

Systém pro automatické detekování a rozpoznávání rostlin

Bc. Matěj Kolář, doc. Ing. Josef Chaloupka, Ph.D.

Abstrakt

Tento příspěvek se zabývá vytvořením systému pro automatické detekování a rozpoznávání rostlin. První část se zabývá detekcí, segmentací a klasifikací. Pro detekci a segmentaci je použita metoda GrabCut. Příznakové vektory obrazů rostlin jsou generovány pomocí metody HOG (histogramy orientovaných gradientů) pro popis tvaru a metodou klasterizace k-means v barevném prostoru HSV pro popis barvy. Systém dále využívá metody PCA (analýza hlavních komponent) pro redukci dimenze příznakových vektorů. Druhou částí systému je webový portál pro správu rostlin vytvořený pomocí NodeJS, HTML a Bootstrap.

Úvod

V rámci této práce byla popsána problematika rozpoznávání obrazu a knihovna OpenCV. Jako vzorový systém pro rozpoznávání byl nalezen systém pro rozpoznávání velkého počtu různých rostlin Velké Británie, vytvořený na Oxfordské univerzitě. Na jeho základě byl vybrán způsob segmentace obrazu a vytvoření příznakových vektorů tvaru a barvy. Jako databáze byla použita databáze 17 rostlin rostoucích ve Velké Británii pro testování a jako webová byla použita databáze bylin z portálu wendys.cz. Velkým problémem databází je rozmanitost obrazů rostlin, zejména různých úhlů focení. Díky tomu může být systém natrénován pro rozpoznávání rostlin v různých úhlech, avšak je snížena úspěšnost. Dále byly vyzkoušeny různé metody klasifikace pro použití v systému a vyhodnoceny na základě rozpoznávacího skóre. Poslední část se zabývá vytvořením webového systému pro správu informací a obrazů rostlin. Zde bylo nutné vytvořit webovou aplikaci, která je schopna komunikovat s rozpoznávací aplikací a splňovat požadavky na uchování dat.

Systém pro rozpoznávání rostlin

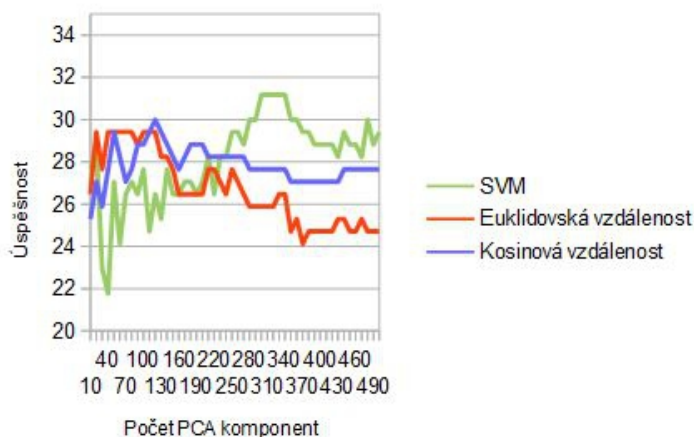
Algoritmus pro rozpoznávání rostlin byl naprogramován v jazyce Python, protože pro něj existuje velké množství knihoven pro práci s obrazem a také knihovna OpenCV. Metody segmentace obrazu byly vyzkoušeny dvě – metoda SRM (Statistical Region Merging) a metoda GrabCut. Jako lepší se ukázala metoda CrabCut, která dokázala oddělit pozadí od popředí rostliny s vysokou úspěšností. Takto vysegmentovaná data jsou použita pro vytvoření příznakových vektorů. Pro popis tvaru rostliny byly použity metody PCA, LARK a HOG. Nejúspěšnější metodou byla metoda HOG, která dokázala rostliny klasifikovat s přesností přes 55 %. Pro tuto metodu bylo použito redukování dimenze pomocí metody PCA pro zrychlení výpočtů a snížení paměťových nároků. Klasifikační metody byly zvoleny dvě – metoda nejbližšího vzorku a klasifikátor SVM. Pro metodu nejbližšího vzorku byla použita euklidovská a kosinová metrika. V porovnání s klasifikátorem SVM však dosahovaly nižší úspěšnosti a tak nebyly dále využívány.

Webová aplikace byla vytvořena pomocí systému Node.js a databáze MongoDB. Aplikace využívá systém šablon Jade a další knihovny pro systém Node.js pro práci se soubory, databází, synchronizací vláken či přihlašování a ověřování uživatelů. Databáze MongoDB byla využita pro ukládání základních informací a popisu rostlin. Obrazy rostlin jsou ukládány pomocí souborového systému.

Výsledky a diskuze

Systém využívá pro klasifikaci metodu SVM, která je neúspěšnější pro tuto konkrétní úlohu klasifikace viz Obrázek 1. Jako příznakový vektor pro popis rostliny byla použita kombinace metody pro popis tvaru HOG a metody pro popis barvy pomocí klasterizace barev k-means v barevném prostoru HSV. Výsledkem práce je systém umožňující automatickou detekci a rozpoznání rostliny z obrazu s úspěšností 57,59 % při použití testovací databáze viz Tabulka 1.

Tabulka 1. Srovnání metod rozpoznávání rostlin



Metoda	Úspěšnost
PCA	31,20%
PCA+GC	32,28%
PCA+GC+LARK	42,60%
PCA+GC+HSV	49,06%
PCA+GC+HOG	55,97%
PCA+GC+HOG+HSV	57,59%

Obrázek 1. Srovnání klasifikačních metod

Závěr

Byl vytvořen systém pro automatické detekování a rozpoznávání rostlin. Nejprve byl vytvořen algoritmus pro rozpoznávání, který umožňuje detekování a klasifikaci rostlin v obraze. Byly porovnány různé metody segmentace a poté vytvořeny příznaky tvaru a barvy. Pro klasifikaci byl vyhodnocen jako nejlepší SVM. Systém dosahuje nejlepší úspěšnosti **57,59 %** při použití testovací databáze.

Následně byl tento algoritmus propojen s webovou aplikací umožňující správu informací a obrazů rostlin. Pomocí databáze rostlin je natrénován algoritmus pro rozpoznávání a je umožněno rozpoznávání z uživatelem nahraného obrazu. Výsledkem je 10 nejlepších shod rostlin seřazených podle pravděpodobnosti rozpoznávání.

Reference

- [1] NILSBACK, M-E. a A. ZISSERMAN. Automated Flower Classification over a Large Number of Classes. 2008 Sixth Indian Conference on Computer Vision, Graphics [online]. IEEE, 2008, , 722-729 [cit. 2016-05-15]. DOI: 10.1109/ICVGIP.2008.47. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4756141>
- [2] ROTHER, C., V. KOLMOGOROV a A. BLAKE. "GrabCut". ACM SIGGRAPH 2004 Papers on - SIGGRAPH '04 [online]. New York, New York, USA: ACM Press, 2004, 309- [cit. 2016-05-08]. Dostupné z: <http://portal.acm.org/citation.cfm?doid=1186562.1015720>
- [3] DALAL, N. a B. TRIGGS. Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) [online]. IEEE, 2005, 886-893 [cit. 2016-05-08]. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1467360>