A Keyword Spotting Approach Using Blurred Shape Model-based Descriptors

A. Fornés, V. Frinken, A. Fischer, J. Almazán, G. Jackson, H. Bunke

Computer Vision Center, UAB, Spain
Institute of Computer Science, UB, Switzerland
Contents

- Introduction
- Blurred Shape Model
- Deformable Blurred Shape Model
- Adaptation to Word Spotting
- Results
- Conclusions
Introduction

- Keyword spotting for historical documents
  - Useful for databases with non-labeled data
    - Query-by-Example
- Two Symbol Recognition methods for hand-drawn symbols
  - Blurred Shape Model
  - Deformable Blurred Shape Model
  - They can cope with variations in handwriting style

- Idea → Adaptation to word spotting?
Blurred Shape Model

The symbol is divided in $n \times n$ cells

- Each cell receives votes
  - The value depends on the distance to the centroid of the neighbouring cells
- Normalize the probability density function

$$d_i = d(x, r_i) = \|x - c_i\|^2$$

$$v(r_i) = v(r_i) + \frac{1}{d_i D_i}, \quad D_i = \sum_{c_k \in N(r_i)} \frac{1}{\|x - c_k\|^2}$$

$$v = \frac{\sum_{i=1}^{n^2} v(i)}{\sum_{j=1}^{n^2} v(j)} \forall i \in [1, ..., n^2]$$

- The number of cells determines the blurred degree allowed
Deformable Blurred Shape Model
Focus representation

Locate over the image $K \times K$ points, denoted as focuses
Their position correspond to the BSM cells centroids
Pixels from the shape will influence nearby focuses
The model used will define:

- Influence area (blue)
- Deformation area (yellow)
Deformable Blurred Shape Model
Image Distortion Model

Defines a $W \times W$ deformation area and a $N \times N$ influence area around every focus
Pixels within the influence area will contribute to the density measure
Focuses are allowed to move inside the deformation area
Influence area will move along with the focus
Two steps for classification: training and matching
The training process consists on, for every image, maximize the value of each one of the $K \times K$ focus
Results a $K^2$ vector $v$ with the final value of each focus and a $2K^2$ vector $p$ with the focuses coordinates
Deformable Blurred Shape Model

Matching

Given a reference image $I$ and a test image $J$, focuses in $J$ are deformed to optimize a matching criterion:

- **IDMmax**: focuses in $J$ will maximize their value

It results in vectors $v_J$ and $p_J$

- **Similarity measure** is the distance between $v_I$ and $v_J$
- **Deformation measure** is the distance between $p_I$ and $p_J$

Final distance between $I$ and $J$ is a weighted combination between similarity and deformation measures
Adaptation to Word Spotting

- Same number of cells?
  - Symbols → Good!
  - Words → Problem!
Adaptation to Word Spotting

- Same number of cells?
  - Symbols → Good!
  - Words → Problem!

- Proposal
  - Template of fix size
  - Word is located in the center
  - Advantages
    - Same number of cells for each character
    - Feature vector short word ≠ long word
    - Center of gravity is robust to noise
Results

- Database
  - George Washington Dataset
  - 20 pages, already preprocessed

Instructions

Instructions.
Results

- Database
  - George Washington Dataset
    - 20 pages, already preprocessed
- Baseline results
  - Rath and Manmatha
    - Features $\rightarrow$ 9 dimensional vector (sliding window)
    - Dynamic Time Warping
Results

- **Database**
  - George Washington Dataset
    - 20 pages, already preprocessed
- **Baseline results**
  - Rath and Manmatha
    - Features → 9 dimensional vector (sliding window)
    - Dynamic Time Warping
- **Metrics**
  - Precision and Recall
  - Average Precision (ap)
    - Average all recall values
  - Mean Average Precision (MaP)
    - Average precision over all queries

\[
\text{Precision} = \frac{TP}{TP + FP} \\
\text{Recall} = \frac{TP}{TP + FN}
\]
Results and Discussion

• Higher resolution $\Rightarrow$ Better performance (on this dataset)
• DBSM $\Rightarrow$ intensity values are more discriminant than positions of focuses
• DBSM obtains slightly better results, but BSM is faster
• BSM and DBSM outperform DTW, with a lower computational cost
Results and Discussion

- The mean average precision is below 60%
- Higher performance for training based systems (e.g. HMMs, RNNs)
- BSM, DTW and DBSM proposals do not require training
  - Suitable for searching in databases without ground-truth
Conclusions

- Conclusions
  - Proposal of a Shape-based keyword spotting (BSM and DBSM)
  - Adaptation to words
  - BSM and DBSM outperform DTW with low complexity

- Future Work
  - Other shape descriptors suitable for handwritten text (e.g. Shape Context)
Thank you !!