A similarity measure between vector sequences with application to handwritten word image retrieval

José A. Rodríguez-Serrano
Loughborough University, UK

Florent Perronnin
Xerox Research Centre Europe, France

Josep Lladós, Gemma Sánchez
Computer Vision Center, UAB, Spain

1 .Goals and motivation

We propose a similarity measure between sequences of vectors

Prior work has dealt with:

• Non-parametric distances, such as dynamic time warping[1] → Limited performance

• Model-based distances between unordered vector sequences (e.g. [2]) → Sequential information lost

• Model based distances using probability product kernels with HMMs (e.g. PPK [3]) → Overfitting

• **We propose PPK between semi-continuous HMMs (SC-HMMs)**
(Gaussians constrained to a "universal vocabulary")

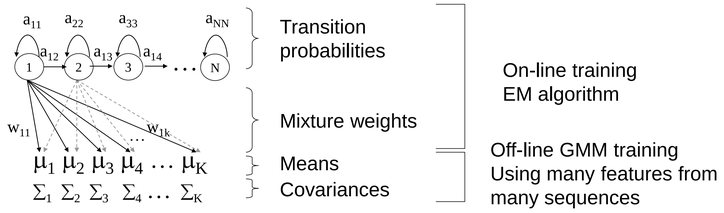
→ Robust to overfitting

→ Computationally efficient

2. SC-HMM training

By using a SC-HMM, we split the HMM parameters into sequence-dependent and sequence-independent

HMM parameters



$$p(x_t|q_t = i) = \sum_{k=1}^{K} w_{ik} p_k(x_t)$$

$X = x_1\ x_2 \ldots x_{Tx}$ → Train SC-HMM with vT_x states → θ_x
v: compression factor

"universal vocabulary" encodes prior information
→ Robust to overfitting

3. Distance between SC-HMMs

Probability product kernel [4] $K^{\rho}_{ppk}(\theta, \theta') = \int_{Z \in \mathcal{Z}} \left[p(Z|\theta)p(Z|\theta')\right]^{\rho} dZ$

Evaluation: for all possible alignments of state sequences, accumulate state-to-state PPK.

Approximation: Consider best path only

➔Evaluation reduces to a DTW between states

DTW: Method for computing an "elastic" dissimilarity between sequences

$$DTW(m,n) = \min \left\{ \begin{array}{c} DTW(m-1,n) \\ DTW(m-1,n-1) \\ DTW(m,n-1) \end{array} \right\} + d(m,n).$$

4. State-to-state approximation

PPK between GMMs

$f = \sum_{i=1}^{M} \alpha_i f_i \quad g = \sum_{j=1}^{N} \beta_j g_j \qquad K^{\rho}_{ppk}(f,g) = \int_x (f(x)g(x))^{\rho}\, dx$

Bhattacharyya similarity $\quad B(f,g) = K^{1/2}_{ppk}(f,g)$

Approximation $\quad B(f,g) \leq \sum_{i=1}^{M}\sum_{j=1}^{N} (\alpha_i \beta_j)^{1/2} B(f_i, g_j)$

Similarity between states → Bhattacharyya distance between mixture weights

$$K^{1/2}_{ppk}(f,g) \approx \sum_{i=1}^{M} (\alpha_i \beta_i)^{1/2}$$

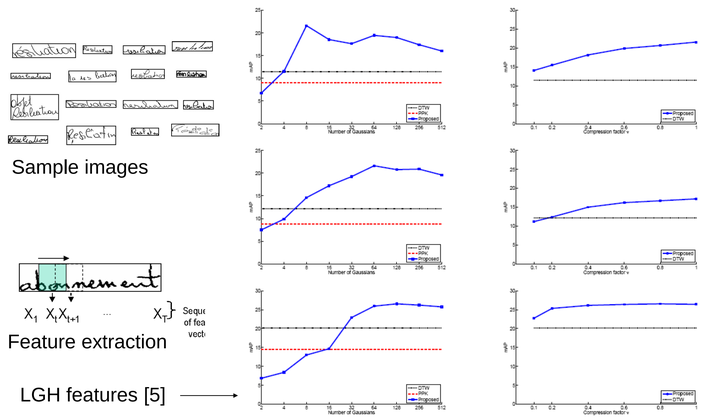
Summary

$X = x_1\ x_2 \ldots x_{Tx}$ ⟶ $W_X = \{w_1^X \ldots w_{N_X}^X\}$ ⟶ DTW
$Y = y_1 y_2 \ldots y_{Ty}$ ⟶ $W_Y = \{w_1^Y \ldots w_{N_Y}^Y\}$

SC-HMM training

5. Experiments: Handwritten CBIR Retrieval accuracy and compression

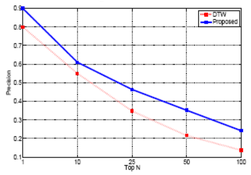
Dataset: Handwritten word images
630 pages, 180K word hypotheses after segmentation
Approx ~10K word images after pruning
3 different feature sets



6. Experiments: Synthesized typed text queries

LGH features [5]

Queries



Retrieval accuracy: querying with typed text samples

Top: DTW retrieval result
Bottom: proposed measure

References

[1] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans on ASSP, 1978.

[2] J. Goldberger, S. Gordon, and H. Greenspan. An efficient image similarity measure based on approximations of KLdivergence between two Gaussian mixtures. In *ICCV*, 2003.

[3] T. Jebara, Y. Song, and K. Thadani. Spectral clustering and embedding with hidden Markov models. In ECML, 2007.

[4] T. Jebara, R. Kondor, and A. Howard. Probability product kernels. JMLR, 2004.

[5] J.A. Rodriguez, F. Peronnin, Local gradient histogram features for word spotting in unconstrained handwritten documents, ICFHR 2008.

IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)