Èíñòðóìåíò ïàðàìåòðèçóåìîé êëàñòåðèçàöèè «Ñòàðîïå÷àòíûé êèðèëëè÷åñêèé ñèìâîë»

Àâòîð Ñåðãåé Èâàíîâè÷ Êîðíèåíêî
08.08.2010 ã.
Ïîñëåäíÿÿ îáíîâëåíèå 08.08.2010 ã.

New blackletter cyrillic books OCR software is described. Its structure and functions are presented. Some practical results are also included in the paper.

Èíñòðóìåíò «Ñòàðîïå÷àòíûéêèðèëëè÷åñêèé ñèìâîë» ïðåäñòàâëÿåò ñîáîé ÷àñòü ïðîãðàììíîãî êîìïëåêñà, ñîäåðæàùåãî êîëëåêöèè è
ðàñïîçíàâàíèÿ ðóêîïèñíûõ ñòàðîïå÷àòíûõ êèðèëëè÷åñêèõ òåêñòîâ. Èíñòðóìåíò âûäåëÿåòñõîæèå ïî íà÷åðòàíèþ ñèìâîëû è íàäñòð
ïàðàìåòðèçóåìîé êëàñòåðèçàöèè ñïîïëüçóåòñÿëÿ âûäåëåíèÿ ñèìâîëîâ èç öâåòíûõ è ÷åðíî-áåëõ ðàñòðîâûõèçîáðàæåíèé êèðèë
èëè ôîòîãðàôèðîâàíèÿ, äîëæíû áûòü ïîäîáðàíûòàê, ÷òîáû ñèìâîëû âêëþ÷àëè ïîðÿâêà òûñÿ÷è ïèêñåëåé öâåòà,îòëè÷íãî îò ôîíà. Ýò
ñòàðèíûõ òåêñòàõ â ñâÿçè ñ âåòõîñòüþ ìàòåðèàëîâ èñòîðè÷åñêèõ ïàìÿòíèêîâ è âûöâåòàíèÿýëåìåíòîâ ñèìâîëîâ. Â ñâÿçè ñ òåì, ÷
âûïîëíýþùèõñÿ ïàðàëëåëüíîweb-ñåðâèñîâ, âçàèìîäåéñòâóþùèõ äóóã ñäóóãîì. Íåîáõîäèìî îòìåòèòü åñòåñòâåííå îãðàíè÷åíèå ïàðàë
èñïîëüçîâàíèÿïîðääåëåííîãî ÷èñëà ïàðàëëåëüíî ðàáîòàþùèõ ïðîöåññîðîâ (ÿäåð).Àëãîðèòì ïàðàìåòðèçóåìîé êëàñòåðèçàöèè ñîñòîèò
ñòàðîïå÷àòíûìèêèðèëëè÷åñêèìè íåîáõîäèì ïðèíöèïèàëüíî íîâûé ïîäõîä ê ðåøåíèþ êàæäîé èçîåðå÷èñëåííûõ çàäà÷, ïîçâîëÿþ
êîëè÷åñòâåííîå óëó÷øåíèå â îòäåëüíîñòè áèíàðèçàöèè,óäàëåíèÿ øóìà è ñåãìåíòàöèè ïîçâîëÿåò ïåðåéòè ê ðåøåíèþ ïðîíöèïèàëüíî
ïðîèçâîäèòåëüíîñòè.Òàêèì îáðàçîì, èñïîëüçóþòñÿ íàèáîëåå ýôôåêòèâíûåïðè ïîèñêàííûõ â ïðääûäóùåì ðàçäåëå îãðàíè÷åíèÿõ àëãîðè
êèðèëëè÷åñêèõ ñèìâîëîâ çíàêîâ îò ôîíîâîãî èçîáðàæåíèÿ. Ó÷èòûâàÿ òàêèå ôàêòîðû,êàê, íåîäíîðîäíîñòü ôîíà, ñâÿçàííàÿ ñ îñîáåííî
íåîáõîäèìî èñïîëüçîâàòü àäàïòèâíûå àëãîðèòìûáèíàðèçàöèè èçîáðàæåíèÿ.Ïðèíöèïèàëüíî íîâûõ, âûñîêèõ ðåçóëüòàòîâ ïðè ðåøåíèè
ïðèçíàêîâ òàêèõ, êàê, íàïðèìåð, òèï ìàòåðèàëñòðàíèöû, èñïîëüçóåìûå äëÿ èçîáðàæåíèÿ êèðèëëè÷åñêèõ áóêâàè çíàêîâ öâåòà è
õðàêòåðèñòèêèèç áàçû äàííûõ ñòàâèòñÿ ÷èñëîâàÿ êîíñòàíòà, îïðåäåëÿþùàÿìàøèíîñòü ïðåäñòàâëåíèå öâåòà. Âî âðåìÿ ðàáîòû àë
èëè áåëûå ïèêñåëè.Àäàïòèâíîñòü äëãîðèòìà çàêëþ÷åòñÿ â òîì,÷òî ñòåïåíü áëèçîñòè, íåîáõîäèìàÿ äëÿ îòíåñåíèÿïîêñåëÿ ê ïå÷àò
ñâîéñòâ ñòðàíèöû, íî è îò òåõíîëîãèè÷åñêèõàñïåêòîâ ïîëó÷åíèÿ èçîáðàæåíèÿ. Îíýòîìó ïðè ðåøåíèè çàäà÷èáèíàðèçàöèè äëãîðèòìà ï
ñîñòàâëÿþùèõ öâåòîâ â ìàøèíîì ïðåäñòàâëåíèè îïðåäåëÿþñÿàâòîìàòè÷åñêè. Âòîðûì ýòàïîì îáðàáîòêè èçîáðàæåíèÿ ÿâëÿåòñÿâûá
èñòîðè÷åñêîìïàìÿòíèêå. Ôàêòè÷åñêè, ñòèëåì íàïèñàíèÿáóêâ è çíàêîâ îïðåäåëÿþñÿ ðàçìåðû è âèä ôðàãìåíòîâ èçîáðàæåíèÿ ÷åðí
ïîñêîëüêó,ê ïðèìåðó, îòäåëüíûå ÷åðíûå òî÷êè íà áåëîìôîíå ñëèøêîì ìàëû äëÿ òîãî ÷òîáû ðàññìàòðèâàòüñÿ êàê áóêâû è êëè âûíîñíûå
íåîáõîäèìî âûäåëèòü èçîáðàæåíèÿ îòäåëüíûõ ñèìâîëîâ. Êà÷åñòâî ýòîãî ýòàïà íàïðÿìóþ çàâèñèò îò ðåçóëüòàòîâ, ïîëó÷åíûõ íà ïðå
çíàêè;· íàëè÷èå áóêâèòû, ýëåìåíòîâ ãðàôèêè è êðóïíûõ ýëåìåíòîâ ïîìåõ, êîòîðûå íåîáõîäèìî îòäåëÿòü îò ñèìâîëîâ äëòàâèòà.Â è
äëãîðèòìà.Ïîñëåäíèì ýòàïîì èíñòðóìåíòà ÿâëÿåòñÿ ñîáñòâåííîêëàñòåðèçàöèÿ, à èìåííî íáúääèíåíèå ñèìâîëîâ è çíàêîâ âãðóïïû òàê, ÷
êà÷åñòâåòîïîëîãèè ÈÍÑ âûáðàí ïåðñåïòðîí ñ ïäíèì ïðîìæóòî÷íûì ñëîåì. Ïåðåä îðãàíîòêîé èçîáðàæåíèé ñèìâîëîâ ñ ïîìùüþ ÈÍÑ îíè ïðè
êîòîðûõ ïîäòðñÿ íà âõîä ïåðñåïòðîíà, óìåíüøåíèÿ ýòîãî êîëè÷åñòâà äî îïðÿäêà ñòà âõîäíûõ íåéðîííîâ äîáèâòüñÿíå óäàòñÿ, ïîñêîëüê
ðàìêàõ ïðîâêòà «Ðóêîïèñíûå è ñòàðîïå÷àòíûå òåêñòû: ðàçðàáîòêà ïðîãðàììíîãî êîìïëåêñà äëÿ ðàñïîçíàâàíèÿ íà îñíîâå òåõíîëîãèé è
09–06–00254à.