE Trans. on PAMI, Vol.6, pp. 386-405, 1984.
14. D. F. Morrison, "Multivariate Statistics methods," McGraw-Hill Int'l Book Co., 07-085469-6.
15. W. Nelson and E. Kishon, "Use of Dynamic Features for Signature Verification," In Proc. IEEE Int. Conf. on SMC, Oct. 1991.
16. W. Nelson, W. Turin and T. Hastie, "Statistical Methods for On-line Signature Verification," Int'l Journal of Pattern Recognition and Artifical Intelligence, Vol.8, No.3, pp.749-770, 1994.
17. W. F. Nemeek and W. C. Lin, "Experimental Investigation of Automatic Signature verification," IEEE Trans. on SMC, Vol.4, pp.121-126, 1974.
18. M. Parizeau and R. Plamondon, "A Comparative Analysis of Regional Correlation, Dynamic Time Warping and Skeletal Tree Matching for Signature Verification," IEEE Trans. 0on PAMI, Vol. 12, No.7, 1990.
19. R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification-The State of the Art," Pattern Recognition, Vol.22, No.2, pp.107, 1989.

www.dbpia.co.kr

20. R. Sabourin and R. Plamondon, "Preprocessing of Handwritten Signature from Image Gradient Analysis," in Proc. 8th Int. Conf. Pattern recog., pp.566-569, 1986.
21. J. T. Tou and R. C. Gonzalez, "Pattern Recognition Principles," Addison-Wesley publishing Co., ISBN 0-201-07587-3.
22. Q. Zhu, "Pattern Classification in Dynamic Environments: Tagged Feature-Class Representation and the Classifiers," IEEE Trans. on SMC, Vol.9, No., 1989.

조 동 욱(Dong Uk Cho) 정회원

1959년 8월 3일생
1983년 2월 : 한양대학교 공대 전자공학과(공학사)
1983년 9월~1985년 8월 : 한양대학교 대학원 전자공학과(공학석사)
1982년~1983년 : (주)신도리코 장학생 겸 기술연구소 연구원
1985년 9월~1989년 2월 : 한양대학교 대학원 전자통신공학과(공학박사)
1987년~1994년 : 한양대학교 공대 강사, 충북대학교 대학원 강사
1991년 3월~현재 : 서원대학교(구·청주사대) 정보통신공학과 부교수(학과정) 겸 전자계산소 소장
1995년~현재 : 한국통신학회 충북지부 임원(감사)
※주관심분야 : 화상처리 및 인식, 퍼지이론, 신경회로

배 영 래(Young Lae J. Bae) 정회원

1953년 3월 29일생
1976년 2월 : 서울대학교 해양물리학과(이학사)
1979년~1980년 6월 : 삼성근무
1986년 2월 : 한양대학교 산업대학원 전자계산학과(공학박사)
1988년 4월~1989년 11월 : 독일 뮌헨대학교 전자계산학과 박사과정
1991년 10월~1995년 7월 : 영국 켄트대학교 전자공학과(공학박사)
1985년 : 최우수 연구원 대통령 표창
1986년 : 우수연구원 한국과학기술연구원 원장 표창
1980년~현재 : 시스템공학연구소 인공지능연구부 책임연구원
※주관심분야 : 화상이해, 인공지능

김 지 영(Ji Yeong Kim) 정회원

1944년 9월 9일생
1971년 2월 : 고려대학교 법학과(법학사)
1971년~1974년 : 대한해운공사
1974년~1977년 : 뉴욕주립대학교 경영학 석사
1977년~1979년 : 뉴욕주립대학교 전산학 석사
1979년~1984년 : 뉴욕주립대학교 전산학 박사
1984년~1985년 : 어번대 전산과 조교수
1986년~1989년 : (주)동양정밀공업 중앙연구소 수석연구원
1989년~현재 : 서원대학교 전자계산학과 교수
※주관심분야 : 화상처리 및 인식, 시스템분석

www.dbpia.co.kr