Biometrics (CSE 40537/60537)

Lecture 6: Handwritten signatures recognition

Adam Czajka

Biometrics and Machine Learning Group
Warsaw University of Technology, Poland

Fall 2014
University of Notre Dame, IN, USA

Lecture 6: Handwritten signatures recognition

- What is a biometric handwritten signature?
- Signature data capture
- Signature data pre-processing
- Feature extraction
- Comparison of signatures

Lecture 6: Handwritten signatures recognition

- What is a biometric handwritten signature?
- Signature data capture
- Signature data pre-processing
- Feature extraction
- Comparison of signatures

“Signature” by the American Heritage Dictionary

1. “One’s name as written by oneself”, or
2. “The act of signing one’s name”, or
3. “A distinctive mark, characteristic, or sound indicating identity”

Off-line vs. on-line signatures

1. Off-line signatures
   - image on the paper, equivalent to a set of points on a plane:

$$S = \{(x, y) : \; x = x(t), \; y = y(t), \; t \in T\}$$

   - order of signature points is not registered

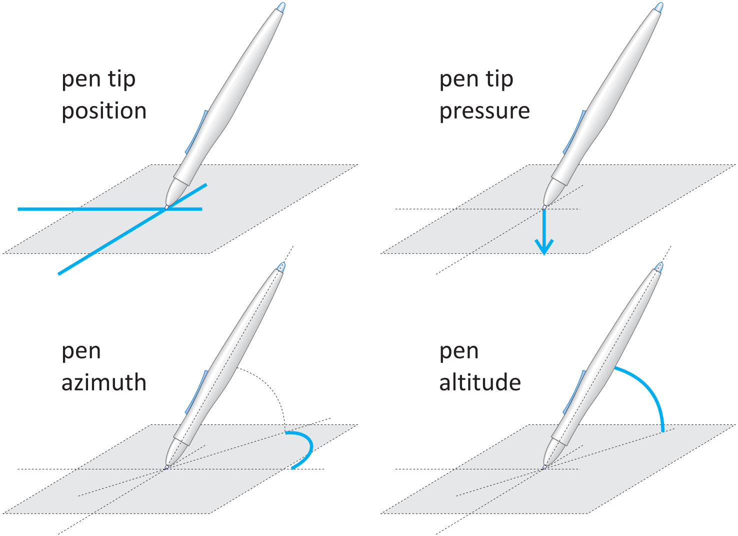
Off-line vs. on-line signatures

2. On-line signatures
   - recording of the signing process
   - observation of the complex, dynamic system output during time T of making the signature

$$s = s(t),\ t \in T$$

   - various signature *components* can be used (i.e. the dimensionality of s varies), depending on the hardware capabilities; typical combinations:
     - 2D: position
     - 3D: position + pen tip pressure
     - 5D: position + pen tip pressure + pen orientation angles

Typical signature components



pen tip position

pen tip pressure

pen azimuth

pen altitude

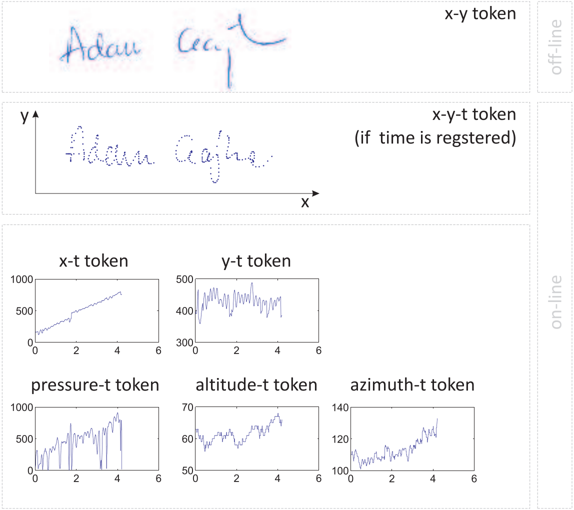
Signature vs. signature token

1. Signature: an abstract class of all possible signature tokens of a single person
    - examples of different signatures: initials, first and last name, just last name
2. Signature token: a particular instance of the signature
    - example: a curve in $\mathbb{R}^6$ when all five signature components are used (pen tip position, pen tip pressure and pen angles)

Signature vs. signature token

3. Signature token graph: any projection of the signature token onto a subspace formed by selected signature components
    - example: x-y-t token – projection of the token onto the position and time subspace
    - example: p-t token – a graph presenting the pen tip pressure as a function of time
    - example: x-y token – projection of the token onto a position plane (e.g. the paper); no time-related information in x-y token

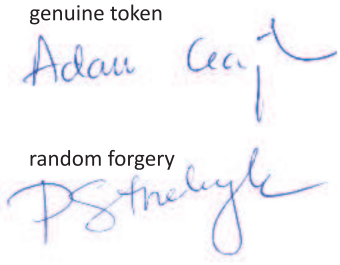
Signature tokens



Signature forgeries

A – a genuine person, B – an attacker

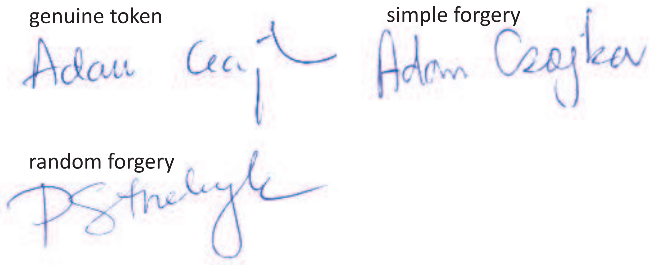
1. Random forgeries
    - B has no information about a signature of A
    - in practice: a different signature used in a forgery



Signature forgeries

A – a genuine person, B – an attacker

2. Simple forgeries
   - B knows elements of the signature of A (e.g. first name, or last name, or both)

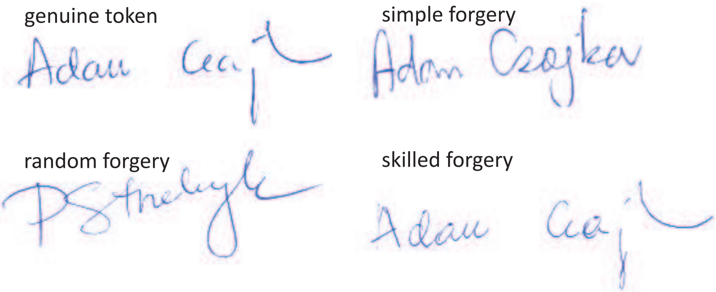


Signature forgeries

A – a genuine person, B – an attacker

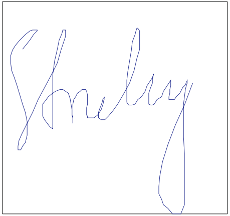
3. Skilled forgeries

- B has an access to signature tokens made by A for training
- B can observe signature tokens of A when making a forgery
- B can observe how A makes the signature
- B can copy of A's signature token (carbon copy)

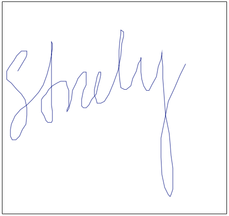


Signature forgeries

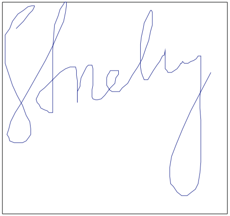
What we can see in off-line and on-line tokens?



token 1



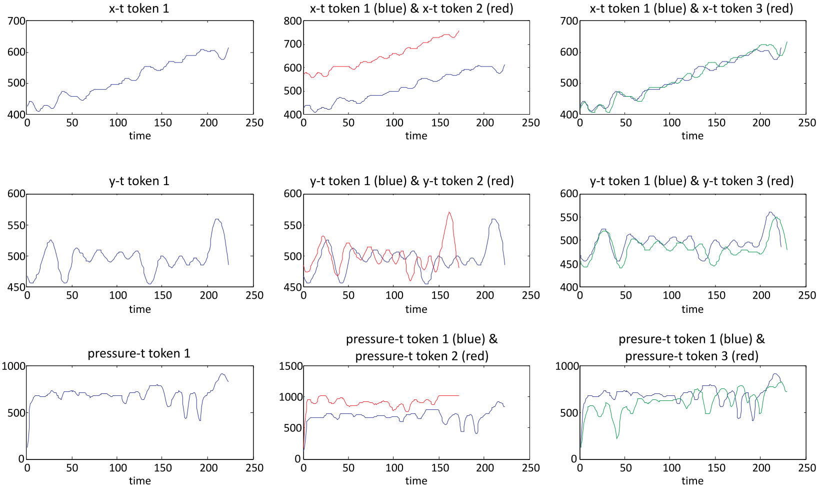
token 2



token 3

Signature forgeries

What we can see in off-line and on-line tokens?



Lecture 6: Handwritten signatures recognition

- What is a biometric handwritten signature?
- Signature data capture
- Signature data pre-processing
- Feature extraction
- Comparison of signatures

Off-line signatures

Scanning

300 dpi

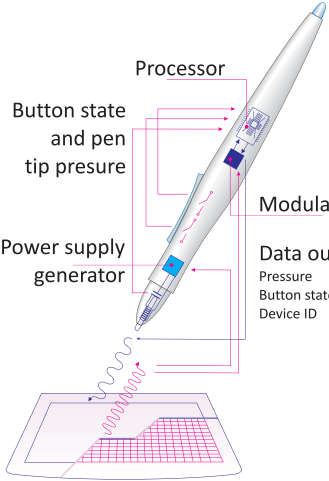
600 dpi

1200 dpi

Scanning resolution impacts the recognition reliability.

On-line signatures

Use of digitizing tablets



Processor

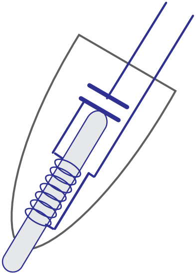
Button state
and pen
tip presure

Modulator

Power supply
generator

Data output

Pressure
Button state
Device ID



Inductor-capacitor
circuit (LC circuit)

based on WACOM documentations

On-line signatures

1. Use of general-purpose graphical tablets
    - signature points registered in order as they appear (typical sampling rate: 100-200 Hz)
    - allow to register other components apart from position (e.g. 2048 levels of the pen tip pressure)
    - can measure the pen tip position when pen does not touch the tablet (even up to 2 cm above the tablet surface)

On-line signatures

2. Use of specialized biometric tablets
   - implement additional elements important in biometric applications (e.g. embedded LCD panels)
   - secure data transmission (cryptography)
   - size adequate to biometric applications (signature does not require large tablets)
   - good developer tools available (e.g. Software Development Kits for popular programming languages)

WACOM

Motion Touch

Topaz Systems

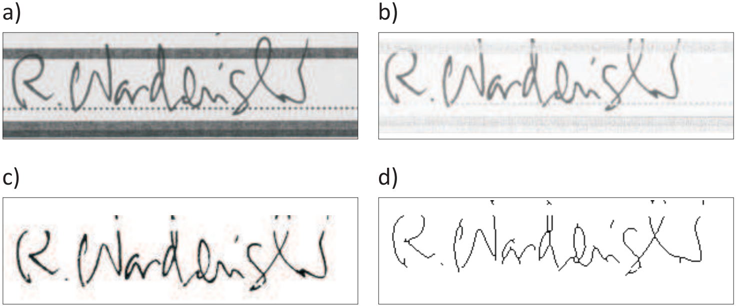
Lecture 6: Handwritten signatures recognition

- What is a biometric handwritten signature?
- Signature data capture
- Signature data pre-processing
- Feature extraction
- Comparison of signatures

Off-line signatures

Segmentation

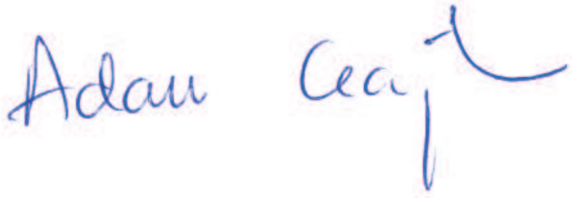
1. Signature extraction
    - detection and localization (a)
    - noise and background removal (b)
    - depending on the matching algorithm: generation of the binary (c) and skeletal (d) images



Off-line signatures

Segmentation

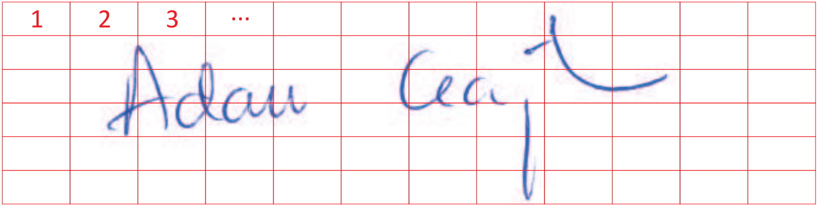
2. Division into segments
    - homogenous: use of a fixed mesh, e.g. rectangular or radial
        - simple but insensitive to signature deformations
        - requires careful alignment of signatures prior processing



Off-line signatures

Segmentation

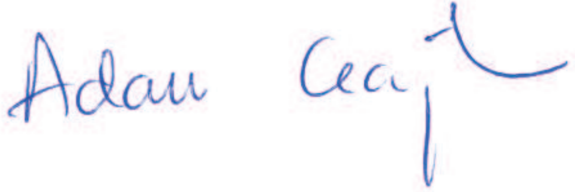
2. Division into segments
   - homogenous: use of a fixed mesh, e.g. rectangular or radial
     - simple but insensitive to signature deformations
     - requires careful alignment of signatures prior processing



Off-line signatures

Segmentation

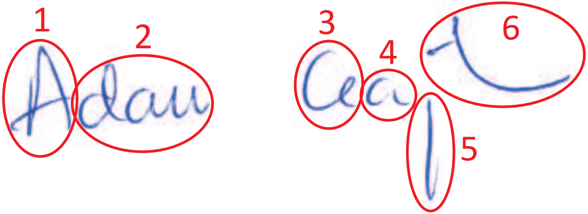
2. Division into segments
    - heterogeneous: finding coherent signature elements (in a sense of selected quality metric), e.g. separate strokes, segments divided by moments when pen does not touch the paper, etc.



Off-line signatures

Segmentation

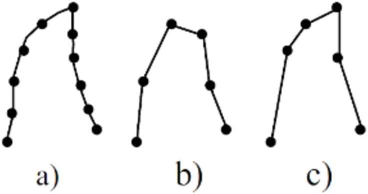
2. Division into segments
    - heterogeneous: finding coherent signature elements (in a sense of selected quality metric), e.g. separate strokes, segments divided by moments when pen does not touch the paper, etc.



On-line signatures

Segmentation

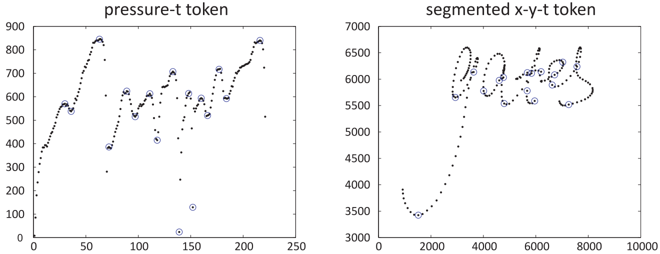
1. Re-sampling
   - homogeneous: skipping every k-th point (b)
   - heterogeneous: skipping points with low curvature index (c)



On-line signatures

Segmentation

2. Division into segments
    - use of static information: pen-up moments, extreme points of single components (popular for the pen tip pressure)
    - use of dynamic information: extreme points of the velocity or acceleration of the signature components



Example on-line signature segmentation using pen pressure information.
Sample originates from SVC 2004 dataset.

Off-line signatures

1. Global features
    - geometrical properties of the signature silhouette
    - number of loops
    - number of intersections
    - number of segments
    - average line thickness
      (should correspond to the pen tip pressure)
    - properties of morphologically altered signature (e.g. multiple dilation with different structural elements): area, direction

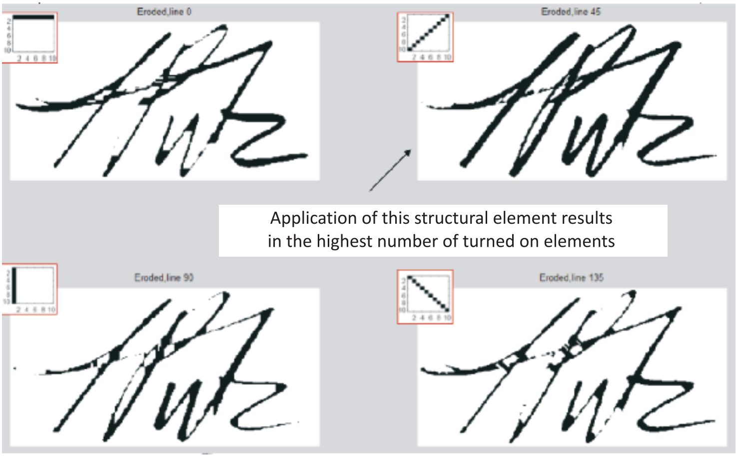
Off-line signatures

Example: multiple dilation with different structural elements

Structural element
used in this multiple dilation

Off-line signatures

Example: multiple erosion to assess the signature direction



Application of this structural element results in the highest number of turned on elements

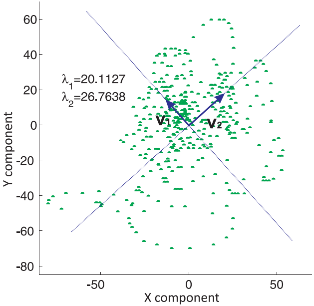
Off-line signatures

2. Local features
    - selected global features (presented in point 1) but calculated within each signature segment
    - local texture descriptors, e.g. LBP, Gabor-based features

On-line signatures

Global features [Pacut, Czajka, 2001]

1. Statistical moments



- For single components: mean (first raw moment), variance (second central moment), skewness (third central moment), kurtosis (fourth central moment)
- Mixed central moments, e.g. eigenvalues and eigenvectors of the covariance matrix

On-line signatures

Global features [Pacut, Czajka, 2001]

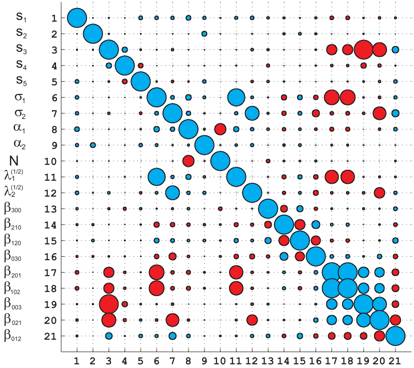
2. Length of the signature (number of points): one of the most important features (!)
3. Linear trends of single components
4. Velocity and acceleration of the pen
5. Wavelet features and zero-crossing points after filtering of the signature components (building a 'signature code')

On-line signatures

Selection of global features

1. Elimination of highly correlated features
2. Selection of most individual features (e.g. use of Fisher ratio)

On the right: s – mean, σ – standard deviation, α – linear trend coefficient, β – third central moments, λ – eigenvalues of the x-y covariance matrix, N – length of the signature



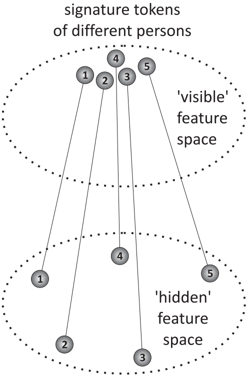
Lecture 6: Handwritten signatures recognition

Signature clustering in the space of global features

1. 'Visible' features
   - easy to be forged by observing off-line signatures (i.e. x-y tokens on the paper)
   - examples: mean and standard deviations of the position components (x,y)
2. 'Hidden' features
   - hard of impossible to be forged when only the off-line signatures can be observed
   - examples: velocity and acceleration of the pen (calculated for x, y or pressure), mean pen altitude, mean pen azimuth

Signature clustering in the space of global features

3. Clustering in 'visible' feature space
    - allows for grouping 'similar' tokens (in a sense of the image on a paper)
4. Classification in 'hidden' feature space
    - allows for building classifiers that are robust for skilled forgeries



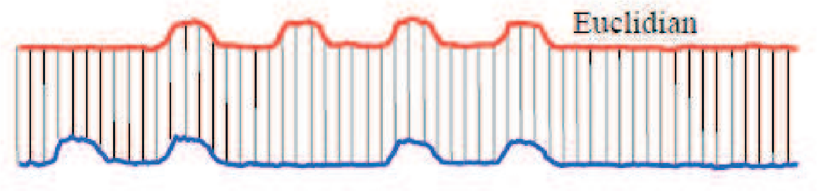
Classification of signatures

Popular approaches

1. Neural classifiers
    - multilayer perceptron
    - radial basis neural networks
    - decision based on the winning neuron ('winner takes all')
    - modification: use of the second-winning neuron output
2. Support Vecotr Machine (SVM)
3. Classification variants
    - single classifier for each signature (appropriate approach in verification tasks)
    - common classifier for clustered signatures (e.g. clustering based on 'visible features')
    - common classifier for all signatures (appropriate approach in identification tasks)

Dynamic Time Warping

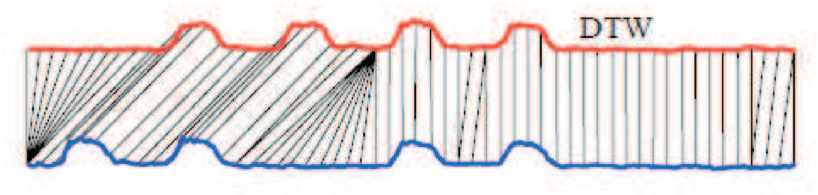
1. Linear (direct) assignment of two components



Source: http://www.markcorbyn.com/work/dissertation/chapter_2.php

Dynamic Time Warping

2. Nonlinear assignment of two components, with time warping



Source: http://www.markcorbyn.com/work/dissertation/chapter_2.php

Dynamic Time Warping

Continuous time case

1. DTW is used to compare functions r and g:

$$r = \{r(t), t \in \langle 1, T_r \rangle\} \quad \rightarrow \quad r(w^t(\ell)), \ell \in \langle 1, L_{\boldsymbol{w}} \rangle$$

$$g = \{g(\tau), \tau \in \langle 1, T_g \rangle\} \quad \rightarrow \quad g(w^\tau(\ell)), \ell \in \langle 1, L_{\boldsymbol{w}} \rangle$$

where $w^t, w^\tau \in \mathcal{W}$ are warping functions, $\mathcal{W}$ is the family of warping functions and $L_{\boldsymbol{w}} = \max(T_r, T_g)$

Dynamic Time Warping

Continuous time case

1. DTW is used to compare functions r and g:

$$r = \{r(t), t \in \langle 1, T_r \rangle\} \quad \rightarrow \quad r(w^t(\ell)), \ell \in \langle 1, L_{\boldsymbol{w}} \rangle$$

$$g = \{g(\tau), \tau \in \langle 1, T_g \rangle\} \quad \rightarrow \quad g(w^\tau(\ell)), \ell \in \langle 1, L_{\boldsymbol{w}} \rangle$$

   where $w^t, w^\tau \in \mathcal{W}$ are warping functions, $\mathcal{W}$ is the family of warping functions and $L_{\boldsymbol{w}} = \max(T_r, T_g)$

2. In general, r and g are curves in multidimensional space, hence they may represent single or multiple signature components

Dynamic Time Warping

Continuous time case

3. Example similarity metrics

$$\inf_{w^t, w^\tau \in \mathcal{W}} \int_1^{L_w} \left\| r(w^t(\ell)) - g(w^\tau(\ell)) \right\|^2 d\ell$$

Dynamic Time Warping

Continuous time case

3. Example similarity metrics

$$\inf_{w^t, w^\tau \in \mathcal{W}} \int_1^{L_w} \left\| r(w^t(\ell)) - g(w^\tau(\ell)) \right\|^2 d\ell$$

4. We often assume (for convenience) that time t in the reference template is identical with 'ideal time' ℓ, that is:

$$\inf_{w \in \mathcal{W}} \int_1^{T_r} \left\| r(t) - g(w(t)) \right\|^2 dt$$

hence time of r is not warped, and time of g is warped by w

Dynamic Time Warping

Continuous time case

3. Example similarity metrics

$$\inf_{w^t, w^\tau \in \mathcal{W}} \int_1^{L_w} \left\| r(w^t(\ell)) - g(w^\tau(\ell)) \right\|^2 d\ell$$

4. We often assume (for convenience) that time t in the reference template is identical with 'ideal time' ℓ, that is:

$$\inf_{w \in \mathcal{W}} \int_1^{T_r} \left\| r(t) - g(w(t)) \right\|^2 dt$$

   hence time of r is not warped, and time of g is warped by w

5. Time should not be warped 'too much', that is:

$$w(t) \approx t, \quad \dot{w}(t) \approx 1$$

Dynamic Time Warping

Discrete time case

1. Discrete domain and values of the functions being compared

$$r = \{r(t), \quad t = 1, 2, .., M_r\}$$

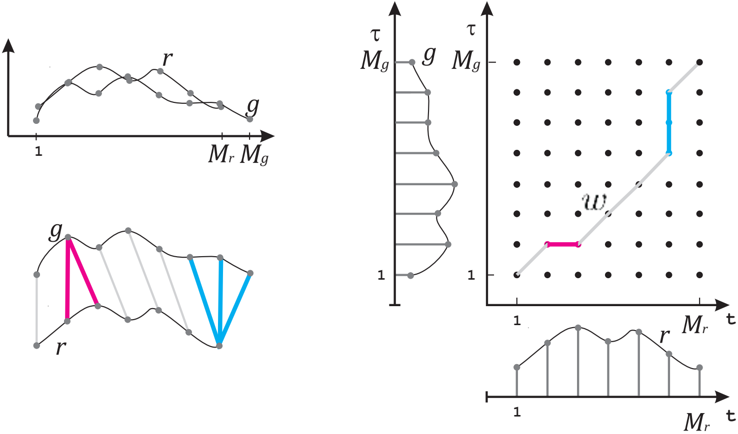
$$g = \{g(\tau), \quad \tau = 1, 2, .., M_g\}$$

2. Warping functions represented by warping paths

$$\boldsymbol{w}(\ell) = [w^t(\ell) \;\; w^\tau(\ell)]^T; \;\; \ell = 1, .., L_w$$
$$\text{where } w^t(\ell) \in \{1, .., M_r\}$$
$$w^\tau(\ell) \in \{1, .., M_g\}$$

Dynamic Time Warping

Discrete time case



Source: Joanna Putz-Leszczynska, Signature Verification – A Comprehensive Study Of The Hidden, International Journal of Applied Mathematics in Computer Science (to be published in 2014)

Dynamic Time Warping

Discrete time case

2a. Diagonal jumps in the warping path correspond to identical time flow in both signature tokens

Dynamic Time Warping

Discrete time case

2a. Diagonal jumps in the warping path correspond to identical time flow in both signature tokens

2b. Vertical or horizontal jumps in the warping path correspond to time constriction or dilation in one of the signature tokens

Dynamic Time Warping

Discrete time case

2a. Diagonal jumps in the warping path correspond to identical time flow in both signature tokens

2b. Vertical or horizontal jumps in the warping path correspond to time constriction or dilation in one of the signature tokens

2c. Warping paths meet the following conditions:

  - boundary: $w^t(1) = w^\tau(1) = 0,\ w^t(L_w) = M_r,\ w^\tau(L_w) = M_g$
    i.e. warping paths connect (1,1) with (M_r, M_g)

Dynamic Time Warping

Discrete time case

2a. Diagonal jumps in the warping path correspond to identical time flow in both signature tokens

2b. Vertical or horizontal jumps in the warping path correspond to time constriction or dilation in one of the signature tokens

2c. Warping paths meet the following conditions:

- boundary: $w^t(1) = w^\tau(1) = 0,\ w^t(L_w) = M_r,\ w^\tau(L_w) = M_g$
  i.e. warping paths connect (1,1) with (M_r, M_g)
- monotonicity: $w^t(\ell+1) \geqslant w^t(\ell),\ w^\tau(\ell+1) \geqslant w^\tau(\ell)$
  i.e. we do not like to 'go back' in time

Dynamic Time Warping

Discrete time case

2a. Diagonal jumps in the warping path correspond to identical time flow in both signature tokens

2b. Vertical or horizontal jumps in the warping path correspond to time constriction or dilation in one of the signature tokens

2c. Warping paths meet the following conditions:

- boundary: $w^t(1) = w^\tau(1) = 0,\ w^t(L_w) = M_r,\ w^\tau(L_w) = M_g$
  i.e. warping paths connect (1,1) with (M_r, M_g)
- monotonicity: $w^t(\ell+1) \geqslant w^t(\ell),\ w^\tau(\ell+1) \geqslant w^\tau(\ell)$
  i.e. we do not like to 'go back' in time
- warping path length: $\max(M_r, M_g) \leqslant L_w \leqslant M_r + M_g - 1$

Dynamic Time Warping

Discrete time case

3. Local distance between points:

$$d\big(r(t), g(\tau)\big) = \|r(t) - g(\tau)\|^2$$

4. Distance along the entire warping path $\boldsymbol{w}$:

$$\mathcal{D}(r, g, \boldsymbol{w}) = \sum_{\ell=1}^{L_w} d\big(r(w^t(\ell)), g(w^\tau(\ell))\big)$$

Dynamic Time Warping

Discrete time case

5. Optimal warping path

$$\hat{\boldsymbol{w}}(r,g) = \arg\min_{\boldsymbol{w}\in\mathcal{W}_{r,g}} \mathcal{D}(r,g,\boldsymbol{w})$$

(provides the solution, i.e. how to warp the time to make two signature tokens as similar as possible, in a sense of $\mathcal{D}(r,g,\boldsymbol{w})$)

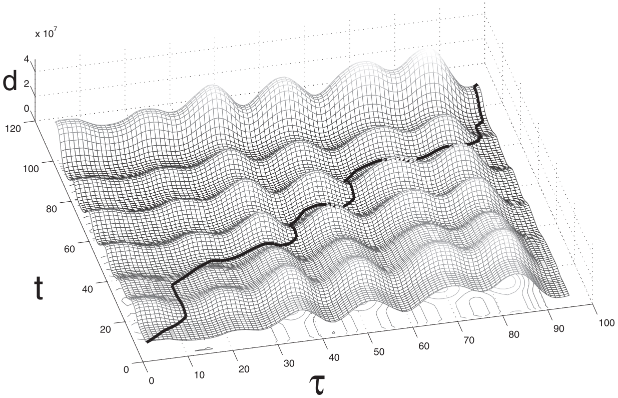
6. Minimum distance

$$\hat{\mathcal{D}}(r,g) = \mathcal{D}(r,g,\hat{\boldsymbol{w}}(r,g))$$

(provides the information about the quality of the optimal solution, i.e. what was the warping intensity; we use this distance in signature comparison)

Dynamic Time Warping

Discrete time case



Source: Joanna Putz-Leszczynska, Signature Verification – A Comprehensive Study Of The Hidden, International Journal of Applied Mathematics in Computer Science (to be published in 2014)

Dynamic Time Warping

Discrete time case

7. Use of dynamic programming to calculate optimal warping path and minimum global distance
   - calculation of the global cost matrix $\mathbf{D}$
     a) first row: $\mathbf{D}(1,j) = \sum_{k=1}^{j} d(r(1), g(k))$; $j = \{1, .., M_r\}$
     b) first column: $\mathbf{D}(i,1) = \sum_{k=1}^{i} d(r(k), g(1))$; $i = \{1, .., M_r\}$
     c) remaining elements:

     $$\mathbf{D}(i,j) = \min\{\mathbf{D}(i-1;j-1), \mathbf{D}(i-1;j), \mathbf{D}(i;j-1)\} + d(r(i), g(j))$$

     where $i = \{1, .., M_r\}$, $j = \{1, .., M_g\}$
   - the last element is the minimum distance, that is:

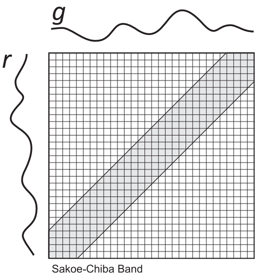
     $$\hat{\mathcal{D}}(r,g) = \mathbf{D}(M_r, M_g)$$

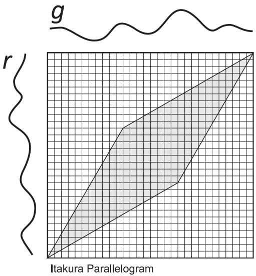
   - we go backwards, starting from (M_r, M_g), to find the optimal warping path $\hat{\boldsymbol{w}}(r,g)$

Dynamic Time Warping

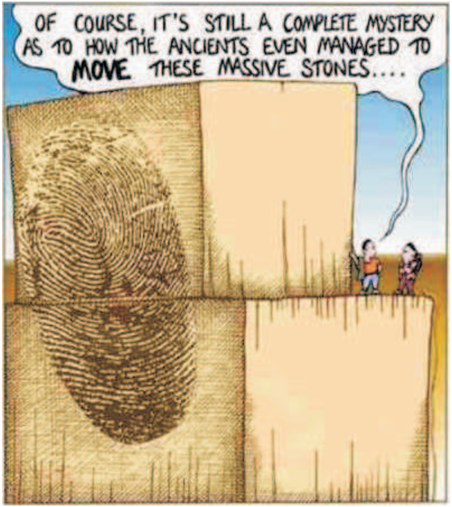
Discrete time case

8. Distance between signature tokens, i.e. the matching score
   - minimum global distance $\hat{\mathcal{D}}(r, g) = \mathbf{D}(M_r, M_g)$, and/or
   - distance between signature tokens after applying optimal warping path
9. Example boundaries speeding up the calculations





Thank you


OF COURSE, IT'S STILL A COMPLETE MYSTERY AS TO HOW THE ANCIENTS EVEN MANAGED TO MOVE THESE MASSIVE STONES....