Abstract

Lists of Figures

Abbreviation

Acknowledgements

1. Introduction (2 pages)
   1. Our Approach
   2. Outline of forthcoming chapters
2. Background
   1. Characteristics of Arabic script
   2. Hand Writing Recognition
      1. Online vs offline
      2. Writer dependent vs. Writer Independent and Closed dictionary vs. open Dictionary
   3. Online Hand Writing Recognition
   4. Databases
3. Our Approach
   1. General Description of the system (Flow)
   2. Data Acquiring
   3. Feature Extraction
   4. Metrics Distances
      1. DTW
      2. ERP
      3. EMD
   5. Features
      1. Shape Context
      2. Multi Angular Descriptor
   6. Fast Retrieval data structure
      1. Kdtree
      2. LSH
   7. Dimensionality Reduction
      1. PCA
      2. LDA
   8. Learning Models
      1. SVM
      2. WaldBoosting
4. Results
5. Conclusions
   1. Contribution to the field
   2. Future Directions
6. Bibliography

# Abstract

Handwriting recognition is a task of transforming a language represented in its spatial form of graphical marks into its symbolic representation. Online Hand Writing Recognition refers to the situation where the recognition is performed concurrently to the writing process.

After a long period of focus on western and East Asian scripts there is now a general trend in the on-line handwriting recognition community to explore recognition of other scripts such as Arabic and various Indic scripts. One difficulty with the Arabic script is the number and position of diacritic marks associated to Arabic characters.

# Abbreviation

WP – WordPart : a single stroke connected component

DTW – Data Time Warping

EMD – Earth Movers Distance

PCA – Principle Component Analysis

LDA – Linear Discrimination Analysis

SVM – Support Vector Machine

# Introduction

## Online Hand Writing Recognition

There is a vast transition in the field of personal computers from the desktop to handheld devices; thereby single hand entry methods are more suited to this shift than a keyboard. Pen or even figures or hand gestures are more natural and convenient interface.

The large number of writing styles and the variability between them makes the problem of writer-independent unconstrained handwriting recognition a very challenging pattern recognition problem.

Nowadays, the primary mode of data input from a human to a computer is still the keyboard. However, the amount of information processed by computers is rapidly increasing. Given this, the time consumption of information exchange between human and computers is becoming a serious bottleneck. In order to be effective, the user interface has to be both effective and natural. Thereby, requiring no learning curve to the user. [1]

[Talk about smartphones/Tablets/etc.]

While much of the today’s data is directly entered into computers using the keyboard, many tasks still exist in which people tend to prefer handwriting over keyboard entry. Note taking (e.g. in classrooms) is a task that can still be done more efficiently by hand for most users. In addition, while people can produce annotated hand sketches very quickly, data entry into a computer using a combination of the mouse and keyboard is relatively time consuming. [1]

Smartphones and tablets are pocket sized consumer devices that can store calendars and address books, provide access to emails, the web, and contain other productivity tools. These devices are too small to have full sized keyboards, or sometimes may be too small for any keyboard at all, requiring pen, hand gestures, figure gestures or voice interface to enter data. [1]

The problem of handwriting recognition has now been a topic of research for over four decades. There are many types of problems (with varying complexity) within handwriting recognition, based on how the data is presented to the recognition system, at what level the data can be unambiguously broke n into pieces (e.g. individual characters or words), and the transcription complexity of the language used. [1]

At the highest level. Handwriting recognition can be broken into two categories: offline and online. Offline handwriting recognition focuses on documents that have been written on paper at some previous point of time. Information is presented to the system in the form of scanned image of the paper document. In contrast, online handwriting recognition focuses on tasks where recognition needs to be performed at the time of writing. This requires the use of special equipment, such touch screen or digitizing tablet, to capture the strokes of the pen as that are being written. The trace of a writer’s pen is stored as a sequence of points sampled at equally spaced time intervals. The information captured for each sample is the  coordinates. While this sequence can be used to construct a static image of the writing, thus allowing offline character recognition techniques to be applied, it has been shown [63] that the information about the pen dynamics can be used to obtain a better recognition accuracies than the static data alone. Therefore, it is beneficial to capture the data in an online form, even if the real-time processing requirements can be relaxed. [1]

Another advantage of online handwritten data over offline data is the availability of the stroke segmentation and the order of writing. Ink in static images must first be separated from the image background, creating a potential source of error. The ability to detect the states of “pen-down” (when the pen touches the tablet or the finger touches the touch screen) and “pen-up” can also be used. A single stroke is defined as the sequence of sample points occurring between consecutive pen-down and pen-up transitions. However, a complication occurs when a stroke is added to a character in a word after the rest of the word has already been written, such as the cross of a ‘t’ or an ‘x’, or the dot of an ‘i’ or a ‘j’. These types are called delayed strokes. [1]

## Characteristics of Arabic Script

The Arabic Aleph bet is widely used for more than twenty different languages such as Farsi, Urdu, Malay, Housa and Ottoman Turkish. [2]

Arabic Scripts consists of 28 basic letters, 12 additional special letters, and 8 diacritics. Arabic script is written from right to left in a semi-cursive manner in both printed and handwritten. Most letters are written in four different letter shapes depending on their position in a word, e.g., the letter ع (Ain) appears as ع (isolated), عـ(initial), ـعـ (medial) and ـع (final). Among the basic letters, six are Disconnective – ا (Alef), د (Dal), ذ (Thal), ر (Reh), ز (Zain) and و (Waw). Disconnective letters do not connect to the following letter and have only two shapes each. The presence of this letters interrupts the continuity of the graphic form of a word. We denote connected letters in a word, as word-part. If the word-part is composed of only one letter, this letter will be in its isolated shape. [3]

The Arabic script is different from the western scripts in that it combines letters into words. [2]

Certain characteristics relating to the obligatory dots and strokes of the Arabic script distinguish it from Roman script, making the recognition of words in Arabic more difficult than in Roman script. First, Most Arabic letters contain dots in addition to the letter body, such as ش (Sheen) which consists of س (Seen) body and three dots above it. In addition to dots, there are stroke that can attach to a letter body creating new letter such as ك, ط and لا. These dots and strokes are called *delayed strokes* since they are usually drawn last in the in handwritten word-part/word. Second, eliminating, adding or moving a dot or stroke could produce a completely different letter and, as a result, produce a word other than the one that was intended (see Table 1). Third, the number of possible variations of delayed strokes is greater than those in Roman scripts, as shown in Figure 2. There are only three such strokes used for English: the cross in the letter *t*, the slash in *x*, and the dots in *i* and *j*. Finally, in Arabic script a top-down writing style called *vertical ligatures* is very common – letters in a word may be written above their consequent letters. In this style, the position of letters cannot be predefined relative to the baseline of the word. [3]

Saabni and El-sana have explored a large collection of Arabic texts and extracted 300,000 different word combinations of 82,000 different word-parts. Ignoring the additional strokes reduced the number of different word-parts to 40,000. [2]

## Literature Review

Automated recognition of text has been an active subject of research since the early days of computers. A 1972 survey cites nearly 130 works on the subject [4]. Despite the age of the subject, it remains one of the most challenging and exciting areas of research in computer science. In recent years it has grown into a mature discipline, producing a huge body of work.

Despite long standing predictions that handwriting, and even paper itself, would become obsolete in the age of the digital computer, both persist. Whilst the computer has hugely simplified the process of producing printed documents, the convenience of a pen and paper still makes it the natural medium for many important tasks.

A brief survey of students in any lecture theatre will confirm the dominance of handwritten notes over those typing on laptops. However, the ease and convenience of having information in digital form provides a powerful incentive to find a way of quickly converting handwritten text into its digital equivalent.

The field of handwriting recognition can be split into two different approaches. The first of these, on-line, deals with the recognition of handwriting captured by a tablet or similar touch-sensitive device, and uses the digitised trace of the pen to recognise the symbol. In this instance the recogniser will have access to the x and y coordinates as a function of time, and thus has temporal information about how the symbol was formed. The second approach concentrates on the recognition of handwriting in the form of an image, and is termed off-line. In this instance only the completed character or word is available. It is this off-line approach that will be taken in this report.

## Dynamic Time Warping

***Dynamic time warping*** (DTW) is an algorithm for measuring similarity between two time serieses. The method is to find an optimal alignment between two given sequences. Intuitively, the sequences are warped in a non-linear fashion to match each other. DTW solves the discrepancy between intuition and calculated distance by achieving the optimal alignment between sample points of two times sequences. Restrictions can be imposed on the matching technique to improve the matching results and the algorithm’s time complexity. The only restriction placed on the data sequences is that they should be sampled at equidistant points in time which can be easily achieved by re-sampling. A naïve approach to calculate the distance between two time serieses can be to resample one of the serieses and to compute the distance sample by sample, the problem with this technique is that it does not yield intuitive results as it might match samples that do not correspond well. See figure 1.

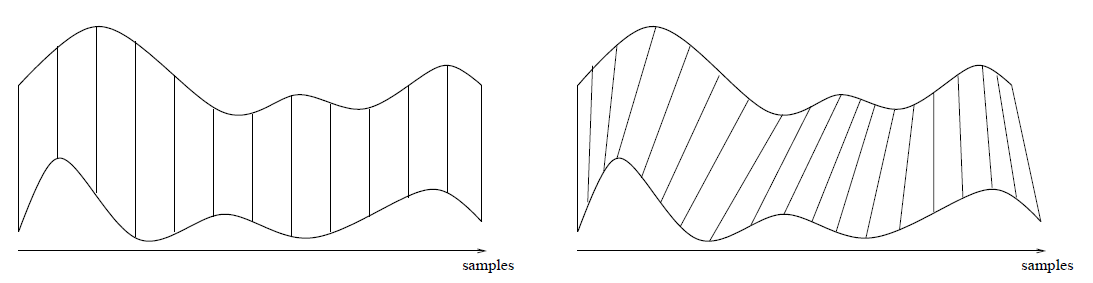


Figure 1 – The right scheme shows sample-by-sample naïve alignment after resampling and in the left scheme the alignment was performed using DTW

Formally, given two time serieses, and , DTW yields an optimal warping path, where  for which satisfies the two following:

1. **Start and End point constrain:** and.
2. **Time preservation of the points:** for each two points in W,,  and the following holds: and .
3. **Local continuity constrain:**

The Distance of the warping path is defined as follows:

|  |  |  |
| --- | --- | --- |
|  |  | Equation 1 |

 is the distance of the two data point indexes (one from X and the other from Y) in the kth element of the warping path.

The warping path which has the minimal  associated with alignment is the optimal warping path which we have previously denoted by.

Using dynamic programing approach, DTW yields the optimal warping path by constructing the **accumulated distance matrix.** We will denote this matrix by the letter, where is the minimum distance warping path that can be constructed from the two time serieses and. Wherefore the value in  will contain the minimum-distance warping path between  and .

The accumulated distance matrix is calculated as follows:

|  |  |  |
| --- | --- | --- |
|  |  | Equation 2 |
|  |  |  |
|  |  | Equation 3 |

Once matrix D is calculated the optimal warping path  is retrieved by backtracking the matrix  from the point  to the point following the greedy strategy of looking for the direction from which the current distance is taken.

It is easy to note that the time and space complexity of DTW is. [5] [6] [7]

### DTW Speedup

One drawback of DTW is its quadratic time and space complexity, thus, many speedup methods have evolved. These methods can be fall in three categories:

1. Constrains: Limiting the amount of calculated cells in the cost matrix.
2. Data Abstraction: Running the DTW on an abstract representation of the data.
3. Lower Bounding: Reducing the times DTW need to run when clustering and classifying time serieses.

Constrains: The most popular two constrains on DTW are the Sakoe-Chuba Band and he Itakura parallelogram, which are shown in figure 2. The grayed out area is the cells of the cost matrix that are filled by the DTW algorithm for each constrain. The warping path is looked for in the constrain window. The width of the window is specified by a parameter. By using such constrains the speedup factor is a constant and the DTW complexity is still (where M and N are the time serieses lengths). Note that if the warping path is does not reside in the constrain window, it will not be found by DTW, thus, such method is usually used when the warping path is expected to be in the constrain window.

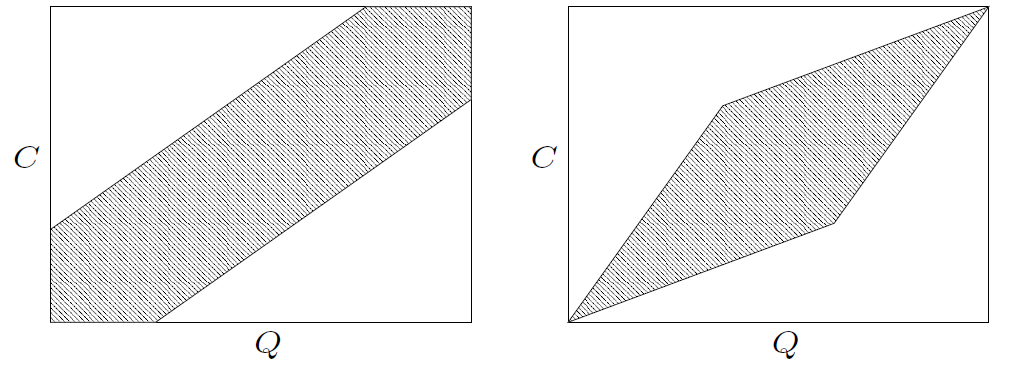


Figure 2 – Cost matrix constrains: Sukoe-Chuba Band (left) and the Itakura Parallelogram (right).

Data Abstraction: Speedup using data abstraction is performed by running DTW on a reduced presentation of the data thus reducing the cell numbers that need to be computed. The warp path is calculated on the reduced resolution matrix and mapped back to the original (full) cost matrix.

Lower bounding: Searching the most similar time series in the database given a template time series can be done more efficiently using lower bound functions than using DTW to compare the template to every series in the database. A lower-bounding is cheap and approximate. However, it underestimates the actual cost determined by DTW. It is used to avoid comparing serieses by DTW when the lower-bounding estimate indicates that the time series is worse match than the current best match.

FastDTW, proposed by Stan Salvador and Philip Chan, approximate DTW in a linear time using multilevel approach that recursively projects a warping path from the coarser resolution to the current resolution and refines it. This approach is an order of magnitude faster than DTW, and also compliments existing indexing methods that speedup time series similarity search and classification. [7]

Ratanamahatana and Kough have shown the following fact about DTW in their work in [8]: Comparing sequences of different lengths and reinterpolating them to equal length produce no statistically significant difference in accuracy or precision/recall.

Although in our work we do make reinterpolating of sequeces to compare equilenth sequences.

## Earth Movers Distance

Earth movers distance (EMD) is a measure of the dissimilarity between two histograms. Descriptively, if the histograms are interpreted as two different ways of piling up a certain amount of dirt, the EMD is the minimal cost of turning one pile to other, where the cost is assumed to be the amount of dirt moved times the distance by which it moved.

EMD has been experimentally verified to capture well the perceptual notion of a difference between images. [9]

[Add a picture]

Computing EMD is based on a solution to the well-known transportation problem. It can be computed as the minimal value of a linear program.

[EMD formal definition]

EMD naturally extends the notion of a distance between single elements to that of a distance between sets, or distributions, of elements. In addition, it’s a true metric if the signatures are equal. [10] A major hurdle to using EMD is its computational complexity (for an N-bin histogram).Various approximation algorithms have been proposed to speed up the computation of EMD.

A histogram can be presented by its signature. The signature of a histogram is a set of clusters where each cluster is represented by its mean (or mode), and by the fraction of the histogram that belongs to that cluster.

### EMD Embedding

The Idea of the Embedding is to compute and concatenate several weighted histograms of decreasing resolutions for a given point set. The embedding of EMD given in [9] provides a way to map weighted point sets A and B from the metric space into the normed space , such that the  distance between the resulting embedded vectors is comparable to the EMD distance between A and B. The motivation of doing so is that the working with normed space is desirable to enable fast approximate Nearest Neighbors (NN) search techniques such as LSH.

[Embedding formal definition]

### Wavelet EMD

A work conducted by Shirdhonkar and Jacobs in [11] has proposed a method for approximating the EMD between two histograms using a new on the weighted wavelet coefficients of the difference histogram. They have proven the ratio of EMD to wavelet EMD is bounded by constants. The wavelet EMD metric can be computed in time complexity.

[Embedding formal definition]

## Features

The selection of valuable features is crucial in pattern recognition.

## Shape context

S. Belongie et al. have defined in [12], a point matching approach named Shape Context. The Shape context is a shape matching approach that intended to be a way of describing shapes that allows for measuring shape similarity and the recovering of point correspondences. This approach is based on the following descriptor: Pick n points on the shape’s contour, for each point  on the shape, consider the  other points and calculate the coarse histogram of the relative coordinates such that is defined to be the shape context of. The bins are normally taken to be uniform log-polar space. This distribution over relative positions is robust and compact, yet highly discriminative descriptor.

To match two points,  from the first shape and from the other shape, define  denote the cost of matching, these two points, use the  test statistic:

.

To perform a one-to-one matching that matches each point  on shape 1 and  on shape 2 that minimizes the total cost of matching: is needed.

This can be calculated in  time using the Hungarian method.

[Add a picture]

## Multi Angular Descriptor

The Multi Angular Descriptor (MAD) is a shape recognition method described in [13], which captures the angular view to multi resolution rings in different heights. Given an Image I of a connected component, calculate the centroid C and the diameter D of the image I. draw a set of rings centered by C with different radius values which are derived from the diameter D. draw k points on each ring taken with uniform distance from each other. Each point in each ring serves as an upper view point watching each contour point in the shape.

## Fast retrieval data structures

Now we will give an overview on 2 methods that tries to answer the second question.

### K dimensional tree

## Support Vector machine

Support Vector Machines were introduced by Vapnik et al in 1992. SVM is widely used in object detection & recognition, content-based image retrieval, text recognition, biometrics, speech recognition, etc. Without Losing generality, we consider the 2 class classification problem. We give a brief introduction for the SVMs basics. SVM paradigm has a nice geometrical interpretation of discriminating one class from the other by a separating hyperplane with maximum margin. SVM can be altered to perform nonlinear classification using what is called the kernel trick by implicitly mapping the inputs into high-dimensional feature spaces. Given a training sample set  where a d-dimensional sample,  is the class label and N is the size of the training set.

Support vector machine first map the data from the sample from the input space to a very high dimensional Hilbet Space.. The mapping  is implemented implicitly by a kernel function  that satisfies Mercer’s conditions. Kernels are functions that give the inner product of each pair of vectors in the feature space, such that. Then in the high dimensional feature space H, we try to find the hyperplane that maximizes the margin between the two classes, and minimizing the number of training error.

The decision function is as follows:

|  |  |
| --- | --- |
|  | Equation 4 |

Where

|  |  |
| --- | --- |
| And | Equation 5 |

Training as SVM is to find  that minimizes the following quadratic cost:

|  |  |
| --- | --- |
|  | Equation 6 |

Where  is a parameter chosen by the user, a large  corresponds to a higher penalty allocated to the training errors. This optimization problem can be solved using quadratic programming. [14]

Many implementation of kernels have been proposed, one popular example is the Gaussian Kernel:

|  |  |
| --- | --- |
|  | Equation 7 |
|  |  |

[Add a picture]

## Gausian dynamic time warping using DTW as a metric

Claus Bahlmann et al. in [15] have proposed combining DTW and SVMs by establishing a new kernel which they have named Gaussian DTW (GDTW) kernel. They have achieved, by incorporating DTW in the Gaussian kernel function, a method to handle online handwriting input which is a variable-sized sequential data. Clearly, the assimilation of the DTW in the Gaussian kernel will result the following kernel function

|  |  |
| --- | --- |
|  | Equation 8 |
|  |  |

We should note that the DTW is not a metric, as the triangle inequality do not hold in many cases, thus the GDTW may miss some important properties as satisfying the Mercer’s condition. This would imply that the optimization problem in 6 may not convergent to the global minimum.

# Our Approach

Recognition of Arabic transcription follows 2 approaches. An analytic approach, wherein the word or word parts are segmented into letters, Recognized and then reassembled into a whole word and a holistic approach where no segmentation is needed, and the words or word-parts are recognized as a whole.

## Arabic Word Parts recognition

* As we have mentioned in the section about the Arabic script. Arabic is a cursive written language.
* Here we use the holistic approach
* A word in Arabic is composed of connected components named Word parts
* We ignore the additional strokes, thus 7 and 5 are the same letter.
* The number of possible word parts is infinite however the number of actual word parts in the language is a finite set which contains about 40k words.
* Our approach is to try to recognize a word part.
* The learning process had several phases:
  1. Letters samples collection
     + The ADAB Database
     + Online Self-maid letter database
  2. Processing – simplification and resampling
  3. A complete WordPart database
  4. Additional Processing on the wordparts
  5. Feature Extraction
     + Sequence
     + Shape Context
     + MAD
  6. EMD Embedding
  7. Dimensionality Reduction
     + PCA
     + LDA
  8. Fast retrieval database
     + Kdtree
     + LSH
* The recognition Process is similar to the learning Process however after we get the processed output, we query the database for the most similar object

## Arabic Progressive Recognition

While the holistic approach is a common approach to recognize handwriting both in online and offline domains, we suggest recognizing the Arabic scribing in a progressive manner in which we try to segment the handwriting and extract the letters from the being written sequence data and recognize the word part being written in an analytic approach. One of the significant benefits of this approach is that, the magnitude of the classes the recognition algorithm is trying to match is much smaller. As was mentioned before, there are about 40,000 valid word parts, with the different main strokes, there are much more if we take into consideration the additional strokes, therefore, the samples space is very large. However, when trying to identify letters, the learning domain is much smaller as there are a small set of letters, thus a better recognition can be performed. We compare several method of learning using the Support Vector Machines (SVMs) with different kernel’s selection.

The process of learning letters in different positions is described below.

1. Segmentation – holistically, segment the given data into sub-sequences that represent letters. This is done by spotting out potential segmentation points.
2. Recognition – Recognizing the given sub-sequences of the Segmentation phase and classify what letter was just completed.

### Segmentation

Segmentation point is a junction point in the written stroke that separates between Arabic letters. Numerous segmentation techniques have been proposed in the literature for Arabic OCR. However correct and efficient segmentation of Arabic text is still considered and a fundamental problem even for offline printed text. Performing such task in an online manner for handwritten text when the segmentation is being done while the word is being written, Is even a more challenging problem.

Over segmentation and under-segmentation are the main challenges such algorithms are facing.

While the word part is being scribed, the system tries to segment the points sequence, by identifying candidate segmentation points. To identify segmentation points, attributes of Arabic segmentation points need to be learned,

# Bibliography

|  |  |
| --- | --- |
| [1] | S. D. Connell, "ONLINE HANDWRITING RECOGNITION USING MULTIPLE PATTERN CLASS MODELS," Michigan, 2000. |
| [2] | R. Saabni and J. El-sana, "Efficient Generation of Comprehensive Database from Online Arabic Script Recognition". |
| [3] | F. Biadsy, R. Saabne and J. El-Sana, "Segmentation-Free Online Arabic Handwriting Recognition," 2010. |
| [4] | L. Harmon, "Automatic recognition of print and script," *Proceedings of the IEEE,* vol. 60, no. 10, pp. 1165 - 1176 , 1972. |
| [5] | P. Senin, "Dynamic Time Warping Algorithm Review," Honolulu, USA, 2008. |
| [6] | T. M. Rath and M. Manmatha, "Word Image Matching Using Dynamic Time Warping". |
| [7] | S. Salvador and P. Chan, "FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space," Melbourne. |
| [8] | C. A. Ratanamahatana and E. Keogh, "Three Myths about Dynamic Time Warping Data Mining". |
| [9] | I. Poitr and N. Thamper, "Fast Image Retrieval via Embeddings," Third International workshop on Statistical and Computional Theories of Vision, Nice, France, 2003. |
| [10] | "The Earth Mover's Distance," The University of EDINBURGH, [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\_COPIES/RUBNER/emd.htm. [Accessed 12 April 2012]. |
| [11] | S. Shirdhonkar and D. W. Jacobs, "Approximate earth mover's distance in linear time," University of Maryland, Maryland, 2008. |
| [12] | S. Belongie, J. Malik and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, p. 509–521, 2002. |
| [13] | R. Saabni and A. Bronstein, "The Multi Angular Descriptor: new binary and gray images descriptor for shape recognition". |
| [14] | J. Sadri, C. Y.Suem and T. D. Bui, "Application of Support Vector Machines for Recognition of Handwritten Arabic/Parsian Digits". |
| [15] | C. Bahlmann, B. Haasdonk and H. Burkhardt, "Online Handwriting Recognition with Support Vector Machines - A Kernel Approach," *Workshop on Frontiers in Handwriting Recognition (IWFHR),* pp. 49-54, 2002. |