Shape: Representation & Recognition

Diedrich Wolter

Shape

- Motivation
- What is Shape?
- Applications
- Recognition & Representation
  - Recognition based on Similarity
  - Similarity of Shapes
  - Review of existing Approaches

Motivation

Shape is probably the most important property that is perceived about objects. It allows to predict more facts about an object than other features, e.g., color (Palmer, 1999).

Thus, recognizing shape is crucial for object recognition. In some applications, it may be the only feature present, e.g., in Logo recognition.

Shape is not only perceived by visual means: tactical sensors can also provide shape information that are processed in a similar way. Furthermore, robots' range sensor provide shape information, too.

Definition of Shape

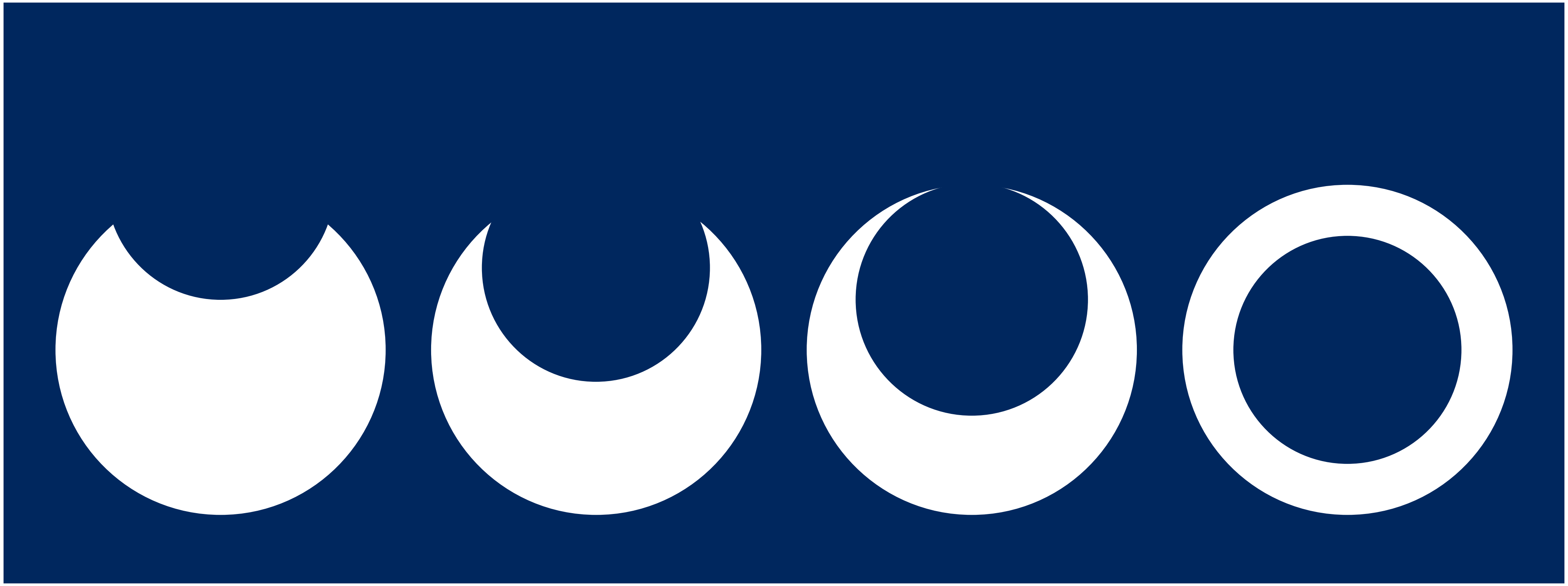
Definition of shape surprisingly complicated. We start with some properties easier to agree on:

- Shape describes a spatial region
- Typically addresses 2D space

Moving on from the naive understanding, some questions arise:

- Is there a maximum size for a shape to be a shape?
- Can a shape have holes?
- Does shape always describe a connected region?
- How to deal with/represent partial shapes?

Shape or Not?

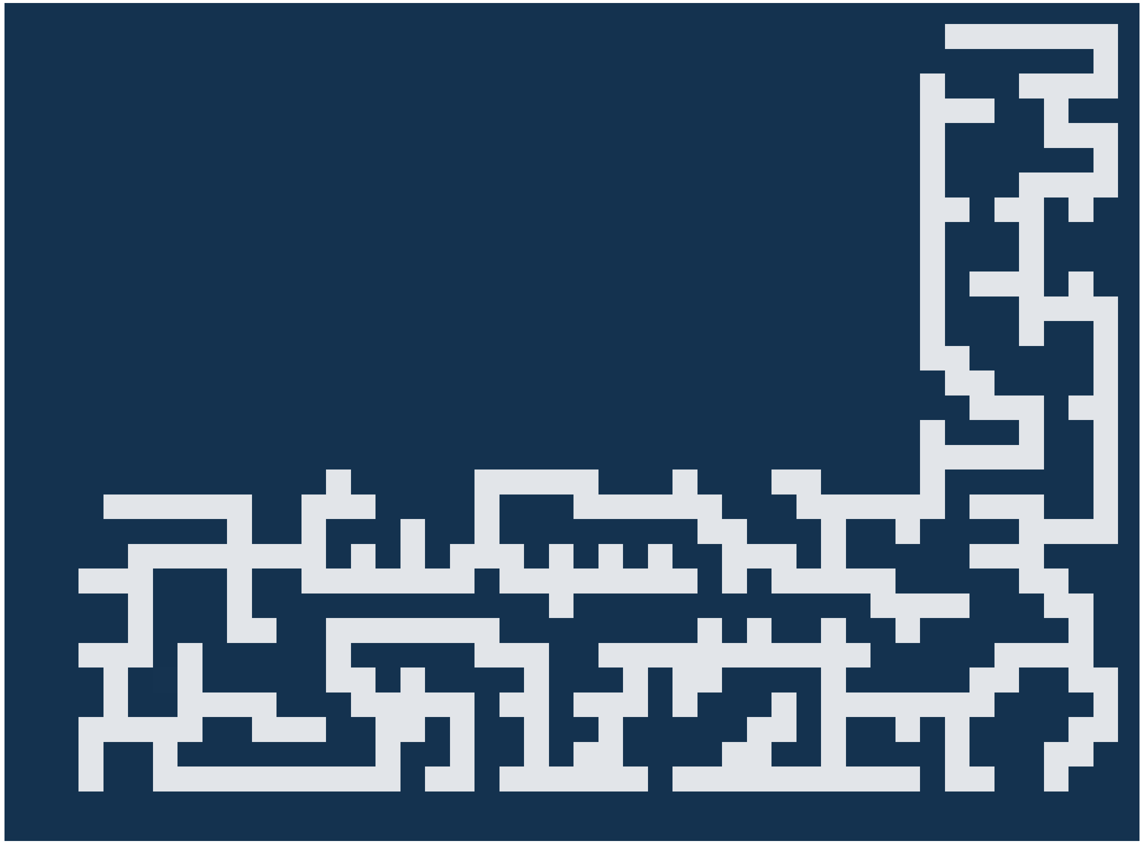


Continuous transformation from shape to no shape: Is there a point when it stops being shape?

Shape or Not?

Continuous transformation from shape to two shapes: Is there a point when it stops being a single shape?

Shape or Not?



This, however, is a shape: one single, connected region.

Applications

Several application areas use shape processing:

- Object recognition
- Image retrieval
- Computer vision
- Image understanding
- Processing of pictorial information
- Video compression (cp. MPEG-7)

This presentation focuses on object recognition and image retrieval.

Recognition & Retrieval

Typically, both applications employ a similarity measure to determine a plausibility that two shapes correspond to each other. However, this similarity seems to differ at a first glance:

Recognition:

- Needs to respect properties of imperfect perception: noise, rotation, shearing, . . .

Retrieval:

- Needs to respect cognitive aspects of similarity: when are two shapes considered similar?

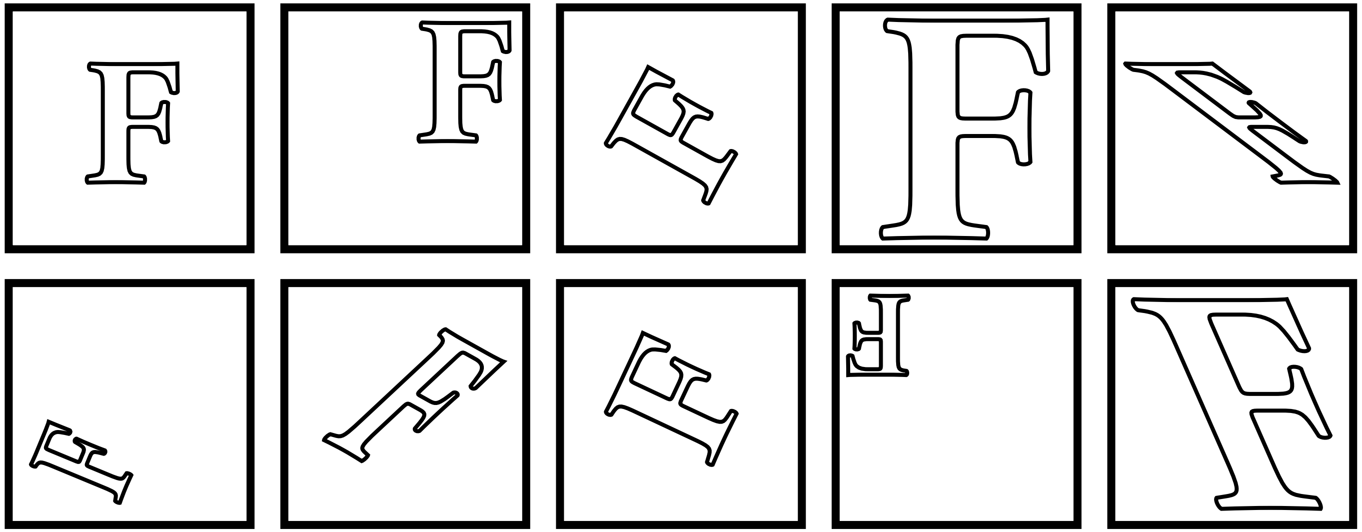
Luckily, a similarity measure useful for retrieval subsumes recognition tasks.

Similarity

Similarity measures are applied to model the likeliness that two shapes correspond. Hereby, various aspects need to be considered.

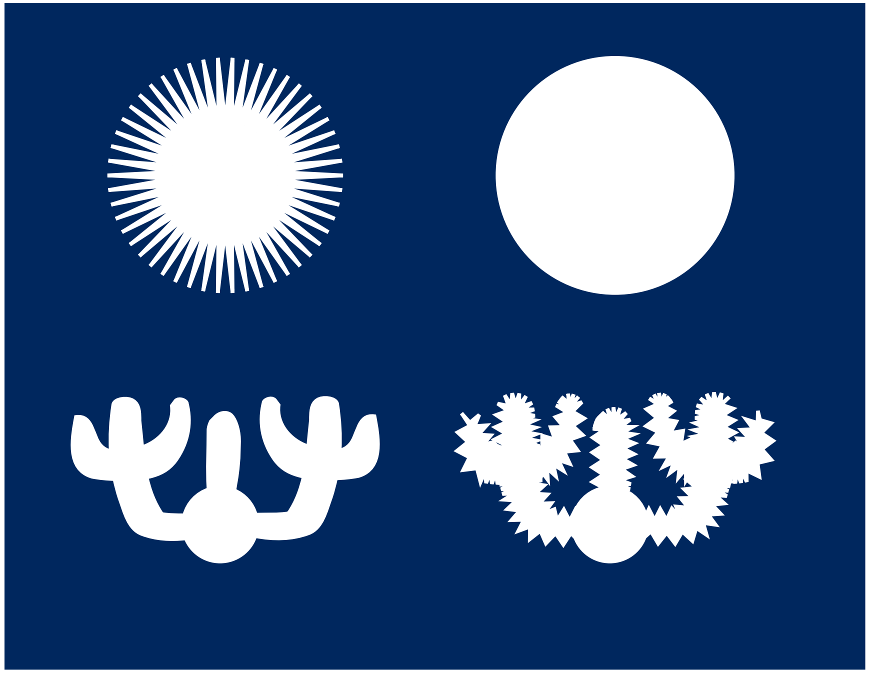
- Noise
- Invariances
- Cognitive adequate similarity of shapes
- Visual parts
- Granularity

Similarity



- Translation invariance
- Scale invariance
- Rotation invariance—but what about □ vs. ◇?
- Reflection invariance?

Similarity II



Some other aspects—I believe to be—worth consideration:

- Similarity of structure
- Similarity of area

Can all these aspects be expressed by a single number?

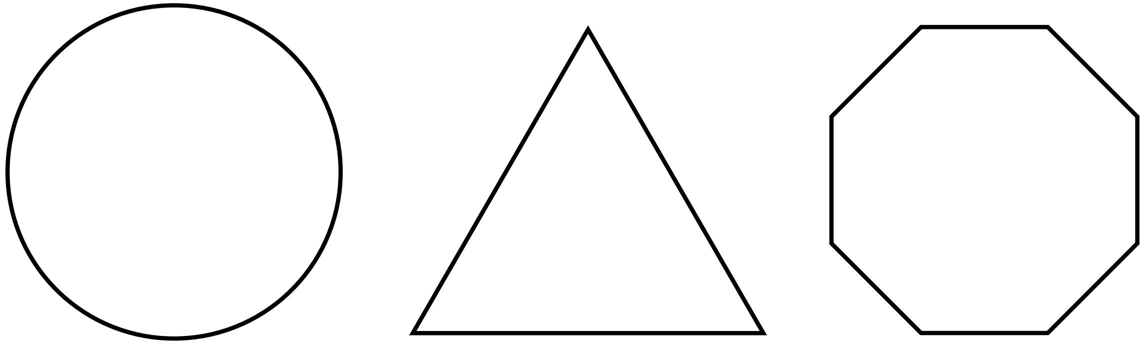
Similarity IV

A similarity measure for shape retrieval in (larger) databases should be a metric (Basri et al., 1998).

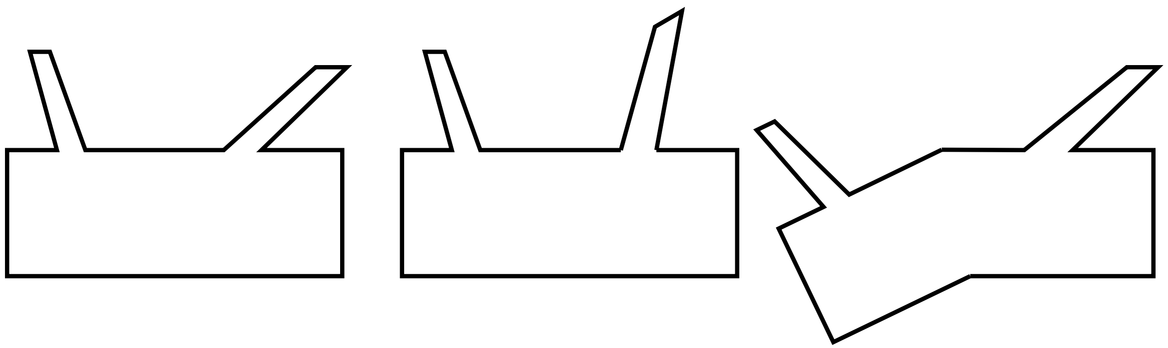
1. $S : SHAPE \times SHAPE \to \mathbb{R}_0^+$
2. $S(A, B) = S(B, A)$
3. $S(A, B) \leq S(A, C) + S(C, B)$
4. S is continuos
5. One major difference should cause a greater dissimilarity than some minor ones.
6. S must not diverge for curves that are not smooth (e.g., polygons).

However, these demands are contradictory (proof is left as an exercise).

Similarity: Example



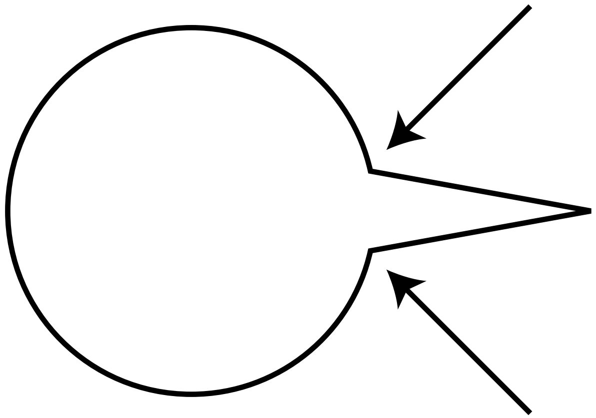
Though all figures have a turning angle of 2π when traversing the contour, the octagon is more similar to the circle than the triangle.

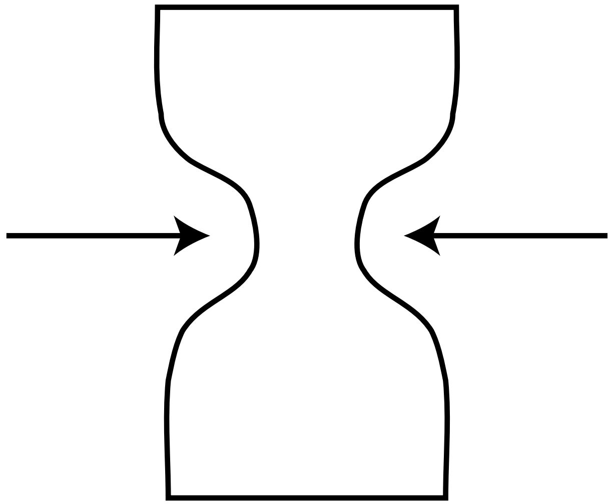


The same amount of bending at places of high curvature does not influence similarity as much as when the curvature is low.

Visual Parts

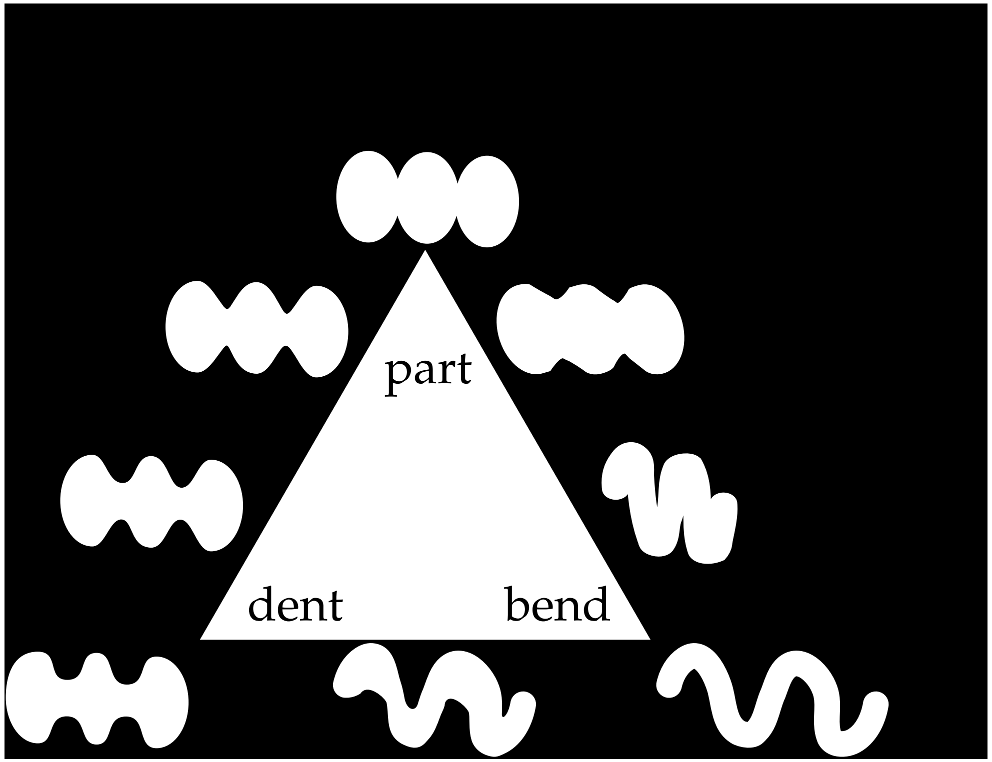
Visual parts play an important role in recognition: from seeing just a hand, a human can be recognized. Either visual parts must be considered in the shape representation explicitly or in the similarity measure implicitly. Visual part are often computationally defined by corresponding points of maximal negative curvature.





Siddiqi's Shape Triangle

Again, defining a visual part precisely is difficult. Ambiguities in understanding shape are illustrated by the so-called shape triangle (Siddiqi, 1996).



Approaches to Shape Recognition

Existing approaches may be classified into 3 categories:

- Feature-based
- Boundary-based
- Structural

Within boundary-based and structural approaches a class of explicitly hierarchical approaches may be defined.

Feature-Based Coding

This category subsumes all approaches that determine a feature-vector for a given shape. Two operations need to be defined: a mapping of shape into the feature space and a similarity of feature vectors.

- $R : SHAPE \rightarrow \mathbb{R}^n, \qquad R(S) \mapsto \vec{F}_S$
- $S : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_0^+$

To measure similarity either the Euclidean or a weighted vector distance is used.

$$S(\vec{F}_A, \vec{F}_B) = \sum_{i=1}^{m} \alpha_i (\vec{F}_{A,i} - \vec{F}_{B,i})^2$$

Fourier Descriptors

Besides rather simple features like for example the quotient of perimeter and area, more sophisticated features may be used. Especially Fourier descriptors and momenta are applied.

Fourier Descriptors

A description of the shape as curve ϕ in polar coordinates is required, the arc length is normalized to 2π.

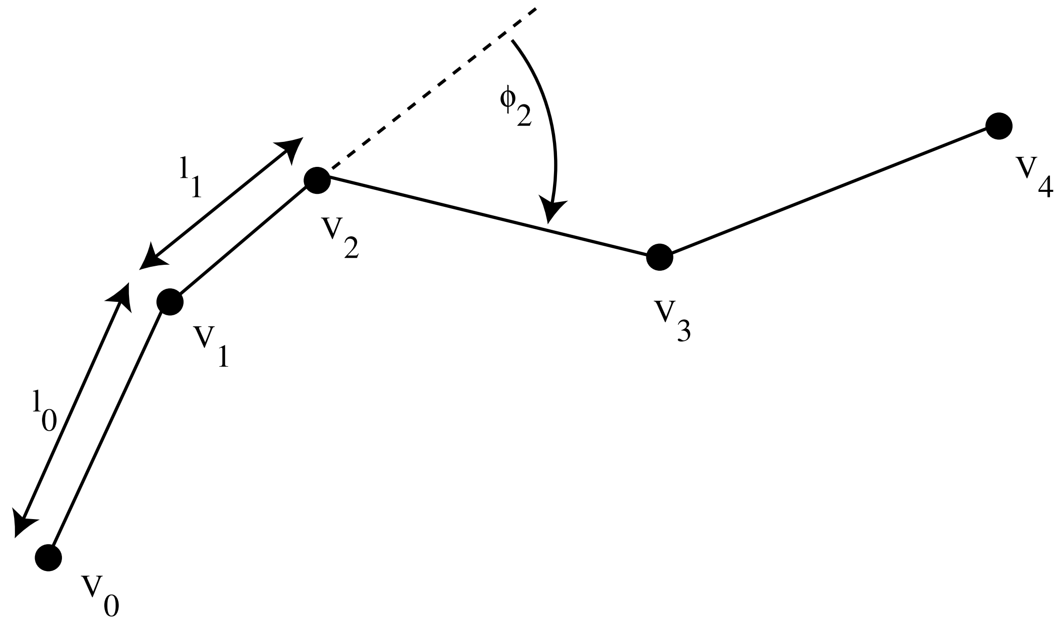
$$\phi^*(t) = \phi\left(\frac{Lt}{2\pi}\right)$$

From the transformation

$$\phi^*(t) = \mu_0 + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt)$$

the first coefficients a_i, b_i are represented in the feature vector $\vec{F}$.

Fourier Descriptors for Polygons



$$
\begin{aligned}
a_n &= -\frac{1}{n\pi}\sum_{k=1}^{m} \Delta\phi_k \sin\frac{2\pi n l_k}{L} \\
b_n &= \frac{1}{n\pi}\sum_{k=1}^{m} \Delta\phi_k \cos\frac{2\pi n l_k}{L}
\end{aligned}
$$

A special property of Fourier descriptors is that a shape's symmetry shows up in the feature vector.

Momenta

Inspired from physics, momenta are applied to compute shape-features. For a natural number n a central moment of degree $n = p+q$ is defined as

$$m_{p,q} = \int\int (x_1 - \bar{x}_1)^p (x_2 - \bar{x}_2)^q g(x_1, x_2) dx_1 dx_2$$

with

$$\bar{x}_i = \frac{\int\int x_i g(x_1, x_2) dx_1 dx_2}{\int\int g(x_1, x_2) dx_1 dx_2},$$

a kind of barycenter. g is hereby a pixel-based feature, e.g., gray-value. For a binary-based image ($g : \mathbb{N} \times \mathbb{N} \to \{0, 1\}$) the formula can be rewritten by sums. The momenta are already translation-invariant. To achieve scale-invariance, a scaling is applied.

$$\tilde{m}_{p,q} = \frac{m_{p,q}}{m_{0,0}^{(p+q+2)/2}}$$

Better results can be achieved by applying Zernike momenta.

Feature-Based Coding

Both approaches, Fourier descriptors and momenta have similar properties.

- Provide a compact representation
- Works for any shape
- Requires complete shapes
- Sensible to noise (except Zernike momenta which are computationally demanding)
- Map dissimilar shapes to similar feature vectors
- Make the choice of a similarity function difficult

Boundary-Based Representation

Shapes are represented by their boundary only. Thus, a similarity for curves is defined.

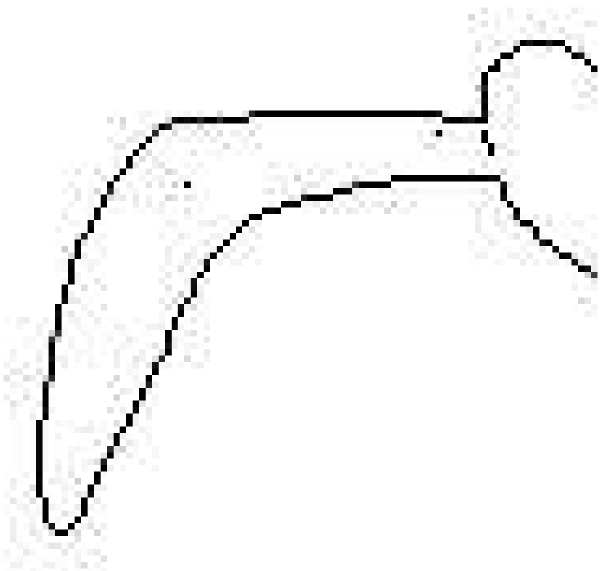
A key features is that a basic similarity measure and a linear matching are combined to make up the similarity measure.

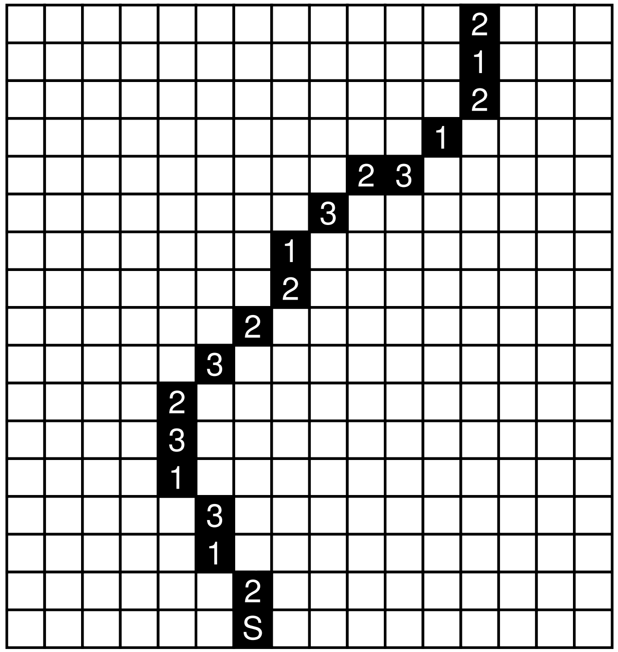
Approaches exist for discrete and continuous curves.

Chain Coding

A binary image can be converted into a so-call chain code representing the boundary. The boundary is traversed and a string representing the curvature is constructed.

1	2	3
4	↑	5
6	7	8





$\Rightarrow$ 2 1 3 1 3 2 3 2 2 1 3 2 3 1 2 1 2

Basic Similarity Measure

For curvature classes, a (dis-)similarity can be defined.

1	2	3
4	↑	5
6	7	8

	1	2	3	4	5	6	7	8
1	0	1	2	1	3	2	3	4
2	1	0	1	2	2	3	4	3
3	...							
⋮								

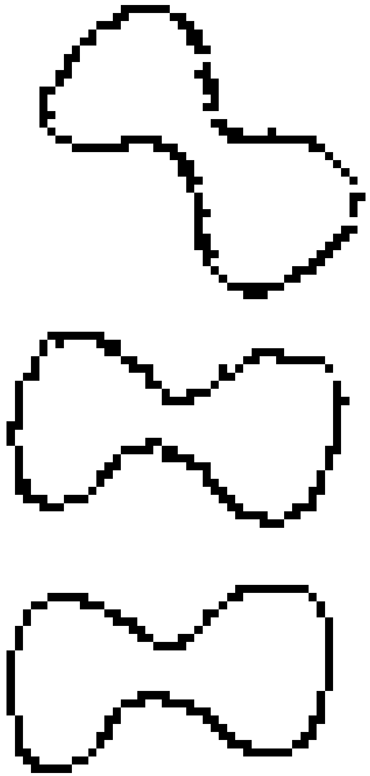
To extend this measure to strings, two steps are carried out.

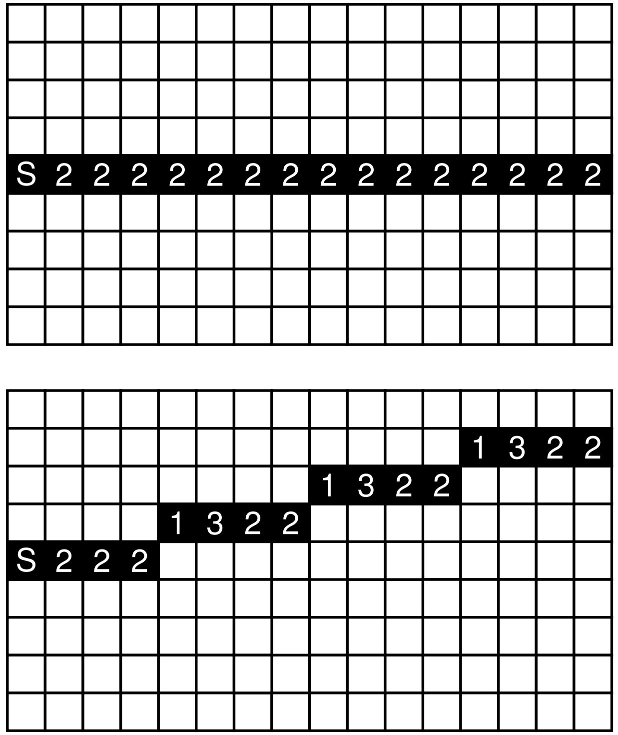
1. Extend the measure to character against string, for example by summing up individual similarity measures.

2. Employing a matching to compute a correspondence of sub-strings. Hereby, the matching constitutes from 1-to-1, 1-to-many, and many-to-1-matchings. It is computed as string matching by means of dynamic programming.

Rotational Effects

Digital curves suffer from effects caused by digitalization.





Dynamic Programming I

Definitions:

- Words $v, w \in X^\star$, sub-strings from position i to j are denoted $v_{i,j}$; $v_{i,i} := v_i$
- Basic similarity measure $S : X \times X^\star \cup X^\star \times X \to \mathbb{R}_0^+$
- Correspondence of strings v and w is denoted by $v \sim v$
- Matrix M of size $|v| \times |w|$, $M_{i,j}$ represents the optimal matching of $v_{1,i} \sim w_{1,j}$. Thus, $M_{|v|,|w|}$ is the overall solution.

Dynamic Programming II

$S(v_{1,1}, w_{1,1})$ $v_1 \sim w_1$	$S(v_{1,1}, w_{1,2})$ $v_1 \sim w_{1,2}$	$S(v_{1,1}, w_{1,3})$ $v_1 \sim w_{1,2,3}$	$\ldots$	$S(v_{1,1}, w_{1,\lvert w \rvert})$ $v_1 \sim w_{1,\lvert w \rvert}$
$S(v_{1,2}, w_{1,1})$ $v_{1,2} \sim w_1$				
$S(v_{1,3}, w_{1,1})$ $v_{1,3} \sim w_1$				
$\vdots$				
$S(v_{1,\lvert v \rvert}, w_{1,1})$ $v_{1,\lvert v \rvert} \sim w_1$				

Empty cells can be computed row by row...

Dynamic Programming III

Three possibilities when computing $M_{i,j}$ need to be considered.

$\nearrow$ Extend the matching stored in $M_{i-1,j-1}$ by $v_i \sim w_j$, add $S(v_i, w_j)$ to the overall similarity.

$\uparrow$ Extend the matching stored in $M_{i-1,j}$ of kind $v_{i',i-1} \sim w_j$, $i' \leq i-1$ to $v_{i',i} \sim w_j$; add to the overall similarity $S(v_{i',i}, w_j) - S(v_{i',i-1}, w_j)$.

$\rightarrow$ Extend the matching stored in $M_{i,j-1}$ of kind $v_i \sim w_{j',j-1}$, $j' \leq j-1$ to $v_i \sim w_{j',j}$; add to the overall similarity $S(v_i, w_{j',j}) - S(v_i, w_{j',j-1})$.

The best alternative, i.e. the one that yields the lowest dissimilarity, is chosen.

Example

- $X = \{A, B\}$
- $S(v, w) = 1 - \frac{\#\ v \text{ occurs as prefix in } w}{|w|}, \quad v \in X, w \in X^\star$

$\Rightarrow$ Compute the similarity of matching "ABA" against "ABBA"

A	0.66 ABA∼A	0.5 A∼A, BA∼B	1 A∼A, B∼B, A∼B	$\nearrow$ 0 $\rightarrow$ 1 $\uparrow$ $n.a.$
B	0.5 AB∼A	0 A∼A, B∼B	0 A∼A, $B \sim BB$	0.33 A∼A, B∼BBA
A	0 A∼A	0.5 A∼AB	0.66 A∼ABB	0.75 A∼ABBA
	A	B	B	A

Alternative Matching-Techniques

As string-matching is not able to model a matching of digital curves adequately (a shortcoming of the chain codes rather than of the matching), more sophisticated matching algorithms are employed in "real applications" using chain codes.

Weighted Levensthein Distance

Defines an edit distance for transforming one string into another. Costs are defined for altering, deleting, or inserting a character.

Extended Distance

Formal translation system with costs assigned to individual production rules.

Deformation Energy

A similarity measure is defined for curves $\Gamma : [0, L] \to \mathbb{R}^2$. Basri et al. propose a measure based on a physical model of deformation.

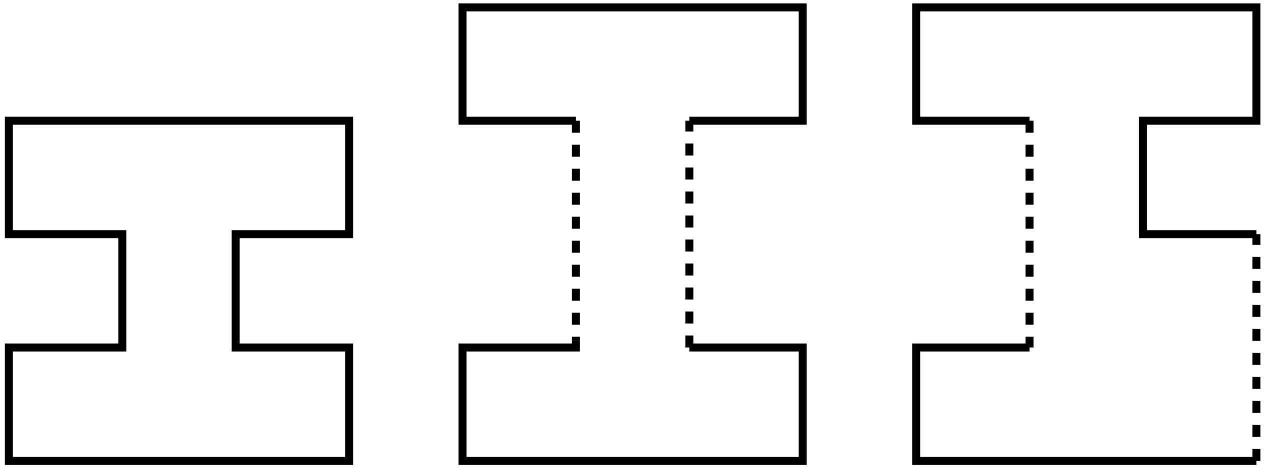
$$C(\Gamma_1, \Gamma_2) = \min_{t(s)} \int_{\Gamma_1} F(k_1(s), k_2(t(s)), \frac{dt}{ds}) ds$$

Hereby $k_i(s)$ denotes the curvature of Γ_i at position s. A generalized model for a spring's energy is utilized as cost function:

$$\begin{aligned} E &= \frac{1}{p} \alpha \Delta x^p \\ E_{strain}(\Gamma_s, \Gamma_t, \frac{dt}{ds}) &= \frac{\alpha}{p} \int_{\Gamma_s} \frac{|\frac{dt}{ds} - 1|^p}{(\frac{dt}{ds} + 1)^{p-1}} ds \\ E_{bend}(\Gamma_s, \Gamma_t, \frac{dt}{ds}) &= \frac{\alpha'}{p} \int_{\Gamma_s} \frac{|k_t t' - k_s|^p}{(|k_t t'| + |k_s|)^{p-1}} ds \end{aligned}$$

Local Similarity Measure

Due to the local nature of the similarity measure, it can be used for partial shapes. Like in chain coding, a matching is employed to compute the best correspondence of arcs. However, global effects cannot be accounted for.



Modifying a shape by stretching one segment by the same amount yields shapes with the same measure of dissimilarity objecting the impression of similarity.

Features

Purely Boundary-based approaches show similar properties. They...

- Handle arbitrary, polygonal shapes
- Can be applied to recognition of partial shapes
- Respect for invariants (except chain coding)
- Constitute of a basis similarity measure and a linear matching
- Cannot account for global features, e.g., symmetry

Hierarchical & Boundary-Based

Two approaches are presented that address hierarchical issues, i.e. a shape's salient parts.

- Explicit representation of hierarchy through curve evolution
- Implicit representation by matching based on maximal arcs

Curvature Scale Space

The curvature of a boundary smoothened by a filter is represented. Given a curve Γ

$$\Gamma : [0, 1] \to \mathbb{R}^2, \quad \Gamma(s) = (x(s), y(s))$$

a convolution is applied to yield obtain smoothened curves Γ_σ:

$$\Gamma_\sigma : [0, 1] \to \mathbb{R}^2, \quad (X(s, \sigma), Y(s, \sigma))$$

$$X(s, \sigma) = x(s) \circledast g(s, \sigma), \quad Y(s, \sigma) = y(s) \circledast g(s, \sigma)$$

$$g(s, \sigma) = \frac{1}{\sigma\sqrt{s\pi}} \mathrm{e}^{\frac{-s^2}{2\sigma^2}}$$

$$X(s, \sigma) = \int_{-\infty}^{\infty} x(u) \cdot g(s, \sigma) du$$

Smoothening by Gaussian Convolution

(Picture removed)

Curvature Scale Space

A shape is represented by the zero set of the array of curves' curvature, the inflection points at a certain level of smoothening.

(Picture removed)

To realize matching algorithm, the positions of the n largest peaks are compared.

Not all shapes have inflection points!

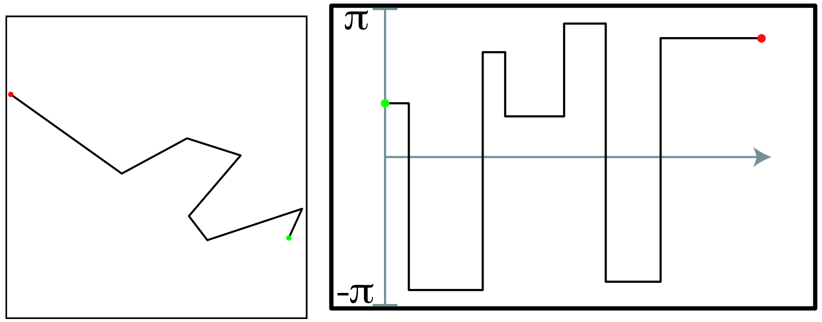
CSS: Example

(Picture removed)

Two different shapes can have the same CSS representation.

Maximal Arc Decomposition I

Similar to the boundary-based approach by Basri et al. a basic similarity measure of arcs is defined. The measure is defined in tangent space. For a curve C the corresponding tangent space representation of the normalized curve is denoted by T_C, $T : [0,1] \rightarrow [0, 2\pi)$.

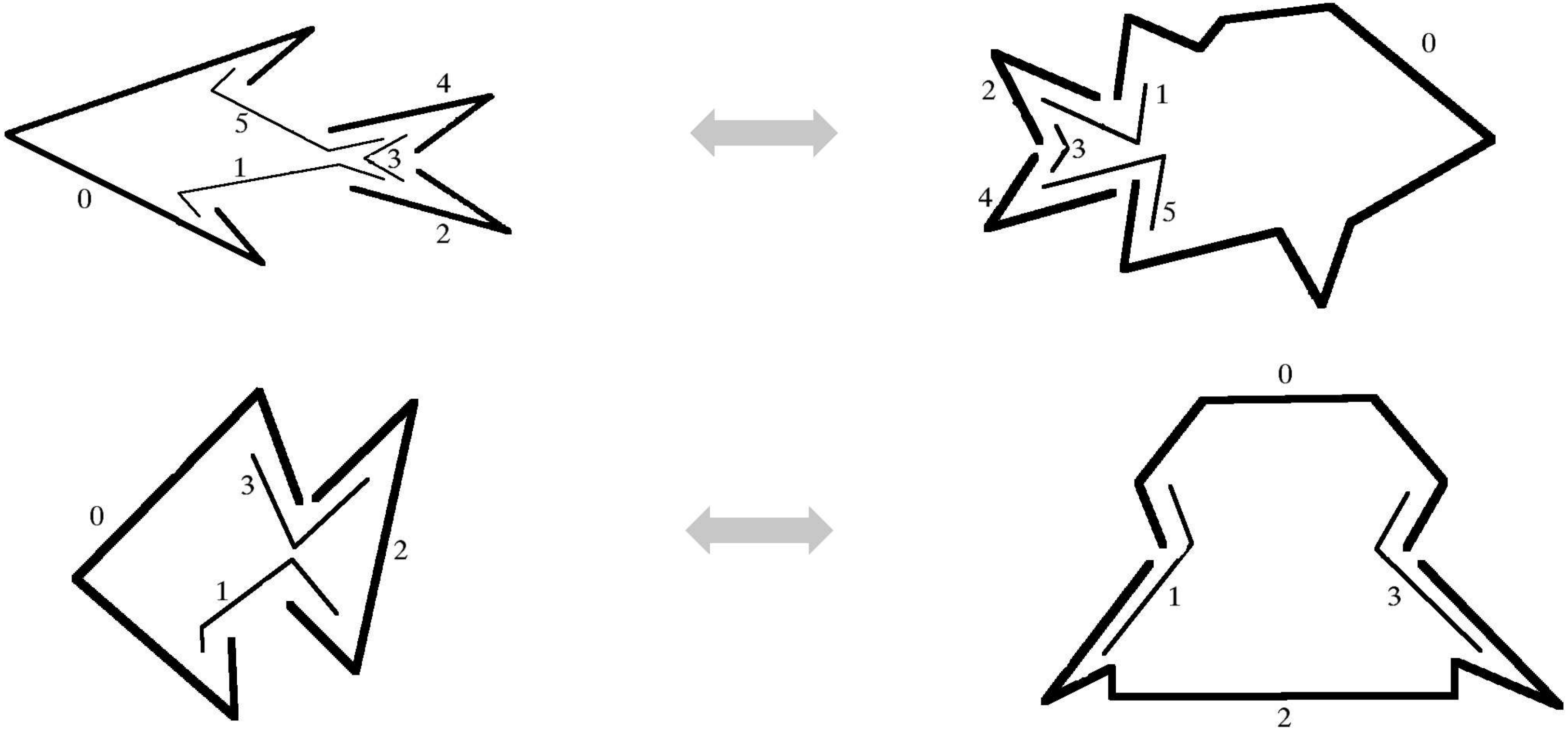


$$S(C,D) = \max\left\{\frac{|C|}{|D|}, \frac{|D|}{|C|}\right\} \cdot \int_0^1 (T_C(s) - T_D(s) + c_{C,D})^2 ds$$

The measure can be viewed as a L_2-distance in tangent space.

Maximal Arc Decomposition II

A polyline is segmented into a sequence of maximal convex or concave arcs which overlap in exactly one line-segment. Matching strings of maximal arcs against each other by means of dynamic programming yields the overall similarity.

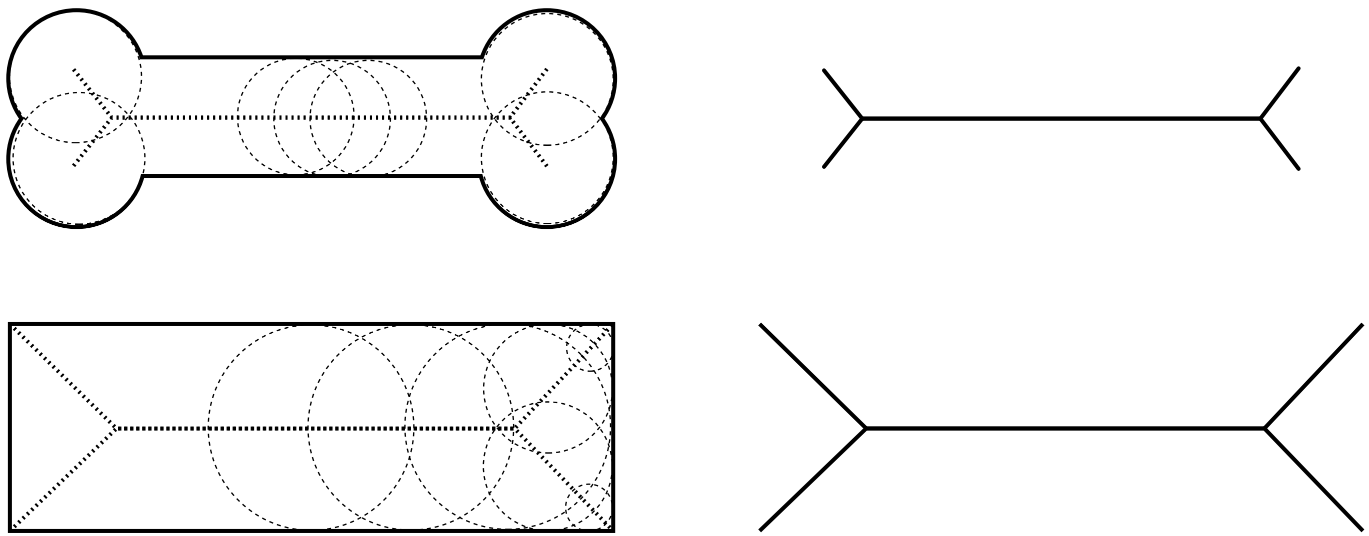


Shapes are pre-processed by means of discrete curve evolution.

Structural Approaches

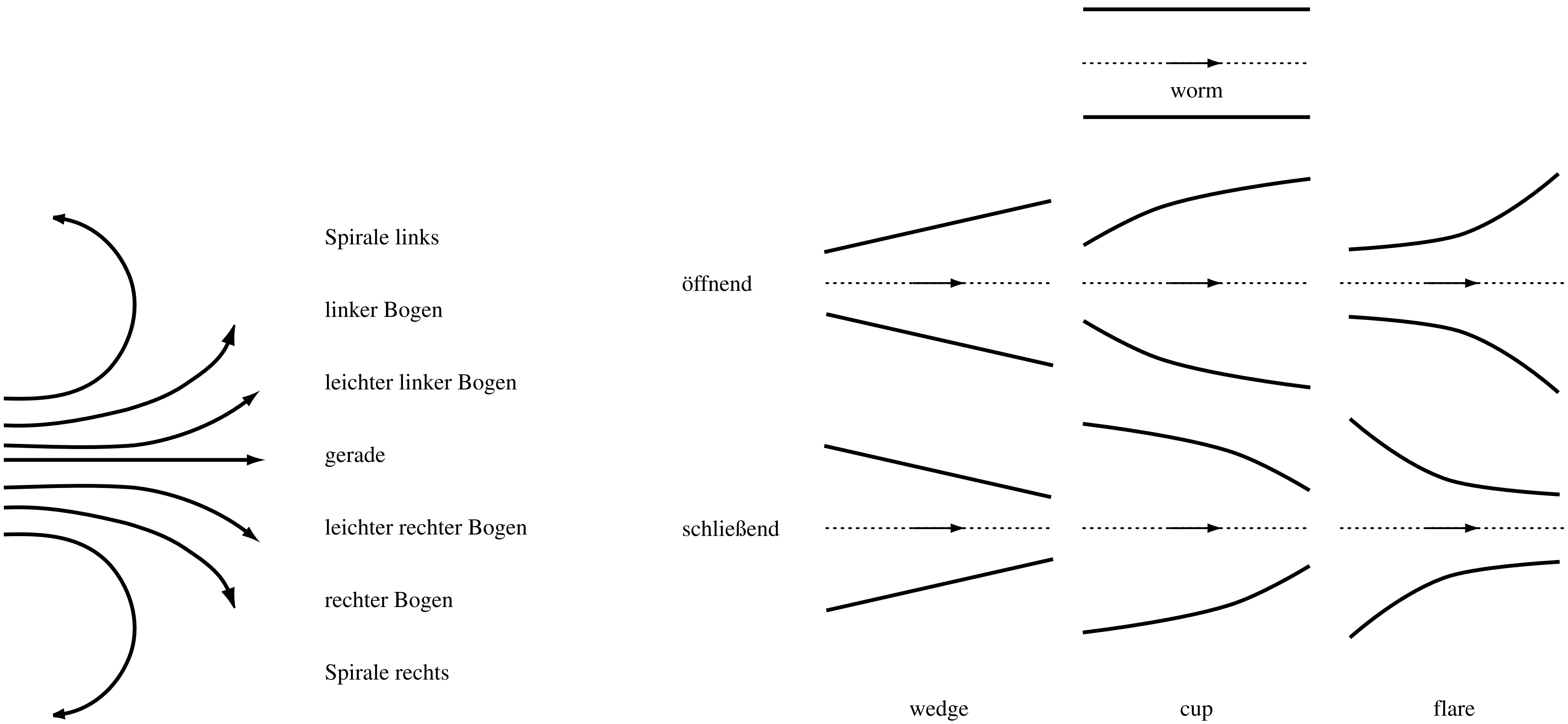
These approaches capture the structure of a shape, typically by representing shape as a graph. Representing 3D-objects by means of *geons*, generalized cylinders, is an example for a structural representation (Biederman, 1987).

To compute a shape's structure, a *skeleton* is computed. The computation can be described as a medial axis transform, a kind of discrete generalized voronoi.



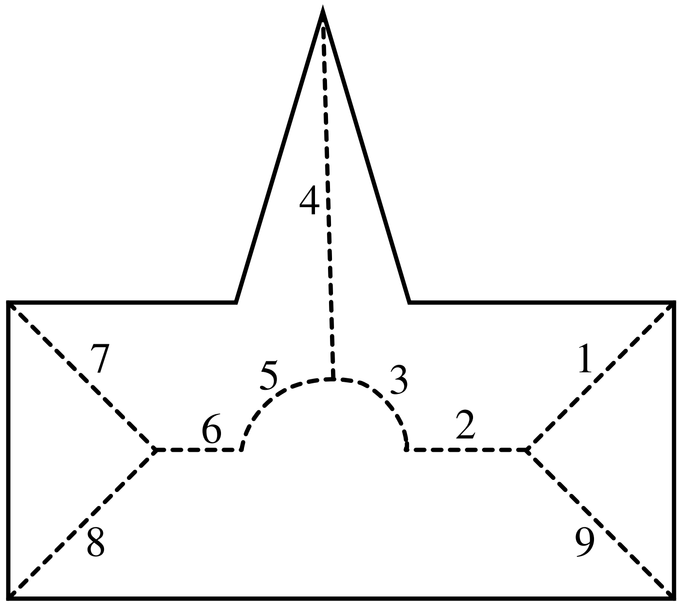
Skeletons I

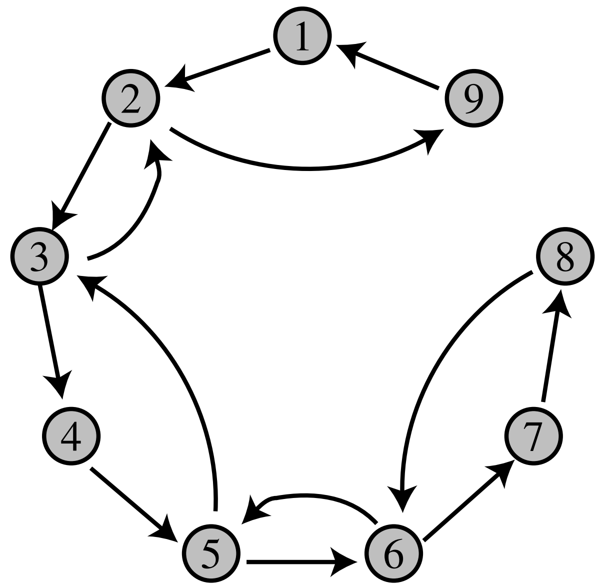
Once the skeleton is computed, its parts are classified qualitatively.



Skeletons II

Finally, the graph is constructed mirroring the adjacency of the skeleton's parts. Edges are labeled according to the qualitative classes.

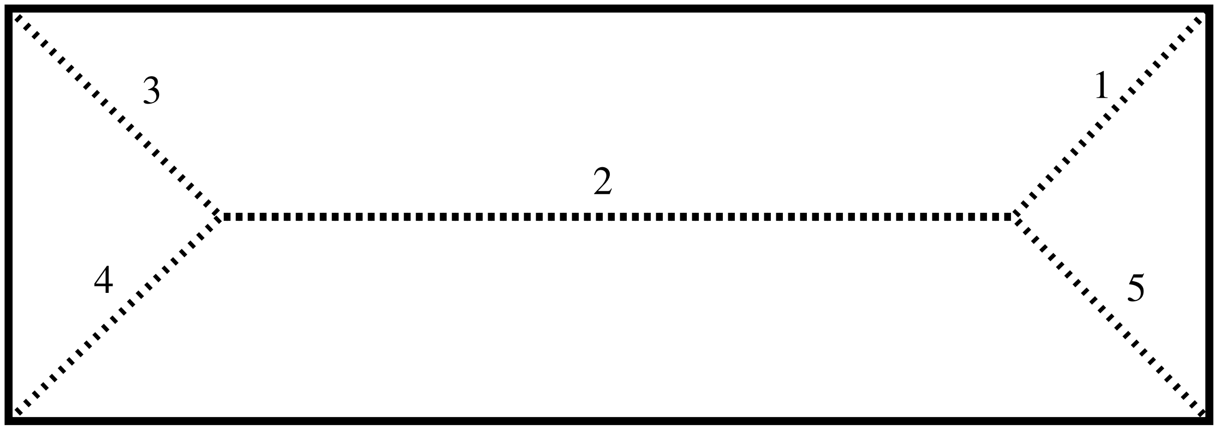


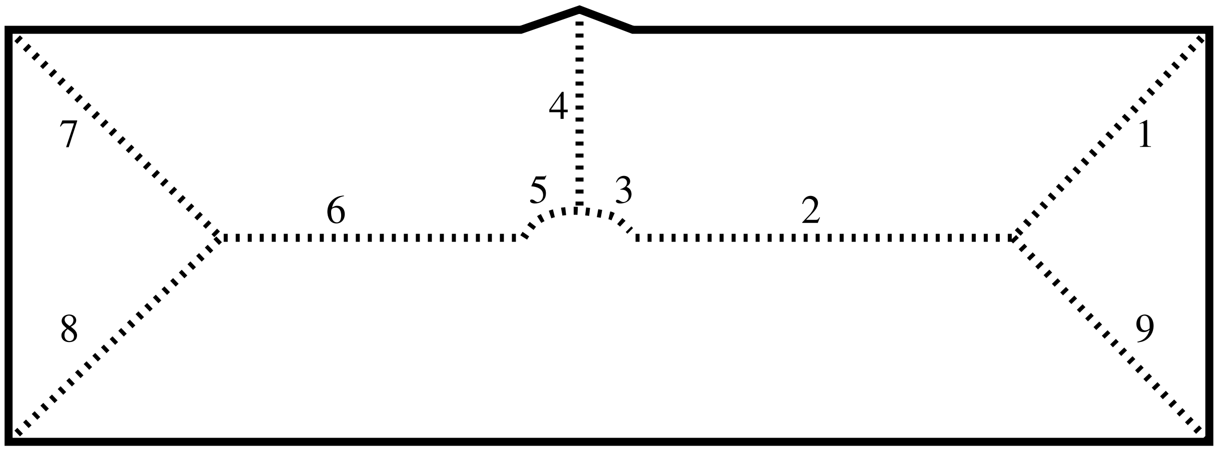


Matching two shapes requires matching two, often different, graphs against each other.

Skeletons III

Even small distortions alter the structure.





Skeletons: Features

Purely structural approaches have not been proven successful yet.

- Incorporate a graph-matching and a basic similarity measure
- Provide more context information than boundary-based approaches
- Can respect for global properties like symmetry
- Graph-matching is computationally demanding
- Computing skeletons is unstable
- Small distortions alter the structure

Structure & Hierarchy: Shock Graphs

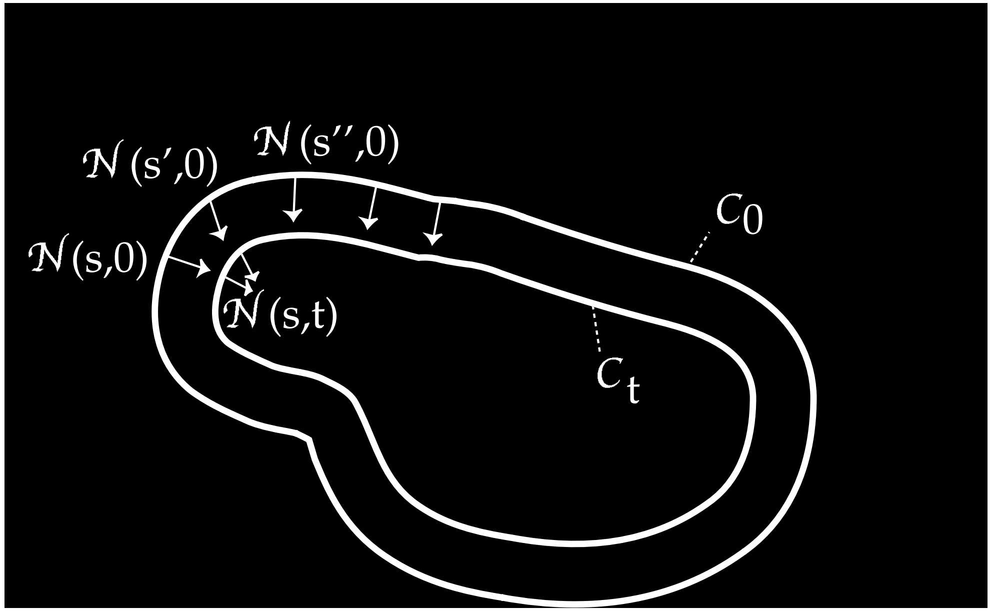
As a structural approach seems promising, work has been carried out to overcome the shortcomings discussed. Four aspects are addressed:

1. Stable computation of the medial axis transform
2. Stable construction of a graph structure
3. Efficient matching of resulting graphs
4. Information to be represented by the nodes & edges

Shock Graphs: MAT

The approach of shock graphs (Siddiqi et al., 1996) employs a curve evolution retracting a curve to its skeleton.

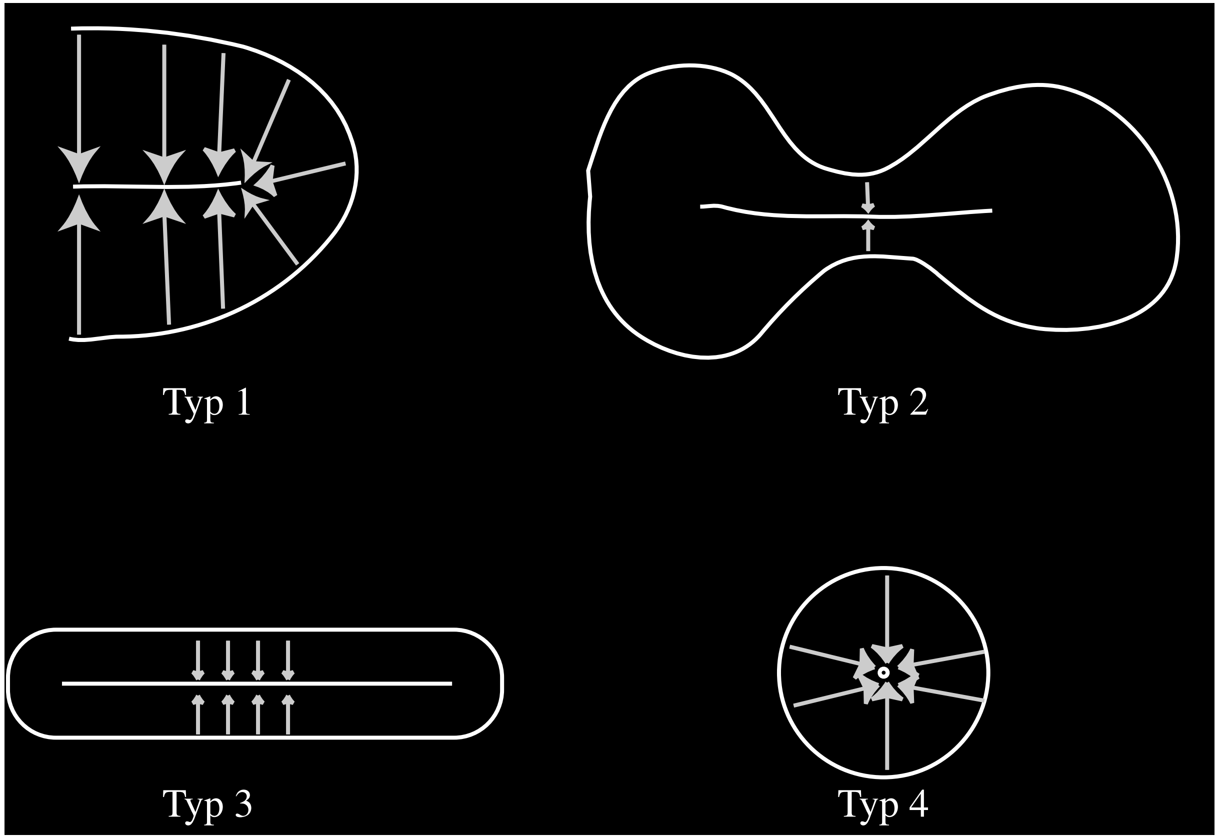
$$\mathcal{C}_t = (1 + \alpha\kappa)\mathcal{N}$$
$$\mathcal{C}(s, 0) = \mathcal{C}_0(s).$$



This computation allow for a computation with sub-pixel accuracy.

Shock Graphs: Labeling

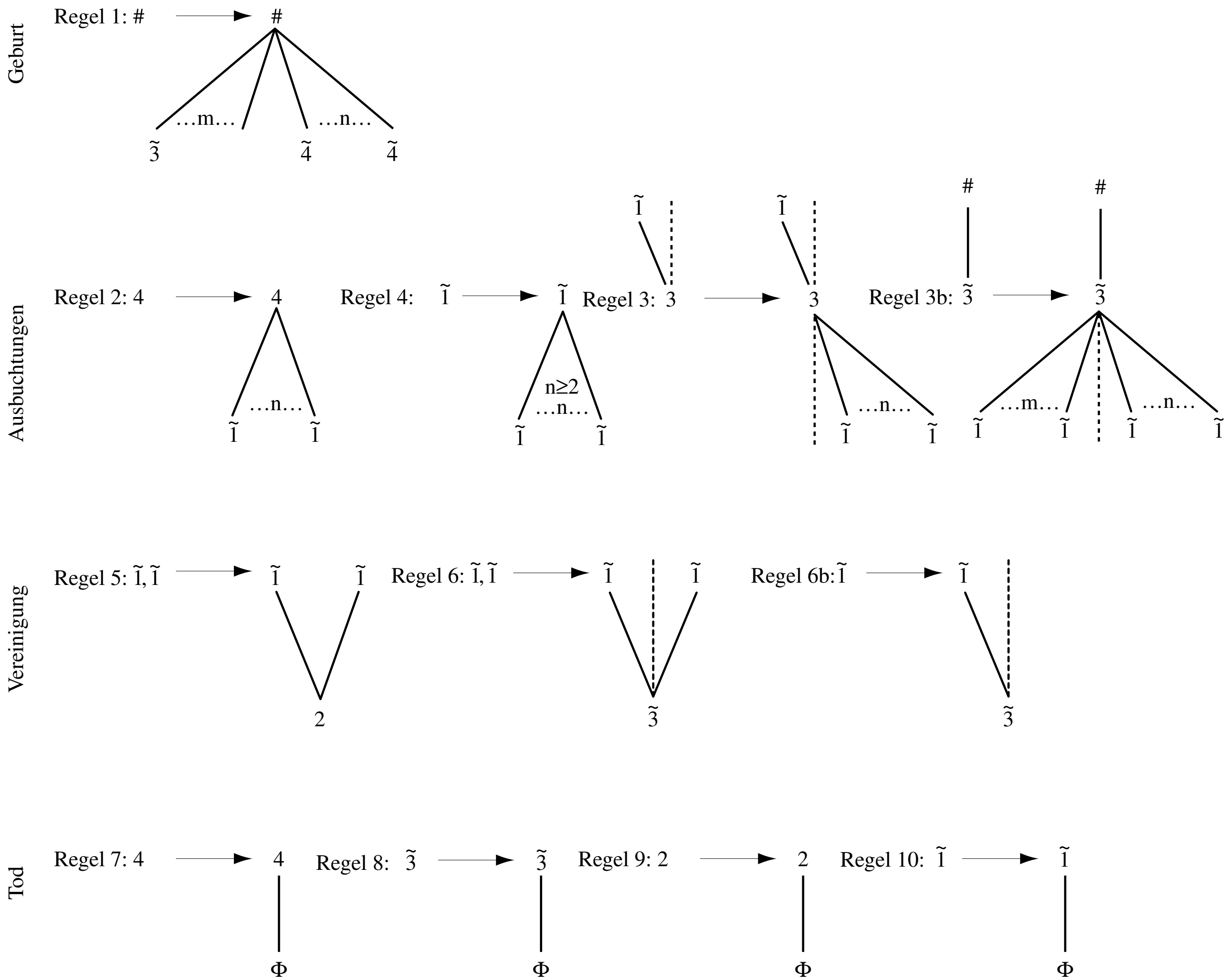
Shock-segments are classified according to 4 classes.



Additionally, the curve an edge corresponds to is used to label the edge. Therefore, it is sub-sampled.

Shock Graphs: Graph Construction

Graphs are constructed according to a formal context-sensitive grammar.

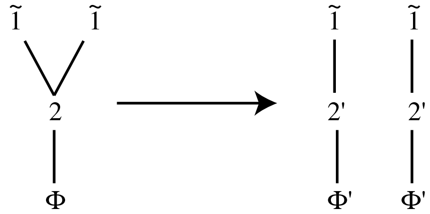


Shock Graphs: Examples

(Picture removed)

Shock Graphs: Graph Matching

To reduce the complexity ($\mathcal{NP} \rightarrow \mathcal{P}$), graphs are transformed into trees.



This problem is equivalent to finding a maximal clique in a derived, so-called association graph (cp. Pellilo, 1999). However, this is quite complicated.

Summary

Feature-based coding is the only approach that does not employ a matching algorithm making it quite fast. However, its is restricted to closed worlds.

Boundary-based representation allows for an efficient linear matching and yields good results. Global properties like symmetry cannot be respected adequately. Respecting the hierarchical nature of shape improves recognition greatly. It can even be accounted for global properties.

Structural approaches offer a high-level interface to shapes. They allow for respecting local and global properties. However, computation of a robust structure has not been solved thoroughly yet; graph matching remains a challenge.

Good night!