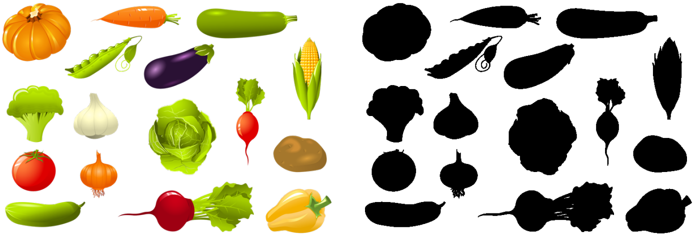
Classification of two-dimensional figures using skeleton-geodesic histograms of thicknesses, distances and directions

Nikita Lomov nikita-lomov@mail.ru
Sergey Sidyakin sersid@gosniias.ru
Yuri Vizilter viz@gosniias.ru

FSUE State Scientific Research Institute of Aviation Systems
Moscow, Russia

IDP-2016
BARCELONA
IDP 2016, Barcelona

Shape recognition problem

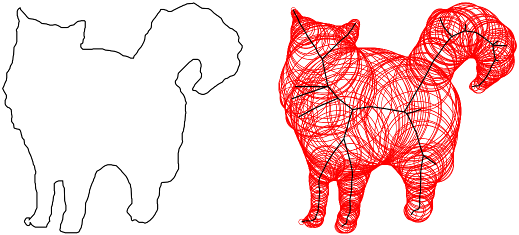


We have no information about the color, texture, brightness, etc.

It is more convenient not to work with an object as with a fragment of the image, but to get an explicit description of its shape.

According to Cambridge Dictionary, **shape** is *the physical form of something made by the line around its outer edge*.
One of the definitions on Dictionary.com states that **shape** is *something seen in outline, as in silhouette*.

Two approaches to shape representation



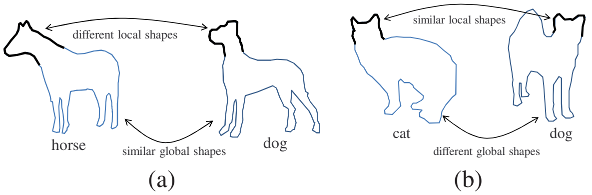
(a) (b)

(a) The *contour* one is a description of the boundaries in the form of a closed curve.
(b) The *medial* one specifies the skeleton and the radial function.

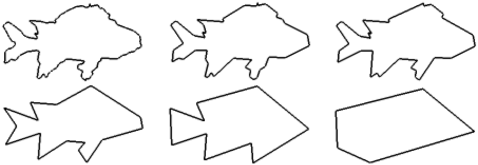
Both representations give a full description of the shape, but emphasize different features of it. Contour representation describes better the outlines of the figure, the medial one — shape bends and its overall structure, as well as thickness of its parts.

Thus, in the problem of shape recognition to use both the contour and the skeleton information is reasonable.

Multiscale analysis

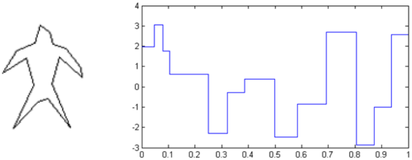


It is necessary to take into account local and global analysis in aggregate to cover all possible situations.
A possible approach is to consider contour fragments of different length. In order to determine ends of these fragments we can use the algorithm of discrete curve evolution (Latecki and Lakämper, 2000).

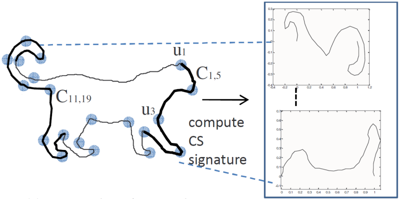


Contour descriptors

1. Tangent Function (Pun and Li, 2009)



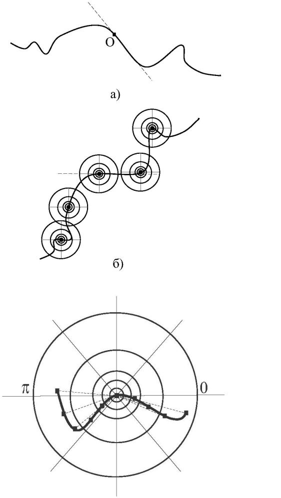
2. Signature-like descriptor
(Sun and Super, 2005)



(c) extracted vertices on the original shape

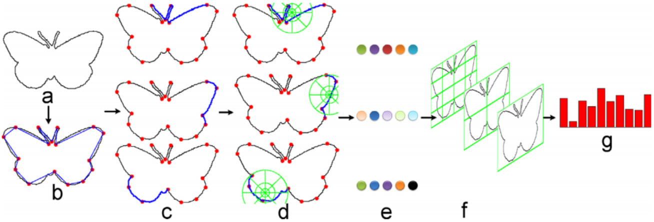
(d) signatures

3. Shape context
(Belongie and Malik, 2000)



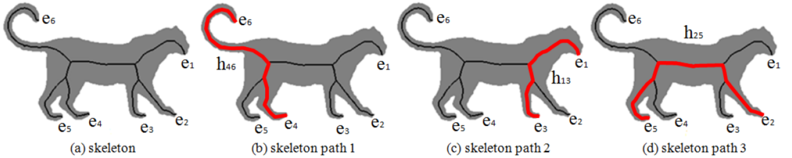
Bag of Contour Fragments

One of the most powerful descriptors extracting contour information proposed by Wand et at. in 2014.



a) Representation of a shape contour by a closed polyline
b) Critical points detection using DCE method
c) Contour fragments extraction
d) Shape context calculation to describe each contour fragment
e) Encoding fragments by local linear embedding (Roweis and Saul, 2000)
f) Spatial pyramid matching
g) Getting the final descriptor by max-pooling at each level of the pyramid

Bag of Skeleton Paths



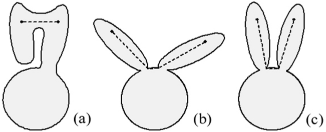
Skeletal path reflects the change in the radial function during the transition from one terminal vertex of the skeleton to another (Bai, Liu and Tu, 2009).

However, the addition of skeletal descriptor, constructed in the similar manner as the contour one, does not give a significant increase in the classification quality (Shen et al., 2014): 85.50% against 83.4% on Animal Dataset and 98.35% against 97.16% on MPEG7).

Therefore, the task of building skeleton-based image descriptor efficiently supplementing the contour part, remains open.

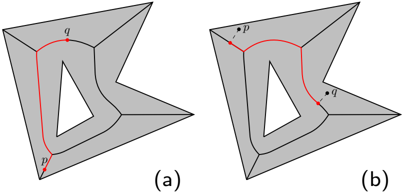
Skeleton-geodesic distances

Geodesic distance $d_{Geod}(X)(p,q)$ is the length of the shortest path between points p and q, lying inside the shape. Distribution of geodesic distances is informative descriptor of the shape, insensitive to articulations (Ling and Jacobs, 2007).



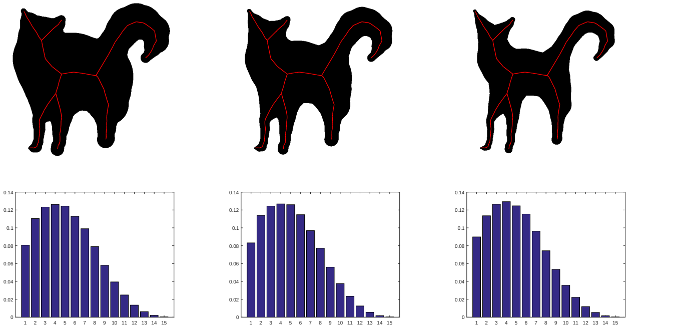
Skeleton-geodesic distance $d_{Geod}(Sk(X))(p,q)$ is the length of the shortest path between points $p, q \in Sk(X)$ anolg the figure skeleton (a). In the general case $d_{Geod}(Sk(X))(p,q)$, $p, q \in X$ is the skeleton-geodesic distance between the projections $p_{Sk}(X)$, $q_S k(X)$ of these points on the skeleton (b):

$$d_{Geod(Sk(X))}(p,q) = d_{Geod(Sk(X))}(p_{Sk(X)}, q_{Sk(X)})$$



Skeleton-geodesic distances (Continued)

Resistant not only to bending fat curves constituting the figure, but also to a systematic change of their width.



In the case of discrete figure and skeleton calculation can be performed according to the formula

$$H_{GS}(d) = \sum_{p,q \in Sk(X),\, d_{Geod(Sk(X))}(p,q)=d} S(p)S(q),$$

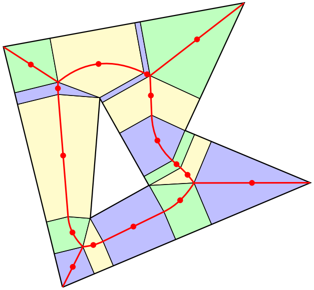
where $S(p)$ and $S(q)$ are the areas of regions of attraction of the corresponding points:

$$S(p) = \#\{q \in X : q_{Geod(Sk(X))} = p\}.$$

Skeleton-geodesic distances: continuous approach

Discrete method of calculation is computationally inefficient. When transition to a continuous representation another problem appears: the number of skeleton points becomes infinite. You must select specific skeleton points and "distribute" to them all the other points of the figure.

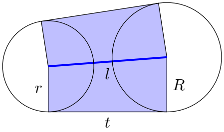
Let us take a point on each edge of the skeleton — let it be the midpoint. As regions of attraction we take proper regions of the edges (Mestetskiy, 2014), which form a partition of the figure. We assume that all the points of the proper region "projected" in the middle of the edge.



Proper regions

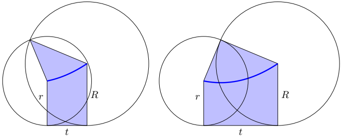
To calculate the area of the region of attraction it is sufficient to know only the radii of the edge end circles and the distance between their centers.

Linear edge



$$S_{lin} = (r+R)t, \quad t = \sqrt{l^2 - (R^2 - r^2)}$$

Parabolic edge



$$S_{par} = \begin{cases} \Phi(R) + \Phi(r) & \text{if } t \geq 2\sqrt{r(R-r)} \\ \Phi(R) - \Phi(r) & \text{if } t < 2\sqrt{r(R-r)} \end{cases}$$

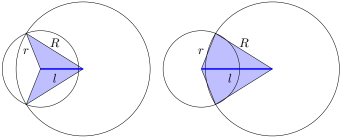
$$\Phi(z) = \frac{1}{2}(z+p)\sqrt{p(2z-p)}$$

$$p = \frac{t^2}{2l^2}\left(R + r + \sqrt{(R+r)^2 - l^2}\right)$$

$$l_{par} = \begin{cases} \Lambda(R) + \Lambda(r) & \text{if } t \geq 2\sqrt{r(R-r)} \\ \Lambda(R) - \Lambda(r) & \text{if } t < 2\sqrt{r(R-r)} \end{cases}$$

$$\Lambda(R) = \frac{1}{2}\left(\frac{\left(z-\frac{p}{2}\right)\sqrt{p^2+\left(z-\frac{p}{2}\right)^2}}{p} + \mathrm{Arsh}\left(z - \frac{p}{2}\right)\right)$$

Hyperbolic edge



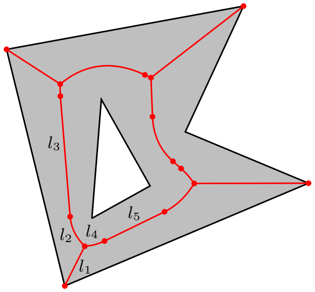
$$S_{hyp} = \begin{cases} \Psi(R) + \Psi(r) & \text{if } l^2 + r^2 \geq R^2 \\ \Psi(R) - \Psi(r) & \text{if } l^2 + r^2 < R^2 \end{cases}$$

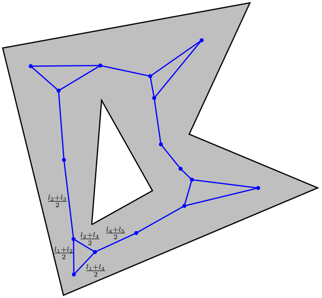
$$\Psi(z) = \frac{q}{2}\sqrt{z^2 - \frac{q^2}{4}}$$

$$q = \frac{1}{l}\left(\sqrt{[(l+r)^2 - R^2]\cdot[R^2 - (l-r)^2]}\right)$$

Skeleton dual graph

Let us build a dual skeleton graph in which the edges corresponding to the vertices in the skeleton are adjacent $\iff$ the edges in the original skeleton are adjacent. The weights of the edges in the new graph is half the sum of lengths of the edge-generators. The shortest way in the dual graph defines the shortest path between the edge midpoints in the original one, and these paths are of the same length.



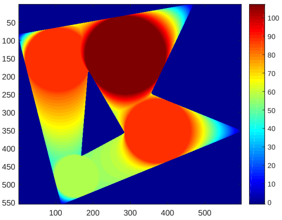


Now all shortest paths in the dual graph can be found by Johnson's algorithm.

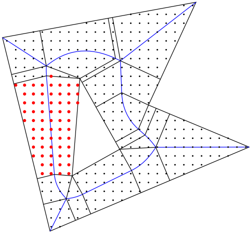
Thickness and direction of the edge

Let us define the other characteristics of the edge e:

- $\alpha(e)$ is the angle to the abscissa determined by the tangent at the midpoint.
- Let $T(q)$ is the thickness of the figure at the point, defined as the radius of the maximum inscribed circle covering the figure (a). Then $T(e)$ is thickness, averaged on the proper region.



(a)

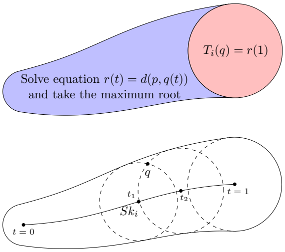


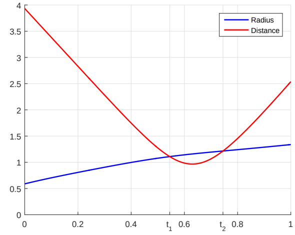
(b)

The fully continuous method of calculating the average thickness is possible, but too difficult to implement. Instead, we calculate $T(q)$ in the desired subset of grid points and average the results (b).

Calculation of the thickness at the point

We need to determine the radius of the maximum circle covering the point. We will seek this circle separately in each edge, which silhouette includes the point. According to the theorem, the maximum covering circle in a skeleton line with monotonic radial function either the largest end circle, or such that the point lies on its boundary.





We solve the equation: the radius of the circle is equal to the distance to the given point. For linear and parabolic edges the equation is reduced to the square one, for hyperbolic — to the linear one. For the purposes of computational efficiency, in the main loop we should look over the edges and consider all points covered by its silhouette.

Feature distributions

To ensure invariance to rotation, we consider the distributions of features between edge pairs.

- t_{ij} is the difference between average thicknesses of the edges e_i and e_j, which is equal to $t_{ij} = |t_i - t_j|$.
- d_{ij} is the skeletal geodesic distance between the edges e_i and e_j (more precise, the distance between their midpoints);
- ϕ_{ij} is the rotation angle between the edges e_i and e_j, calculated as $\max(|\alpha_i - \alpha_j|, \pi - |\alpha_i - \alpha_j|)$;

The features are normalized to ensure invariance to scaling:

$$d^*_{ij} = \frac{d_{ij}}{\max_{i,j} d_{ij}}, \quad \phi^*_{ij} = \frac{\phi_{ij}}{\frac{\pi}{2}}, \quad t^*_{ij} = \frac{t_{ij}}{r_{max}}.$$

Skeleton-geodesic histogram of thicknesses-distances-direction is defined as

$$H_{GS}(a,b,c) = \sum_{\substack{i,j:\lfloor k_t t^*_{ij} \rfloor = a,\\ \lfloor k_d d^*_{ij} \rfloor = b,\\ \lfloor k_a \phi^*_{ij} \rfloor = c}} s_i \cdot s_j \quad / \quad \sum_i s_i.$$

and includes $k_t k_d k_a$ cells.
We can consider thickness as an independent feature and build a thickness spectrum with n cells:

$$PS(a) = \#\{p \in \mathbb{Z} \cap X : \lfloor n \frac{T(p)}{r_{max}} \rfloor = a\}.$$

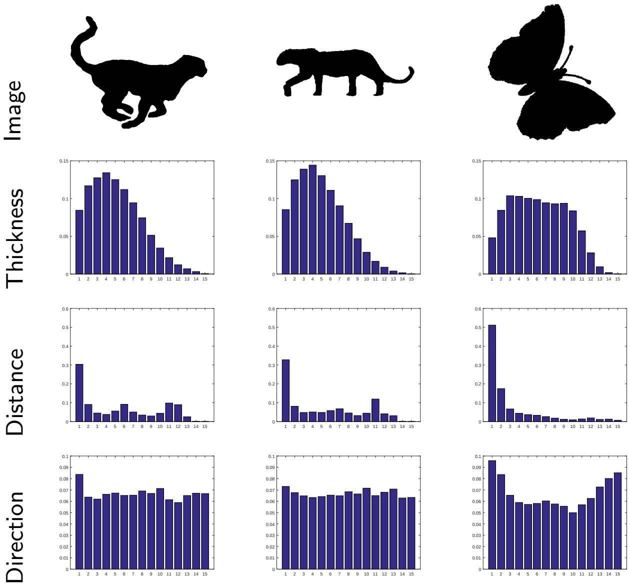
Combine them all together

The final descriptor is called CST (contour-skeleton-thickness) and consists of the following components:

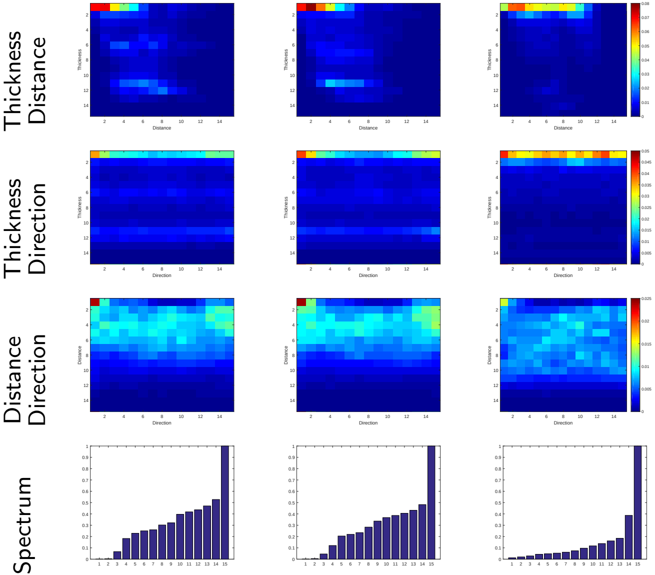
- Bag of contour fragments descriptor containing m features and responsible for contour information
- Skeleton geodesic histograms:
  - histogram of pairwise thicknesses (k_t features)
  - histogram of pairwise distances (k_d features)
  - histogram of pairwise directions (k_a features)
  - histogram of thicknesses-distances ($k_t k_d$ features)
  - histogram of thicknesses-directions ($k_t k_a$ features)
  - histogram of distances-directions ($k_d k_a$ features)
  - histogram of thicknesses-distances-directions ($k_t k_d k_a$ features)
- Thickness spectrum containing n features

In total, there is $m + (k_t + 1)(k_d + 1)(k_t a + 1) - 1 + n$ features in the descriptor.

How histograms look like



How histograms look like (Continued)



Determining the number of bins in histograms

Quality estimation is the result obtained by support vector machine classification with a tenfold partition the sample in half, and averaging the results. In some cases, we use a leave-one-out cross-validation for the SVM too.

Dataset	**3**	**4**	**6**	**8**	**10**	**15**	**25**	**50**
Animal	88.59	88.83	88.92	88.84	88.53	88.50	88.29	87.87
MPEG7	98.01	98.07	98.13	98.06	97.94	97.83	97.69	97.50
Swedish leaf	98.08	98.15	98.09	98.04	97.96	97.84	97.68	97.41

Varying the number of bins in pairwise thicknesses histogram along with BCF.

Dataset	**3**	**4**	**6**	**8**	**10**	**15**	**25**	**50**
Animal	90.38	90.47	90.98	90.99	91.11	91.04	90.92	90.42
MPEG7	97.90	98.23	98.30	98.27	98.26	98.13	98.11	97.99
Swedish leaf	97.55	97.67	97.84	98.17	98.48	98.60	98.72	98.72

Varying the number of bins in pairwise distances histogram along with BCF.

Dataset	**2**	**3**	**4**	**5**	**6**	**8**	**12**
Animal	87.64	87.71	87.85	87.80	87.72	87.68	87.70
MPEG7	97.54	97.59	97.80	97.89	97.90	97.84	97.69
Swedish leaf	97.35	97.35	97.43	97.47	97.44	97.36	97.31

Varying the number of bins in pairwise directions histogram along with BCF.

Contribution of descriptor components

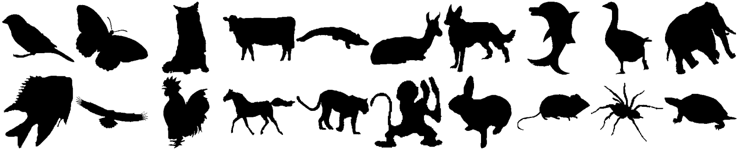
Dataset	1D	2D	3D	1D+2D	1D+3D	2D+3D	1D+2D+3D
Animal	92.40	92.96	90.59	93.29	92.79	93.19	93.50
MPEG7	97.61	98.53	97.49	99.03	98.70	98.64	99.06
Swedish leaf	99.03	98.95	98.03	99.21	99.12	99.05	99.25

Using histograms of various dimensions and BCF descriptor together.

Dataset	4	5	7	10	15	20	30	50	100
Animal	93.47	93.61	93.87	93.75	93.57	93.88	93.95	93.87	93.68
MPEG7	99.27	99.31	99.27	99.36	99.31	99.30	99.37	99.31	99.30
Swedish leaf	99.08	99.11	99.15	99.17	99.17	99.17	99.19	99.16	99.15

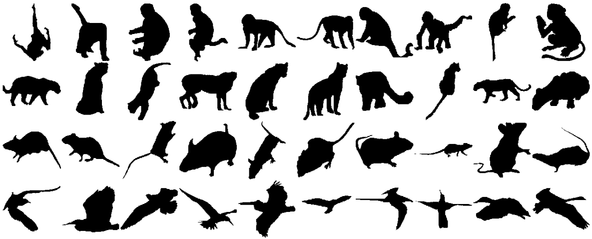
Varying the number of bins of thickness spectrum along with the previous parts.

Animal (20 classes of 100 instances)



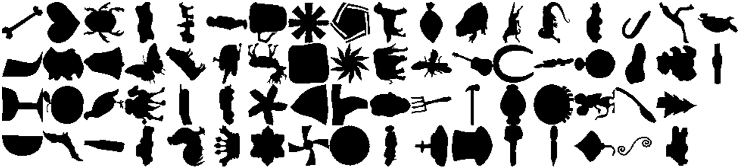
Method	**Accuracy**
Class Segment Sets	69.7%
Inner-Distance Shape Context	73.6%
Contour Segments	71.7%
Skeleton Paths	67.9%
Contour Segments & Skeleton Paths	78.4%
Bag of Contour Fragments	83.4%
Shape Tree	80.0%
Local and Global Features	80.4%
Shape Vocabulary	84.3%
Bag of Contour Fragments + Bag of Skeleton Paths	85.5%
Multi Component-2	85.9%
Contour-Skeleton-Thickness	**93.9%**

Results for each class of Animal



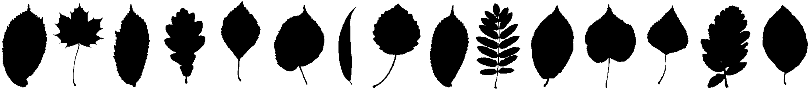
Method	Bird	Butterfly	Cat	Cow	Crocodile	Deer	Dog
BCF	87.6%	92.2%	73.8%	77.4%	76.8%	90.4%	82.6%
CST	95.2%	99.6%	91.0%	88.8%	83.8%	94.4%	98.4%
Method	**Dolphin**	**Duck**	**Elephant**	**Fish**	**Flyingbird**	**Hen**	**Horse**
BCF	89.0%	87.0%	95.2%	79.8%	72.0%	94.2%	95.4%
CST	97.6%	97.2%	96.6%	90.4%	89.0%	96.0%	99.4%
Method	**Leopard**	**Monkey**	**Rabbit**	**Rat**	**Spider**	**Tortoise**	
BCF	66.4%	58.4%	85.8%	70.6%	99.2%	93.6%	
CST	85.2%	92.6%	96.4%	90.4%	99.8%	97.2%	

MPEG7 (70 classes of 20 instances)



Method	Accuracy (half testing)	Accuracy (leave one out)
CSS	90.9%	97.93%
CS	91.1%	-
SP	86.7%	-
ICS	96.6%	-
PMR	-	97.57%
SoS	-	97.36%
RS	-	98.57%
KD	-	98.93%
BCF	97.03%	98.86%
BCF+BSP	98.35%	-
CST	**99.27%**	**99.71%**

Swedish leaf (15 classes of 75 instances)



Method	Accuracy
MAC	82.00%
Fourier	89.60%
CS + DP	88.12%
IDSC + DP	94.13%
MDM	93.60%
IDSC + MS	94.80%
RS	95.47%
Shape Tree	96.28%
BCF	97.52%
CST	**99.25%**

Conclusion

- Skeleton-geodesic histograms are informative characteristic of the shape that is resistant to changes in a flexible shape and thickness changes.
- They complement contour description better than other known features extracted from the skeleton, allowing to achieve results that outperform all existing counterparts.
- Growth in the classification quality is achieved by adding a relatively small number of features (some hundreds against tens of thousands).
- Efficient computation of histograms is made possible by the continuous mathematical morphology methods.
- The main drawback of the method is instability of the skeleton to accident holes occurring in the figure.

Thank you for attention!