

Čítanie textu na pneumatike

Andrej Lúčny

ME-Inspection & Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského
Mlynská Dolina, 842 48 Bratislava
lucny@fmph.uniba.sk

Abstrakt

Pre rôzne účely je potrebné vedieť z pneumatiky prečítať záznam o testovaní. Pre variabilnosť záznamu od množstva výrobcov i charakter materiálu ide o náročnú úlohu OCR. Riešenie tejto úlohy nie možné dosiahnuť použitím jednej metódy, čo dáva priestor na zakomponovanie sady rôznych a často protichodných trikov, pomocou ktorých číta taký záznam človek a to bez toho, že by si ich použitie plne uvedomoval. Prezентujeme naše riešenie dosahujúce slušnú úspešnosť ako aj návrh jeho ďalšieho zlepšenia.

1 Úvod

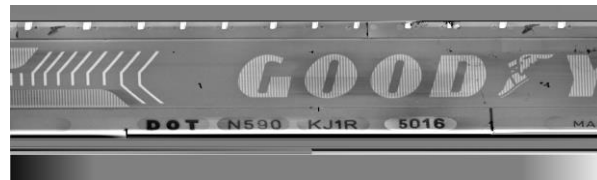
Našu metódu spracovania snímku získaného rotovaním pneumatiky pod 2D laserovým snímačom, sme implementovali pomocou open-source knižnice OpenCV (spracovanie obrazu) a Tesseract (OCR). Avšak pokiaľ sme ich použili priamočiaro, úspešnosť ktorú dosiahli nepresiahla 70%. Žiadna binarizačná metóda neodseparuje text dokonale. Tesseract zase nemôžeme pretrénovať, lebo nemáme k dispozícii fonty používané výrobcami pneumatík.

2 Riešenie

Rozhodli sme sa preto kombinovať viaceré metódy binarizácie (obr. 2) vstupného obrazu (obr. 1) a viaceré kandidátov na písmeno kódu triediť podľa toho, ako k sebe dobre susedné písmená a skupiny písmen pasujú. Pritom sa snažíme využívať vedomosť o tom, čo sa už podarilo rozpoznať (spracovanie prebieha potenciálne v niekoľkých iteráciách) a prikladaním „pravítka“ k rozpoznanému (obr. 3) pomáhame separácii písmen, ktoré boli dosiaľ zliate s ostatnými prvkami na obraze (nepríjemné sú hlavne oválne obkruženia časti textu). Nebezpečenstvo pritom predstavuje rozpad písmen, pretože pri ňom môže vzniknúť fantomické písmeno (napríklad dve I z jedného H, I z T a podobne). Tieto charakteristické rozpady premietame do váhy rozpoznaného písmena a hľadáme riešenie s najvyššou kvalitou (obr. 4).

Keď už máme písmeno odseparované, prebieha na ňom OCR. Jadrom tohto OCR je bežne používaný

klasifikátor Tesseract, snažiaci sa pokryť písmeno z pripravených kúskov, pracujúci na princípe algoritmu A*. Zvyšujeme pritom jeho schopnosť vrátiť viacero možností s definovanou pravdepodobnosťou tým, že písmenu pridáme dobre rozoznateľný kontext. Binárny obraz vstupujúci do OCR upravujeme viacerými úpravami tak, aby sme vstupné písmeno modifikovali do tvaru, v ktorom má OCR najvyššiu účinnosť. Dôležitá je tu hlavne šírka línií písmena, ktorú sa snažíme optimalizovať, pričom musíme dávať pozor, aby sa touto modifikáciou písmeno nerozpadlo alebo nestratilo zásadné charakteristiky svojho tvaru. Písmeno potom normalizujeme na zvolenú štandardnú veľkosť a detegujeme na ňom sadu príznakov ako počet veľkých dier v stredovej vertikálnej línii, prítomnosť vonkajších a vnútorných oblúkov a rohov, zúženie v strede zľava a sprava, záseky v obvodovej línii zľava v dolnej polovici a sprava v dolnej polovici, spojitosť vnútra s ľavou či pravou stranou a podobne.



Obr. 1: Vstupný obraz



Obr. 2: Jedna z možných binarizácií obrazu



Obr. 3: Použitie pravítka



DOT N590 KJ1R 5016

Obr. 4: Rozpoznaný kód

Vďaka tomu, že je veľkosť pri tomto procese normalizovaná, vieme nastaviť do algoritmov určujúcich tieto príznaky, pevné a všeobecne správne konštanty, doladené na veľkej vzorke dát. Pomocou týchto príznakov potom eliminujeme niektoré možnosti, ktoré ponúka OCR (U s dierou je neprípustné), zatiaľ čo iné transformujeme (B s vonkajšími oblúkmi a stredovým zúžením zľava je 8, 6 bez diery, ktorej vnútro je spojené s ľavou stranou a má vľavo hore roh je 5), čím opravujeme charakteristické chyby zvoleného klasifikátora.

3 Záver

Pri testovaní na 377 vzorkách (kód má 10 až 16 písmen) sme zaznamenali 6 nesprávne učených písmen. Úspešnosť rozpoznania kódu bola teda 98,4%. V prepočítaní na písmená je to 99,998%. Žiadna vzorka nebola odmietnutá ako nerozpoznaná.

Kombináciou viacerých metód s nízkou účinnosťou sme tak získali použiteľný systém. Predstavujeme si, že podobným spôsobom funguje čítanie textu v ľudskom mozgu. Hoci nám sa text číta ľahko, za touto ľahkosťou sa skrýva konglomerát trikov, ktoré si neuvedomujeme. Pri implementácii analogického umelého systému ich musíme namáhavo odhaľovať a zakomponovávať do nášho riešenia. Náš prístup by sa dal ešte výrazne zlepšiť, keby sme nepracovali s jedinou snímkom obrazu a snímkovanie prispôbovali potrebám spracovania (senzo-motorický prístup).

PodĎakovanie

Tento príspevok vznikol v spolupráci so spoločnosťou ME-Inspection.

Literatúra

Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.

Smith, R.: An Overview of the Tesseract OCR Engine. ICDAR '07 Proceedings of the Ninth International Conference on Document Analysis and Recognition, Volume 02, pp. 629-633