Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Bakalářská práce

Automatická anotace obrázků

Plzeň 2017 Kateřina Kratochvílová

Místo této strany bude zadání práce.

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 21. dubna 2017

Kateřina Kratochvílová

Abstract

The text of the abstract (in English). It contains the English translation of the thesis title and a short description of the thesis.

Abstrakt

Text abstraktu (česky). Obsahuje krátkou anotaci (cca 10 řádek) v češtině. Budete ji potřebovat i při vyplňování údajů o bakalářské práci ve STAGu. Český i anglický abstrakt by měly být na stejné stránce a měly by si obsahem co možná nejvíce odpovídat (samozřejmě není možný doslovný překlad!).

Obsah

1 Úvod **6**

2 Úvod **7**

3 JEC Joint Equal Contribution **8**
3.1 Příznaky 8
3.1.1 Barva 8
3.1.2 textura 10
3.2 Vzdálenosti 10
3.3 Textura 10
3.4 Kombinace vzdáleností 11
3.5 Přenesení klíčových slov 11

4 Testovací databáze **13**
4.1 iaprtc12 13
4.2 ČTK 13
4.3 ESP 14

5 Návrh systému **15**

6 Použité programové prostředky **16**
6.1 OpenCV 16

7 Vyhodnocení výsledků **17**

8 Závěr **18**

9 Uživatelská dokumentace **19**

Literatura **20**

1 Úvod

V souboru `literatura.bib` jsou uvedeny příklady, jak citovat knihu [4], článek v časopisu [2], webovou stránku [3].

2 Úvod

V dnešní době, kdy je svět přesycen obrázky v digitální podobě, není vůbec snadné nalézt obrázek zobrazující požadovaný obsah. Naneštěstí počítače nedokáží vnímat obraz jako lidé, vnímají totiž obrazy jako sérii binárních informací. Přitom počítače a jejich práce s obrazy by se dala využít v mnoha oborech jako je lékařství nebo doprava. Na základě toho vyplouvá na povrch problém jak spravovat digitální obrázky a efektivně mezi nimi vyhledávat. Prostřednictvím klíčových slov přiřazených k obrázkům se dá problém vyhledávání zjednodušit. Přiřazení klíčových slov probíhá pomocí procesu automatické anotace obrázků. Klíčová slova přiřazená k obrázku by měla vyjadřovat jeho obsah (například les, strom). Při reálném použití můžeme ovšem narazit na problém při zadávání abstraktních slov, například šťastná rodina.

Pro automatickou anotaci obrázků se používá strojové učení. Můžeme ji rozdělit na dvě části. V první části získáme klíčové příznaky ve druhé už je samotná anotace, tedy přidělení klíčových slov. Abychom tento postup mohli provést v praxi, musíme nejdřív klasifikátor natrénovat pomocí trénovací množiny. Trénovací množina je množina obrázků, která již má ke každému obrázku přidána metadata s klíčovými slovy připravenými od lidí. Vybrané obrázky v trénovací množině musí být různorodé, aby anotace probíhala správně. Pojem automatická anotace obrázků je jednoduše řečeno proces, při kterém jsou k obrázku automaticky přiřazena metada, která obsahují klíčová slova.

Cílem práce je navrhnout a implementovat software umožňující automatickou anotaci obrázků. Konkrétně se bude zabývat metodou JEC. Práce bude využívat nízkoúrovňové příznaky, konkrétně barvu a texturu. Metodu budeme zkoušet na standardních datech, následně se budeme snažit výsledky vylepšít, a porovnat s další metodou a literaturou.

# 3 JEC Joint Equal Contribution

Tato metoda je založena na hypotéze, že podobné obrázky mají podobná klíčová slova. Pomocí metody hledání nejbližších sousedů (dále jen KNN) najdeme K nejpodobnějších obrázků. Přičemž klíčová slova od jednotlivých sousedů jsou posuzována odlišně a to právě na základě toho o kolik se s testovaným obrázkem liší. Metoda je postavena na dvou typech příznaků - barevných a texturových. [1]

3.1 Příznaky

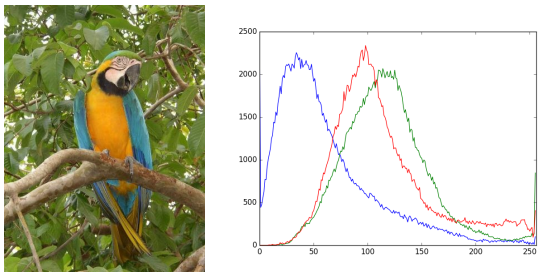
Barva a textura jsou považovány za dva nejdůležitější nizkoúrovňové příznaky pro obrázkovou reprezentaci. Nejběžnější barevné deskriptory jsou barevné histogramy, které jsou často využívány pro porovnávání a indexování obrázků, zejména z důvodu jejich efektivnosti a snadného výpočtu. K vytvoření texturových příznaků se používají Haarovy a Gaborovy wavelety a to především z důvodu že jsou efektivní při vytváření rozptýlených diskriminativních obrázkových rysů.

3.1.1 Barva

U digitálního obrazu je barva reprezentovaná n-rozměrným vektorem. Jeho velikost a význam jednotlivých složek (tzv. barevných kanálů) zavisí na příslušném barevném prostoru. Počet bitů použitých k uložení buď celého vektoru nebo jeho jednotlivých složek se nazývá barevná hloubka (totožně bitová hloubka). Obvykle se můžeme setkat s hodnotami 8, 12, 14 a 16 bitů na kanál.
V použité metodě získáme vlastnosti z obrázků ve třech rozdílných barevných prostorech: RGB, HSV a LAB. RGB (Red, Green, Blue) je nejpoužívanější barevný prostor pro zachycení obrázu nebo jeho zobrazení. Oproti tomu HSV (Hue, Saturation and Value) se snaží zachytit barevný model tak jak ho vnímá lidské oko, ale zároveň se snaží zůstat jednoduchý na výpočet. Hue znamená odstín barvy (měří se jako poloha na standartním barevném kole $0° - 360°$), saturation je systost barvy (množství šedi v poměru k odstínu 0% šedá barva - 100% plně sytá barva) a value je hodnota jasu nebo také množství bílého světla (relativní světlost nebo tmavost barvy). Některé

kombinace hodnot H, S a V mohou dávat nesmyslné výsledky. RGB je závislý na konkrétním zařízení, nemůže dosáhnout celého rozsahu barev, které vidí lidské oko, zatímco barevný model LAB je shopen obsáhnout celé viditelné spektrum a navíc je nezávislý na zařízení. L (ve zkratce LAB) značí Luminanci (jas dosahuje hodnot 0 - 100, kde 0 je černá a 100 je bílá). Zbylé A a B jsou dvě barvonosné složky, kdy A je ve směru červeno/zeleném a B se pohybuje ve směru modro/žlutém.

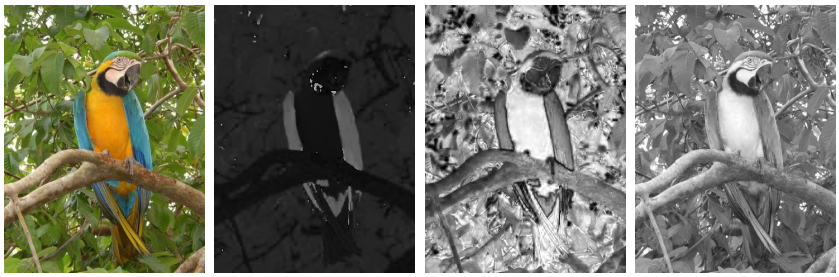
Pro RGB, HSV i LAB použijeme barevnou hloubku 16 bitů na kanál histogramu v jejich příslušném barevném prostoru.



Obrázek 3.1: RGB histogram - zastoupení jednotlivých složek v obrázku

Obrázek 3.2: Barevný prostor RGB a jeho jednotlivé složky v pořadí R, G, B

Obrázek 3.3: Barevný prostor LAB a jeho jednotlivé složky v pořadí L, A, B



Obrázek 3.4: Barevný prostor HSV a jeho jednotlivé složky v pořadí H, S, V

3.1.2 textura

Jako reprezentace textur budou použity Gabor a Haar wavelety. Každý obrázek bude filtrován s Gabor wavelet na třech škálách a čtyřech orientacích.

3.2 Vzdálenosti

K určení příslušné vzdálenosti se můžeme setkat se čtyřmi měřítky vzdálenosti pro histogramy a rozdělení Kullaback-Leibler divergence KL - divergence, χ^2 statistika, L1 - vzdálenost a L2 - vzdálenost. Na RGB a HSV je nejlépší použít L1 zatímco pro LAB je nejvhodnější KL - divergence.

Problém s KL - divergencí nastává pouze tehdy, když se histogramy nebudou shodovat v nulách. Jeden předpoklad pro fungování tohoto vzorce je totiž že když je $Q(i) = 0$ tak zároveň musí být i $P(i) = 0$.

$$D_{KL}(P||Q) = \sum_i P(i) \log_e(\frac{P(i)}{Q(i)}) \tag{3.1}$$

3.3 Textura

Gabor filter je lineární filter používaný pro detekci hran. Frekvence a orientace reprezentující Gabor filter je podobná lidskému vnímání a jsou zvláště vhodné pro reprezentaci textury a rozlišování.

Gabor filtr reaguje na hrany a texturové změny.

Obrázky jsou filtrovány za použité reálné části z různých odlišních Gabor filtrů jader. Průměrný a rozptyl filtrovaných snímků jsou pak použity jako funkce pro klasifikaci, která je založena na nejméně kvadratické chyby pro jednoduchost.

TODO Dohledat Haar a Gabor wavelety, přidat vzorečky a zase klidně i obrázky

3.4 Kombinace vzdáleností

Nejrozumnějším přístupem ke zkombinování vzdáleností od různých desktiptorů je aby jednotlivé vzdálenosti přispívali rovnoceně. Z tohoto důvodu je potřeba vzdálenosti přeškálovat na jednotné měřítko.
Označme si I_i jako i-tý obrázek a řekněme, že máme N jeho příznaků $f_i^1, ..., f_i^N$. Nadefinujme si $d_{(i,j)}^k$ jako vzdálenost mezi příznaky f_i^k a f_j^k. Chtěli bychom zkombinovat všechny vzdálenosti příznaků mezi obrázky I_i a I_j tedy $d_{(i,j)}^k$, $k = 1, ..., N$. Vzdálenosti nám ale v praxi nevyjdou tak aby měli stejný poměr na výsledku, proto předtím než vzdálenosti zkombinujeme musíme je normalizovat do jednotné formy. Získáme maximální a minimální hodnotu pro každý příznak a na základě toho hodnotu přeškálujeme na interval od 0 do 1. Jestliže označíme přeškálovánou vzdálenost jako $\tilde{d}_{(i,j)}^k$ následně můžeme označit kompletní vzdálenost mezi obrázky I_i a I_j jako (3.2) Joint Equal Contribution (JEC).

$$JEC = \sum_{k=1}^{N} \frac{\tilde{d}_{(i,j)}^k}{N} \tag{3.2}$$

3.5 Přenesení klíčových slov

Pro přenesení klíčových slov používáme metodu, kdy přeneseme n klíčových slov k dotazovanému obrázku $\tilde{I}$ od K nejbližších sousedů z trénovací sady. Mějme $I_i, i = 1, ..., K$,těchto K nejbližší sousedů seřadíme podle vzrůstající vzdálenosti (tzn. že I_1 je nejvíce podobný obrázek). Počet klíčových slov k danému I_i je označen jako $|I_i|$. Dále jsou popsány jednotlivé kroky alogoritmu na přenesení klíčových slov.

1. Seřadíme klíčová slova z I_1 podle jejich frekvence výskytu v trénovací sadě.

2. Ze všech $|I_1|$ klíčových slov z I_1 přeneseme n nejvýše umístěná klíčová slova do dotazovaného $\tilde{I}$. Když $|I_1| < n$ pokračujte na krok 3.

3. Seřadíme klíčová slova sousedů od I_2 do I_K podle dvou faktorů

   (a) výskytu v trénovací sadě s klíčovými slovy přenesených v kroku 2

   (b) místní frekvence (tj. jak často se vyskytují jako klíčová slova u obrázků I_2 až I_K). Vybereme nejvíce vyskytující $n - |I_1|$ klíčových slov převedených do $\tilde{I}$.

Tento algoritmus pro přenos klíčových slov je poněkud odlišný od algoritmů, které se běžně používají. Jeden z běžně užívaných funguje na principu, že klíčová slova jsou vybrána od všech sousedů (se všemi sousedy je zacházeno stejně bez ohledu na to jak jsou danému obrázku podobní), jiný užívaný algoritmus k sousedům přistupuje váženě (každý soused má jinou váhu)a to na základě jejich vzdálenosti od testovaného obrázku. Při testování se ovšem ukázalo, že tyto přímé přístupy přináší horší výsledky v porovnání s použitým dvoufaktorovým algoritmem pro přenos klíčových slov.
V souhrnu použitá metoda je složenina ze dvou složenin a to obrázkové vzdálenosti (JEC) a výše popsaným algoritmem na přenášení klíčových slov.

4 Testovací databáze

Pro natrénování a následné testování byla použita data z databází ČTK, ESP a IAPRC. Kolekce obrázků na natrénování muís být pečlivě vybrána aby zahrnovala co možná největši okruh z různých témat.

4.1 iaprtc12

Data iaprtc12 obsahují 20 000 obrázků ve formátu jpg s celkovým počtem 291 klíčových slov. Ke každému obrázku jsou přiložena metadata ve formátu XML, která obsahují informace o obrázku v různých jazycích. Kromě angličtiny je tam i například španělština nebo němčina. V metadatatech ovšem nenajdeme klíčová slova tak jak by jsme si je představovali, ale v různých tagách nalezneme například titulek obrázku, který může vypadat například The Plaza de Armas, a v tagu description je například a woman and a child are walking over the square.

Obrázek 4.1: Ukázka obrázku s klíčovými slovy: front lake man mountain rock sky summit

4.2 ČTK

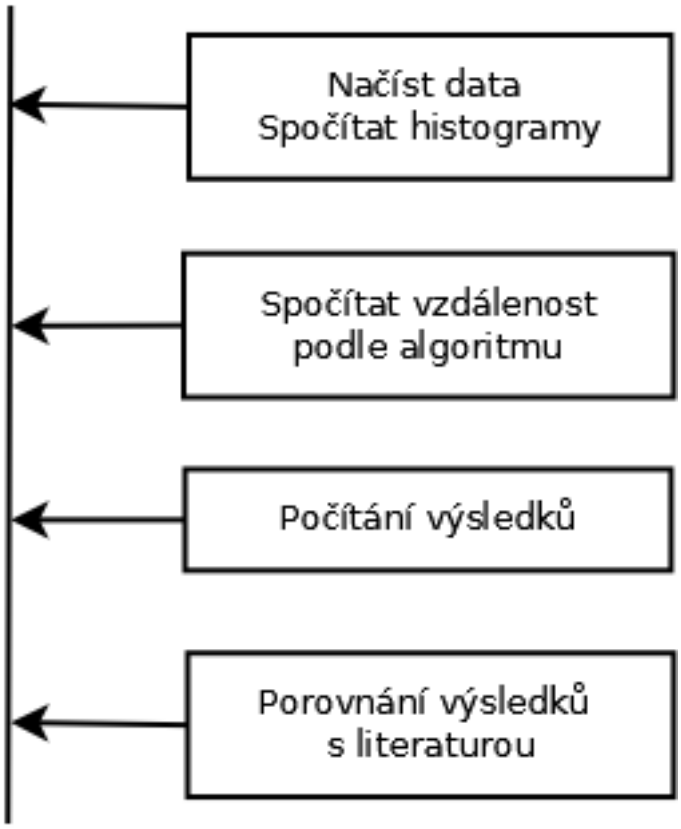
Data od ČTK obsahují 3 383 obrázků. Ke každému obrázku jsou přidána metadata ve formátu XML, která obsahují klíčová slova a další informace o obrázku.

4.3 ESP

Data od ESP obsahují 67 796 obrázků ve formátu jpg. Ke každému obrázku je přiřazen soubor ve formátu desc, který obsahuje anglické anotace.

5 Návrh systému

Systém byl navržen jako modulový a to z důvodu snadné obměny některé z částí, což je výhodné zejména pokud bychom potřebovali například spočítat vzdálenosti vektorů podle jiného algoritmu.



Obrázek 5.1: Návrh systému

6 Použité programové prostředky

Program byl navržen na operační systému Linux. Jako programovací jazyk byl zvolen Python a to z důvodu jeho jednoduchého použití, což je na prototyp, jako je tento velice výhodné na časovou náročnost. Program využívá knihovnu OpenCV 3.1.

6.1 OpenCV

OpenCV (Open source computer vision) je knihovna vydávána pod licencí BSD a je volně k dispozici jak pro akademické účely, tak pro komerční použití. Je vhodná pro použití v C++, C, Python a Javě. Podporuje operační systémy Windows, Linux, Mac OS, iOS a Android.
Knihovna byla navrhnuta pro výpočetní efektivitu v oblasti počítačového vidění a zpracování obrazu se zaměřením na zpracování obrazu v reálném čase. Z důvodu optimalizace byla napsána v C/C++.
Knihovnu OpenCV je možné stáhnout na adrese: http://opencv.org/

7 Vyhodnocení výsledků

Zpracování výsledků probíhá jako porovnání anotací přidělených člověkem s anotacemi přidělenymi klasifikátorem. Označme si w_{auto} jako počet obrázků, kterým bylo dané slovo přiřazeno klasifikátorem, w_{human} počet obrázků, kterým bylo dané slovo přiřazeno člověkem a $w_{corrrectly}$ jako počet obrázků, kterým bylo slovo přiřazené správně. U klasifikátorů se počítá precision (přesnost) a recall (úplnost) pro každé slovo v testovací sadě.
Recall (7.1) je počet obrázků správně anotovaných s daným slovem děleno počtem obrázků, kterým bylo toto slovo přiděleno v anotaci člověkem. Precision (7.2) je počet správně anotovaných obrázků s tímto slovem děleno celkovým počtem anotovaných obrázků s tímto slovem (správně nebo ne). [5]

$$Rec = \frac{w_c}{w_h} \tag{7.1}$$

$$Prec = \frac{w_c}{w_a} \tag{7.2}$$

8 Závěr

V teoretické části byly popsány nízkoúrovňové příznaky barva a textura. Byla rozebrána metoda JEC, která bude v bakalářské práci implementována. Seznámili jsme se s knihovnou OpenCV, prostudovali obrázky a přiložená metadata od ČTK, ESP a iaprtc12.

9 Uživatelská dokumentace

popsani jak vypada zdrojovej soubor kterej to zere, nejdriv cesta k souboru a pak jeho klicovy slova

Literatura

[1] Ameesh Makadia, S. K. V. P. A new baseline for image annotation. *Commun. ACM*. July 1961, 4, 7, s. 321. ISSN 0001-0782. doi: 10.1145/366622.366644. Dostupné z: http://doi.acm.org/10.1145/366622.366644.

[2] Hoare, C. A. R. Algorithm 64: Quicksort. *Commun. ACM*. July 1961, 4, 7, s. 321. ISSN 0001-0782. doi: 10.1145/366622.366644. Dostupné z: http://doi.acm.org/10.1145/366622.366644.

[3] *Class Graphics2D* [online]. Oracle, 2016. [cit. 2016/03/09]. Java SE Documentation. Dostupné z: https://docs.oracle.com/javase/7/docs/api/java/awt/Graphics2D.html.

[4] Knuth, D. E. *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*. Addison-Wesley Longman Publishing Co., Inc., 1997. ISBN 0-201-89684-2.

[5] V. Lavrenko, J. J. R. A model for learning the semantics of pictures. *Commun. ACM*. July 1961, 4, 7, s. 321. ISSN 0001-0782. doi: 10.1145/366622.366644. Dostupné z: http://doi.acm.org/10.1145/366622.366644.