

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

*Aa Bb Cc Dd Ee Ff
Gg Hh Ii Jj Kk Ll
Mm Nn Oo Pp Qq
Rr Ss Tt Uu Vv Ww
Xx Yy Zz*

*ABCDEFGHIJKLM
NOPQRSTUVWXYZ
abcdefghijklm
nopqrstuvwxyz*

*abcdefghijkl
klmnopqr
stuvwxyz*

**ABCDEFGHI
JKLMNOPQ
RSTUVWXY
Z
1234567890**

**AaBbCcDdEeFf
GgHhIiJjKkLlMm
NnOoPpQqRrSsTt
UuVvWwXxYyZz
1234567890@%&!?**

**ABCDEF
GHIJKLMN
OPQRSTU
vwxyz**

FreeFonts.Xyz

GHOULISH FRIGHT

**ABCDEFGHIJ
KLMNOPQR
STUVWXYZ
ABCDEFGHIJKLM
NOPQRSTUVWXYZ
0123456789**

**ABCDEF
GHIJKLMN
OPQRSTU
vwxyz**

FreeFonts.Xyz

abcdefghijkl
mnopqr
stuvwxyz
hamburgerfont

**ABCDEFGHI
JKLMNOPQR
STUVWXYZ**

**ABCDEF
GHIJK
LMNOP
QRSTU
vwxyz**

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

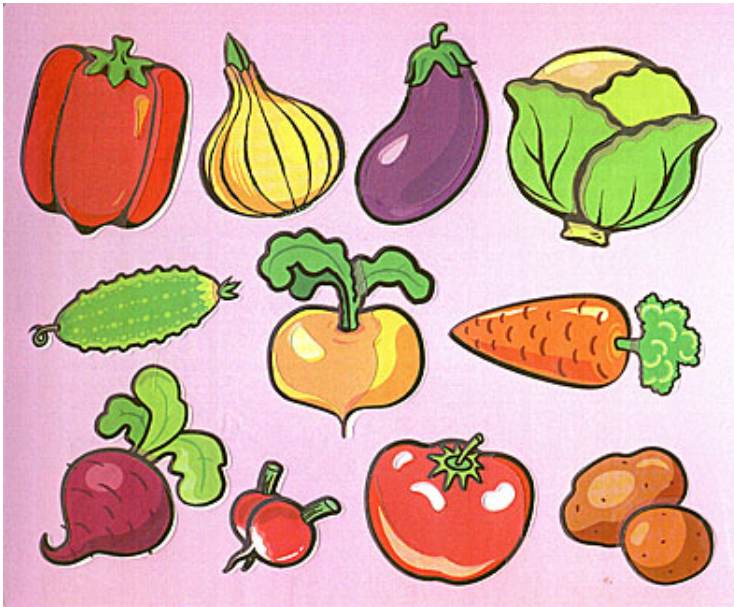
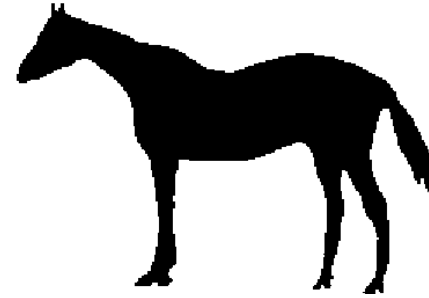


Wanted a measure of the similarities and differences for digital fonts

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



Covering of the object by inscribed disks of different sizes

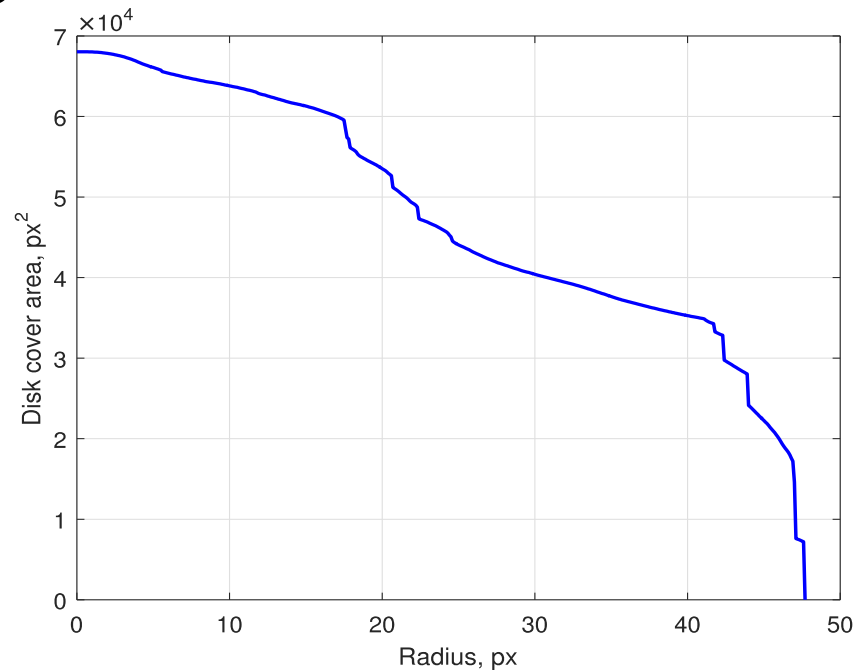
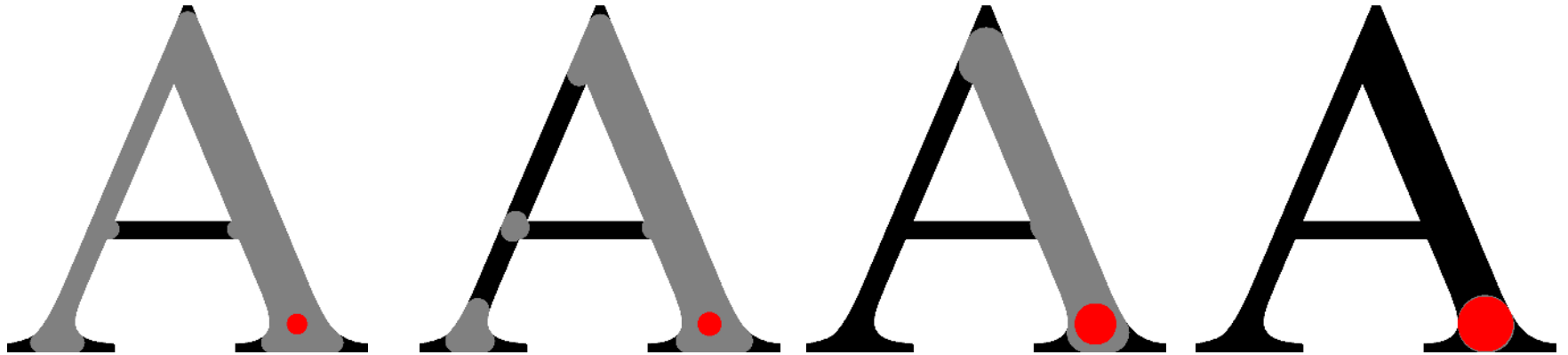
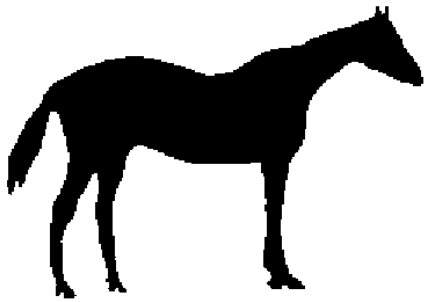


Diagram of disk cover area depending on the size of covering disc

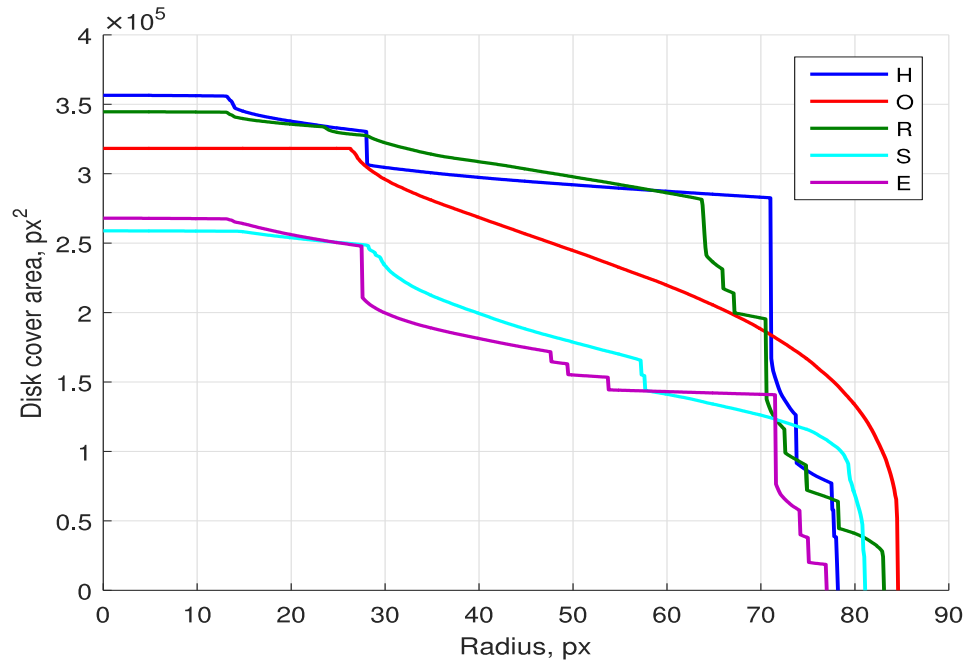
Region of predetermined width



Morphological width of the image is an informative feature



HORSE

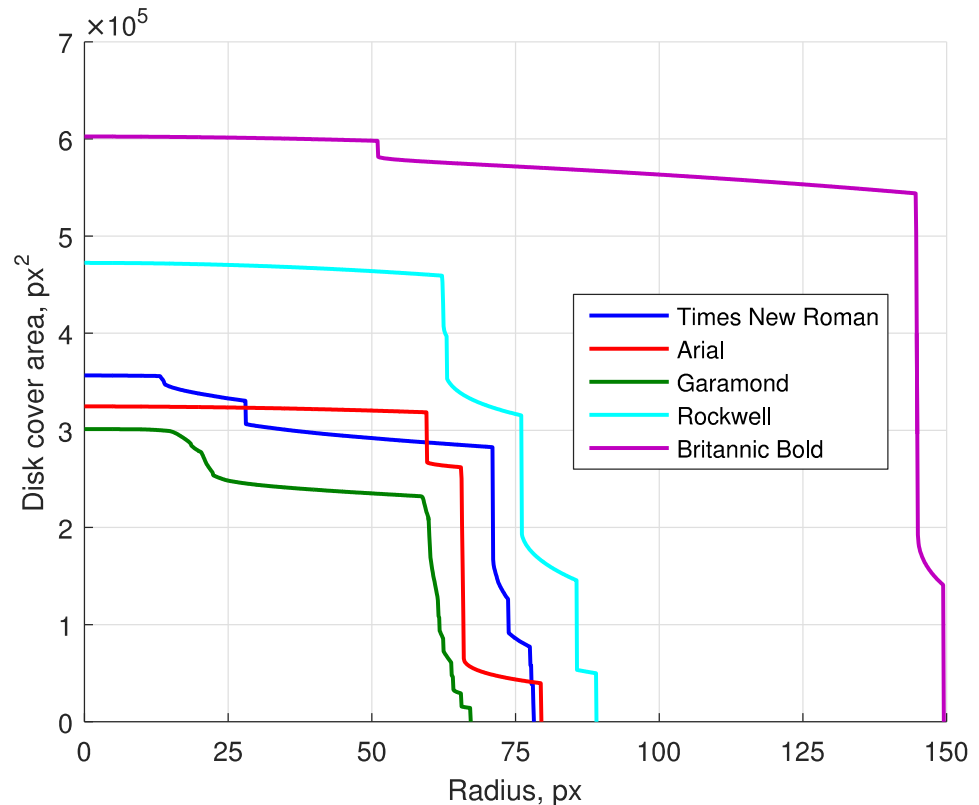


Morphological width of different characters of the same font

Morphological width of fonts is the basis for comparison

HORSE
HORSE
HORSE
HORSE
HORSE

- Times New Roman
- Arial
- Garamond
- Britannic Bold
- Rockwell



Morphological width differentiates between fonts

Font similarity measure based on morphological width

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

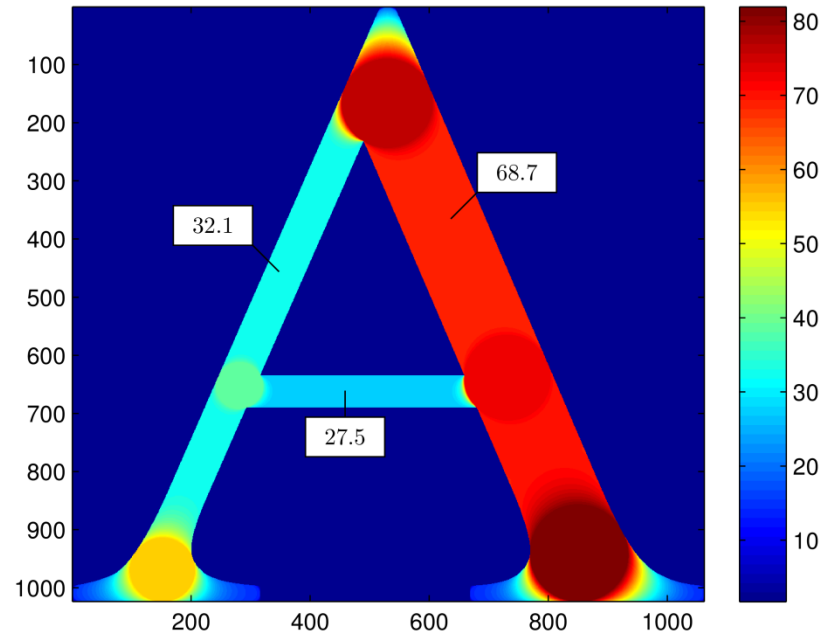
c_j – код j -го символа шрифта

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

\mathbf{p}^k – шрифт

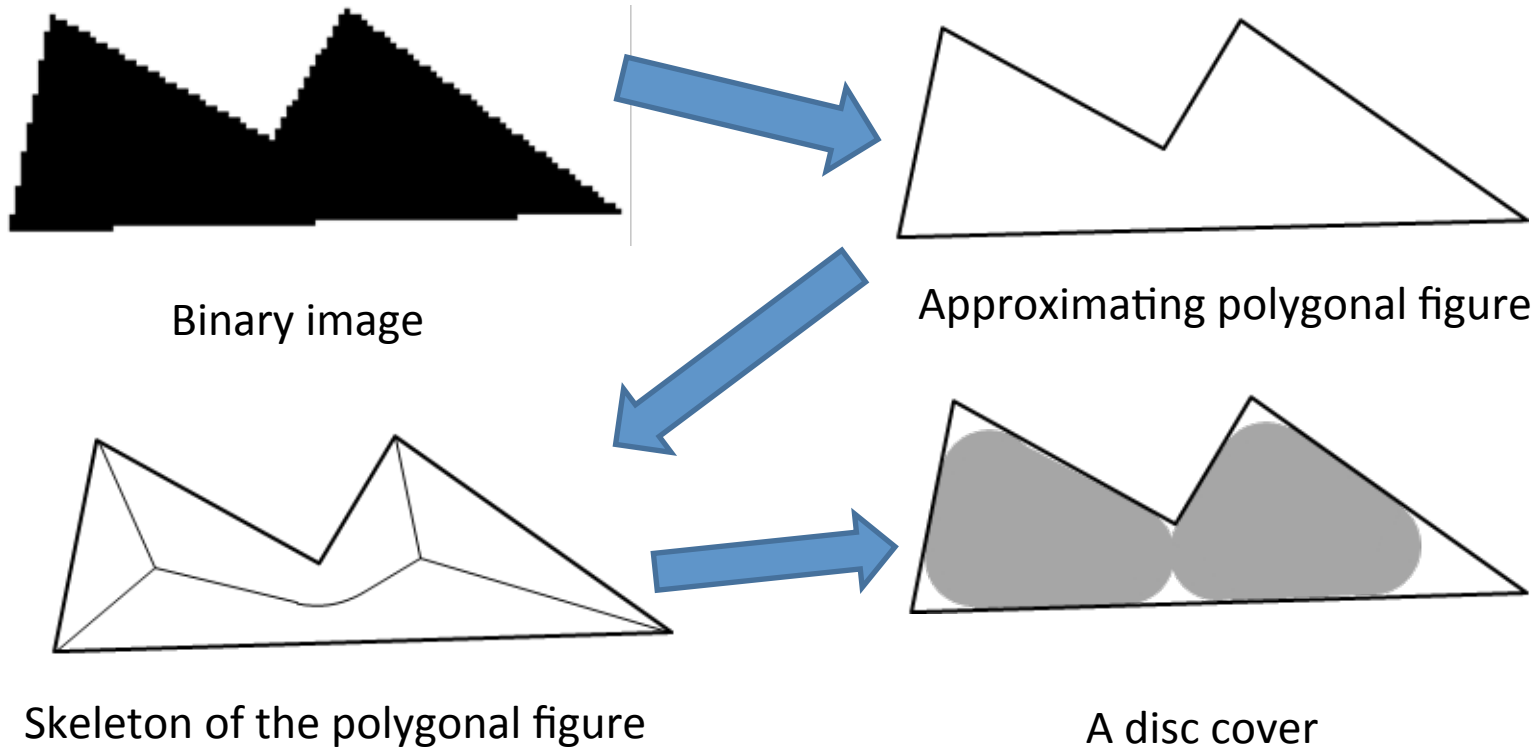
$\mathbf{q} = \{q_1, \dots, q_n\}$ – символы шрифта

The problem of calculating of the disk cover area



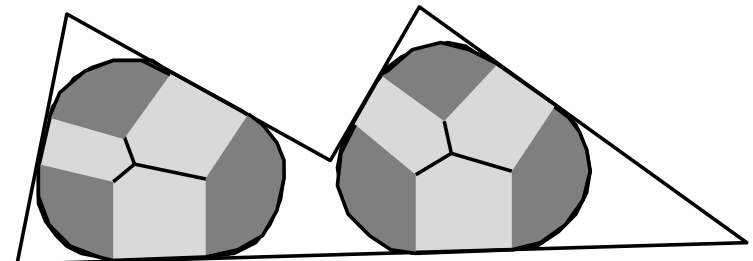
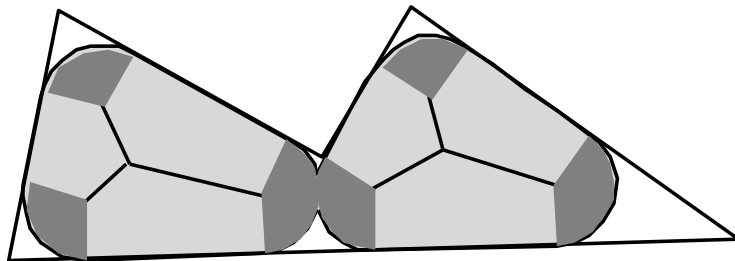
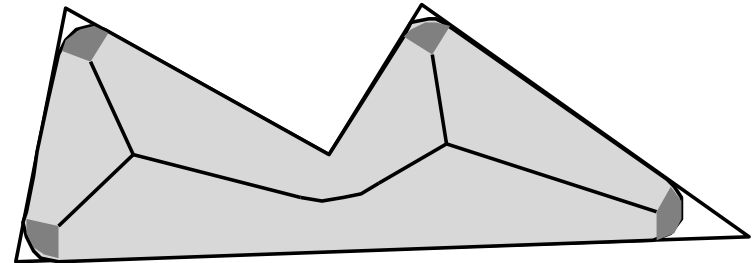
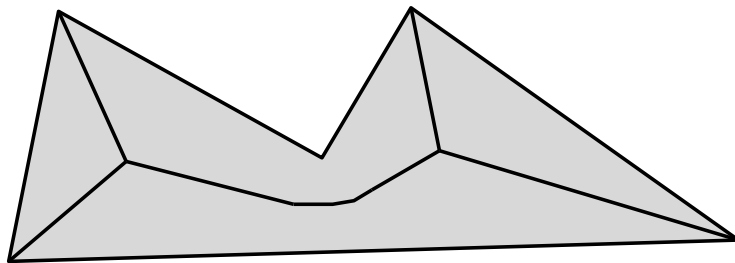
The main challenge - the development of efficient and accurate algorithm

The idea of the proposed approach



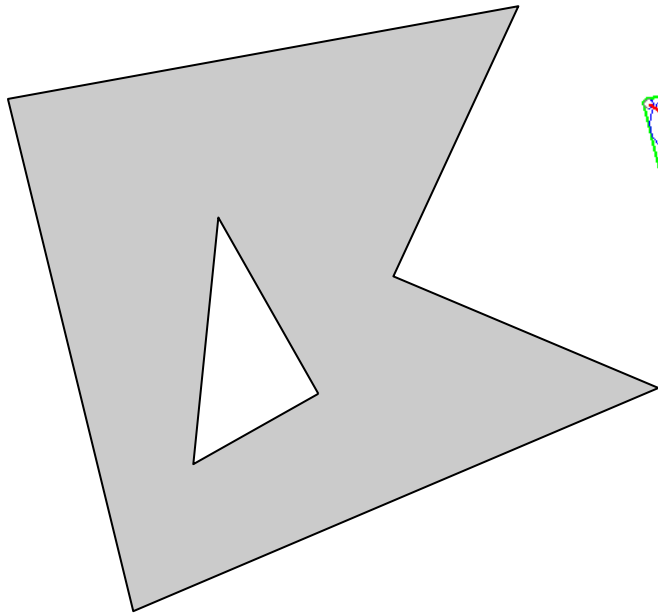
Continuous model of the disk cover for binary image

Disk cover of polygonal figure

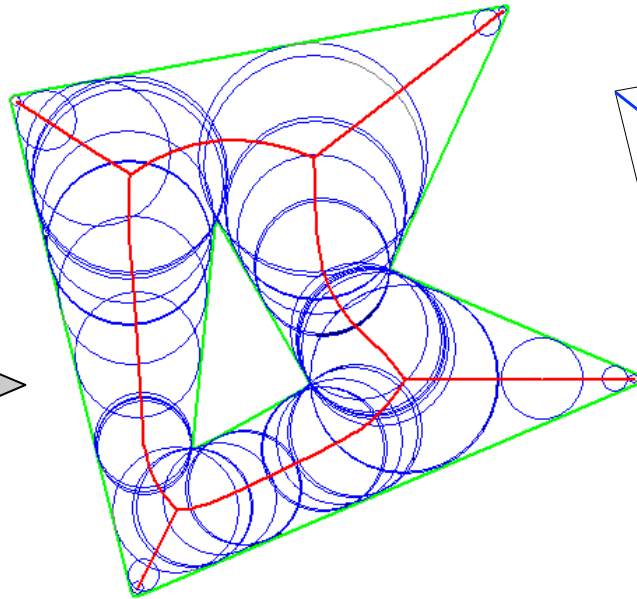


Changing of disk cover with increasing of disk size

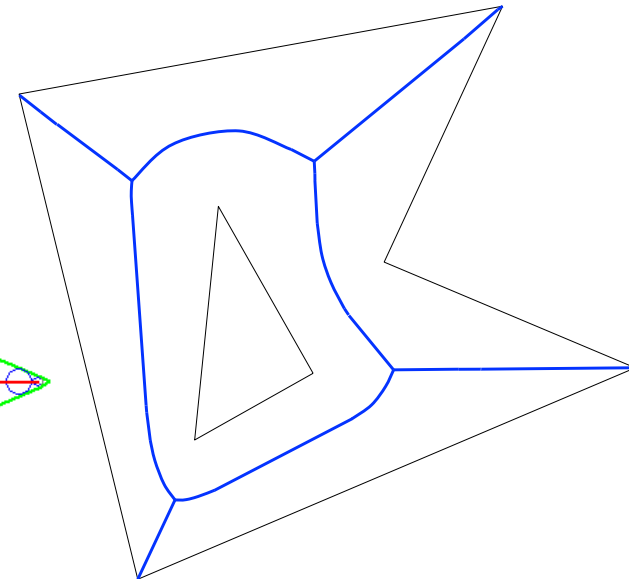
Continuous medial representation of the polygonal figure



Polygonal figure



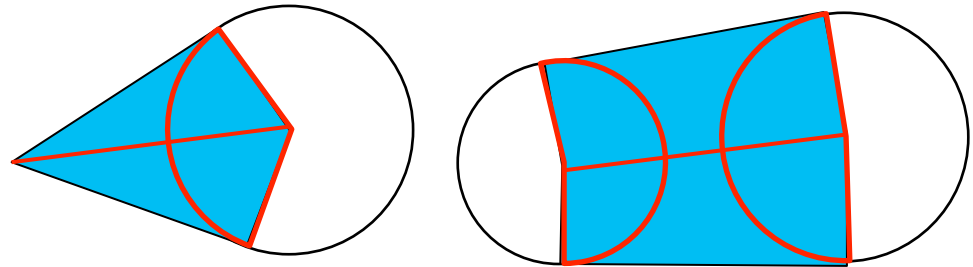
Medial representation
(medial axis & radial function)



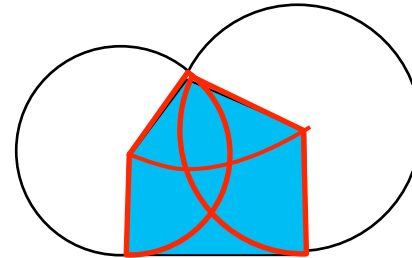
Skeleton

Polygonal figures and bicircles

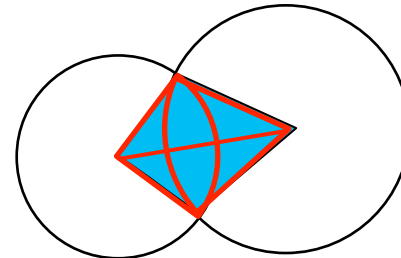
Lineare bicircles



Parabolic bicircles

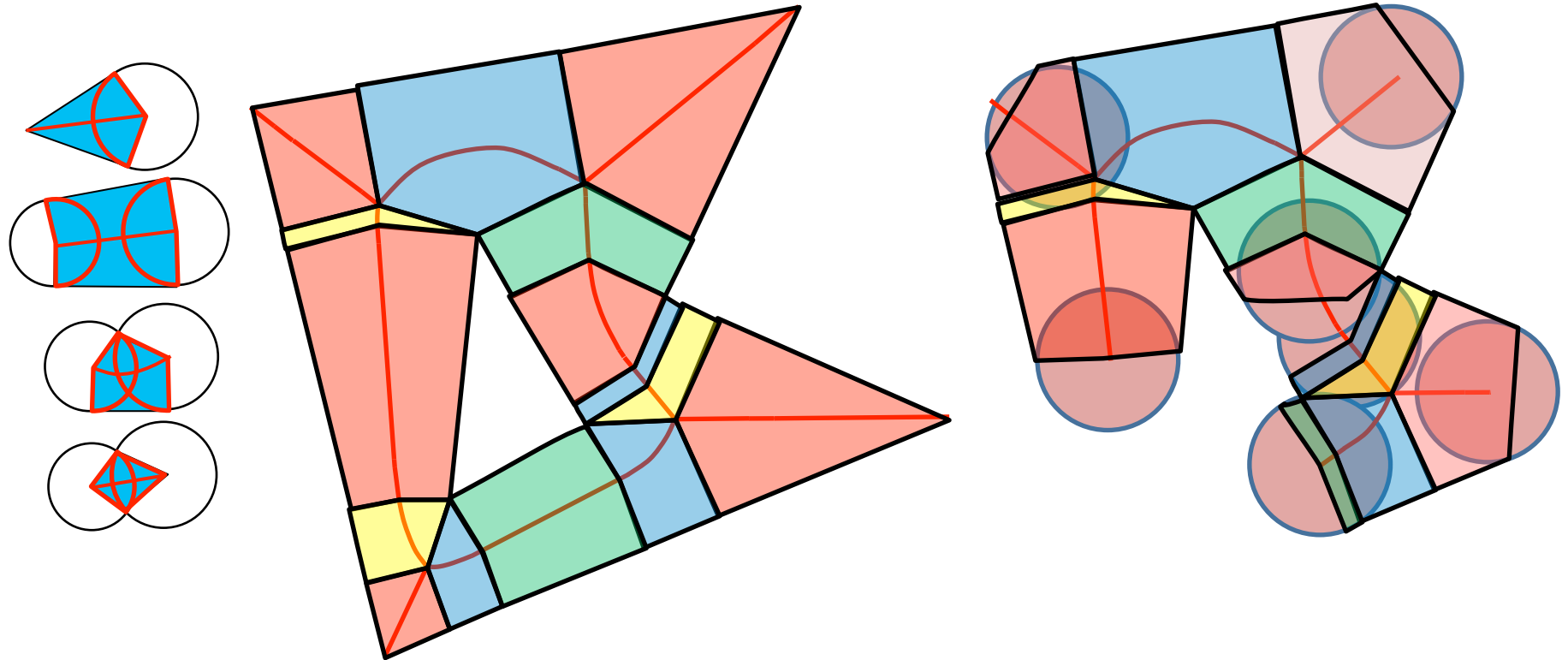


Hyperbolic bicircles



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

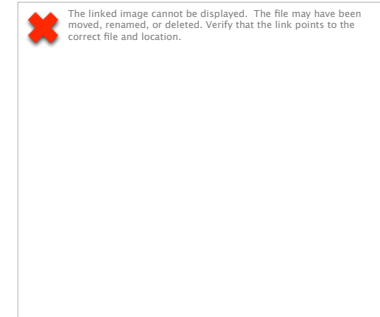
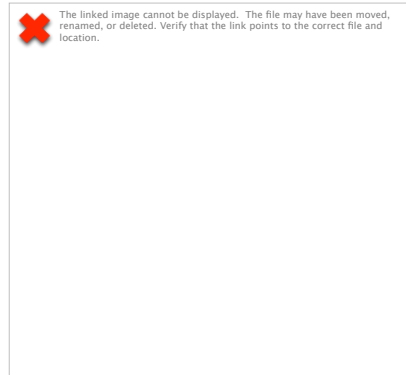
Coverage of the figure by bicircles



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

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

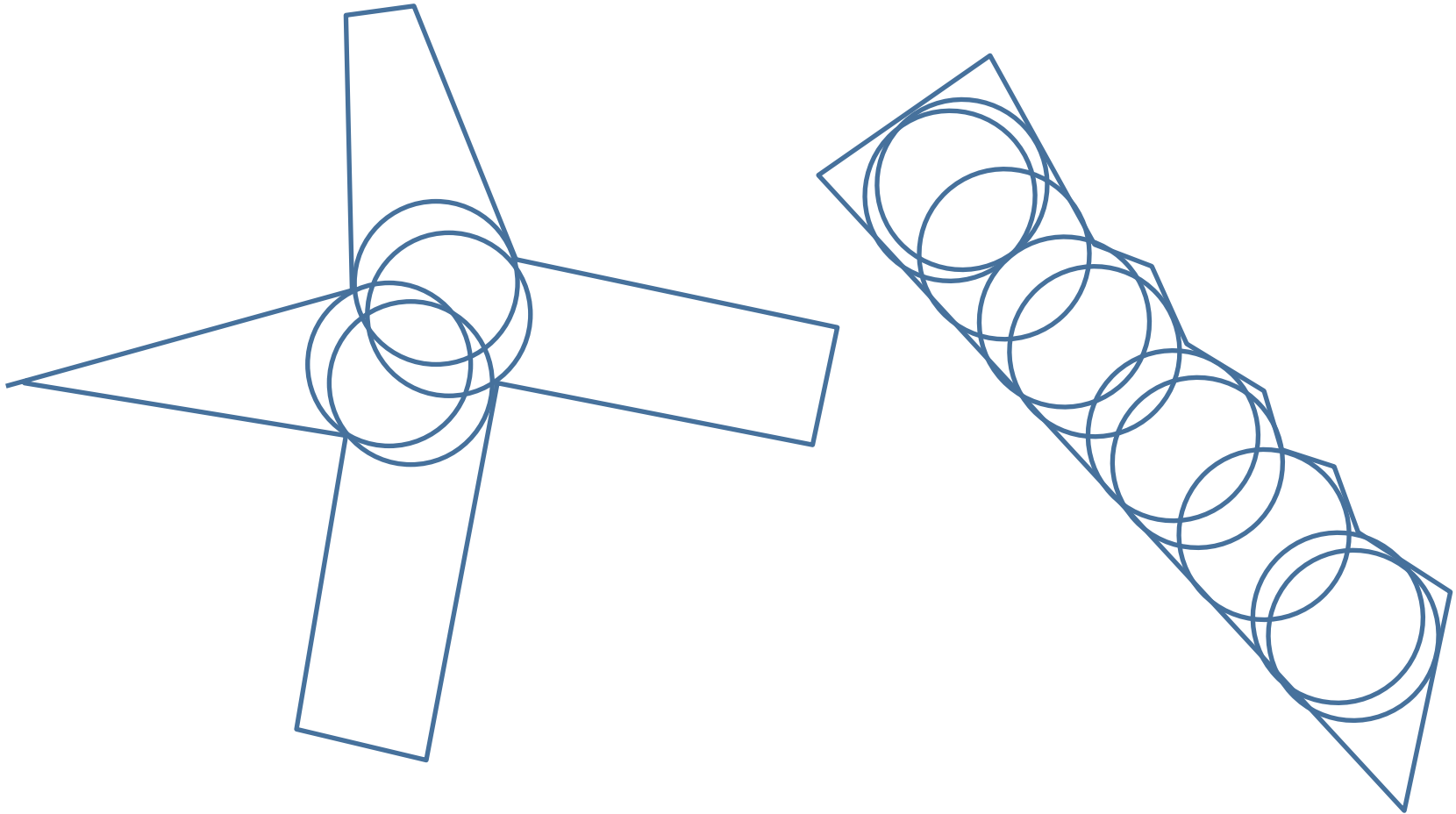
$$\varphi_{hyp} = 2\arcsin \frac{q}{2r}$$

}

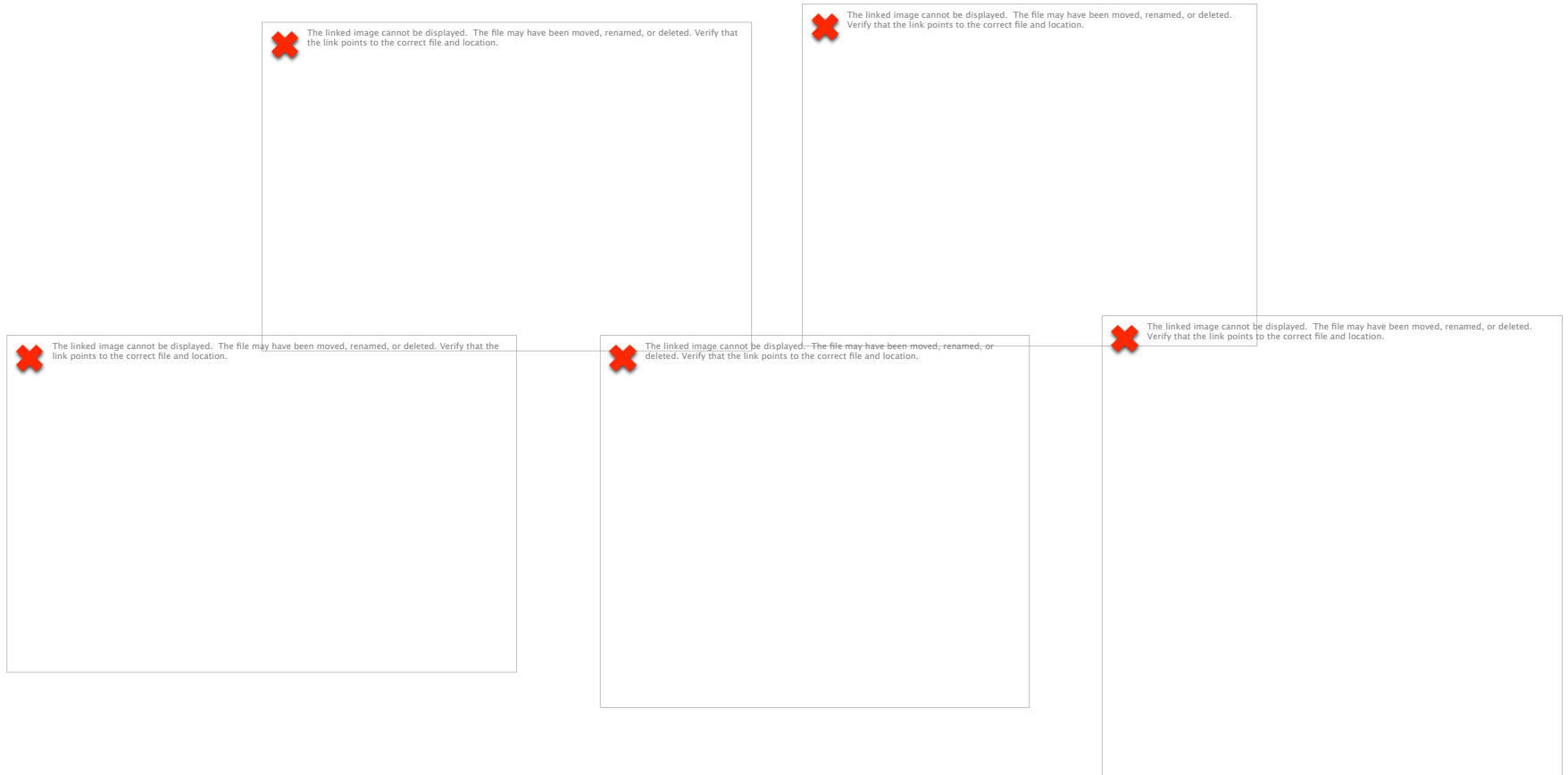
Hyperbolic bicircle

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

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
PT Addendum on CD (282 шрифта)

ParaType

■ ACSIOMA	DA ACS1	■ Coventry	DA CVN	■ Гармонд	DA GARMOND
■ ACSIOMA NEXT	DA ACS4	■ Cricket Light	TM CRC46	■ Гармонд Итальян	DA GARMOND ITA
■ Flksent	TM AKS	■ Cricket Book	TM CRC56	■ Гармонд Седм	DA GARMOND SEMI
■ Alliance Regular	DA ALL58	■ Cricket Bold	TM CRC76	■ SPSL Garland	SP GR1
■ Alliance Italic	DA ALL56	■ Cricket Heavy	TM CRC76	■ GLASTIEN	TM GLT
■ Alliance Bold	DA ALL76	■ Debby	SU D88	■ GRANIT	TM GRN
■ Alliance Bold Italic	DA ALL76	■ Dikovina	TM DK048	■ Gymnasia	TM GYN
■ Anastasia Script	TM ANS	■ Dikovina B	TM DK048	■ SIS ROM Light	TM HRV
■ ANGELICA NORMAL	DA ANG55	■ Dots	SU D75	■ Heinrich	TM HNR
■ ANGELICA Italic	DA ANG56	■ Drop	DA DRP	■ Helios Thin	TM HLS36
■ Apical Light	DA AP148	■ Dynar Medium	SU DNR93	■ Helios Thin Oblique	TM HLS36
■ Apical Roman	DA AP169	■ Dynar Bold	SU DNR93	■ Helios Light	TM HLS46
■ Apical Medium	DA AP169	■ EdgeLine Book	SU EDL53	■ Helios Light Oblique	TM HLS46
■ Azbat	TM ARB	■ EdgeLine Demi Bold	SU EDL53	■ Helios	TM HLS56
■ Arvis	SJ ARS	■ Electronic Normal	DA ELCS6	■ Helios Oblique	TM HLS56
■ Arthur	TM ART	■ Electronica Oblique	DA ELCS6	■ Helios Bold	TM HLS76
■ BARRY GOTHIC	TM BRG58	■ SPSL Elegant	SP ELG	■ Helios Bold Oblique	TM HLS76
■ BARRY GOTHIC	TM BRG76	■ EskaOne Light	DA ES146	■ Helios Black	TM HLS86
■ Bastion Kontrast Regular	DA BS073	■ EskaOne Normal	DA ES166	■ Helios Black Obli	TM HLS86
■ Bastion Kontrast Oblique	DA BS076	■ EskaOne Bold	DA ES166	■ Helios Condensed Thin	TM HLS93
■ BBBIT	TM BBT	■ EskaTwo Light	DA ES246	■ Helios Condensed Thin Obliq	TM HLS93
■ Bruskovaya Book	TM BR575	■ EskaTwo Normal	DA ES246	■ Helios Condensed Light	TM HLS47
■ Bruskovaya Narrow Book	TM BR577	■ EskaTwo Oblique	DA ES266	■ Helios Condensed Light Ob	TM HLS48
■ Calipso	DA CLP	■ EskaTwo Bold	DA ES266	■ Helios Condensed	TM HLS57
■ Camerton 1	DA CMR4	■ EskaTwo Bold Oblique	DA ES266	■ Helios Condensed Obliq	TM HLS77
■ Caston Book	TM CS155	■ Europe Book	TM ERP93	■ Helios Condensed Bold	TM HLS77
■ Caston Italic	TM CS156	■ Europe Book Oblique	TM ERP96	■ Helios Condensed Bold	TM HLS77
■ Caston Bold	TM CS176	■ Europe Bold	TM ERP76	■ Helios Condensed Bold	TM HLS77
■ Caston Bold Ital	TM CS176	■ Europe Bold Obliq	TM ERP76	■ Helios Condensed Bold	TM HLS77
■ С250 00230300	SP CK1	■ Europe Demi Bold	TM ERP66	■ Helios Condensed Blac	TM HLS87
■ С250 00230300	SP CK2	■ Europe Condensed	TM ERP67	■ Helios Condensed Blac	TM HLS87
■ SPSL Clarendon	SP CLR66	■ Europe Condensed Obliq	TM ERP68	■ Helios Compressed	TM HLC75
■ SPSL Clarendon	SP CLR66	■ Europe Condensed Bold	TM ERP77	■ Helios Extra Compressed	TM HLC77
■ SPSL Clarendon	SP CLR76	■ Europe Extend	TM ERP93	■ Helios Ultra Compressed	TM HLC79
■ SPSL Clarendon	SP CLR76	■ Europe Extend Oblique	TM ERP94	■ Helios Extended T	TM HLS33
■ Coliseum Roman	DA CU959	■ Europe Extend	TM ERP63	■ Helios Extended T	TM HLS34
■ Coliseum Italic	DA CU956	■ Europe Extend	TM ERP64	■ Helios Extended L	TM HLS43
■ Coliseum Bold	DA CU976	■ Everest Demi Bold	TM EVR76	■ Helios Extended L	TM HLS44
■ Coliseum Bold Italic	DA CU976	■ Everest Ultra	TM EVR93		
■ Cotlin Book	SJ CT156	■ Finist	DA FNS		
■ Cotlin Italic	SJ CT156	■ Fonta Can	TM FTV		
■ Cotlin Bold	SJ CT166	■ Fonta Polygraf	TM FTP		
■ Cotlin Bold Italic	SJ CT166				
■ GroundDown	TM CNT				

■ Helios Extended	TM HLS53	■ AZ Newspaper Book	AZ NWP45	■ Rodeo Book	TM RD055
■ Helios Extended	TM HLS54	■ AZ Newspaper Bold	AZ NWP95	■ Rodeo Book Italic	TM RD056
■ Helios Extended	TM HLS73	■ Newspaper Sans	TM NWS	■ Rodeo Bold	TM RD075
■ Helios Extende	TM HLS74	■ NEW ZELEK	TM NZK	■ Rodeo Bold Italic	TM RD076
■ Helios Exten	TM HLS83	■ Office Type Sans N	TM OTS55	■ Rodeo Extra Bold	TM RD085
■ Helios Exten	TM HLS84	■ Office Type Sans It	TM OTS56	■ Reges Script	TM RGR
■ AZ Highway	AZ HGW	■ Office Type Sans	TM OTS75	■ ROMUL	TM RHL
■ SPSL HILL MEDIUM	SP HLL	■ Office Type Sans	TM OTS76	■ Rostislav	DA RST
■ Industry	TM IND	■ SPSL Old King Regu	SP OLK55	■ Rubrica Medium	DA RBK55
■ ITALIANSHY A	DA ITLA	■ SPSL Old King Bo	SP OLK75	■ Rubrica Bold	DA RBK66
■ ITALIANSHY C	DA ITLC	■ Oliver Book	TM OLVS5	■ Secession Light	SJ SC545
■ Jatran	TM JTR	■ Oliver Book Italic	TM OLVS6	■ Secession Demi Bold	SJ SC565
■ Josephine	DA JSP	■ Oliver Bold	TM OLVS7	■ Secession Book	SJ SC566
■ КЮРДЯШЕВ	TM KBR	■ Oliver Bold Italic	TM OLVS7	■ Secession Bold	SJ SC575
■ Kisty-C	DA KSTC	■ Opium New Normal	TM OPN55	■ Secretary Normal	TM SCR66
■ Kudryashev Book	TM KOR45	■ Opium New Italic	TM OPN56	■ Secretary Bold	TM SCR75
■ Kudryashev Italic	TM KOR46	■ Opium New Bold	TM OPN75	■ Simetiz Book	SJ SH256
■ Kudryashev Bold	TM KOR65	■ Opium New Bold It	TM OPN76	■ Simetiz Italic	SJ SH256
■ Kudryashev Bold	TM KOR66	■ Optimus	TM OPT	■ Simetiz Bold	SJ SH286
■ Kudryashev Sans	TM KOS	■ Oswald Light	TM OSV45	■ Simetiz Bold Italic	SJ SH286
■ Laguna Plain	DA LON55	■ Oswald Light Italic	TM OSV46	■ Sistemnyi	DA SST
■ Laguna Italic	DA LON56	■ Oswald Medium	TM OSV65	■ Sonet Serif Regular	TM SNS55
■ Laguna Bold	DA LON75	■ Oswald Medium Ita	TM OSV66	■ Sonet Serif Italic	TM SNS56
■ AZ Latin	AZ LTW	■ Oswald Black	TM OSV95	■ Sonet Serif Bold	TM SNS56
■ Liana	TM LIA	■ Oswald Black Ital	TM OSV96	■ Sonet Serif Bold	TM STR
■ LIDIA	TM LDI	■ OTTISK REGULAR	DA OTT55	■ Surpriz Book	DA SRP55
■ Luga Extra Light	SU LUG35	■ OTTISK SEMI BOLD	DA OTT65	■ Surpriz Italic	DA SRP66
■ Luga Extra Light Lig	SU LUG36	■ Palladium Normal	TM PLD55	■ Surpriz Bold	DA SRP75
■ Luga Book	SU LUG45	■ Palladium Italic	TM PLD56	■ Surpriz Bold Italic	DA SRP76
■ Luga Oblique	SU LUG46	■ Palladium Bold	TM PLD75	■ SPSL Swordsman	SP SWR
■ Luga Demi Bold	SU LUG65	■ Palladium Bold Ita	TM PLD76	■ Tavrida Book	SJ TVR55
■ Luga Demi Bold Ob	SU LUG66	■ AZ Paragon Nord Bo	AZ PRN65	■ Tavrida Bold	SJ TVR65
■ Luga Medium	SU LUG55	■ AZ Paragon Nord It	AZ PRN66	■ Tavrida AD Book	SJ TVR65
■ Luga Medium Obli	SU LUG56	■ AZ Paragon Nor	AZ PRN95	■ Tecno2B	TM TCH
■ Luga Bold	SU LUG75	■ Penta Light	TM PNT45	■ ttalera	TM TRF
■ Luga Bold Oblique	SU LUG76	■ Penta Demi Bold	TM PNT65	■ Traktor Book	TM TRK
■ Luxor	TM LXR	■ Penta Bold	TM PNT75	■ troover	SJ TRR
■ Mueriro	TM MAE	■ Perfo Normal	SU PRF55	■ Unicum Condensed Thin	TM UNC29
■ Maqic	TM MGS76	■ Perfo Bold	SU PRF65	■ Unicum Condensed Light	TM UNC49
■ Maqic Bold	TM MGS76	■ plain	TM PLN	■ VACANSIA ROMAN	DA VCN55
■ Manuscript	DA MNS55	■ AZ Poligon	AZ PLG	■ VACANSIA BOLD	DA VCN75
■ Metropol Normal	DA MTP	■ Regular Script	SU PPS	■ VACANSIA EXTE	DA VCN73
■ Morris	TM MRR	■ SPSL ProFont Bold	SP PRF	■ Verbena Normal	DA VRB55
■ Motter Tektura	SJ MTT	■ Rader	SU RDR	■ Verbena Bold	DA VRB65
■ Mysl Narrow	TM MNS57	■ Ralenta Extra Bold	SJ RLE	■ Vla	TM VLA
■ Mysl Narrow Italic	TM MNS58	■ Regata	TM REG	■ Vodvile	DA VDV
■ Mysl Narrow Bold	TM MNS67	■ ROMAN LIGHT	DA RRL55		
■ Mysl Narrow Bold Ita	TM MNS68	■ ROMAN NORMAL	DA RRL56		
■ SPSL New Serif Condensed	SP NSR	■ Rodeo Light	TM RD045		
■ SPSL New Skoryna Bo	SP NSK	■ Rodeo Light Italic	TM RD046		

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)

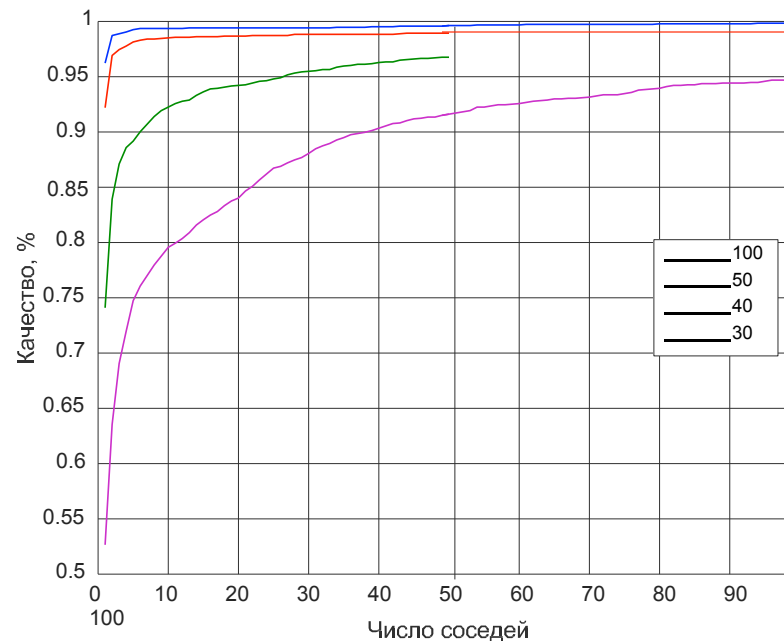
$1884 \times 1885 \times 52 = 184\,571\,712$ distances font-font (70 min – 2.25 sec/request)

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

Размер	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