Feature Weighted Support Vector Machines for Writer-independent On-line Signature Verification

14th International Conference on Frontiers in Handwriting Recognition, Crete, Greece, 2014

Jacques Swanepoel and Johannes Coetzer

Department of Mathematical Sciences, Stellenbosch University, South Africa

1. OVERVIEW

We investigate the feasibility of constructing an effective *writer-independent* model for the purpose of *on-line* handwritten signature verification. The proposed strategy utilises a dynamic time warping (DTW) based dichotomy transformation and a writer-specific dissimilarity normalisation strategy, in order to convert the initial writer-dependent signature representation in feature space into a robust writer-independent signature representation in dissimilarity space (see Figure 1). This dissimilarity-based representation is used to train a support vector machine (SVM).

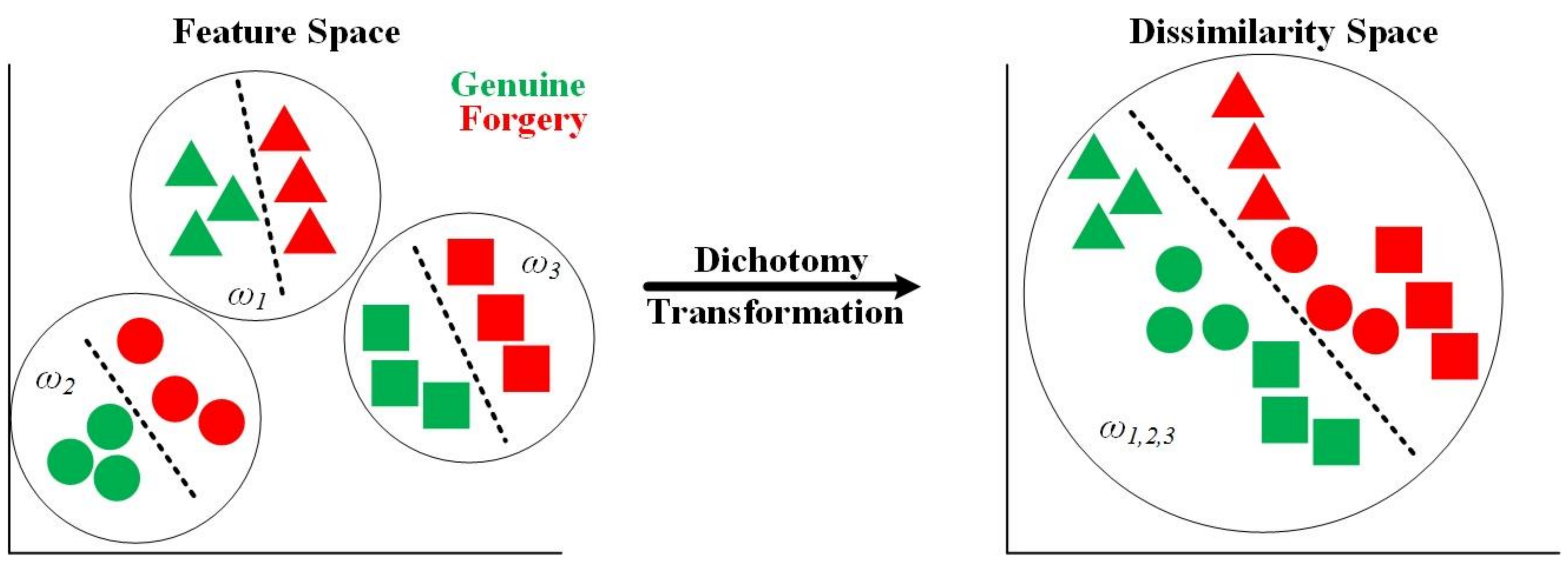

Figure 1 – Conceptualisation of (left) the writer-dependent and (right) the writer-independent approach to signature modelling when three writers are considered.

Furthermore, we investigate the potential gain in system proficiency resulting from the incorporation of *feature weighting* into the SVM kernel function. The purpose of feature weighting is to maximally exploit the discriminative potential of superior signature descriptors, whilst the role of features that are found to be less discriminative is minimised. We present our initial findings as a proof of concept.

2. SIGNATURE REPRESENTATION

Since on-line signatures are captured using specialised hardware, several features are already recorded during the signature acquisition process. These include the pen tip coordinates (x and y), the axial pen pressure p, as well as the azimuth and elevation angles (θ and φ) between the pen and the writing surface (see Figure 2). Furthermore, using the pen tip coordinates, we derive several additional temporal features, namely the pen velocity (v_x and v_y) and pen acceleration (a_x and a_y). Any signature is therefore described by a feature set

$$\boldsymbol{X} = [\boldsymbol{p}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{v}_x, \boldsymbol{v}_y, \boldsymbol{a}_x, \boldsymbol{a}_y, \boldsymbol{\theta}, \boldsymbol{\varphi}]$$

that contains $D = 9$ feature vectors, where each d-dimensional vector comprises the entire sequence of d measurements associated with a specific feature.

In order to obtain a writer-independent representation, such a feature set $\boldsymbol{X}^{(q)}$ is subsequently converted into a dissimilarity vector (DV) by means of a dichotomy transformation. This process quantifies the dissimilarity between $\boldsymbol{X}^{(q)}$ and another feature set $\boldsymbol{X}^{(k)}$ that was extracted from a known genuine sample belonging to the same writer, ultimately yielding the DV $\boldsymbol{z}^{(k,q)}$ such that

$$\boldsymbol{z}^{(k,q)} = \bigcup_{d=1}^{D} \delta(\boldsymbol{x}_d^{(k)}, \boldsymbol{x}_d^{(q)}),$$

where $\delta(\boldsymbol{x}_d^{(k)}, \boldsymbol{x}_d^{(q)})$ denotes the distance between the d^{th} pair of vectors $\boldsymbol{x}_d^{(k)} \in \boldsymbol{X}^{(k)}$ and $\boldsymbol{x}_d^{(q)} \in \boldsymbol{X}^{(q)}$. We utilise a DTW algorithm for this distance calculation, since it enables the non-linear alignment of the feature vectors prior to matching (see Figure 3), and consequently compensates for reasonable intra-class variability.

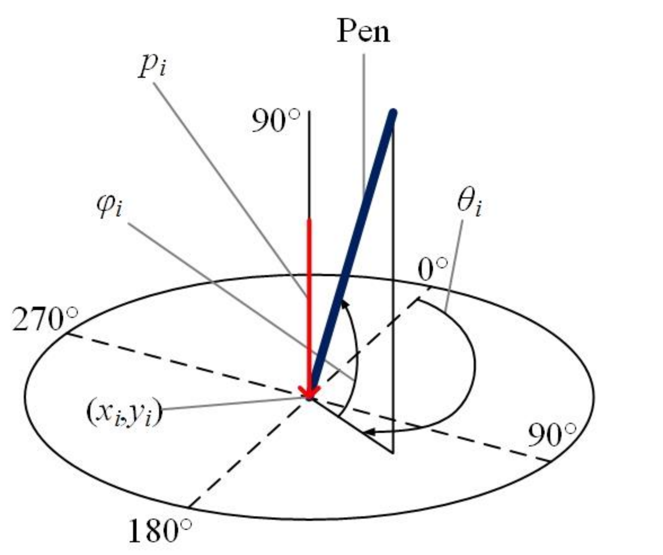

Figure 2 – Several measurements recorded during on-line signature acquisition at time i. These descriptors are combined with additional derived features for the purpose of feature set construction.

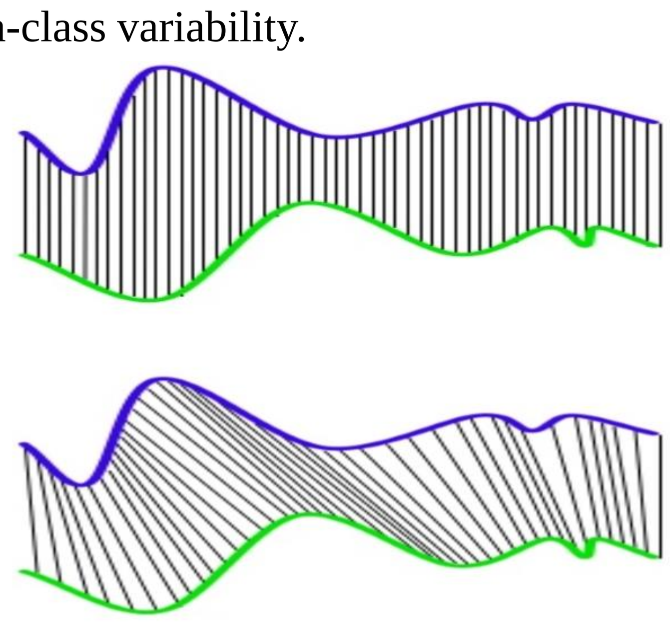

Figure 3 – Conceptual comparison of the feature correspondences considered during dissimilarity vector construction when either (top) the Euclidean distance or (bottom) a DTW-algorithm is utilised.

Finally, each DV belonging to writer ω, denoted by $\boldsymbol{z}^{(\omega)}$, is normalised using a modified logistic function, such that

$$\overline{\boldsymbol{z}}^{(\omega)} = \bigcup_{d=1}^{D} [1 + \exp(\mu_d^{(\omega)} + \sigma_d^{(\omega)} - \frac{6 z_d}{\mu_d^{(\omega)} + \sigma_d^{(\omega)}})],$$

where $\mu_d^{(\omega)}$ and $\sigma_d^{(\omega)}$ denote the mean and standard deviation of the feature-specific dissimilarities obtained when each of the K reference samples belonging writer ω are compared to each other. When many different writers are represented in the training set, this writer-specific strategy ensures improved class separation in dissimilarity space, since only strictly relevant information is considered during the normalisation process.

3. MODELLING & VERIFICATION

The entire set of positive and negative DVs, obtained from all the writers in the training set, is used to train a SVM classifier with membership function

$$f(\overline{\boldsymbol{z}}) = \boldsymbol{w}'\phi(\overline{\boldsymbol{z}}) + b,$$

where $\boldsymbol{w}$ and b denote the weight vector and bias of the optimally separating hyperplane. We consider the conventional linear (LIN) and radial basis function (RBF) SVM kernels, as well as two *weighted* RBF (WRBF) kernels. The WRBF kernel is defined as

$$\phi(\boldsymbol{z}^{(i)})'\phi(\boldsymbol{z}^{(j)}) = \exp(-\gamma \textstyle\sum_{d=1}^{D} \alpha_d (z_d^{(i)} - z_d^{(j)})^2),$$

where α_d denotes the weight associated with the d^{th} feature. We consider two feature weighting strategies, namely the Fisher score (FS) and linear support vector (LSV) methods. The FS-based feature weights equal the inter-to-intra-class-variability-ratio of each feature, whilst the LSV-based weight vector is obtained from the $\boldsymbol{w}$-parameter in the membership function of a trained LIN-SVM.

During system deployment, any questioned signature presented for authentication is first converted into a set of K normalised DVs, which is subsequently presented to the trained SVM. The signature is accepted as genuine if the final confidence score

$$s^* = \frac{1}{K}\sum_{k=1}^{K} [1 + \exp(-f(\overline{\boldsymbol{z}}_{(q,k)}^{(\omega)}))]^{-1}$$

is greater than a specific threshold $\tau \in [0,1]$.

4. EXPERIMENTS

We consider the well-known Philips signature database for system evaluation. This data set contains 1530 genuine signatures and 3000 amateur skilled forgeries obtained from 51 writers. In order to ensure a comprehensive and unbiased performance estimation, the experimental protocol incorporates both 3-fold cross-validation and 10-fold writer randomisation. Each result reported (see Table 1) therefore represents the average performance achieved for 30 system evaluations.

Table 1 – Average EERs (%) obtained when the Philips evaluation set is considered.

μ_{EER}	K							*AVE*
	3	5	7	9	11	13	15	
LIN	8.16	7.08	6.13	5.82	5.17	4.78	4.52	*5.95*
RBF	4.55	3.67	2.62	2.38	1.89	1.47	1.29	*2.55*
FS-WRBF	4.42	3.54	2.60	2.26	1.77	1.37	**1.26**	*2.46*
LSV-WRBF	4.34	**3.52**	2.59	2.23	1.74	1.36	**1.26**	*2.44*

It is clear that, as expected, there is a definite correlation between K (the number of available reference signatures) and μ_{EER} (the verification proficiency of each system). It is also clear that the linear kernel is significantly outperformed by its RBF-based counterparts. Furthermore, both WRBF-kernels also consistently, albeit marginally, outperform the conventional RBF-kernel. Interestingly, the FS-WRBF and LSV-WRBF kernels achieve practically identical performance, despite the fact that they utilise dissimilar feature weights (see Figure 4).

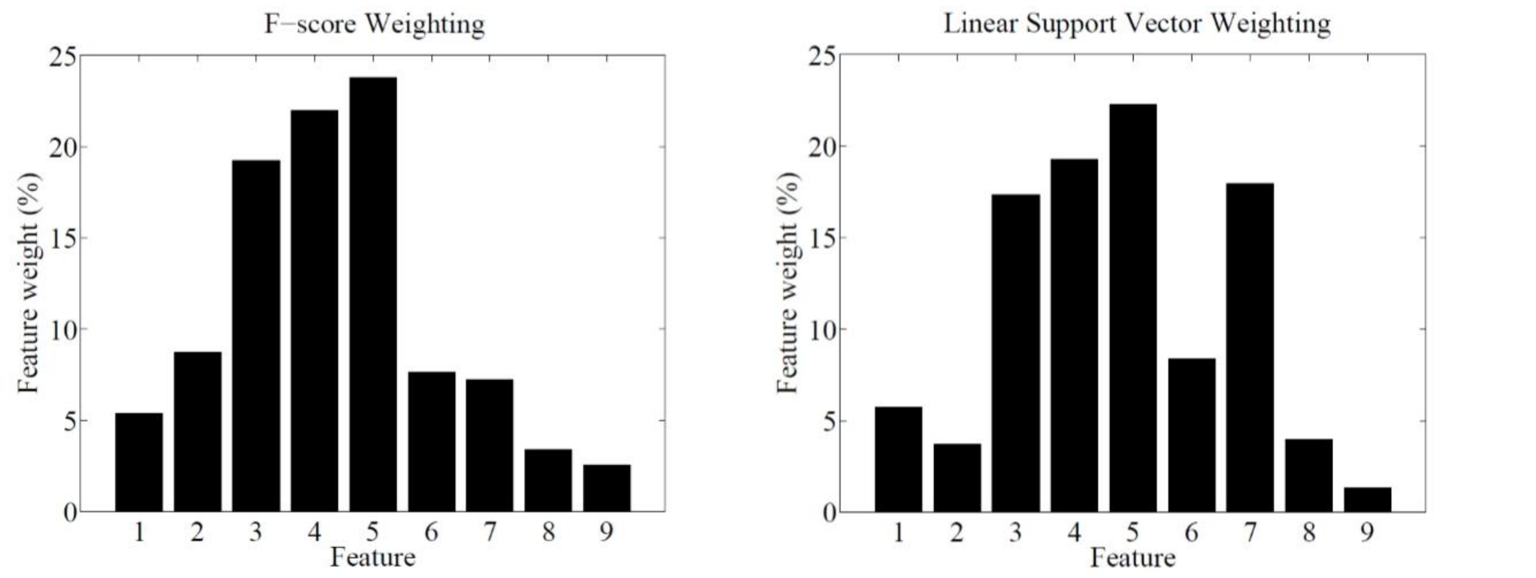

Figure 4 – Average feature weights indicated by the (left) FS and (right) LSV feature weighting methods. The feature indices 1-9 correspond to the feature set column indices i.e. p, x, y, v_x, v_y, a_x, a_y, θ and φ respectively.

Ultimately, the LSV-WRBF kernel is identified as the most proficient candidate. This system achieves an average EER of 4.34%-1.26% when 3-15 genuine reference signatures are available for model construction.

5. CONCLUSION & FUTURE WORK

We showed that a DTW-based dichotomy transformation is able to effectively convert a writer-dependent on-line handwritten signature feature set into a writer-independent dissimilarity-based representation. We also showed that the incorporation of feature weights into the SVM kernel function is able to consistently improve verification proficiency. Our initial findings are promising and the proof of concept is therefore deemed successful.

Future work includes the construction of a significantly expanded feature set, as well the use of more advanced feature weighting strategies. It is reasonable to expect that the inclusion of many additional features should further exploit the benefits resulting from feature weighting and, in all likelihood, lead to improved system proficiency.