The 16^{th} *International Conference on Frontiers in Handwriting Recognition (ICFHR-2018), Niagara Falls, USA*

CONSTRAINED AND PARAMETRIC DYNAMIC PROGRAMMING FOR WORD IMAGE RETRIEVAL

TANMOY MONDAL, MICKAËL COUSTATY

L3I, UNIVERSITÉ DE LA-ROCHELLE, FRANCE.

INTRODUCTION

- ⇒ Dynamic Time Warping (DTW) based similarity matching has been highly used for word image matching
- ⇒ Constrained DTW performs better than classical DTW, hence can avoid pathological matching
- ⇒ Several variants of DTW can perform better compared to classical DTW

CONTRIBUTION

- ⇒ Demonstrated the combination of these variants of DTW improves results
- ⇒ Propose to combine Pseudo local DTW (LDTW) with Itakura Parallelogram
- ⇒ Several other sequence matching techniques are combined to study their potential advantages
- ⇒ Propose technique to compute weighting parameter to combine two matching techniques

DYNAMIC TIME WARPING (DTW)

- ⇒ Measures similarity between two time series
- ⇒ Two time series $X = x_1, x_2, x_3, ..., x_p$ and $Y = y_1, y_2, y_3, ..., y_q$

$$DTW(X,Y) = \min_{L} \sum_{L} \mathfrak{D}(l_k);$$

$$\mathfrak{P}_{(i,j)} = \mathfrak{D}_{(i,j)} + min \begin{cases} \mathfrak{P}_{(i,j-1)} \\ \mathfrak{P}_{(i-1,j-1)} \\ \mathfrak{P}_{(i-1,j)} \end{cases} \Bigg|_{i=1,2,....,p; j=1,2,....,q}$$

- ⇒ Best warping path is computed by dynamic programming from ($\mathfrak{P}$)
- ⇒ Elements in $\mathfrak{P}$ is $l_k \in L; l_k = (i,j)$ represents best alignment between x_i and y_j; $1 \le i \le p; 1 \le j \le q$
- ⇒ Distance between x_i and y_j is given by $\mathfrak{D}_{l_k} = \mathfrak{D}_{(i,j)} = (x_i - y_j)^2$.
- ⇒ $\mathfrak{P}_{(i,0)} = \mathfrak{P}_{(0,j)} = \infty$
- ⇒ Optimal distance ($\mathfrak{V}$) is stored in the cell $\mathfrak{P}_{(p,q)}$
- ⇒ L : represents all possible warping paths in $\mathfrak{P}$

CONSTRAINTED LDTW

- ⇒ Applied *Pseudo Local DTW*, constrained by *Itakura parallelogram*
- ⇒ Reduces time and space complexity of LDTW and avoid pathological matching
- ⇒ It improves the result to a good extend (3 to 4%)

Algorithm 1: LDTW + ITAKURA PARALLEOGRAM

```
Input: p, q, 𝔇
Output: 𝒟
for i ← 1 to p do
    for j ← 1 to q do
        bool = ITAKURA(i, j, p, q);
        if bool then
            𝔔(i, j) ← LDTW(i, j, 𝔇);
𝒟 = 𝔇(p, q)/|w_k|                ▷ The final distance
return;
procedure ITAKURA(i, j, p, q)
    bool = 0;
    if (j < 2 × i & i ≤ 2 × j) & (i ≥ (p − 1) − (2 × (q − j))
        & j > (q − 1) − (2 × (p − i))) then
        bool = 1;
    return bool
end procedure
```

PRUNNING TECHNIQUES

- ⇒ The target images satisfies Criterion-1 and Criterion-2 or Criterion-3 are considered as relevant

$$\frac{1}{2} \ge \frac{query\ aspect\ ratio}{target\ aspect\ ratio} \ge 2 \quad \frac{1}{2} \ge \frac{query\ area}{target\ area} \ge 2 \quad \frac{1}{2} \ge \frac{\mathscr{K} \text{ of query}}{\mathscr{K} \text{ of target}} \ge 2$$

REFERENCES

1. T. Mondal, N. Ragot, J.-Y. Ramel, and U. Pal, "Performance Evaluation of DTW and its Variants for Word Spotting in Degraded Documents," in ICDAR, Aug. 2015.
2. P. Wang, E. Veronique, L. Christine, L. Josep, and F. Alicia, "A Novel Learning-free Word Spotting Approach Based On Graph Representa- tion," in DAS, 2014, pp. 207–211.
3. T. Gorecki and M. Luczak, "Multivariate time series classification with parametric derivative dynamic time warping," Expert Systems with Applications, vol. 42, no. 5, pp. 2305–2312, 2015.

PARAMETRIC COMBINATION

- ⇒ Parametrically combine different dynamic programming based sequence matching techniques
- ⇒ Introduced for time series matching but adapted here for word image matching

Parametric Distance Measure

- ⇒ Weighting based distance measure ($\mathfrak{D}_{ab}$) can be calculated
- ⇒ By linear combination of $\mathfrak{D}_1$ and $\mathfrak{D}_2$, based on real parameters $a, b \in [0,1]$

$$\mathfrak{D}_{ab}(X,Y) = a \times \mathfrak{D}_1(X,Y) + b \times \mathfrak{D}_2(X,Y)$$

Parameters (a,b) can be chosen between points $(0,1)$ and $(1,0)$

$$a = (1-\alpha); b = \alpha; \alpha \in [0,1]$$

α-optimization Technique

- ⇒ α-value is tuned during the traing phase
- ⇒ Select query image $\mathfrak{Q}_1$ from training dataset and compute all distances between target set (T^{s^1})
- ⇒ For all $\alpha(\{\alpha \in [0,1]\})$, compute $d^1_{\{T^{1 \dots s^1}\},\{1 \dots p\}} = \mathfrak{D}_{ab}(\mathfrak{Q}_1, T^{1 \dots s^1})$ for $\alpha_{1,\dots,p}$ and put in 2D matrix $\mathfrak{D}$ (s^1 rows and p columns)
- ⇒ we sort (descending order) each column of $\mathfrak{D}$, as each column represents the matches for different α values
- ⇒ As we have already the ground truth for this query image $\mathfrak{Q}_1$, we calculate the precision and recall (PR) values at each rank for every columns.
- ⇒ Repeat the procedure for other query images in training set
- ⇒ Same process of PR values calculation is followed for each column of $d^2_{\{T^{1 \dots s^2}\},\{1 \dots p\}}$ matrix
- ⇒ The best α is computed by mean average precision for each column (each $\alpha_{1,\dots,p}$) for all the query images

PSEUDO LOCAL DTW (LDTW)

- ⇒ DTW is modified to perform pseudo-local alignment using specific DP-path
- ⇒ Applies different DP paths at different location of path cost matrix
- ⇒ Helps to handle stretching and compression

EXPERIMENTAL RESULTS

Word Spotting Framework

- ⇒ 8 column based features are chosen and saved offline
- ⇒ Saved features are matched by accessing it from memory

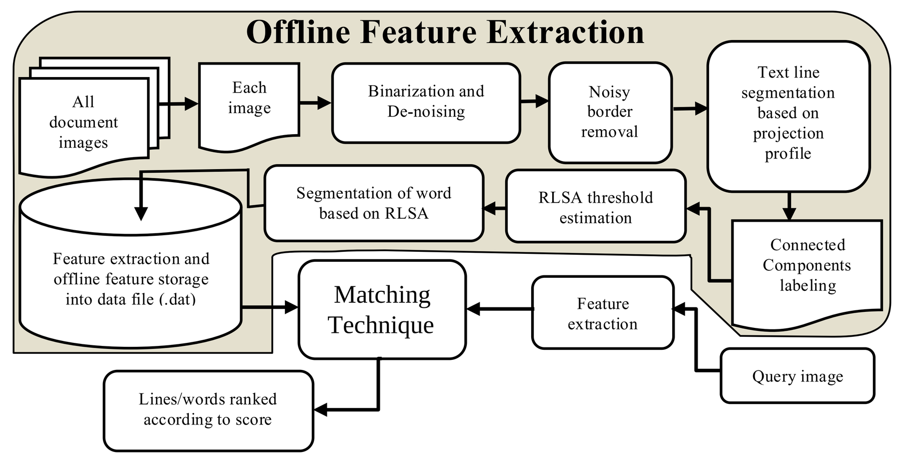

Fig : The block diagram of word image retrieval system

Table IV: Comparative word image matching accuracy on GW dataset.

Technique	mAP
Proposed (combination of **LDTW** and Itakura)	**0.28**
See [2]	0.175
BOVW (see [2])	**0.422**
Pseudo-Struct (see [2])	0.072
Structural (see [2])	0.028

- ⇒ **DTW** = Itakura Paralleogram
- ⇒ **LDTW** = LDTW + Itakura Paralleogram

Table I: mAP of different algorithms, applied on Bentham (1^{st} row) and GW (2^{nd} row) dataset.

DTW	PDTW	CDP	SSDTW	**DTW**	LDTW	**LDTW**
0.45	0.43	0.44	0.33	**0.52**	**0.499**	**0.524**
0.178	0.157	0.171	0.118	**0.229**	**0.229**	**0.276**

- ⇒ Learning and testing steps repeated 10 times
- ⇒ 2 queries are randomy seletected as queries
- ⇒ Optimized α value is obtained
- ⇒ Testing is performed on remaining queries

Table II: mAP of parameterized matching techniques for GW dataset

	DTW	CDP	LDTW	SSDTW	PDTW	**LDTW**	Itakura
DTW		0.214	0.232	0.185	0.181	0.262	0.221
CDP	0.205		0.237	0.162	0.194	**0.279**	0.238
LDTW	0.239	0.244		0.229	0.234	0.277	0.254
SSDTW	0.188	0.168	0.223		0.168	0.264	0.211
PDTW	0.186	0.188	0.227	0.171		0.265	0.214
LDTW	0.273	0.271	0.272	0.270	0.264		0.272
Itekura	0.220	0.236	0.250	0.214	0.217	**0.280**	

Table III: mAP of parameterized matching techniques for Bentham dataset

	DTW	CDP	LDTW	SSDTW	PDTW	**LDTW**	Itakura
DTW		0.468	0.496	0.436	0.433	0.529	0.503
CDP	0.461		0.504	0.412	0.443	**0.540**	0.520
LDTW	0.492	0.495		0.467	0.501	0.515	0.523
SSDTW	0.410	0.410	0.477		0.416	0.508	0.511
PDTW	0.434	0.462	0.489	0.422		0.533	0.506
LDTW	**0.543**	0.519	0.514	0.518	0.552		0.536
Itekura	0.503	0.529	0.522	0.511	0.510	**0.539**	

Acknowledgement

This work is supported by the Region Nouvelle Aquitaine and European Union under the project "Sécurisation et Authentification des Diplômes" in the "programme opérationnel FEDER/FSE 2014-2020" (grant number P2017- BAFE-46).