Recognition of digital fonts from images based on the disk cover

Mestetskiy Leonid (mestlm@mail.ru)
Lomov Nikita (*nikita-lomov@mail.ru*)

Moscow State University
Computational Mathematics and Cybernetics

A lot of fonts

- The total number of Latin fonts - a few tens of thousands, and Cyrillic - twenty times less.

Very wide variety of fonts

Fonts recognition problem

- Font Copyright Holder wants to control its use

Fonts comparing problem

Designers want to choose a font similar to the sample available

Wanted a measure of the similarities and differences for digital fonts

Requirements:

- The adequacy determining similarities and differences

- Universality for a variety of types, sizes, character sets

- The high computational efficiency for processing thousands of fonts

A common problem - a comparison of shapes of objects
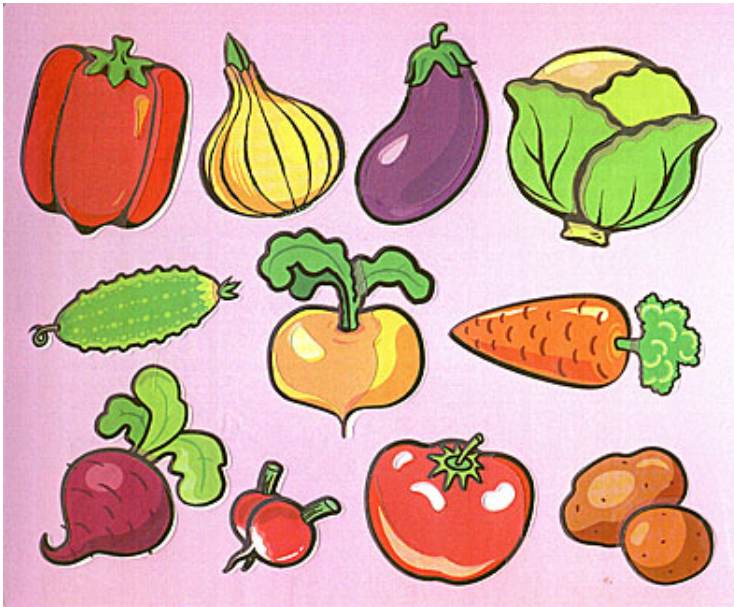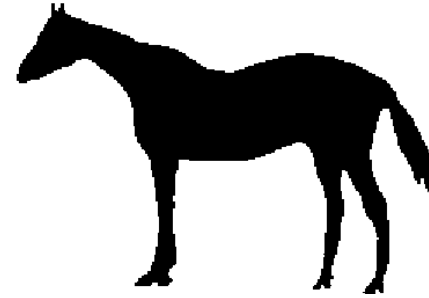
The shape - an external outline, external appearance of the object.

The basic idea - to evaluate the area of the disk cover of the object

Morphological width

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Covering of the object by inscribed disks of different sizes
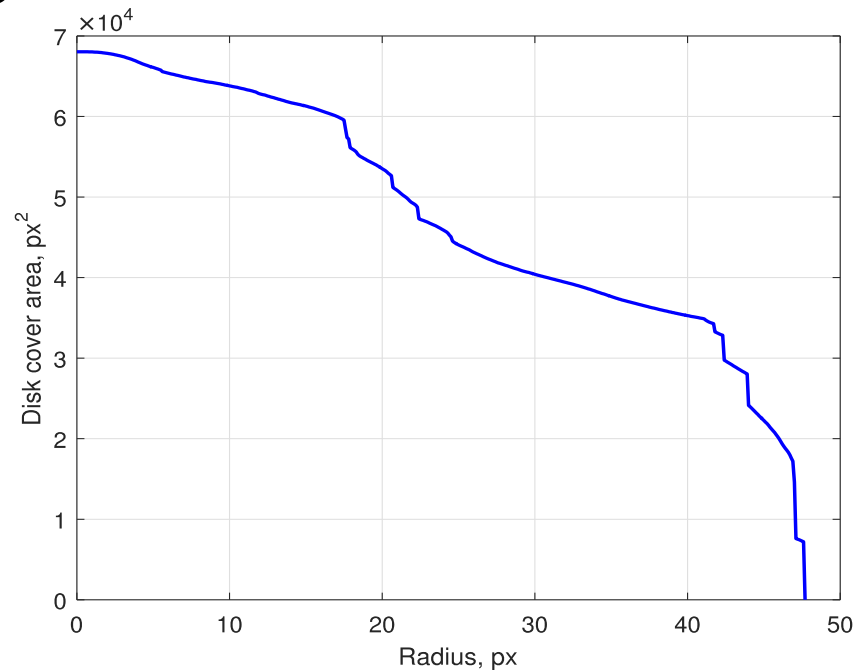
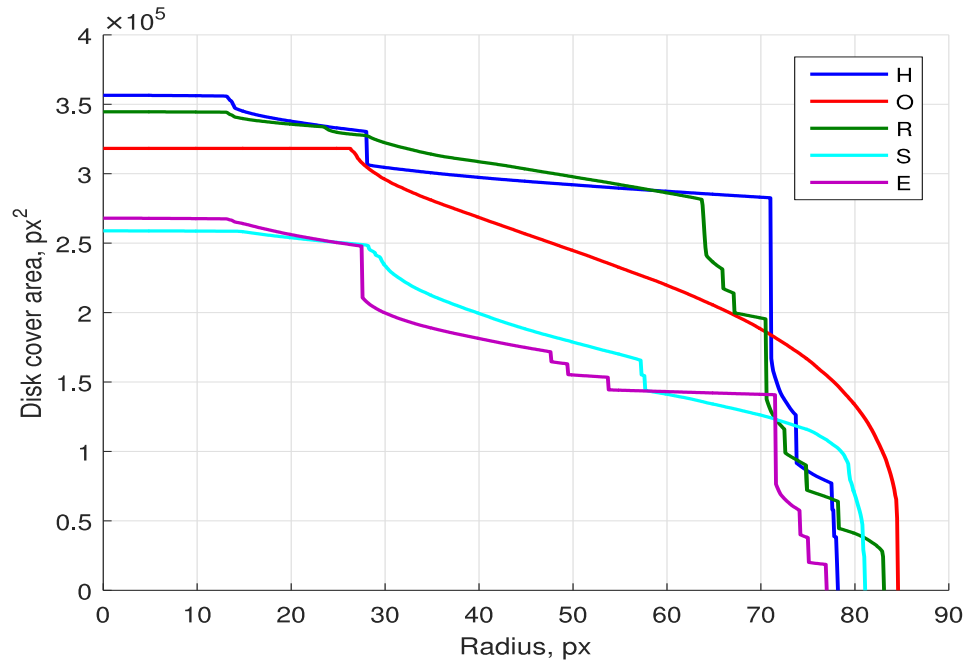Diagram of disk cover area depending on the size of covering disc

Region of predetermined width
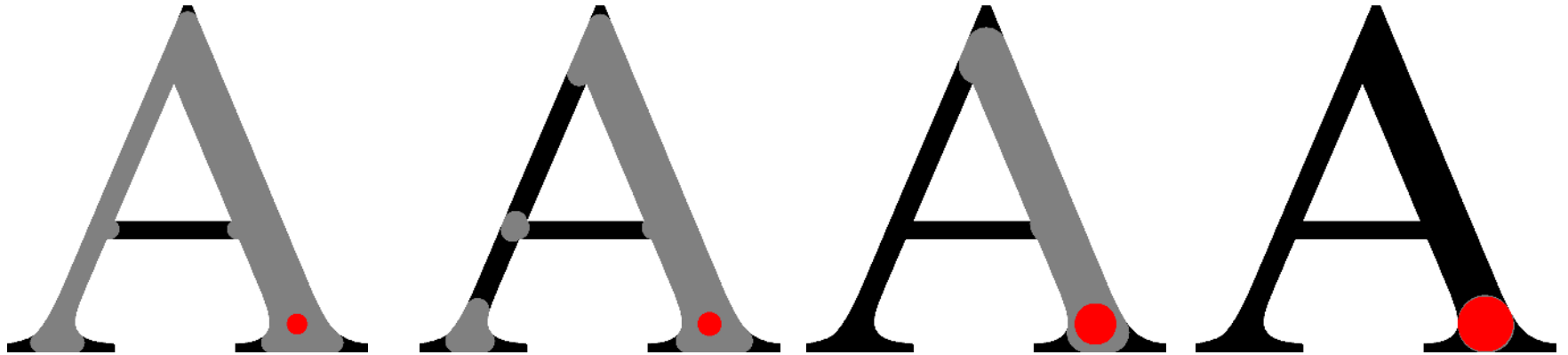
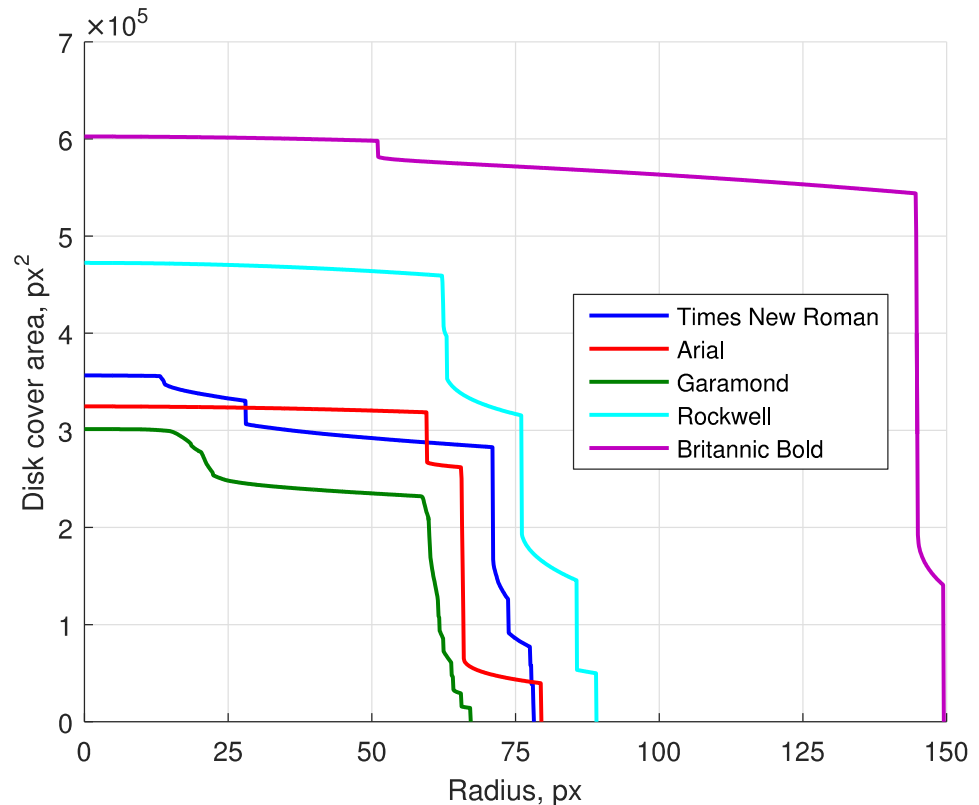# Morphological width of the image is an informative feature
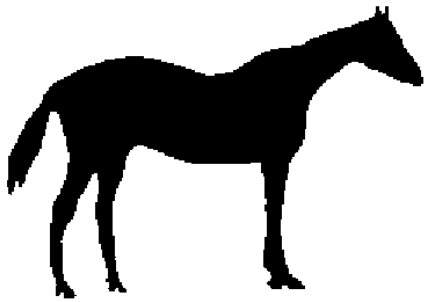
HORSE

Morphological width of different characters of the same font

Morphological width of fonts is the basis for comparison

Morphological width differentiates between fonts

Font similarity measure based on morphological width

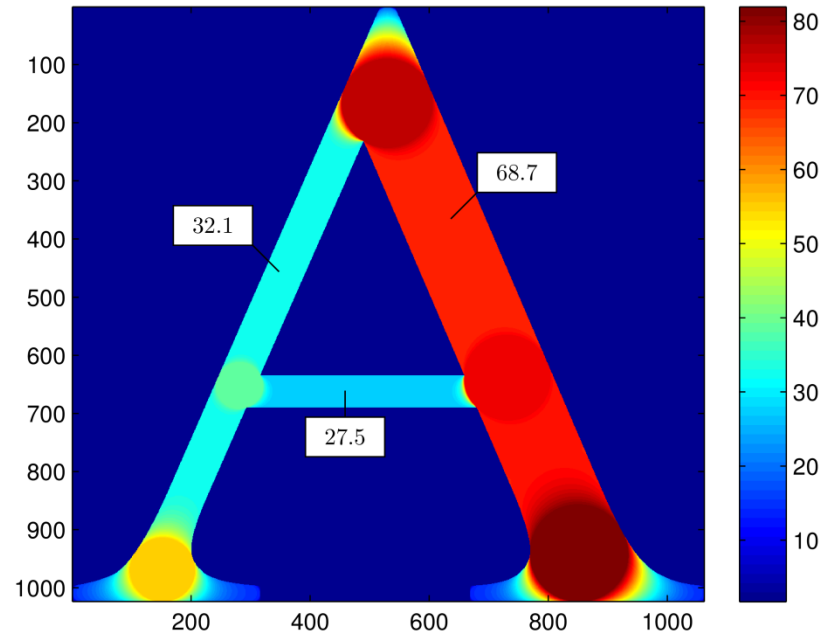$$d(\mathbf{p}^k, \mathbf{q}) = \sum_{j=1}^{n} \alpha_j d_1(p_{c_j}^k, q_j).$$

c_j – код j-го символа шрифта

$p_{c_j}^k$ – эталонный дескриптор j-го символа шрифта $\mathbf{p}^k$

$\mathbf{p}^k$ – шрифт

$\mathbf{q} = \{q_1, \ldots q_n\}$ – символы шрифта
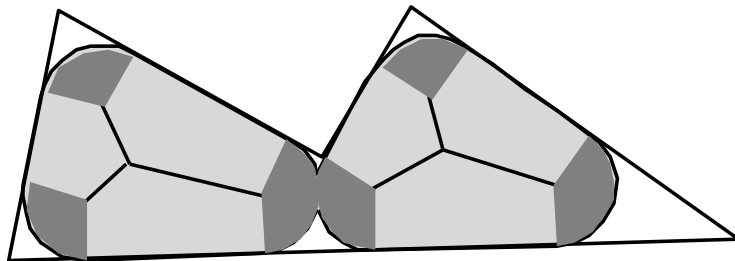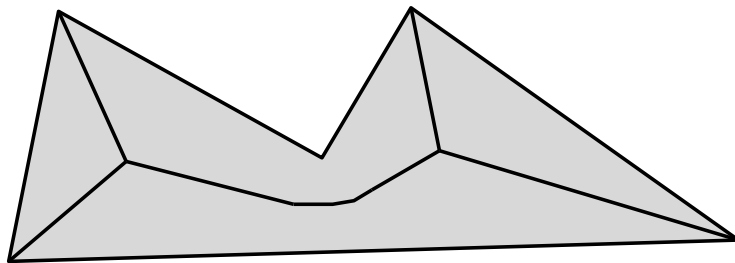
The problem of calculating of the disk cover area

The main challenge - the development of efficient and accurate algorithm

The idea of the proposed approach

Binary image

Approximating polygonal figure

Skeleton of the polygonal figure
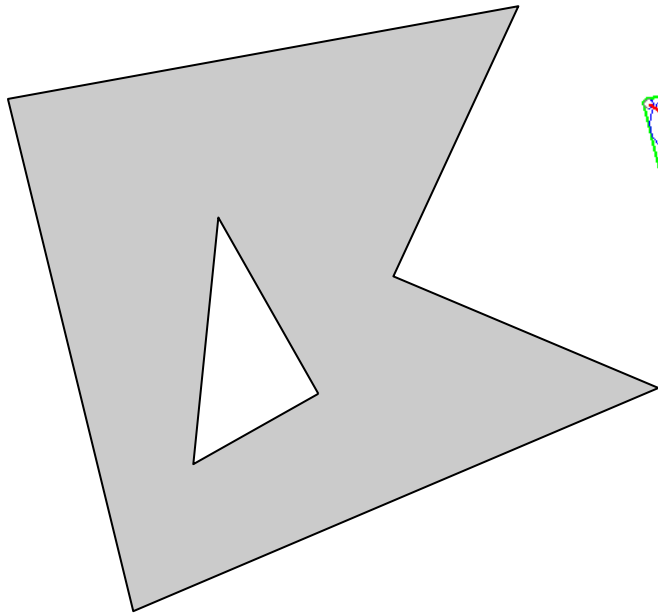
A disc cover
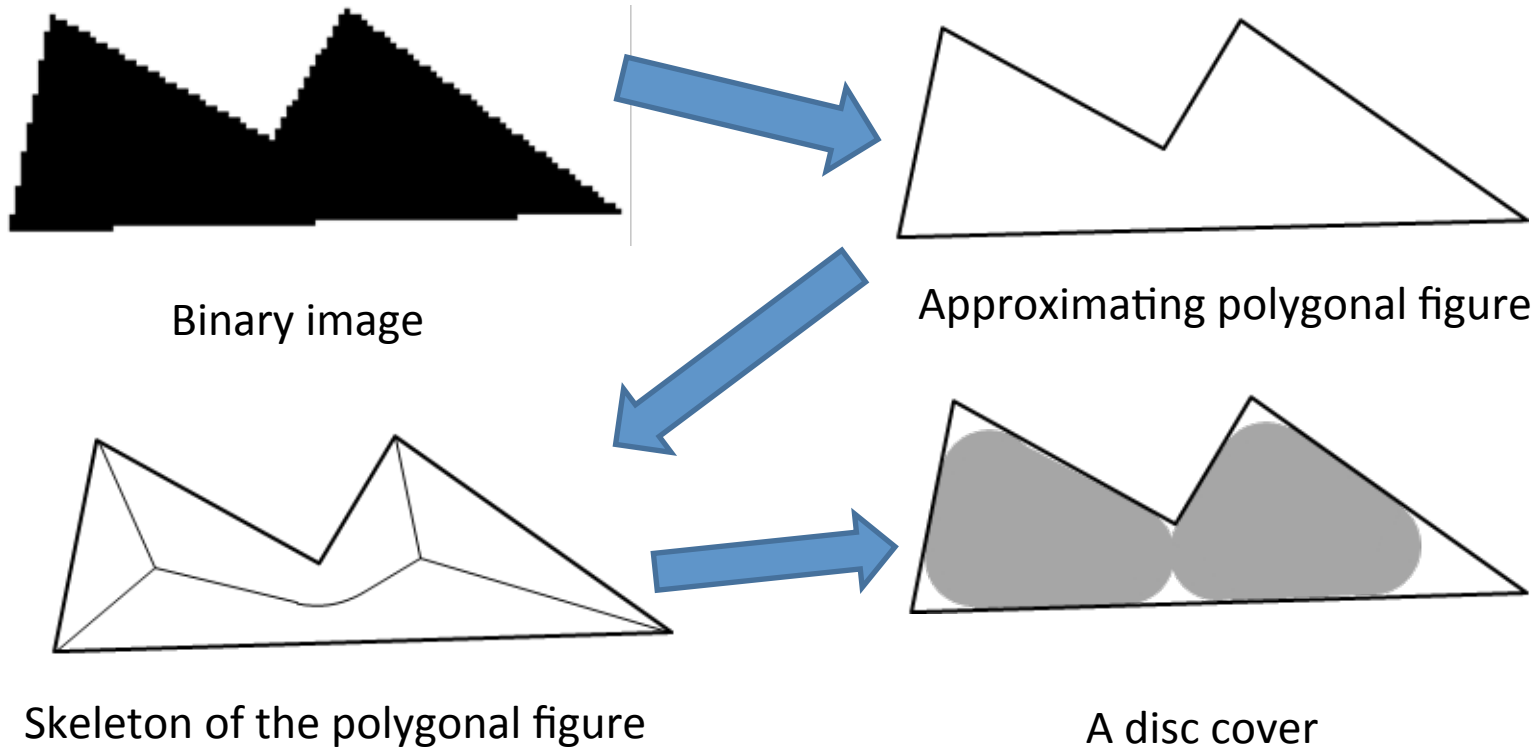
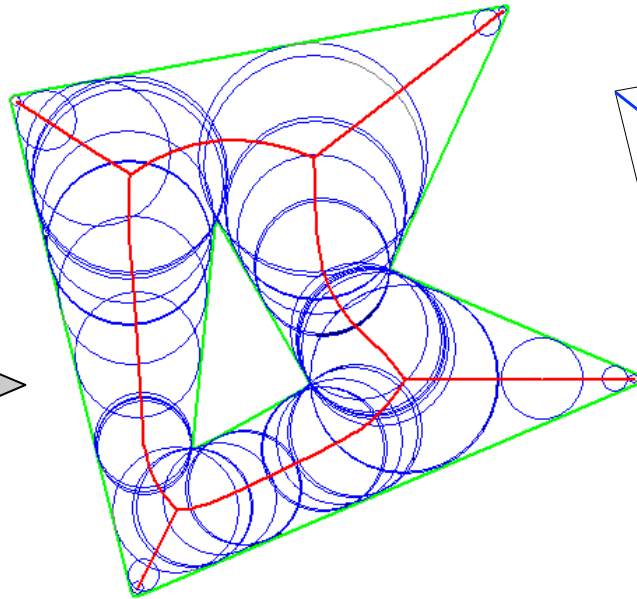## Continuous model of the disk cover for binary image

Disk cover of polygonal figure
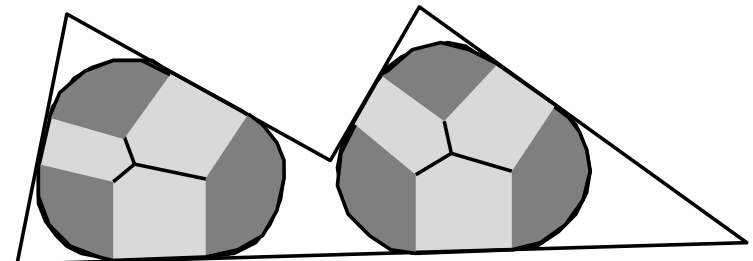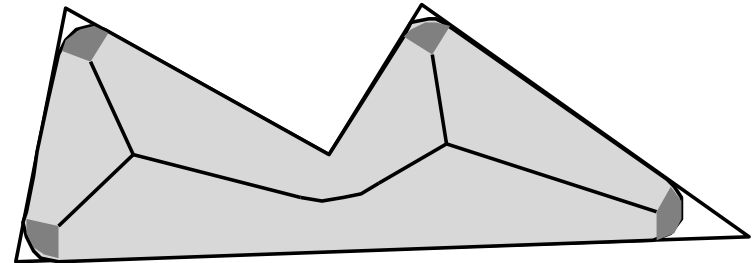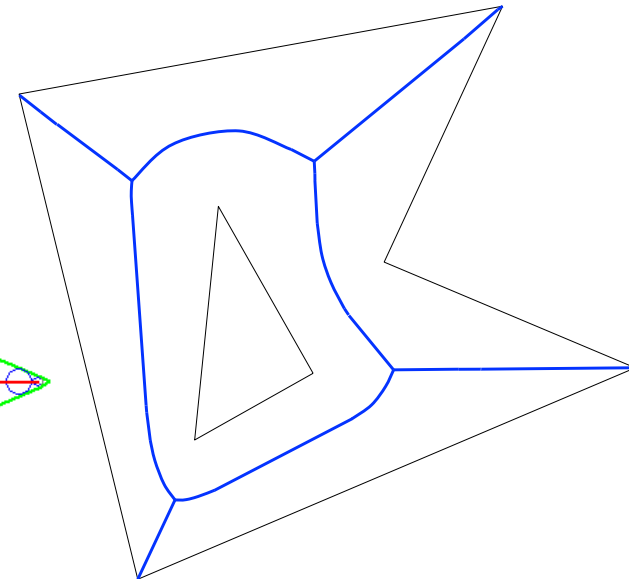

Changing of disk cover with increasing of disk size

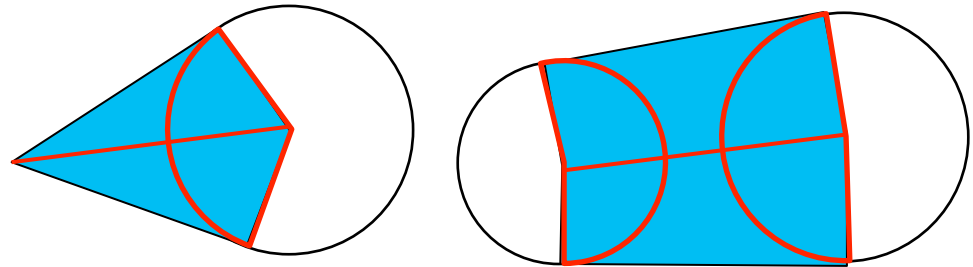# Continuous medial representation of the polygonal figure

Polygonal figure

Medial representation
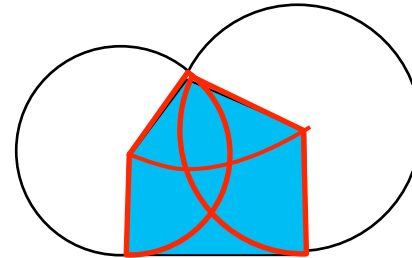(medial axis & radial function)
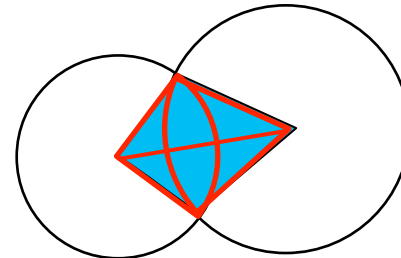
Skeleton

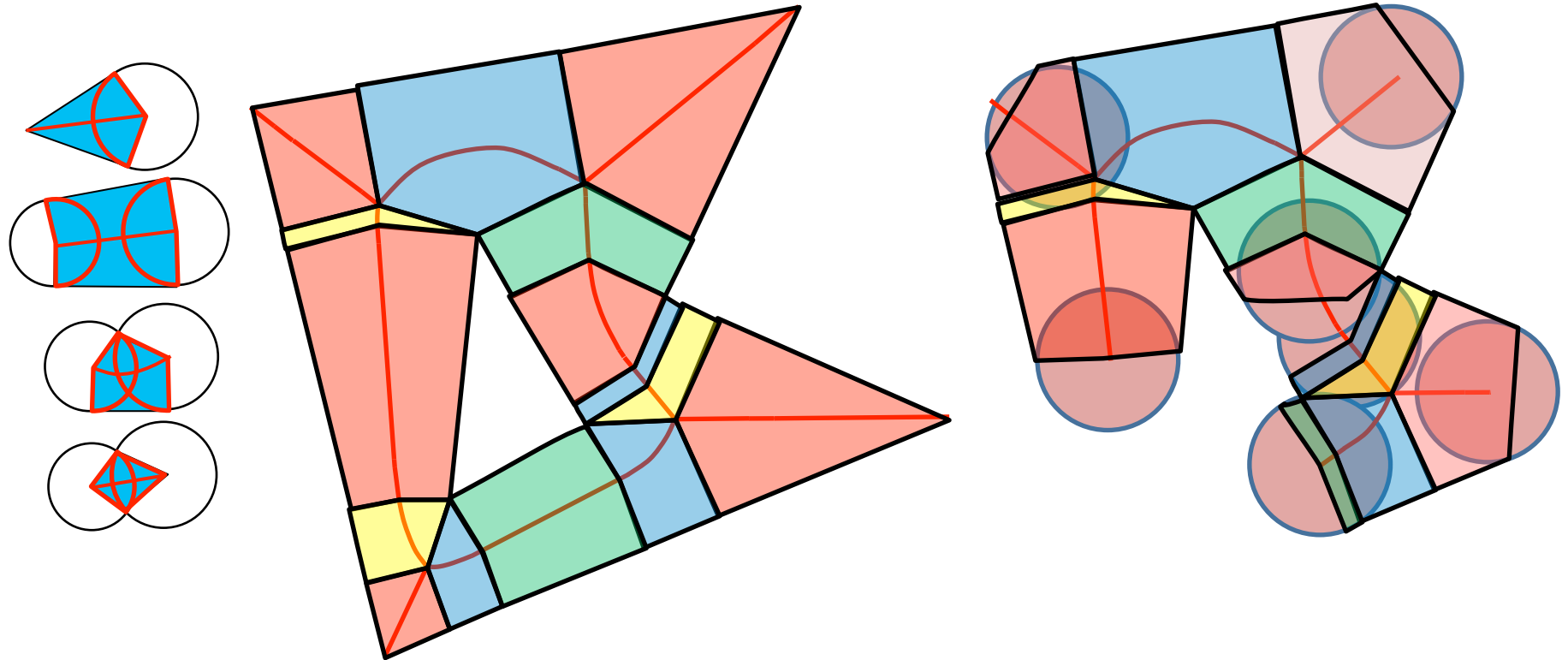# Polygonal figures and bicircles

Lineare bicircles

Parabolic bicircles

Hyperbolic bicircles

A bicircle is the union of all inscribed circles
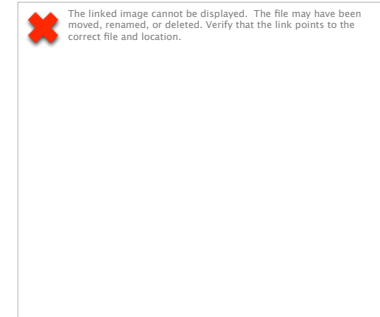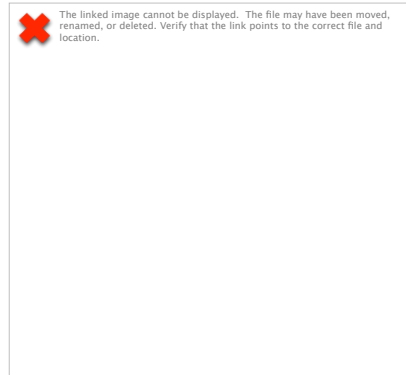centered on one edge of the skeleton.

Coverage of the figure by bicircles

All of the bicircles

Bicircles of given size

Proper region and external sector

r , R – the radii of the smaller and larger end circles,

l – distance between centers of end circles,

t – the length of the projection of the bicircle axis to the segment-site,

S – area of the proper region of bicircle,

φ – the angular size of the external sector of bicircle.

The calculation of the bicircle parameters

$$S_{lin} = 2 \cdot \frac{R+r}{2} \cdot t = (R+r) \cdot t$$

$$\varphi_{lin} = 2\alpha = 2 \cdot arcsin\frac{t}{l}$$

} Linear bicircle

$$S_{par} = \frac{R+r}{2} \cdot t + \sqrt{P(P-R)(P-r)(P-l)}$$

$$\varphi_{par} = \frac{\pi}{2} + \alpha = \frac{\pi}{2} + arcsin\frac{p-r}{r}$$

} Parabolic bicircle

$$S_{hyp} = 2\sqrt{P(P-R)(P-r)(P-l)}$$

$$\varphi_{hyp} = 2\,arcsin\frac{q}{2r}$$

} Hyperbolic bicircle

$$P = (R+r+l)/2$$

Bicircle overlapping problem
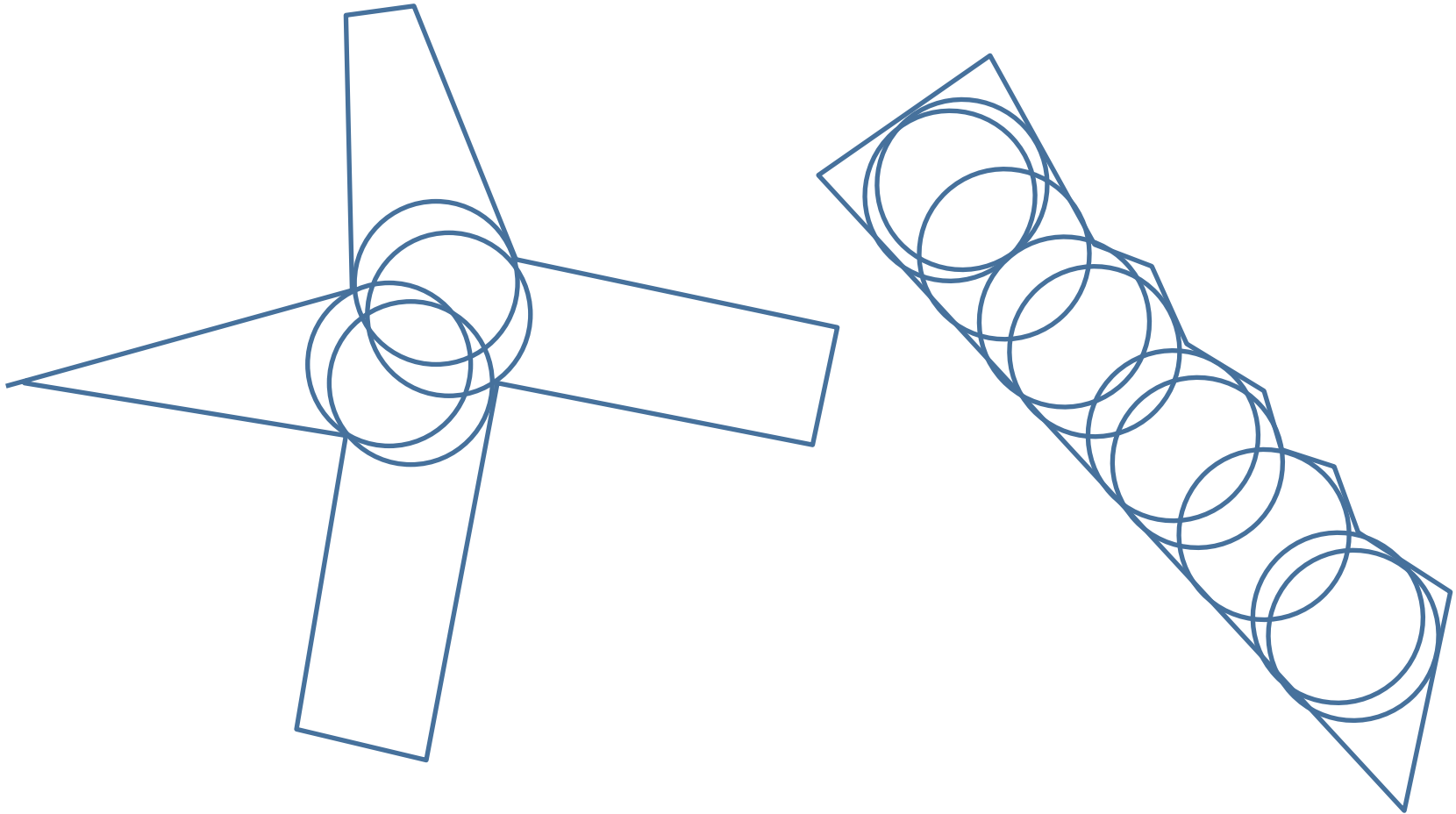
Overlapping of bicircles
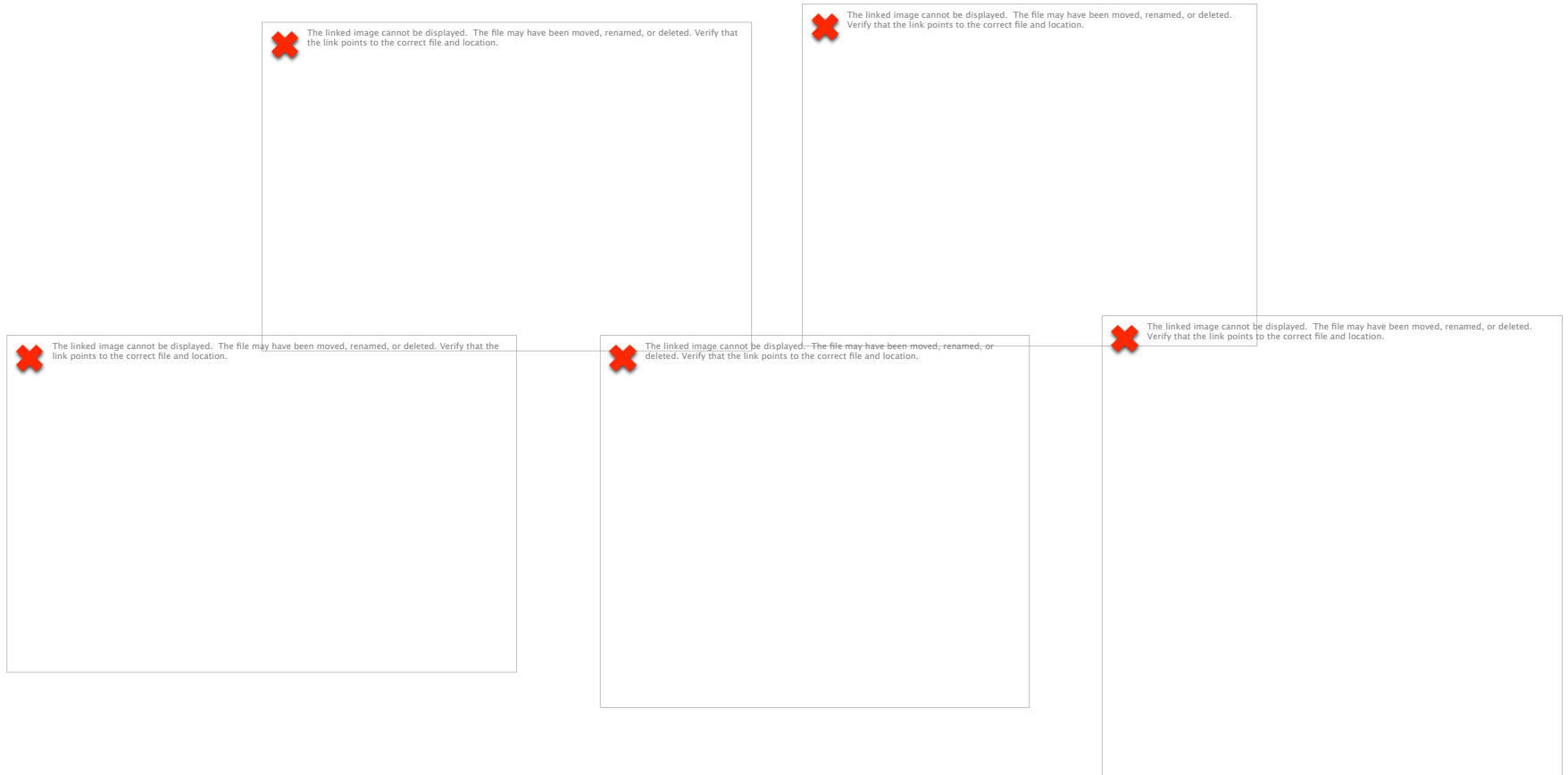
Theorem. *When we calculate the area of disk cover only pair crossing bicyclic should be considered* (!!!)

Font Library of ParaType company

Библиотека шрифтов
ParaType
PT Addendum on CD (282 шрифта)

Windows
Macintosh
PostScript

Total 1884 fonts

Reference histograms - etalons

1884 fonts

52 glyphs (symbols)

5 font size (1000, 100, 50, 40, 30)

$1884 \times 52 \times 5 = 489840$ descriptors (4.2 hour, 11 min – polygon+skeleton+width)

$1884 \times 1885 \times 52 = 184\,571\,712$ distances font-font (70 min – 2.25 sec/request)

Font Height	Running time, min	Time per glyph, sec
1000	192	0.118
100	25	0.015
50	14	0.009
40	12	0.007
30	9	0.006

Experiment results

Р а з м е р	1	3	5	7
100	96.23%	98.89%	99.26%	99.36%
50	92.20%	97.45%	98.14%	98.41%
40	74.10%	87.10%	89.17%	90.71%
30	52.65%	69.06%	74.73%	77.07%

The percentage of hits in the number of nearest neighbors
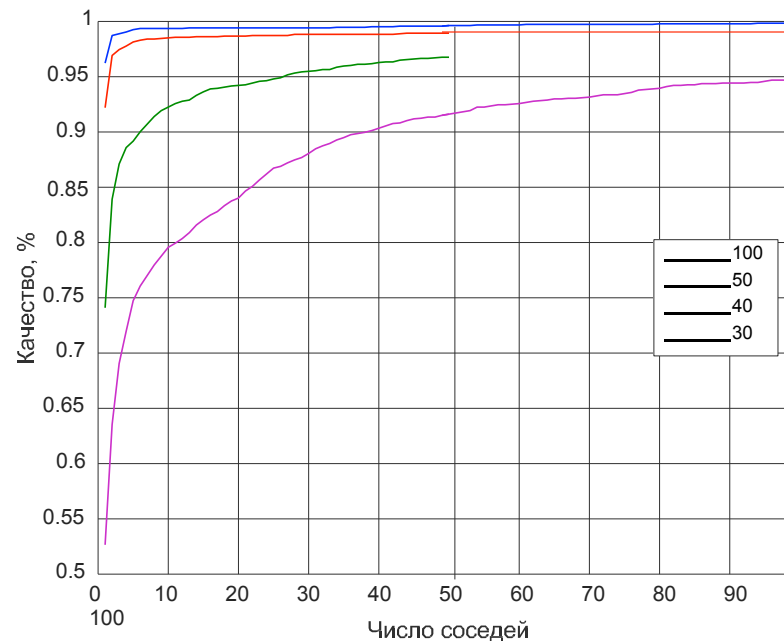
Conclusion

- The proposed descriptor and the method of its calculation open up new possibilities for the use of highly efficient computational geometry algorithms in image analysis and shape recognition.

- The proposed continuous model of width of polygonal figures on the basis of the disk cover allowed to make the decomposition of the original problem and reduce the computation to simple geometric calculations.

- The developed algorithm is the first to provide accurate analytical representation of the width distribution function of a polygonal figure.

- The high efficiency of the proposed method allows to compare and measure the similarity of figures by their width in real-time systems.

- The method successfully solves the problem of individual identification of the font on the real images of sufficient quality.

Full description of the algorithm

Ломов Н.А., Местецкий Л.М.

Площадь дискового покрытия – дескриптор формы изображения

в журнале *Компьютерная оптика*, том 40, № 4, с. 516-525