

[HA-12434] CNN based fonts with rejection class do not take the aspect ratio into account Created: 30.09.2016 Updated: 13.10.2016 Resolved: 10.10.2016	
Status:	Closed
Project:	HALCON
Component/s:	HLib -> OCR
Affects Version/s:	nightly-13.0
Fix Version/s:	13.0
Security Level:	Accessible by Reporter (The issue is accessible by MVTec employees and the reporter.)

Type:	Bug	Priority:	Major
Reporter:	Markus Ulrich	Assignee:	Unassigned
Resolution:	Fixed	Votes:	0
Labels:	None		
Remaining Estimate:	Not Specified		
Time Spent:	Not Specified		
Original Estimate:	Not Specified		

Attachments:	 test_universal_fonts.hdev		
Issue Links:	Defect		
	defect	QA-1364	Test new feature - Improved Identific... Pass
Release Note Status:	Not Needed		
Affected Platforms:	All		

Description

The CNN fonts with rejection class (Universal*_Rej.occ) often mix up characters that can only be distinguished by their aspect ratios. For example, '-' is confused with 'l' or 'I' and ':' is confused with '='.

Comments

Comment by Markus Ulrich [30.09.2016]
After discussion with Michael Klostermann it turned out that only the fonts without rejection class take the aspect ratio into account. See the attached example program test_universal_fonts.hdev to reproduce the problem.
Comment by Michael Klostermann [07.10.2016]
I fixed it with new classifiers.
The new classifiers use a different preprocessing which preserves the aspect ratio of characters. This results in the wanted effect of discriminating chars like : against =.
The problem is fixed and this shown in the results of our aocr benchmark where the error of the signs is reduced:
1 Old Classifier:
2 Error on 'u:': 15.30 % of 863
3 Error on 'u=:': 38.89 % of 72
4
5
6 New Classifier:
7 Error on 'u:': 7.53 % of 863
8 Error on 'u=:': 5.56 % of 72
Note that, the remaining errors are made due to different problems or mislabeling.
Comment by Michael Klostermann [07.10.2016]
I shared the issue with Markus Ulrich and David Sattlegger . One of them should be able to review the fix.
Comment by David Sattlegger [10.10.2016]
I re-evaluated all subsets of the Universal CNN OCR Classifier, the results can be found at: https://confluence.mvtec.com/display/EN/Evaluation+of+the+Universal+CNN+OCR+Classifier
The errors have the same magnitude as before. In particular, the fonts with a large number of special characters have lower error due to the new preprocessing. In some cases they are not directly comparable since the new classifier has more classes as before.

Generated at Fri May 26 12:47:24 CEST 2017 by Juan-Pablo De La Cruz Gutierrez using JIRA 7.3.4#73015-sha1:a262b3457b3605f12635df4b0a0c3dc71d631a1e.