

# LOCAL ENVIRONMENT AS HYPERLINK

**Marek Mešár**

Master Degree Programme (3), FIT BUT

E-mail: xmesar00@stud.fit.vutbr.cz

Supervised by: Vítězslav Beran

E-mail: beranv@fit.vutbr.cz

**Abstract:** This paper describes selected techniques and approaches to problem of text detection and recognition on modern mobile devices. It also describes their proper presentation to the user interface and their conversion to hyperlinks as a source of information about surrounding world. The paper outlines text detection and recognition technique based on MSER detection and also describes the use of image features tracking method for text motion estimation.

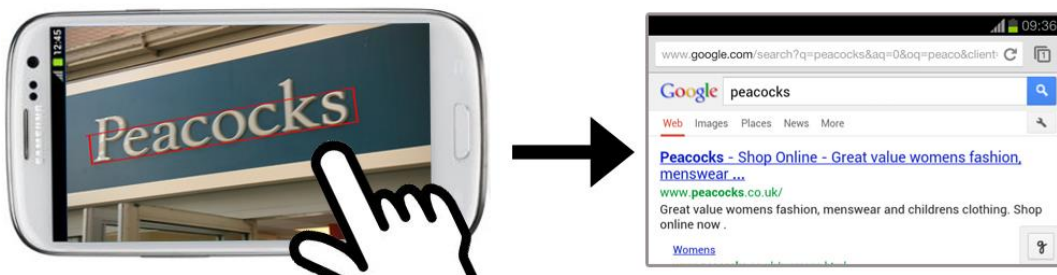
**Keywords:** Android, MSER, text detection, text recognition, OCR, text digitalization, tracking

## 1. ÚVOD

Bežnému užívateľovi slúži sieť internet okrem iného aj ako nevyčerpatelný zdroj informácií a faktov o okolitom svete. Ďalším z fenoménov dnešnej doby sú moderné mobilné telefóny a tablety. Kombináciou mobilnosti telefónov a obrazu z kamery ako zdroja dát z okolitého sveta sa naskytá výborná možnosť ako priniesť užívateľom rýchly, ľahký a pohodlný prístup k informáciám z internetu kdekoľvek a kedykoľvek.

V praxi si ľudia nezanedbateľnú časť informácií o svete vymieňajú v textovej podobe. Sú to napríklad názvy obchodov vystavené na ulici, rôzne reklamné slogany, plagáty, značky a podobne. Pri riešení môjho projektu som sa primárne zamerlal na túto formu informácií.

Predmetom projektu je teda vytvorenie mobilnej aplikácie využívajúcej obraz z kamery zariadenia pre efektívne, interaktívne a rýchle vyhľadávanie textu a textových častí obrazu na internete. Základnú ideu prezentuje obrázok č. 1. V nasledujúcich kapitolách popíšem zvolené princípy a postupy navrhnuté s ohľadom na implementáciu real-time aplikácie v mobilnom zariadení.



Obrázok 1: Ukážka použitia aplikácie.

## 2. POPIS RIEŠENIA

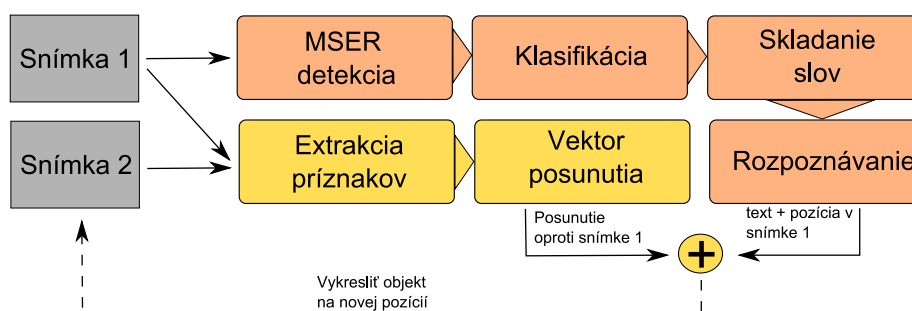
Aby bola práca s aplikáciou pre užívateľa čo najjednoduchšia a najkomfortnejšia, rozhodol som sa implementovať spracovanie obrazu z kamery v reálnom čase. Práca aplikácie sa skladá z viacerých častí. Základná a najdôležitejšia z nich je spracovanie a získanie textu z dodaného obrazu. V druhej časti spracovania je potrebné získaný text užívateľovi vhodne prezentovať a umožniť mu výber slov a riadkov, ktoré ho ďalej zaujímajú. Nakoniec je potrebné vybrané heslá vyhľadať v sieti internet. Blokovú schému návrhu riešenia ukazuje obrázok č. 2.

## 2.1. DETEKCIA A ROZPOZNÁVANIE TEXTU

Použitie bežných postupov a prostriedkov OCR je pre moju aplikáciu nevhodné, keďže spracovávam obrázky z reálnych scén. Tie sa často vyznačujú množstvom šumu, rôznymi odleskami, presvietením scény a podobne. Tieto metódy tiež počítajú s pomerne uniformnými vlastnosťami textov (napr. konštantná veľkosť písmen, použitie štandardných fontov), čo sa pri obrazoch z reálnych scén nedá vždy predpokladať. Pri implementácii rozpoznávania textu v obraze som vychádzal z prác [1] a [2]. V riešení som z nich využil najmä postupy detekcie, klasifikácie a spájania písmen do slov. Aby sa zvýšila rýchlosť spracovania na mobilnom zariadení, musel som niektoré postupy mierne pozmeniť. Spracovanie obrazu na text pozostáva z niekoľkých základných krokov.

**Detekcia oblastí** obrazu potenciálne predstavujúcich text je postavená na detekcii *maximálne stabilných extrémnych oblastí* (MSER) [3]. Samotnú detekciu spúšťam nad kanálmi *saturation* a *value* obrázku vo formáte HSV, pre zvýšenie odolnosti metódy voči šumu a farebnému gradientu textu. Výstupom detekcie sú oblasti reprezentované ako množiny bodov. Týmto krokom je tiež zaručená extrakcia oblastí od zdrojového obrazu (segmentácia).

**Klasifikácia nájdených oblastí** na písmena a oblasti nepredstavujúce písmená predstavuje ďalší krok spracovania. Na klasifikáciu nájdených MSER oblastí používam *support vector machine* (SVM) klasifikátor s jadrom RBF (*radial basis function*). Pri tréovaní som použil dáta získane MSER detekciou z 269 fotografií nápisov v reálnych scénach obsahujúcich približne 6650 písmen. Tréovacie aj testovacie dáta pochádzajú zo stránky súťaže ICDAR2003<sup>1</sup>.



Obrázok 2: Bloková schéma spracovania.

**Zhlukovanie oblastí do slov** prebieha v dvoch fázach. Najskôr sa pre každé písmeno vytvorí hypotéza slova, ktorú rozširujem o ďalšie oblasti (aj nepísmená) na základe ich špecifických vlastností (napr. ich vertikálna a horizontálna vzdialenosť, výška a šírka oblasti, aproximácia riadku a pod.). V druhom kroku sú hypotézy ďalej validované na základe vlastností ako napr. počet zahrnutých oblastí, pomer počtu písmen k počtu ostatných oblastí, zaplnenie obsahu hypotézy, počet oblastí na línii riadku a pod. Validáciou hypotéz obsahujúcich aj oblasti neklasifikované ako písmena sa snažím o zlepšenie výsledkov klasifikácie.

**Rozpoznávanie textu** z obrazu vytvorených slov sa deje pomocou externej knižnice Tesseract<sup>2</sup>.

## 2.2. UŽÍVATEĽSKÉ ROZHRIANIE

Nájdené a rozpoznané textové oblasti je nakoniec potrebné zobrazit' v užívateľskom rozhraní zvýraznením detekovanej oblasti v obraze z kamery. Z bloku spracovania textu je do hlavnej časti aplikácie predaná informácia o digitálnej podobe jednotlivých slov a okrajoch ich ohraničujúcich obdĺžnikov v obraze zo začiatku detekcie textu. Polohu rozpoznávaných textov je potrebné pri prezentácii na displej skorigovať podľa posunutia obrazu počas jeho spracovania. Je preto potrebné do aplikácie zahrnúť systém na extrakciu a tracking význačných bodov obrazu.

<sup>1</sup> <http://algoval.essex.ac.uk/icdar/Datasets.html>

<sup>2</sup> <http://code.google.com/p/tesseract-ocr/>

**Proces analýzy pohybu obrazu** sa začína ešte pred začatím každého cyklu vyhľadávania a rozpoznávania textových oblastí. Vo vstupnom snímku sa hľadajú význačné body obrazu vhodné na sledovanie. Na základe ich posunutia v nasledujúcich snímkach kamery sa pri ďalšom spracovaní určuje posunutie pohľadu kamery. Nájdený vektor posunutia obrazu sa následne použije pri výpočte polohy vykreslenia oblastí textu v danom snímku. Navrhnutý spôsob spracovania znamená, že výsledky rozpoznávania sú aktualizované v pravidelných intervaloch nezávislých na ich zobrazení v užívateľskom rozhraní. Počas procesu detekcie a rozpoznávania textov sú užívateľovi na displeji prezentované výsledky získané v predošlom cykle s aktualizovanou polohou.

**Nájdené slová sú v užívateľskom rozhraní** prezentované ako farebné obdĺžniky okolo nájdeného textu. Po zvolení textovej oblasti, klepnutím do vnútra daného obdĺžnika, aplikácia na základe jej korešpondujúceho digitálneho reťazca vytvorí požiadavku do vyhľadávania Google vo forme URL adresy. Nasleduje spustenie nainštalovaného internetového prehliadača s vytvorenou adresou.

### 3. IMPLEMENTÁCIA

Mobilná aplikácia bola implementovaná pre operačný systém Android. Hlavné prvky užívateľského rozhrania a základná kostra aplikácie je naprogramovaná v programovacom jazyku Java. Časti kódu, ktoré sú výkonnostne najviac náročné sú implementované v jazyku C++ pomocou nástrojov Android NDK. Operácie spracovania obrazu a klasifikácie som zrealizoval pomocou knižnice OpenCV. Na rozpoznávanie textu z extrahovaných slov bol použitý OCR systém Tesseract.

Vývoj aplikácie prebiehal na mobilnom zariadení Samsung Galaxy SII I9100 (Dual-core 1.2 GHz, 1GB RAM). Keďže je spracovanie obrazu na text navrhnuté sériovo, jeho rýchlosť závisí najmä na množstve nájdených MSER oblastí. Ani pri ich menšom počte a zmenšení vstupného obrazu však proces nedokázal bežať v reálnom čase. Pokúsil som sa to kompenzovať zavedením spracovania vo viacerých vláknoch. Prvé vlákno obsluhuje užívateľské rozhranie a prezentáciu výsledkov užívateľovi spolu s trackingom polohy nájdených reťazcov. V druhom prebieha spracovanie vstupného obrazu a extrakcia textov.

### 4. ZÁVER

Popísaná aplikácia si kladie za cieľ interaktívne a zaujímavo sprístupniť užívateľom moderných mobilných zariadení informácie o ich bezprostrednom okolí. Aplikácia vyhľadáva a rozoznáva textové reťazce v obraze z kamery mobilného zariadenia a výsledky následne interaktívne užívateľovi prezentuje vo forme odkazu do siete internet. Hlavnou výhodou riešenia je spracovávanie obrazu a prezentácia výsledkov vo videu v reálnom čase. Pre zrýchlenie procesu rozpoznania textov bolo potrebné metódy z ktorých som čerpal modifikovať a zjednodušiť, čím sa znížila ich presnosť a odolnosť voči šumom. V budúcnosti by bolo potrebné riešenie vylepšiť najmä v tejto oblasti.

### POĎAKOVANIE

Tento príspevok vznikol s podporou grantu FIT-S-11-2 a výskumného zámeru MSM 0021630528.

### REFERENCIE

- [1] Neumann, L.: Vyhledání a rozpoznání textu v obrazech reálných scén. [diplomová práce], České vysoké učení technické v Praze, Fakulta elektrotechnická, 2010.
- [2] Neumann, L., Matas, J.: Real-Time Scene Text Localization and Recognition. CVPR 2012 Providence, Rhode Island, USA.
- [3] Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing, 761-767, September 2004.