Àâòîìàòè÷åñêàÿ êëàññèôèêàöèÿ øðèôòà ñòàðîïå÷àòíûõ òåêñòîâ

Àâòîð Âëàäèìèð Ñåðãååâè÷ Þæèêîâ
25.09.2009 ã.
Ïîñëåäíåå ìçìåíåíèå 05.10.2009 ã.

The article describes
an approach of font classification for old-printed texts and manuscripts. This
is necessary if we have an unknown antique book and want to determine the
printing press, place and time of issue, etc. The proposed approach is based on
partitioning the page into rows and letters, and then makes a consistent
comparison with the available font samples. Also the technique for the
automatic font samples generating is described.

Ââåäåíèå

Êîãäà â ðóêè ÷åëîâåêó, çàíèìàþùåìóñÿ
ñòàðîïå÷àòíûìè òåêñòàìè, ïîïàäàåò íåèçâåñòíàÿ ñòàðèííàÿ êíèãà, ëèáî åå
ôðàãìåíòû, âîçíèêàåò ïðîáëåìà - êàê îïðåäåëèòü âðåìÿ è ìåñòî åå ïå÷àòè, åñëè
ýòà èíôîðìàöèÿ ïî êàêèì-òî ïðè÷èíàì íåäîñòóïíà. Îïûòíûé èññëåäîâàòåëü, íåñêîëüêî,
ìîæåò ïðèìåíèòü îïðåäåëèòü ýòî, îñíîâûâàÿñü íà òåõ êíèãàõ è ðóêîïèñÿõ, ÷òî îí
âèäåë ðàíåå, íî äåòàëüíûé àíàëèç øðèôòà, ñòèëÿ íàïèñàíèÿ òðåáóåò âåñüìà
êðîïîòëèâîé ðàáîòû è íåìàëûõ çàòðàò âðåìåíè.

Ïîìî÷ü â ðåøåíèè òàêîé çàäà÷è îêàçà áû
ñïåöèàëüíàÿ êîìïüþòåðíàÿ ñèñòåìà, êîòîðàÿ óìåëà ñðàâíèâàòü áóêâû íåèçâåñòíîé
ñòðàíèöû ñ ìíîãî÷èñëåííûìè îáðàçöàìè øðèôòîâ èç ðàçíûõ èñòî÷íèêîâ, êîòîðûå
õðàíÿòñÿ â åå áàçå. Èñõîäÿ èç ýòîãî, ìîæíî ñôîðìèðîâàòü îñíîâíûå òðåáîâàíèÿ ê
òàêîé ñèñòåìå:

.
Êàæäûé îáðàçåö øðèôòà äîëæåí èìåòü ìåòàèíôîðìàöèþ, òàêóþ
êàê: òèïîãðàôèÿ, ãîðîä/ñòðàíà, ãîä èçäàíèÿ è ò.ä.

.
Äîëæíà áûòü âîçìîæíîñòü ëåãêîãî çàíåñåíèÿ â áàçó äàííûõ
íîâîãî îáðàçöà øðèôòà.

.
Âûäà÷à ðåçóëüòàòîâ äîëæíà áûòü ðàíæèðîâàííîé ïî êàæäîìó
ïóíêòó ìåòàèíôîðìàöèè. Íàïðèìåð:

o
Òèïîãðàôèÿ - «Êðàêîâ» (87%), «Êèðèëë è Ìåôîäèé» (75%);

o
Ãîä èçäàíèÿ - 1764 (95%), 1795 (84%); ...

.
Êëàññèôèêàöèÿ øðèôòà íå äîëæíà çàíèìàòü ìíîãî âðåìåíè.

Îáçîð
ëèòåðàòóðû

Îïèñûâàåìàÿ ïðîáëåìà îòíîñèòñÿ ê êëàññó çàäà÷ ðàñïîçíàâàíèÿ îáðàçîâ, à èìåííî - OFR (Optical Font Recognition). Ýòà îáëàñòü ñåé÷àñ àêòèâíî ðàçâèâàåòñÿ è íàõîäèò ñâîå ïðèìåíåíèå âî ìíîãèõ êîììåð÷åñêèõ ïðîäóêòàõ, òàêèõ êàê FineReader, CunieForm è ìíîãèõ äðóãèõ. Ê ñîæàëåíèþ, êëàññèôèêàòîð øðèôòîâ âõîäÿùèé â ñîñòàâ ýòèõ ñèñòåì, íå ïîäõîäèò äëÿ ðåøåíèÿ íàøåé çàäà÷è, ò.ê. «çàòî÷åí» ïîä âåñüìà óçêóþ îáëàñòü - ïîäîáðàòü íàèáîëåå ïîäõîäÿùóþ ãàðíèòóðó è ñòèëü íà÷åðòàíèÿ èç î÷åíü îãðàíè÷åííîãî íàáîðà. Îáû÷íî ýòî: ìîíîøèðèííûé øðèôò (Courier), øðèôò ñ çàñå÷êàìè (Times New Roman), øðèôò áåç çàñå÷åê (Arial), ïëþñ èõ ïîäìíîæåñòâà: bold è italic. Äëÿ çàäà÷è ðàñïîçíàâàíèÿ òåêñòà (OCR) áîëüøåãî è íå òðåáóåòñÿ, ïîòîìó ÷òî ïîäàâëÿþùåå ÷èñëî ñîâðåìåííûõ äîêóìåíòîâ èñïîëüçóþò âûøåïåðå÷èñëåííûå øðèôòû.

Îáðàòèìñÿ ê ëèòåðàòóðå - â ïóáëèêàöèÿõ âñòðå÷àþòñÿ äâà îñíîâíûõ ïîäõîäà ê ðàñïîçíàâàíèþ øðèôòà:

Âûäåëåíèå
ãëîáàëüíûõ ïðèçíàêîâ èç òåêñòîâûõ áëîêîâ, òàêèõ êàê ñëîâà, ñòðîêè è èõ äàëüíåéøèé àíàëèç. Ýòî ïîçâîëÿåò îïðåäåëÿòü ñòèëü íà÷åðòàíèÿ øðèôòà (bold, italic), åãî ðàçìåð, íàëè÷èå çàñå÷åê. Êàê ïðàâèëî, òàêîé ïîäõîä íå òðåáóåò áîëüøîãî îáðàçöà, ìîæåò îïðåäåëÿòü íåñêîëüêî îñíîâíûõ êëàññîâ øðèôòà è íå çàâèñèò îò ÿçûêà (â áîëüøèíñòâå ñëó÷àåâ), ÷òî äåëàåò åãî îïòèìàëüíûì äëÿ èñïîëüçîâàíèÿ â ñèñòåìàõ ðàñïîçíàâàíèÿ òåêñòà OCR.

Ê ïðèìåðó, â ðàáîòå Shi, Pavlidis [1] äëÿ ðàñïîçíàâàíèÿ øðèôòà èñïîëüçóåòñÿ ðÿä ãëîáàëüíûõ äëÿ âñåé ñòðàíèöû ïðèçíàêîâ, òàêèõ êàê: ãèñòîãðàììà ðàñïðåäåëåíèÿ äëèí ñëîâ (â ïèêñåëÿõ), âåëè÷èíà ìåæñòðî÷íûõ, ìåæáóêâåííûõ èíòåðâàëîâ. Yong Zhu [2] ïðåäëàãàåò ðàñïîçíàâàíèå øðèôòà, îñíîâàííîå íà àíàëèçå òåêñòóðû èçîáðàæåíèÿ âñåé ñòðàíèöû. Äëÿ èçâëå÷åíèÿ ýòèõ ïðèçíàêîâ èñïîëüçóåòñÿ ìíîãîêàíàëüíûé ôèëüòð Ãàáîðà.

Âûäåëåíèå
ëîêàëüíûõ ïðèçíàêîâ. Çäåñü óæå àíàëèçèðóþòñÿ îòäåëüíûå áóêâû è èõ ýëåìåíòû. Òàêîé ïîäõîä ÷óâñòâèòåëåí ê íåáîëüøèì èçìåíåíèÿì â øðèôòå, ÷òî ïîçâîëÿåò õîðîøî ðàçäåëÿòü áëèçêèå ïî íà÷åðòàíèþ øðèôòû è áîëüøå âñåãî ïîäõîäèò äëÿ íàøåé çàäà÷è.

Â ðàáîòå Cooperman'à [3] èññëåäóþòñÿ îöåíêè ñâîéñòâ øðèôòà äëÿ ñèñòåì OCR. Â êà÷åñòâå äåòåêòîðîâ ýòèõ ñâîéñòâ èñïîëüçóþòñÿ ëîêàëüíûå ïðèçíàêè, òàêèå êàê: íàëè÷èå çàñå÷åê, ïëîòíîñòü è ò.ä. Â ðàáîòå Zramdini, Ingold [4] ïðåäëàãàåòñÿ ñòàòèñòè÷åñêèé ïîäõîä äëÿ ðåøåíèÿ çàäà÷è êëàññèôèêàöèè øðèôòà, îñíîâàííûé íà âûäåëåíèè ëîêàëüíûõ ïðèçíàêîâ. Ñõîæèé ìåòîä èñïîëüçóåòñÿ è â ðàáîòå [5]

Ïðåäâàðèòåëüíàÿ
îáðàáîòêà

Ñòàðîïå÷àòíûé è, â îñîáåííîñòè, ñòàðèííûé ðóêîïå÷àòíûé (hand-writing) òåêñò èìååò ðÿä îñîáåííîñòåé, êîòîðûå íå ïîçâîëÿþò íàïðÿìóþ ïðèìåíèòü ìåòîäû, ïðåäëàãàåìûå â ïóáëèêàöèÿõ ïî OFR. Â ÷àñòíîñòè, ìåæáóêâåííûå è ìåæñòðî÷íûå ðàññòîÿíèÿ âàðüèðóþòñÿ â äîñòàòî÷íî øèðîêèõ ïðåäåëàõ äàæå íà îäíîé ñòðàíèöå. Êðîìå òîãî, òåêñò èìååò áîëüøîå êîëè÷åñòâî âàæíûõ äëÿ êëàññèôèêàöèè øðèôòà äèàêðèòè÷åñêèõ ýëåìåíòîâ, íà ñòðàíèöàõ ÷àñòî åñòü ðèñóíêè, ïîìåòêè. Òàêæå èçîáðàæåíèå ñîäåðæèò äåôåêòû, âûçâàííûå äëèòåëüíûì õðàíåíèåì, êîòîðûå ìîæíî óñëîâíî ðàçäåëèòü íà 2 òèïà. Ê ïåðâîìó òèïó îòíåñåì äåôåêòû ñàìèõ ñòðàíèö êíèãè, ïîÿâèâøèåñÿ â ðåçóëüòàòå äîëãîãî õðàíåíèÿ, äåéñòâèÿ âëàæíîñòè, òåìïåðàòóðû, ïîðàæåíèÿ ãðèáêîì îòäåëüíûõ ñòðàíèö, âûöâåòàíèå áóêâ, íåðàâíîìåðíûé öâåò áóìàãè, êðóïíûå è ìåëêèå ïÿòíà è ò.ä. Êî âòîðîìó òèïó ìîæíî îòíåñòè äåôåêòû, âîçíèêøèå ïðè îöèôðîâêå, ýòî: íåðàâíîìåðíàÿ ÿðêîñòü è êîíòðàñòíîñòü èçîáðàæåíèÿ (÷àñòî ïðîÿâëÿåòñÿ ïðè ñúåìêå öèôðîâûì ôîòîàïïàðàòîì), ïðîñâå÷èâàíèå íàäïèñåé ñ îáðàòíîé ñòîðîíû ëèñòà, öèôðîâîé øóì. Ïðèìåð òàêîãî èçîáðàæåíèÿ ïîêàçàí íà Ðèñóíêå 1.

Ðèñóíîê 1. Ôðàãìåíò ñòàðèííîãî òåêñòà (Ñèíîäèê Çèëàíòîâà ìîíàñòûðÿ, Íàöèîíàëüíûé àðõèâ Ðåñïóáëèêè Òàòàðñòàí)

Âñå ýòè îñîáåííîñòè ìîãóò ïîìåøàòü ïðîöåññó êëàññèôèêàöèè øðèôòà. Â ñâÿçè ñ ýòèì, èçîáðàæåíèå íåîáõîäèìî ïðåäâàðèòåëüíî îáðàáîòàòü è ïîäãîòîâèòü. Äëÿ ýòîãî âîñïîëüçóåìñÿ àëãîðèòìàìè, îïèñàííûìè â [6] - â ðåçóëüòàòå èõ ðàáîòû èìååì áèíàðíîå èçîáðàæåíèå (îíî ñîñòîèò òîëüêî èç ÷åðíûõ è áåëûõ ïèêñåëåé), î÷èùåííîå îò øóìà è ïîìåõ, à òàêæå ïîâåðíóòîå íà íåîáõîäèìûé óãîë, åñëè ñòðàíèöà ïðè îöèôðîâêå áûëà íå ïàðàëëåëüíà êðàþ ñêàíèðóþùåãî óñòðîéñòâà. Äàëåå ñ ïîìîùüþ ìåòîäà, ïðåäëîæåííîãî â [7], âûäåëÿåì òåêñòîâûé áëîê. Íåáîëüøèå íåòî÷íîñòè â îïðåäåëåíèè ãðàíèö òåêñòà, êîòîðûå âîçíèêàþò íà ñëîæíûõ èçîáðàæåíèÿõ, ìîæíî íå ïðèíèìàòü âî âíèìàíèå, ïîòîìó ÷òî íàøà öåëü - ýòî íå ðàñïîçíàâàíèå òåêñòà, ïîýòîìó ïîòåðÿ íåêîòîðûõ áóêâ íà ãðàíèöàõ ïðàêòè÷åñêè íå ïîâëèÿåò íà êîíå÷íûé ðåçóëüòàò. Èòîã ïðåäâàðèòåëüíîé îáðàáîòêè ïîêàçàí íà Ðèñóíêå 2.

Ðèñóíîê 2. Ïðåäâàðèòåëüíàÿ
îáðàáîòêà èçîáðàæåíèÿ

Ñîïîñòàâëåíèå
øðèôòà

Îáîçíà÷èì ïèêñåëè áèíàðíîãî èçîáðàæåíèÿ êàê A_{ij}, ãäå i=1..x, j=1..y, ãäå x è y - øèðèíà è âûñîòà ñîîòâåòñòâåííî. Ïèêñåëè âíå òåêñòîâûõ áëîêîâ ïðèìåì çà áåëûå, ò.ê. ðèñóíêè è êàðòèíêè äëÿ íàøåé çàäà÷è íå âàæíû. Îáîçíà÷èì dres êàê ðàçðåøåíèå ñêàíèðîâàíèÿ â dpi.

Ïóñòü îáðàçåö øðèôòà - ýòî íàáîð øàáëîíîâ
áóêâ, âõîäÿùèõ â åãî ñîñòàâ, â êîëè÷åñòâå ïðèìåðíî ðàâíûì ÷èñëó áóêâ äàííîãî
àëôàâèòà (ïî÷åìó «ïðèìåðíî», áóäåò îáúÿñíåíî íèæå). Ïèêñåëè øàáëîíà áóêâû èç
îáðàçöà øðèôòà îáîçíà÷èì êàê , z=1..k, ãäå k - êîëè÷åñòâî
øàáëîíîâ â îáðàçöå øðèôòà, i=1..mz, j=1..nz, ãäå mz è nz - øèðèíà è âûñîòà øàáëîíà
ñîîòâåòñòâåííî (Ðèñóíîê 3).

Ðèñóíîê 3. Âèä øàáëîíà áóêâû

Îáðàçöîâ øðèôòîâ òàêæå ìîæåò áûòü ìíîãî.
Äëÿ òîãî ÷òîáû êëàññèôèöèðîâàòü øðèôò íà íåèçâåñòíîé ñòðàíèöå, íåîáõîäèìî íàéòè
íàèëó÷øåå ñîîòâåòñòâèå ñ îäíèì èç îáðàçöîâ øðèôòà. Äëÿ ýòîãî êàæäóþ áóêâó
ñòðàíèöû áóäåì ñðàâíèâàòü ñ êàæäûì øàáëîíîì áóêâû âñåõ îáðàçöîâ øðèôòà ñ
ïîìùüþ ôóíêöèè êîððåëÿöèè, ïîêàçàííîé â [8].

,

ãäå i=1, 2,..., mz, j=1, 2,..., nz, - ñðåäíåå çíà÷åíèå
ïèêñåëåé â øàáëîíå (âû÷èñëÿåìîå òîëüêî îäèí ðàç), - ñðåäíåå çíà÷åíèå ýëåìåíòîâ
èçîáðàæåíèÿ A â îáëàñòè, ñîâïàäàþùåé ñ òåóùèì ïîëîæåíèåì B, à ñóììèðîâàíèå âåäåòñÿ ïî âñåì ïàðàì
êîîðäèíàò, îáùèì äëÿ A è B. Êîýôôèöèåíò
êîððåëÿöèè ìåíÿåòñÿ îò -1 äî 1 è íå çàâèñèò îò èçìåíåíèÿ ìàñøòàáà àìïëèòóä A è B.

Èòîãîâàÿ êîððåëÿöèÿ äëÿ îáðàçöà øðèôòà
áóäåò ïîñëåäîâàòåëüíîñòüþ êàê:

Îáðàçåö øðèôòà (èëè íåñêîëüêî îáðàçöîâ) ñ
ìàêñèìàëüíûì êîýôôèöèåíòîì êîððåëÿöèè áóäåò ÿâëÿòüñÿ íàèëó÷øèì ñîîòâåòñòâèåì ñî
øðèôòîì íåèçâåñòíîé ñòðàíèöû.

Ñåãìåíòàöèÿ
èçîáðàæåíèÿ íà ñòðîêè è áóêâû

Èòàê, ìû ïîðåäåëèëè ìåòîä äëÿ íàõîæäåíèÿ
íàèëó÷øåãî îáðàçöà øðèôòà. Íî äëÿ òîãî, ÷òîáû åãî ïðèìåíèòü, íåîáõîäèìî íà
èçîáðàæåíèè ñòðàíèöû (íà êîòîðîì èìåþòñÿ òîëüêî òåêñòîâûå áëîêè áåç ðèñóíêîâ è
êàðòèíîê) âûäåëèòü ñòðîêè, à çàòåì è îòäåëüíûå áóêâû.

Äëÿ ñåãìåíòàöèè ñòðîê âîñïîëüçóåìñÿ
ñëåäóþùèì ïîäõîäîì: ïîñòðîèì âåðòèêàëüíóþ ïðîåêöèþ áèíàðíîãî èçîáðàæåíèÿ ïî
ñëåäóþùåé ôîðìóëå:

, ãääå j=1..y

Ïîñëå ýòîãî ñãëàäèì ïîëó÷åííóþ ïðîåêöèþ
ïóòåì óñðåäíåíèÿ êàæäîãî ýëåìåíòà âåêòîðà ñ åãî ñîñåäíèìè çíà÷åíèÿìè. Ðåçóëüòàò
ïîêàçàí íà Ðèñóíêå 4.

Ðèñóíîê 4. Âûäåëåíèå ñòðîê

Äàëåå âûäåëÿþòñÿ ñòðîêè ïî òåì íåïðåðûâíûì
ïîëîñàì, ãääå çíà÷åíèå ωj > λline, ãääå λline - íåêîòîðûé êîýôôèöèåíò. Çàòåì èç ïîëó÷åííûõ ñòðîê
àíàëîãè÷íûì îáðàçîì âûäåëÿþòñÿ îòäåëüíûå áóêâû. Äëÿ ÷àñòè ñèìâîëîâ, êîòîðûå
ñîäåðæàò íåïäíîñâÿçíûå îáëàñòè (íàïðèìåð, áóêâà «Û») ìîãóò áûòü íåïðàâèëüíî
îïðåäåëåíû ãðàíèöû. Íî ýòî ìàëî âëèÿåò íà êîíå÷íûé ðåçóëüòàò, ò.ê. äëÿ
êëàññèôèêàöèè øðèôòà äîñòàòî÷íî èñïîëüçîâàíèÿ òîëüêî ÷àñòè áóêâ.

Îáó÷åíèå

Òåïåðü îñòàåòñÿ åùå îäíà ïðîáëåìà - êàê
ïîäãîòîâèòü îáðàçöû øðèôòà. Ìîæíî ñäåëàòü ýòî âðó÷íóþ, âûðåçàÿ øàáëîíû áóêâ èç
èçîáðàæåíèÿ, íî ýòî áóäåò âåñüìà òðóäîåìêî è çàéìåò ìíîãî âðåìåíè, ïîýòîìó
íåîáõîäèìà àâòîìàòèçàöèÿ äàííîãî ïðîöåññà.

Ïðåäëàãàåìûé ñëåäóþùèé ïîäõîä - íà âõîä
ïîäàåòñÿ íåñêîëüêî ñòðàíèö ñ òåì øðèôòîì, êîòîðûé ïðåäïîëàãàåòñÿ çàíåñòè â
êà÷åñòâå îáðàçöà. Âñå ñòðàíèöû, êðîìå îäíîé, èñïîëüçóþòñÿ äëÿ ôîðìèðîâàíèÿ îáðàçöîâ
áóêâ, íàçíâàì èõ îáó÷àþùåé âûáîðêîé. À îñòàâøàÿñÿ îäíà ñòðàíèöà - êîíòðîëüíàÿ, äëÿ
óòî÷íåíèÿ ñîîòíîøåíèé îáðàçöîâ.

Äëÿ ñòðàíèö îáó÷àþùåé âûáîðêè äåëàþòñÿ
ýòàïû ïðåäâàðèòåëüíîé îáðàáîòêè è âûäåëåíèÿ ñòðîê/ñèìâîëîâ ïîèñàííûå âûøå. Âûäåëåííûå
áóêâû çàïîìèíàþòñÿ êàê øàáëîíû. Ïîñêîëüêó ñðåäíÿÿ ñòðàíèöà ñòàðîíå÷àòíîãî
òåêſòà ñîäåðæèò, êàê ïðàâèëî, 200-1000 ñèìâîëîâ, ñëåäîâàòåëüíî, ñòîëüêî æå ìû
ïîëó÷èì è øàáëîíîâ áóêâ. Ñîîòâåòſòâåííî, äëÿ êàæäîãî ñèìâîëà, â ñðåäíåì, ìû
èìååì 6-30 øàáëîíîâ â êàæäîì ſòðàíèöå. Òàêîå áîëüøîå êîëè÷åſòâî íåæåëàòåëüíî,
ïîýòîìó íåîáõîäèìî ñîêðàòèòü èõ ÷èſëî. Ýòî áóäåì äåëàòü â äâà ýòàïà. Íà ïåðâîì

ìû ïîìàðíî ñðàâíèâàåì âñå âûäåëåííûå øàáëîíû ñ ïîìîùüþ ôóíêöèè êîððåëÿöèè,
èñïîëüçîâàííé âûøå. Åñëè äëÿ î÷åðåäíîé ïàðû çíà÷åíèé êîýôôèöèåíòà ìåíüøå λequal, òî îäíó èç êîïèé óäàëÿåì.

Ïîñëå òàêîé ïðîöåäóðû óäàëåíèÿ äóáëèêàòîâ ó
íàñ èìåþòñÿ óæå äîñòàòî÷íî ïðèãîäíûé êîëëåêöèÿ øàáëîíîâ ñèìâîëîâ øðèôòà. Íî òàì
ìîãóò ñîäåðæàòüñÿ ïîñòîðîííèå ýëåìåíòû, íàïðèìåð ñëèïøèåñÿ èëè ïîðâàæäåííûå
áóêâû, ýëåìåíòû ðèñóíêîâ, êëÿêñ è ïðî÷åå. Äëÿ òîãî ÷òîáû åùå áîëüøå ïîâûñèòü
êà÷åñòâî øàáëîíîâ, ìû ïðîèçâîäèì ïðîöåäóðó ðàñïîçíàâàíèÿ øðèôòà ñ ïîìîùüþ ýòèõ
øàáëîíîâ íà îäíé îñòàâøåéñÿ êîíòðîëüíîé ñòðàíèöå. Ïîñêîëüêó øðèôò íà îáèõ ñòðàíèöàõ
èäåíòè÷íû, òî òå øàáëîíû, êîòîðûå íå ñîâïàäàòñÿ íà êîíòðîëüíîé, ñòðàíèöå íå
îòíîñÿòñÿ ê ñèìâîëàì, ñëåäîâàòåëüíî, èõ ìîæíî óäàëÿòü. Ïîñëå òàêîé ïðîöåäóðû ìû
ïîëó÷èëè âåñüìà õîðîøóþ áèáëèîòåêó îáðàçöîâ äàííîãî øðèôòà.

Çàêëþ÷åíèå

Áûëà ïîêàçàíà ìåòîäèêà äëÿ àâòîìàòè÷åñêîé
êëàññèôèêàöèè øðèôòà ñòàðîîå÷àòíîãî òåêſòà, ïðåäëîæåíû àëãîðèòìû è ìåòîäû äëÿ
ñåãìåíòàöèè èçîáðàæåíèÿ íà ñòðîêè è ñèìâîëû. Ïîêàçàíà ìåòîäèêà äëÿ
àâòîìàòè÷åñêîãî ôîðìèðîâàíèÿ îáðàçöà øðèôòà.

Àëãîðèòìû è ïîäõîäû áûëè ðåàëèçîâàíû â âèäå
ïðîãðàììíîé ñèſòåìû. Íà êîíôåðåíöèè òàêæå áóäåò ïðîäåìîíſòðèðîâàíà ðàáîòà ñàìîé
ſèſòåìû è ïîëó÷åííûå ïðàêòè÷åſêèå ðåçóëüòàòû.

Ëèòåðàòóðà

[1] H. Shi
and T. Pavlidis, "Font Recognition and Contextual Processing for More Accurate
Text Recognition," ICDAR'97, pp. 39-44, Ulm, Germany, Aug. 1997.

[2] Yong Zhu,
Font Recognition Based on Global Texture Analysis. IEEE Transactions PAMI. October 2001 (vol. 23 no. 10) pp. 1192-1200.

[3] R.
Cooperman, "Producing Good Font Attribute Determination Using Error-Prone
Information," SPIE, vol. 3,027, pp. 50-57, 1997.

[4] A.
Zramdini and R. Ingold, "Optical Font Recognition Using Typographical
Features," IEEE Trans. Pattern Anal. Machine Intell., vol. 20, no. 8,
pp.877-882, 1998.

[5] A.
Schreyer, P. Suda and G. Maderlechner, "Font Style Detection in Documents Using
Textons", Proc. of 3rd IAPR Document Analysis Systems Workshop, Nagano, Japan,
1998.

[6] Ñîëîâüåâ Â.Ä., Þæèêîâ Â.Ñ.
Àâòîìàòèçèðîâàííàÿ ñèñòåìà îáðàáîòêè è ðåñòàâðàöèè èçîáðàæåíèé ñòàðîïå÷àòíûõ òåêſòîâ è ðóêîïèñåé. Âåðſèíèê ÊÃÒÓ èì. Òóïîëåâà. Âûï.3. - Êàçàíü: Èç-âî ÊÃÒÓ, 2006. - ſ.28-30

[7] Â.Ñ. Þæèêîâ. "Ñåãìåíòàöèÿ èçîáðàæåíèé ſòðàíèö äðåâíèõ ðóêîïèſåé". Ñáîðíèê òðóäîâ êîíôåðåíöèè "RCDL-2007". Ïåðåſëàâëü-Çàëåſſêèé. Òîì 1, c. 236-240, 2007

[8] Ð. Ãîſſàëſ. Öèôðîâàÿ îáðàáîòêà èçîáðàæåíèé. Ìſſêâà: Òåõſíſôåðà, 2005. - ſ. 996-997.