

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-5220-17452

Bc. Eduard Kuric

AUTOMATICKÉ ANOTOVANIE
FOTOGRAFIÍ ZALOŽENÉ NA ANALÝZE
VIZUÁLNEHO OBSAHU

Diplomová práca

Vedúci práce:
prof. Ing. Mária Bieliková, PhD.

máj 2011

ANOTÁCIA

Slovenská technická univerzita v Bratislave
FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ
Štúdijný program: SOFTVÉROVÉ INŽINIERSTVO

Autor: Bc. Eduard Kuric

Diplomový projekt: Automatické anotovanie fotografií založené na analýze vizuálneho obsahu

Vedenie diplomového projektu: prof. Ing. Mária Bieliková, PhD.
máj 2011

S narastajúcou popularitou digitálnych fotoaparátov a mobilných telefónov sú kladené vyššie nároky na rýchle a presné vyhľadávanie medzi fotografiami. Extrakcia obsahu z fotografií je omnoho zložitejšia ako pri textových dokumentoch, pretože neobsahujú jednotky akými sú slová. Vyhľadávanie sa spravidla realizuje pomocou anotácií, ktoré sú k fotografiám pridané používateľmi. Ručné vytváranie anotácií je však časovo náročné a výsledky sú často subjektívne, a preto je automatické anotovanie fotografií výzvou.

Úlohou automatických anotačných systémov je nájsť najviac podobné fotografie v korpuse k cieľovej fotografii, ktorý obsahuje dobre anotované fotografie. Podobnosť fotografií je vyhodnotená porovnaním lokálnych alebo globálnych vlastností, ktoré sú z nich extrahované. Z vyhľadaných najlepších kandidátov sú extrahované kľúčové slová, ktoré sú následne priradené cieľovej fotografii.

V našej práci sme navrhli novú metódu pre automatické anotovanie fotografií. Na extrakciu lokálnych vlastností sme použili metódu SIFT (angl. *Scale Invariant Feature Transform*) a na extrakciu globálnych vlastností JCD (angl. *Joint Composite Descriptor*) v kombinácii s Lokálne senzitívnym hašovaním. Lokálne vlastnosti sú oveľa viac diskriminujúce ako globálne vlastnosti. Pri vyhľadávaní špecifických objektov je táto vlastnosť vítaná, avšak pri vyhľadávaní všeobecných kategórií je prekážkou. V našej metóde kombinujeme lokálne a globálne vlastnosti na získanie najlepších výsledkov. Pri hľadaní podobných kandidátov na extrakciu kľúčových slov sa zameriavame na analýzu fotografií z pohľadu pravdepodobnosti, že obsahujú správne slová pre cieľovú fotografiu. Napríklad, preferujeme fotografie, na ktorých sú extrahované objekty záujmu z cieľovej fotografie dominantnejšie, alebo je ich frekvencia výskytu vyššia. Pri takýchto fotografiách je väčšia pravdepodobnosť, že kľúčové slová opisujú práve objekty záujmu. Výsledkom je, že sme schopní pomenovať objekty priamo na cieľovej fotografii.

Metódu sme overili a vyhodnotili na korpuse Corel5K, ktorý obsahuje 5000 fotografií, rozdelených na tréningovú a testovaciu množinu a každej fotografii je priradených 1 – 5 slov.

ANNOTATION

Slovak University of Technology in Bratislava

FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Degree Course: SOFTWARE ENGINEERING

Author: Bc. Eduard Kuric

Diploma project: Automatic Photo Annotation Based on Visual Content Analysis

Supervisor: prof. Ing. Mária Bieliková, PhD.

2011, May

With increasing popularization of digital and mobile phone cameras, there occurs a need of quick and exact searching. Content based indexing of photos is more difficult than text documents because they do not contain units like words. Searching is based on annotations and semantic keywords that are entered by a user and associated with photos. However, manual creating of annotations is very time-consuming and results are often subjective. Therefore, automatic photo annotation is most challenging task.

The aim of automatic annotation system is to find the most similar photos for a target photo in a corpus which contains well-annotated photos. Similarity of photos is evaluated of comparing local or global features which are extracted from them. After a retrieval process, related keywords are extracted from the best candidates and associated with the target photo.

In our work, we propose a novel method for automatic annotating photos. To capture local information that is essential in object retrieval, we use *Scale Invariant Feature Transform* and to capture global features, we use *Joint Composite Descriptor* in combination with a hash-based method known as *Locality Sensitive Hashing*. Local features are much more precise and discriminating than global descriptors. By searching a specific object, this feature is welcome, but by searching general categories it is an obstacle. In our method, we combine local and global features to retrieve the best results for the target photo. By searching similar candidates for extraction of keywords, we focus on photos analysis in terms of probability, that the retrieved photos contain the right keywords for the target photo. For example, we prefer photos where extracted objects of interest from the target photo are dominant in retrieved photos or their frequency of occurrence is greater, because it is more likely, that such photos contain the right keywords. The result is that we are able to name objects directly on the target photo.

We have evaluated our method on the Corel5K dataset which contains 5000 photos. The dataset is divided to training and testing set and each photo is associated with 1-5 tags.

ZAD

ZAD – NAVRH OLD

ZAD – NAVRH OLD

ZAD – NAVRH NEW

ZAD – NAVRH NEW

Obsah

1	Úvod.....	1
2	Vyhľadavanie fotografií.....	3
2.1	Súčasn� trendy	3
2.1.1	Vyhľadavanie podľa kateg�rie	4
2.1.2	Vyhľadavanie podľa zameru.....	5
2.1.3	Vyhľadavanie na z�klade asoci�ci�.....	6
2.2	Deskriptory pou�ivan� v obrazovom vyhľadan�	7
2.2.1	Priemer farieb	7
2.2.2	Deskriptory MPEG-7	8
2.2.3	CCD	9
2.2.4	Farebn� korelogramy	10
2.2.5	Tamurove text�rov� pr�znaky	10
2.2.6	SIFT	12
3	Automatick� anotovanie fotografi�	15
3.1	Pravdepodobnostn� modely automatick�ho anotovania fotografi�	15
3.1.1	Co-occurrence model	15
3.1.2	Translation model	17
4	Zhodnotenie s�časn�ho stavu a ciele pr�ce	19
5	N�vrh met�dy pre automatick� anotovanie fotografi�	21
5.1	Extrakcia vlastnost�	22
5.1.1	V�po�et lok�lnych vlastnost�.....	22
5.1.2	V�po�et glob�lnych vlastnost�.....	23
5.2	Indexovanie vlastnost� do LSH �lo�iska.....	23
5.3	Vyhľadanie podobn�ch vlastnost� k dopytu.....	25

5.4	Vyhodnotenie a priradenie slov	26
6	Overenie metódy	29
7	Zhodnotenie.....	33
	Literatúra	35
	Príloha A – Technická dokumentácia	1
	A.1 Dátový model LSH úložiska	1
	A.2 Ukážka zdrojového kódu.....	3
	A.3 Nástroj ANNO ^R	5
	Príloha B – Ukážka výsledkov anotovania	1
	Príloha C – Obsah dátového nosiča.....	1
	Príloha D – Návrh článku na konferenciu SMAP 2011	1

1 Úvod

Fotografie, či už v digitálnej podobe alebo klasickej, sa radia medzi najpopulárnejšie médium zachytávajúce život a udalosti – náš ľudský život. S narastajúcou obľúbenosťou a kvalitou digitálnych fotoaparátov sa zvyšuje počet digitálnych fotografií, ktoré sa akumulujú v on-line úložiskách, resp. osobných fotogalériách (fotoalbumoch). Tieto fotografie zachytávajú každodenné zážitky ľudí, rôznorodé okamihy, príbehy jednotlivcov, skupín, komunít a pod.

Väčšina úložísk pre fotografie umožňuje vyhľadávať medzi fotografiami pomocou kľúčových slov - anotácií, ktoré sú s nimi asociované (*text-based photo retrieval*). Vytváranie anotácií pre fotografie z ich vizuálneho obsahu je podstatne zložitejšie ako pre textové dokumenty, pretože neobsahujú jednotky akými sú slová. Fotografie sú vo väčšine prípadov preto anotované samotným používateľom ručne – pripojením kľúčových slov. Takýto prístup je časovo veľmi náročný a je v istej miere subjektívny. Často sa stáva, že anotácie k fotografiám chýbajú, sú nepostačujúce, alebo priradené slová neopisujú fotografiu korektne. Dvaja ľudia môžu rovnakú fotografiu opísať rôzne.

Automatické anotovanie fotografií v súčasnosti priamo závisí od kvality metód na vyhľadávanie podobných fotografií a od trénovacej množiny. K výskumu metód pre vyhľadávanie podobných fotografií sa zapájajú nové poznatky z digitálneho spracovania obrazu a počítačového videnia. Na získanie opisov k novým fotografiám je spravidla potrebné nájsť v úložisku najskôr podobné fotografie, ktoré už disponujú anotáciami. Z daných fotografií sa následne extrahujú prislúchajúce kľúčové slová, ktoré sa priradia k novým fotografiám. Rozdiel medzi automatickým a poloautomatickým procesom anotácie je v tom, že poloautomatický proces zahŕňa aj spätnú väzbu od používateľa, tzn. výstupom procesu sú návrhy kľúčových slov k novej fotografii a tieto následne používateľ potvrdí, odmietne, príp. upraví.

V našej práci navrhujeme novú metódu pre automatické anotovanie fotografií. Na extrakciu lokálnych vlastností používame metódu *Scale Invariant Feature Transform* a na extrakciu globálnych vlastností *Joint Composite Descriptor* v kombinácii s Lokálne senzitívnym hešovaním. Lokálne vlastnosti sú oveľa viac diskriminujúce ako globálne vlastnosti. Pri vyhľadávaní špecifických objektov je táto vlastnosť vítaná, avšak pri vyhľadávaní všeobecných kategórií je prekážkou. V našej metóde kombinujeme lokálne a globálne vlastnosti na získanie najlepších výsledkov. Pri hľadaní podobných kandidátov na extrakciu kľúčových slov sa zameriavame na analýzu fotografií z pohľadu pravdepodobnosti, že obsahujú správne slová pre cieľovú fotografiu. Napríklad, preferujeme fotografie, na ktorých sú extrahované objekty záujmu z cieľovej fotografie dominantnejšie, alebo je ich frekvencia výskytu vyššia. Pri takýchto fotografiách je väčšia pravdepodobnosť, že kľúčové slová opisujú práve objekty záujmu. Tiež sa zameriavame na identifikáciu menej dominantných objektov. Výsledkom je, že sme schopní pomenovať objekty priamo na cieľovej fotografii.

2 Vyhľadávanie fotografií

2.1 Súčasné trendy

Súčasný výskum sa koncentruje prevažne na vyhľadávanie fotografií podľa ich obrazovej informácie, tzn. vizuálneho obsahu (*content-based image retrieval*). Vyhľadávanie fotografií sa zvykne deliť do troch scenárov: podľa kategórie, zámeru a asociácie [29]. Pri vyhľadávaní podľa kategórie sa analyzuje, či daná fotografia obsahuje požadovaný objekt – prebieha identifikácia objektov záujmu. Cieľom vyhľadávania podľa zámeru je lokalizovať fotografie, ktoré majú náležitú podobnosť. Tomu predchádza dopyt v podobe vzorovej fotografie, príp. skice vytvorenej používateľom. Pri vyhľadávaní na základe asociácií, používateľ nedisponuje konkrétnou predstavou. Vyhľadávanie prebieha prehliadaním úložiska (priestoru) samotným používateľom a pri výbere fotografie sú mu ponúknuté podobné fotografie pomocou rôznych vizualizačných techník.

Analýza podobnosti fotografií na základe obrazovej informácie sa zvykne deliť na dve úrovne. Na nízkej úrovni sa spracovávajú vlastnosti akými sú farba, textúra, tvar a pod. – jedná sa o tzv. globálne vlastnosti. Pri vysokej úrovni sa vychádza z globálnych vlastností a identifikuje sa celková štruktúra fotografie, tzn. samotné objekty. Vyhľadávanie podobných fotografií je založené na porovnaní deskriptorov, ktoré opisujú vlastnosti (objekty). Vizuálna identifikácia objektov strojom je netriviálna, časovo náročná a často je možné rozoznať s uspokojivými výsledkami len určitú množinu objektov, pre ktorú je algoritmus navrhnutý. Plne spoľahlivé a univerzálne riešenia, ktoré by jednoznačne identifikovali objekty na fotografiách v reálnom čase v súčasnosti prakticky neexistujú.

Vyhľadávanie vizuálne podobných fotografií len na základe globálnych vlastností je rýchle, avšak pre automatické anotovanie je väčšinou nepostačujúce – nespoľahlivé. Tieto vlastnosti a ich kombinácie nie sú schopné opísať (zachytiť) vysoko-úrovňové koncepty „v mysli“ používateľa. Ten vníma na fotografii scénu (pozadie) a objekty (popredie) osobitne. Svoj dopyt formuluje spôsobom „nájdi fotografie, na ktorých je západ slnka“ a nie spôsobom „nájdi fotografie, na ktorých sú červené a žlté farby“. Snahou súčasného výskumu sú preto nové modely založené na sémantickej reprezentácii fotografií. Takáto reprezentácia je vytvorená procesom, ktorý na základe extrakcie globálnych vlastností vytvorí zhluky segmentov s podobnými vlastnosťami. Z týchto sú vytvorené oblasti, ktoré predstavujú kandidátov na identifikované objekty. K daným objektom sa následne (polo)automatickým procesom priradia kľúčové slová. Takáto reprezentácia umožňuje používateľovi vytvárať dopyty, ktoré sú intuitívne, presnejšie a sú schopné zachytiť jeho predstavu (strom vľavo, auto na ceste vpravo).

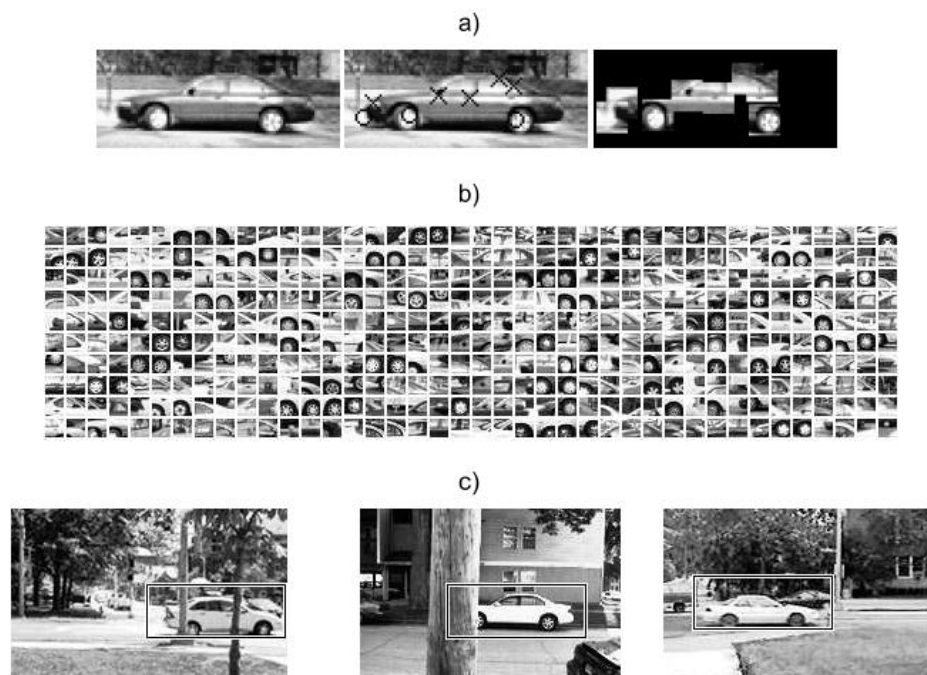
Ako sme uviedli, vyhľadávanie fotografií na základe podobnosti sa zvykne deliť do troch scenárov, a to podľa kategórie, zámeru a asociácie. V nasledujúcej časti uvedieme pre každý zo scenárov niekoľko zaujímavých prístupov.

2.1.1 Vyhľadávanie podľa kategórie

Pri vyhľadávaní podľa kategórie sa analyzuje, či daná fotografia obsahuje požadovaný objekt – identifikácia objektov záujmu. Zvykne sa deliť do dvoch základných tried, a to modelovo orientované prístupy a štatisticky orientované.

Modelovo orientované prístupy sú založené na geometrických obmedzeniach objektov, ktoré sú zahrnuté priamo v navrhnutých algoritmoch. Môže sa jednať o charakteristické vlastnosti objektu vo forme šablóny, s ktorou sú navzájom porovnávaní kandidáti pri samotnom vyhľadávaní. Prvý výskum v roku 1973, ktorý sa zaoberal hľadaním objektov bol zameraný na lokalizáciu jednoduchých tvárových čít na fotografiách [18]. Výsledky neboli nijak zvlášť okázalé. Od tých čias však výskum značne pokročil a v roku 2001 bol publikovaný prielom v identifikácii tvári [31]. Publikované výsledky boli výborné. Daná metóda však disponuje významným obmedzením. Môžeme pomocou nej úspešne identifikovať len tváre zachytené čelne.

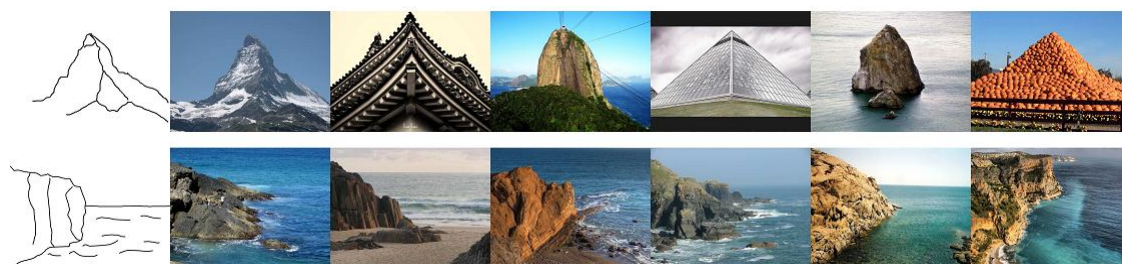
Štatisticky orientované prístupy sa sústreďujú na zostavenie klasifikátorov objektov, pre ktoré je vstupom samotná fotografia. Klasifikácia je proces, ktorý hľadá modely, ktoré opisujú známe triedy údajov tak, aby bolo možné klasifikovať aj neznáme objekty. Z tohto dôvodu nie je všeobecne použiteľná, tzn. len v prípadoch, kde sme schopní určiť znalostnú bázu v podobe tried. Klasifikátory sú trénované na rôznych fotografiách (ich častiach, blokoch), ktoré zachytávajú jednak rôzne objekty, jednak rovnaké objekty v iných scénach, resp. polohách. Jedným z pomerne nových prístupov je SVM (*Support Vector Machine*), pomocou ktorého sú dosahované výborné výsledky pri rozpoznávaní tvári a postáv. Medzi ďalšie známe metódy patria napr. Skryté Markovove modely a Kalmanov filter. V roku 2004 Agarwal a kol. [2] publikoval zaujímavý prístup, v ktorom automaticky skonštruoval množinu charakteristických častí objektov z triedy vzorových fotografií, ktoré zachytávajú objekty záujmu. V tomto prípade objekty tvorili osobné automobily odfočené z profilu v mestskom prostredí. Trénovacia množina bola vytvorená pomocou Förstnerovho Interest operátora, ktorý vráti množinu zaujímavých bodov (*interesting points*, pozri obr. 1a). Pomocou daných bodov sú vytvorené bloky predstavujúce objekty záujmu (pozri obr. 1b). Klasifikátor bol natrénovaný pomocou vytvorených blokov, ktoré tvoria časti automobilu. Súčasne boli klasifikátoru pri trénovaní predložené bloky, na ktorých sa časti automobilu nevyskytovali. Fotografia, ktorá má byť vyhodnotená je podobne ako trénovacia fotografia rozdelená do blokov, pomocou ktorých je následne klasifikovaná.



Obrázok 1 – a, Vzorová fotografia, na ktorej sú identifikované zaujímavé body pomocou Förstnerovho Interest operátora. Pomocou daných bodov sú vytvorené bloky reprezentujúce objekty záujmu; b, Trénovacia množina v podobe blokov (400 blokov); c, Korektná klasifikácia objektu [2].

2.1.2 Vyhľadávanie podľa zámeru

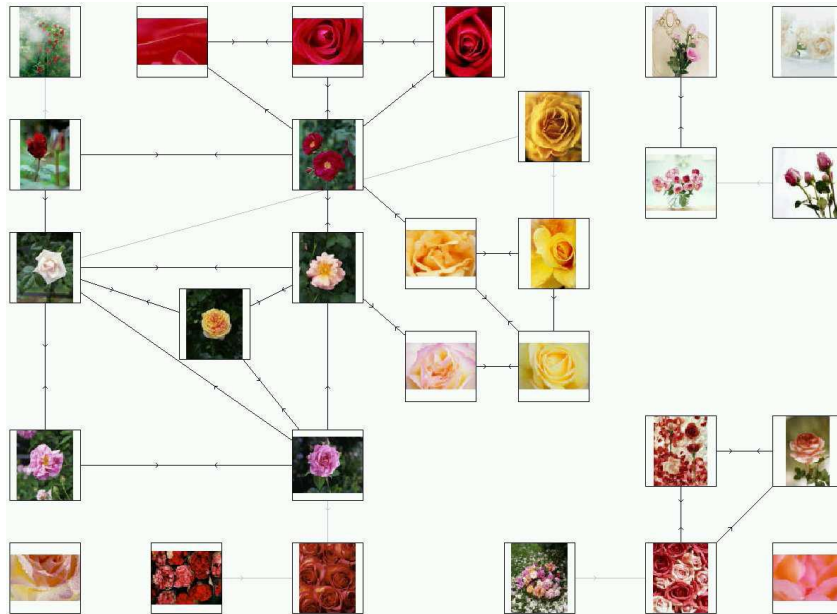
Cieľom vyhľadávania je lokalizovať fotografie, ktoré majú náležitú podobnosť. Tomu predchádza dopyt v podobe vzorovej fotografie, príp. skice vytvorenej používateľom. Fotografie a skice sú typicky reprezentované použitím „štandardných“ deskriptorov, ktoré sú vytvorené z globálnych, alebo lokálnych histogramov jasu, farieb, alebo smerovania (*directionality*). Nedostatkom týchto deskriptorov je, že pri dopytoch v podobe skíc môžu byť použité, len ak sa používateľom vygenerovaný vstup podobá fotografii vo farbe, intenzite, resp. smerovaní. Používateľom vytvorená skica, resp. jej línie transformované na vektory nie sú kompatibilné s danými deskriptormi. V článku [14] navrhli deskriptor, ktorý elegantne rieši asymetriu medzi vstupnou skicou vygenerovanou používateľom a plne farebnou fotografiou. Na obr. 2 je vyobrazený výsledok dopytu do databázy s 1,5 mil. fotografiami. Daný dopyt bol vytvorený použitím navrhnutého deskriptoru.



Obrázok 2 – vľavo skica vytvorená používateľom, ktorá predstavuje dopyt, vpravo sú vyobrazené výsledky pre daný dopyt [14].

2.1.3 Vyhľadávanie na základe asociácií

Vyhľadávanie na základe asociácií prebieha prehliadaním úložiska (priestoru) samotným používateľom. Pri výbere fotografie sú mu ponúknuté podobné fotografie pomocou rôznych vizualizačných techník. Heesch [17] navrhol metódu organizovania fotografií do grafu (pozri obr. 3). Tá je skonštruovaná tak, že fotografie, ktoré zdieľajú určité spoločné vlastnosti (napr. farbu, textúru) sú navzájom prepojené. Myšlienka siete je inšpirovaná Internetom, resp. webom. Medzi fotografiami sú vytvorené asociácie. Podobné fotografie sú prepojené a používateľ môže „objavovať“ časti siete s množstvom fotografií s požadovanými vlastnosťami, resp. požadovaného typu.



Obrázok 3 –Fotografie, ktoré zdieľajú spoločné vlastnosti sú navzájom prepojené, Heesch [17].

2.2 Deskriptory používané v obrazovom vyhľadávaní

Vyhľadávanie podobných fotografií je založené na porovnávaní deskriptorov jednotlivých objektov - fotografií. Deskriptory tvoria príznaky, ktoré predstavujú merateľné veličiny. Každý objekt je teda opísaný príznakom, resp. vektorom príznakov, ktoré je možné vypočítať. Skalár, resp. vektor týchto príznakov (deskriptor) potom umožňuje odlíšiť objekty (celé fotografie) pri vyhľadávaní podobných kandidátov. Na deskriptory sú vo všeobecnosti kladené požiadavky ako efektivita, robustnosť, invariancia a diskriminačná schopnosť [34]. Pri efektivite sa kladie dôraz na čo najlepšiu detekovateľnosť, rýchly výpočet, malé pamäťové nároky (rozmer deskriptora). Mal byť robustný, tzn. odolný voči aditívnemu šumu, príp. k chybám predspracovania. Ďalej by mal byť invariantný voči operáciám s obrazom ako je jeho zväčšenie, zmenšenie, posun, 2d rotácia, zmena kontrastu, jas, osvetlenia, príp. čiastočne invariantný voči priestorovej zmene uhlu pohľadu. Diskriminačná schopnosť – schopnosť odlišovať navzájom rôzne fotografie (rôzne hodnoty príznakov pre rôzne objekty). Podobné fotografie by mali byť opísané teda podobnými deskriptormi. Pri návrhu však niektoré požiadavky stoja oproti sebe, a preto je často potrebné voliť kompromisy.

2.2.1 Priemer farieb

Jedná sa o triviálnu metódu, ktorá sa používa najmä pri jednoduchých vizualizáciách, keď si vystačíme s príbuznosťou fotografií na základe obsiahnutých farieb (v angl. literatúre pod názvom *Average color*, [21]). Jej podstata spočíva vo výpočte euklidovskej vzdialenosti (skaláru) z „výsledných farieb“ dvoch skúmaných objektov (obrázkov). Pre výpočet sa uvažuje model farebného priestoru HSV¹ (farebný tón – *Hue*, sýtosť – *Saturation*, hodnota jas - *Value*).

Označme výslednú farbu obrázku I ako I_c . Potom jednotlivé zložky výslednej farby h_{I_c} , s_{I_c} , v_{I_c} pre obrázok I v modeli HSV vypočítame takto:

$$h_{I_c} = \frac{\sum_{k=1}^{cp} h_k}{cp}; s_{I_c} = \frac{\sum_{k=1}^{cp} s_k}{cp}; v_{I_c} = \frac{\sum_{k=1}^{cp} v_k}{cp},$$

kde h_k , s_k , v_k sú zložky k -tého pixelu obrázku I vo farebnom priestore HSV a cp je počet obrazových bodov (pixelov) obrázku I .

Farebný tón H v modeli HSV reprezentuje uhol v kružnici z intervalu $\langle 0^\circ, 360^\circ \rangle$, napr. rýdzo červená farba 0° uhol, rýdzo zelená farba 120° uhol a rýdzo modrá 240° uhol. Sýtosť farby S je reprezentovaná vzdialenosťou od stredu kružnice, ktorá sa zvyšuje s narastajúcim polomerom od jej stredu. Geometrická interpretácia modelu HSV je šesťboký ihlan. Podstavu ihlanu tvoria farebný tón a sýtosť, zatiaľ čo výška udáva jas. Čiernu farbu reprezentuje vrchol ihlanu. Na základe daného rozloženia sú rýdze farby na obode podstavy. Pretože priestor je tvorený ihlanovitým tvarom, musí byť použitý nasledujúci vzťah pre výpočet vzdialenosti d (podobnosti) medzi dvomi navzájom skúmanými obrázkami I, J [21]:

¹ Niekedy sa zvykne chybné označovať ako HSL, kde L - Lightness sa nesprávne interpretuje ako ekvivalentná zložka k V - value (HSV). V skutočnosti je HSL synonymum pre model farebného priestoru HLS.

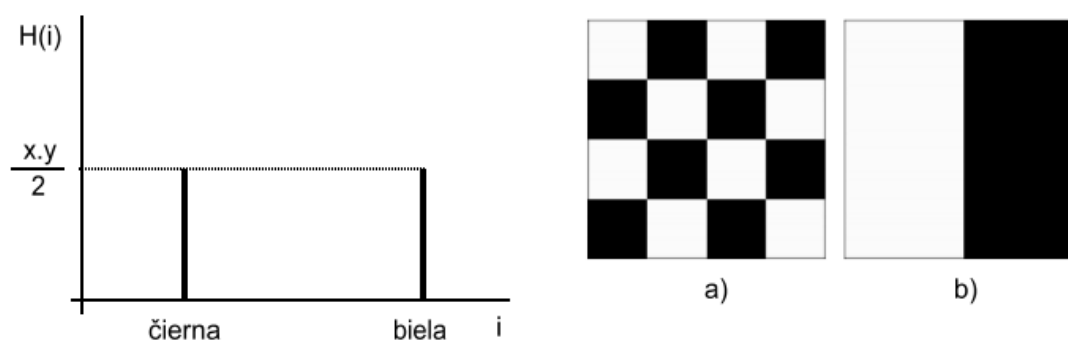
$$d(I, J) = 1 - \frac{1}{5} \sqrt{(v_{I_c} - v_{J_c})^2 + (s_{I_c} \cosh_{I_c} - s_{J_c} \cosh_{J_c})^2 + (s_{I_c} \sinh_{I_c} - s_{J_c} \sinh_{J_c})^2},$$

kde h_{I_c} , s_{I_c} , v_{I_c} , h_{J_c} , s_{J_c} , v_{J_c} sú výsledné farby obrázkov I a J v modeli HSV.

2.2.2 Deskriptory MPEG-7

Farebné histogramy

Farebný histogram [36] kvantifikuje množstvo a frekvenciu farieb obsiahnutých v obraze (distribúciu farieb). Z matematického hľadiska je histogram vektor absolútnych početností hodnôt zastúpených v obraze, resp. hodnota histogramu H pre index i hovorí, koľko pixelov v obraze má intenzitu i . Pokiaľ je pôvodný obraz zložený zo zložiek jasu, potom je histogram jednorozmerný vektor. V prípade RGB zložiek je zložený z troch vektorov. Obrazy, ktoré reprezentujú farbu pomocou viacfarebných kanálov, majú pre každý kanál osobitný histogram. Nakoľko reprezentuje absolútne početnosti jednotlivých zložiek, platí, že hodnota histogramu H pre index i je rovná počtu pixelov tejto hodnoty. Z toho plynie, že súčet hodnôt histogramu sa rovná počtu pixelov v obraze. Histogram je štatistickou veličinou. Môžeme ho tiež chápať ako pravdepodobnosť výskytu pixelu farby v obraze. Kvantifikuje len jasové pomery v obraze a nenesie žiadnu informáciu o ich plošnom rozložení. Histogram na obr. 4 je získaný z obrazu, ktorý je zložený z dvoch farieb (biela, čierna). Obidve sú zastúpené rovnakým pomerom. Daný histogram je totožný pre obrazy a) aj b).

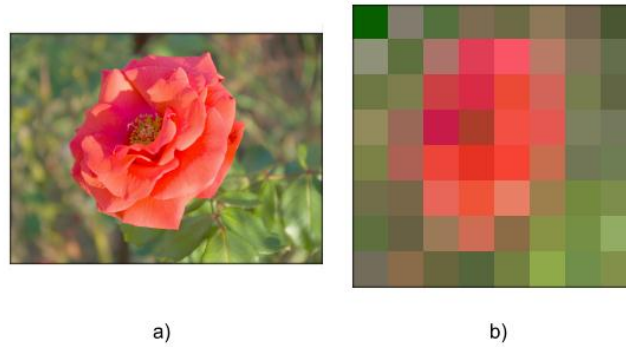


Obrázok 4 – Histogram vľavo je získaný z obrazov vpravo a) a b) a je pre ne totožný, pretože nenesie informáciu o plošnom rozložení [36].

Štandard MPEG-7 [28] definuje deskriptor farebný histogram (*Scalable Color Descriptor*) vo farebnom priestore HSV, ktorý je zakódovaný Haarovou transformáciou. Deskriptor je použiteľný pri vyhľadávaní podobných fotografií na základe distribúcie farieb, tzn. nie ich plošnom rozložení.

Farebné rozloženie

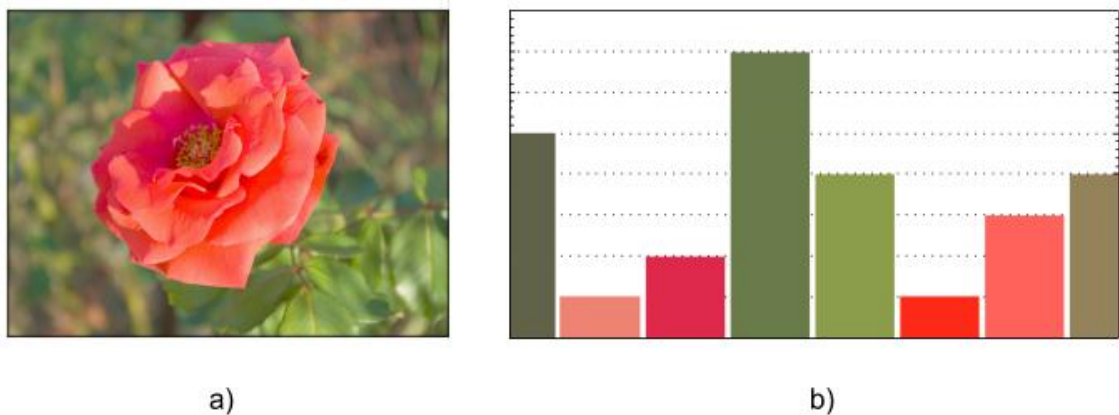
Deskriptor farebného rozloženia (*Color Layout Descriptor* [28], pozri obr. 5) zachytáva informáciu o priestorovom (plošnom) rozložení farieb v obrázku. Jeho výhodou je, že nie je závislý od formátu, rozlíšenia a bitovej hĺbky obrazu. Výpočtové nároky sú veľmi malé a je blízky ľudskému vnímaniu obrazu.



Obrázok 5 – Deskriptor farebného rozloženia: a) vstupná fotografia, b) plošné rozloženie farieb.

Dominantné farby

Deskriptor dominantných farieb (*Dominant Color(s) Descriptor* [28], pozri obr. 6) je obzvlášť vhodný pre reprezentáciu objektov, alebo častí obrázkov (*image regions*), kde malý počet farieb dostatočne vystihuje analyzovanú oblasť. Príkladom pre použitie sú vlajky štátov.



Obrázok 6 – Deskriptor dominantných farieb: a) vstupná fotografia, b) histogram dominantných farieb.

Histogram hrán

Deskriptor histogramu hrán (*Edge Histogram Descriptor* [28]) reprezentuje priestorové rozloženie piatich typov hrán (štyri smerové, jeden nesmerový). Hrany hrajú dôležitú úlohu pri vnímaní obrazu, a preto použitím tohto deskriptoru dokážeme nájsť obrázky s podobným motívom. Pre zvýšenie výkonu sa často kombinuje s inými deskriptormi, napr. farebným histogramom.

2.2.3 CCD

Compact Composite Descriptors (CCD) je množina deskriptorov, ktoré kombinujú informáciu o farbe a textúre v jednom histograme. Boli navrhnuté so zreteľom na veľkosť (dimenziu), avšak tak, aby súčasne zabezpečili čo najlepšiu diskriminačnú schopnosť. Sú navrhnuté pre rôzne typy obrázkov ako napr. rádiologické obrázky v medicíne (BTDH - *Fuzzy rule based scalable composite descriptor* [7]), umelecké obrázky – skice (SpCD - *Spatial Color Distribution Descriptor* [6]) a fotografie

všeobecne (CEDD - *Color Edge Directivity Descriptor* [9], FCTH - *Fuzzy Color and Texture Histogram* [10], JCD – *Joint Composite Descriptor* [8]). Táto, pomerne nová množina deskriptorov je predstavená v systéme Anaktisi² [35].

2.2.4 Farebné korelogramy

Farebné korelogramy (ďalej len korelogramy) zachytávajú lokálnu priestorovú koreláciu farieb v obrázku (fotografii). Svojou podstatou sa jedná o farebné histogramy, ktoré sú však rozšírené o priestorovú informáciu.

Formálna definícia: Nech I je obrázok o dimenzii $x \times y$, p je pixel v obrázku I ($p = (x, y) \in I$) a I_c je množina pixelov farby c . Nech vzdialenosť d dvoch pixelov $p_1 = (x_1, y_1)$ a $p_2 = (x_2, y_2)$ je definovaná ako $|p_1 - p_2| = \max\{|x_1 - x_2|, |y_1 - y_2|\}$. Potom korelogram C pre vzdialenosť d a pár farieb c_i, c_j udáva pravdepodobnosť Pr , že sa pixel p_2 farby j nachádza vo vzdialenosti d od pixelu p_1 farby i :

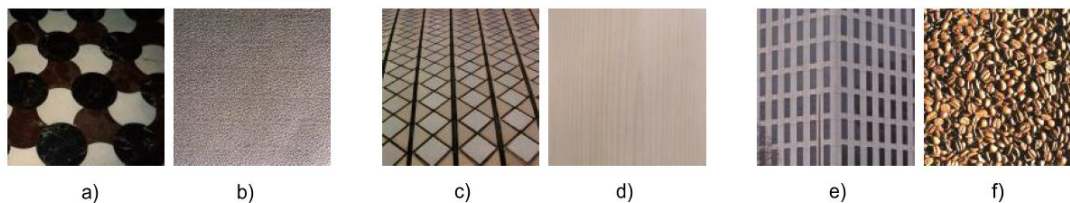
$$C^{(d)}(c_i, c_j) = \Pr_{p_1 \in I_{c_i}, p_2 \in I} [p_2 \in I_{c_j} \mid |p_1 - p_2| = d]$$

V spojitosti s korelogramom sa zavádza pojem autokorelogram (AC), ktorý zachytáva informáciu len medzi identickými farbami (c_i):

$$AC^{(d)}(c_i) = C^{(d)}(c_i, c_i)$$

2.2.5 Tamurove textúrové príznaky

Tamurove textúrové príznaky vychádzajú z princípov ľudského vnímania obrazu. Tamura a kol. [12] navrhli šesť príznakov: granularita (*coarseness*), kontrast (*contrast*), smerovosť (*directionality*), podobnosť línií (*line-likeness*), pravidelnosť (*regularity*), drsnosť (*roughness*). Experimentom overili, že prevažne prvé tri z uvedených príznakov silne korelujú so spôsobom ľudského vnímania obrazu. Príklady, ktoré ilustrujú význam jednotlivých príznakov sú vyobrazené na obr. 7.



Obrázok 7 – Príklady obrázkov, ktoré ilustrujú vlastnosti textúr z prvých troch Tamurových príznakov:
a) vysoká granularita, b) nízka granularita, c) vysoký kontrast, d) nízky kontrast,
e) vysoká smerovosť, f) nízka smerovosť.

Granularita

Daný príznak nesie informáciu o veľkosti elementov (hrubosť zŕn), ktoré sa v textúre pravidelne opakujú (vzory). Čím je vypočítaná hodnota vyššia, tým je veľkosť zŕn väčšia. V prípade, že sa v textúre vyskytuje viacero vzorov, tak sa uvažuje väčší z nich. Výpočet príznaku sa realizuje v nasledujúcich krokoch:

² <http://www.anaktisi.net/>

1. pre každý obrazový bod (p_x, p_y) obrázku I sa vypočíta priemerná hodnota v jeho okolí. Uvažujú sa okolia veľkosti mocniny 2, tzn. $2^k \times 2^k$ (pre $k = 0, 1, \dots, 5$), teda $1 \times 1, 2 \times 2, \dots, 32 \times 32$:

$$A_k(p_x, p_y) = \frac{1}{2^{2k}} \sum_{i=1}^{2^{2k}} \sum_{j=1}^{2^{2k}} X(p_x - 2^{k-1} + i, p_y - 2^{k-1} + j)$$

2. pre každý bod (p_x, p_y) sa vypočíta rozdiel priemerných hodnôt neprekrývajúcich sa okolí bodu (p_x, p_y) , ktoré sa nachádzajú na opačných stranách voči zvolenému bodu v horizontálnom a vertikálnom smere:

$$E_k^h(p_x, p_y) = |A_k(p_x + 2^{k-1}, p_y) - A_k(p_x - 2^{k-1}, p_y)|$$

$$E_k^v(p_x, p_y) = |A_k(p_x, p_y + 2^{k-1}) - A_k(p_x, p_y - 2^{k-1})|$$

3. v každom bode (p_x, p_y) sa vyberie okolie s maximálnou hodnotou rozdielu:

$$S(p_x, p_y) = \arg \max_{k=1, \dots, 5} \max_{d=h, v} E_k^d(p_x, p_y)$$

4. granularita je potom hodnota cez priemer 2^S :

$$F_{crs} = \frac{1}{P_x P_y} \sum_{p_x=1}^{P_x} \sum_{p_y=1}^{P_y} 2^{S(p_x, p_y)}$$

Kontrast

Vyjadruje dynamický rozsah hodnôt jasu v textúre, polarizáciu distribúcie bielej a čiernej farby v histograme, ostrosť hrán a periódu opakovania sa vzorov. Kontrast obrázku je vypočítaný takto:

$$F_{con} = \frac{\sigma}{\alpha_z}, \alpha_4 = \frac{\mu_4}{\sigma^4},$$

kde $\mu_4 = \frac{1}{P_x P_y} \sum_{p_x=1}^{P_x} \sum_{p_y=1}^{P_y} (X(p_x, p_y) - \mu)^4$ je štvrtý moment hodnoty jasu, σ^2 je rozptyl

hodnôt jasu a parameter z bol experimentálne zvolený na $\frac{1}{4}$.

Smerovosť

Zachytáva mieru existencie orientácie v textúre, tzn. nie orientáciu samotnú ale jej prítomnosť. V každom bode textúry $X(p_x, p_y)$ sa najskôr aplikujú jednoduché konvolučné masky pre výpočet derivácií vo vertikálnom a horizontálnom smere ($\Delta H, \Delta V$):

$$\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Pre každý obrazový bod sa vypočíta hodnota $\theta = \frac{\pi}{2} + \tan^{-1} \frac{\Delta V(p_x, p_y)}{\Delta H(p_x, p_y)}$. Z hodnôt θ následne zostavíme histogram. Príznakom je suma druhých momentov z hodnôt okolo vrcholov histogramu.

Modifikácia vlastností pri vyhľadávaní

Pre použitie uvedených príznakov vo vyhľadávaní sa pristupuje k ich modifikácii. Pre každý z príznakov sa požaduje výpočet konkrétnych hodnôt pixelu. Pre výpočet granularity sa uvažujú len kroky 1 – 3 pre každý pixel. Pri kontraste je vypočítané okolie (13×13) pre každý pixel osobitne, rovnako pre smerovosť. Namiesto derivácii je však použitý Sobelov filter a hodnota θ sa vypočíta pre každý pixel označujúci smer oblasti okolia. Výsledkom sú tri hodnoty. Tieto hodnoty sú uvažované ako zložky RGB obrázku, z ktorých sa vytvára trojrozmerný histogram.

2.2.6 SIFT

Scale Invariant Feature Transform (SIFT) je v súčasnosti jedna z najpopulárnejších metód pre rozoznávanie obrazových vzorov typu lokálnych vlastností, resp. metóda pre automatické nájdenie korešpondencií medzi dvojicou obrázkov (rozoznávanie objektov). Autorom metódy je David Lowe. Prvýkrát ju publikoval v roku 1999 [19] a v roku 2004 [25] opísal kompletný postup, akým sa dá použiť na rozoznávanie objektov. V súčasnosti existuje viacero odvodených metód ako napr. PCA-SIFT [20], Fast approximated SIFT [16], či SIFT-CCH [5]. Pôvodná metóda extrahuje lokálne vlastnosti z obrázku v stupnici šedej (*grayscale*). Jej nevýhodou je teda náchylnosť na farebné obrázky, a práve SIFT-CCH kombinuje pôvodnú SIFT s CCH (*Color Co-occurrence Histograms*), čo sú histogramy používané pri rozoznávaní objektov.

Metóda pozostáva z detektora a deskriptora. Je založená na identifikovaní lokálnych vlastností, tzv. bodov záujmu (angl. *feature points, keypoints*). Každý bod je opísaný deskriptorom, ktorý pozostáva zo súradníc (x, y), veľkosti, orientácie a 128-rozmerného vektora príznakov. Je invariantný voči operáciám s obrazom ako: zmena veľkosti (škálovateľnosť), posun, rotácia, aditívny šum, zmena kontrastu, jas, osvetlenia a čiastočne voči priestorovej zmene uhlu pohľadu. V nasledujúcej časti opíšeme metódu (jej fázy) podrobnejšie, pretože je v súčasnosti populárna (aj napr. pri vytváraní panorám – *autopano-sift*³) vďaka pomeru výkonu a kvality (presnosti). Podrobné informácie o danej metóde je možné nájsť v článkoch Davida Lowe [19] [25].

Vyhľadanie bodov záujmu

V tejto fáze sa hľadajú body záujmu pomocou detektora, z ktorých sú v konečnej fáze vytvorené deskriptory (vektory príznakov). Zo vstupného obrazu je zostrojený priestor škál (*scalespace*), aby sa dosiahla invariancia voči zmene veľkosti. Vstupný obraz je v škále šedej (*grayscale*) a hodnoty pixelov sú normalizované na interval $\langle 0, 1 \rangle$. Priestor škál sa definuje ako funkcia $L(x, y, \sigma)$, ktorá je tvorená konvolúciou variabilnej Gaussovskej funkcie so vstupným obrazom $I(x, y)$ nasledovne:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y),$$

kde $*$ je konvolúcia na pozícii x, y a

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

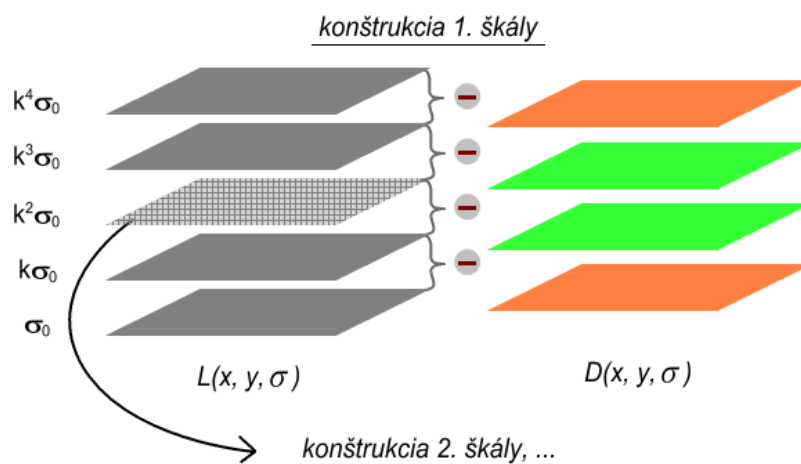
Na zistenie bodov záujmu používa metóda lokálne extrémny vo funkcii rozdielu Gausiánov $D(x, y, \sigma)$ konvolovaných s obrazom, ktorý je vypočítaný pomocou

³ autopano-sift (making panoramas fun): <http://user.cs.tu-berlin.de/~nowozin/autopano-sift/>

rozdielu dvoch susedných rozostrených obrazov L so škálami, ktoré sú odlišené násobiacim faktorom k :

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma).$$

Na obr. 8 sú šedou farbou znázornené rozostrené obrázky, pričom ak od niektorého z nich odčítame obrázok pod ním, dostaneme rozdiel Gausiánov (znázornený zelenou farbou, resp. oranžovou farbou). Zelenou farbou sú znázornené rozdiely, v ktorých budeme hľadať lokálne extrémny. Uvedeným spôsobom skonštruujeme jednu škálu. Ďalšiu škálu vytvoríme tak, že zoberieme v poradí tretí vyhľadovaný obraz zhora a vynecháme v ňom každý druhý stĺpec a riadok, čím zmenšíme jeho rozmer štvornásobne. Takto upravený a rozostrený obraz tvorí vstup pre vytvorenie ďalšej škály.



Obrázok 8 – Ilustrácia konštrukcie množiny škál.

Na obr. 9 je priamo znázornená konštrukcia rozdielu dvoch Gaussovsky rozostrených obrázkov. Ak z obrázku c odčítame obrázok b , dostaneme obrázok d - rozdiel Gausiánov. Ako bolo uvedené, v takto skonštruovanom obrázku nás zaujímajú lokálne extrémny. Nezaujímajú nás hrany a veľké homogénne oblasti, ktoré sú nevýrazné (čierne - nulový rozdiel), ale naopak, zaujímajú nás oblasti v strede malých kruhov a vedľa hrán (biele - veľký rozdiel). V rozdieloch Gausiánov sa následne hľadajú lokálne extrémny. Bod, ktorý je kandidátom na extrém sa porovnáva so svojimi ôsmimi susedmi (obrazovými bodmi) na svojom rozdiel Gausiánov, ako aj v okolí 3x3 na rovnakom mieste v susedných obrázkoch (nad a pod). Ak je hodnota kandidáta väčšia (menšia) ako hodnoty všetkých susedov, tak sme na danej pozícii (v danom rozdiel Gausiánov) našli bod záujmu.

Určenie polohy, orientácie a veľkosti bodu záujmu

Po nájdení stabilných bodov sa ich poloha spresní na desatinné čísla. Na spresnenie je použitá 3D kvadratická funkcia na ich okolie. Pre každý extrém vezmeme jeho 3x3 okolie. Týmito hodnotami preložíme trojrozmernú kvadratickú funkciu, nájdeme minimum a maximum funkcie, čím zistíme polohu bodu so sub-pixelovou presnosťou. Podľa tvaru funkcie odstránime nestabilné body, tzn. s nízkym kontrastom alebo ležiace pozdĺž hrán. Bodom záujmu priradíme orientáciu a veľkosť gradientu. Orientácia sa

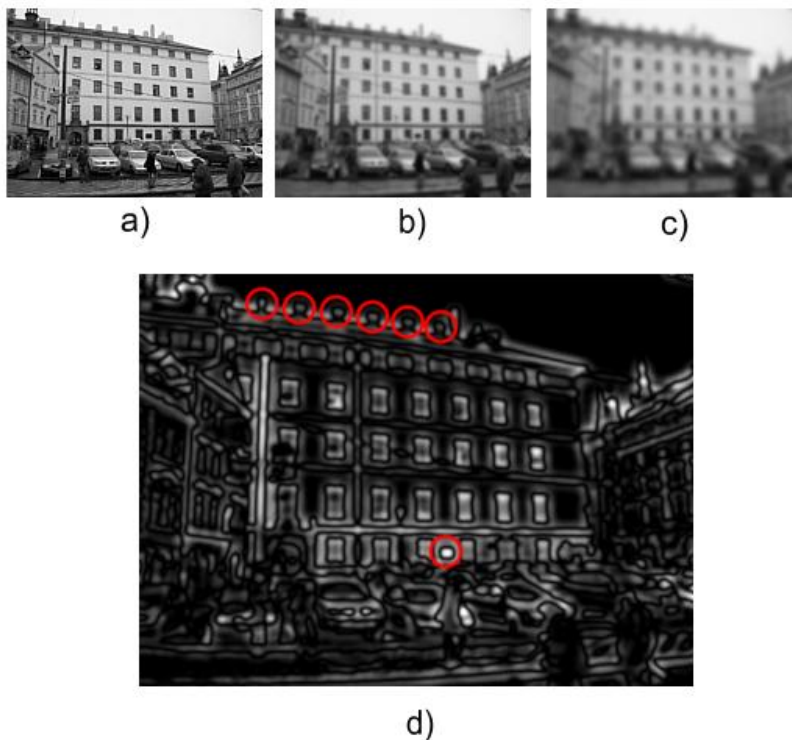
priradí bodu záujmu na základe jeho okolia z rozostrených obrazov, pričom jeho škála poslúži na výber vhodného rozostreného obrazu. Na základe uvedeného postupu na priradenie orientácie je potrebné, aby mal každý rozostrený obraz $L(x, y, \sigma)$ vypočítanú orientáciu a veľkosť gradientu pre všetky svoje body. V rozostrenom obraze vypočítame prvú deriváciu v mieste bodu (v smere osí x a y , d_x , d_y).

Pre veľkosť: $m = \sqrt{d_x^2 + d_y^2}$, resp.:

$$m(x, y, \sigma) = \sqrt{\left(\frac{1}{2}(L(x+1, y, \sigma) - L(x-1, y, \sigma))\right)^2 + \left(\frac{1}{2}(L(x, y+1, \sigma) - L(x, y-1, \sigma))\right)^2}$$

Pre orientáciu: $\theta = \arctan\left(\frac{d_y}{d_x}\right)$, resp:

$$\theta(x, y, \sigma) = \tan^{-1}\left(\frac{L(x, y+1, \sigma) - L(x, y-1, \sigma)}{L(x+1, y, \sigma) - L(x-1, y, \sigma)}\right)$$



Obrázok 9 - Konštrukcia rozdielu dvoch Gaussovsky rozostrených obrázkov ($d = c - b$) [26].

Deskriptor

Pomocou deskriptoru chceme vhodne opísať okolie bodu záujmu. Výpočet deskriptoru sa realizuje relatívne vzhľadom k orientácii θ z predchádzajúceho kroku. Tým sa docielí invariancia voči orientácii obrázku. Deskriptor je vypočítaný na základe gradientov v okolí bodu záujmu, resp. je histogramom týchto gradientov.

3 Automatické anotovanie fotografií

Automatické anotovanie fotografií predstavuje riešenie najmä pre rozsiahle úložiská fotografií, ktoré nedisponujú opismi. Ručné anotovanie je časovo náročné. Ak fotografie nedisponujú opismi v podobe kľúčových slov, je prakticky nemožné medzi nimi „presne“ vyhľadávať. Takéto fotografie sa zvyknú v angl. označovať ako *unseen photos*. Automatické anotovanie fotografií chápeme ako proces modelovania práce anotátora (používateľa) pri priradovaní kľúčových slov k fotografiám na základe jej vizuálneho obsahu, tzn. identifikovaných objektov a ich vlastností.

Základom automatických anotačných systémov je robustný model, ktorý je navrhnutý tak, že dokáže v istej miere zastúpiť používateľa v role anotátora a pre cieľovú fotografiu dokáže patričný opis vytvoriť. Je to výzva, ktorou sa už viac ako dve desaťročia zaoberá množstvo výskumníkov. Niektoré modely pre automatické anotovanie opíšeme v nasledujúcej časti.

Modely automatického anotovania sa zvyknú deliť do troch tried, a to pravdepodobnostné modely, modely založené na klasifikátoroch a na grafových štruktúrach [24]. Tieto spravidla vyžadujú robustnú množinu fotografií pre ich natréňovanie. V súvislosti s automatickým anotovaním fotografií existujú tiež modely (systémy), ktoré takúto množinu pri svojom tréňovaní priamo nevyžadujú, ale disponujú mechanizmom, ktorý ju dokáže v reálnom čase „emulovať“. Využívajú na to napr. externý on-line vyhľadávač (*Google Images*). Tieto systémy sa označujú ako automatické anotačné systémy s neobmedzenou tréňovacou množinou (*unlimited vocabulary*) [33].

3.1 Pravdepodobnostné modely automatického anotovania fotografií

Medzi známe publikované pravdepodobnostné modely patria: Co-occurrence [27] model, Translation model [13], Continuous Relevance Model [22], Global Scene Modelling [30], Latent Variable Models [3]. Tieto modely sú spravidla všeobecnejšie ako napr. modely založené na klasifikátoroch, a preto im budeme venovať väčšiu pozornosť.

3.1.1 Co-occurrence model

Mori a kol. [27] navrhli v roku 1999 metódu, ktorá je založená na štatistickom učení sa z fotografií, ktoré už disponujú kľúčovými slovami, tzn. predpokladom tejto metódy je prítomnosť tréňovacej množiny.

Predpokladajme fotografiu, ktorá disponuje kľúčovými slovami „obloha“ a „hory“ (objekty sa vyskytujú na fotografii). Po rozdelení fotografie na bloky, každý zdedí obidve kľúčové slová, tzn. s každým blokom sú asociované slová „obloha“ a „hory“. Medzi blokmi sú však také, ktoré skutočne reprezentujú „oblohu“ a „hory“ sa na nich

nevyskytujú (platí aj opačne). Zoberme ďalšiu fotografiu, s ktorou sú asociované kľúčové slová „obloha“ a „rieka“ (objekty sa vyskytujú na fotografii). Danú fotografiu rovnako rozdelíme na bloky a s každým asociujeme obidve kľúčové slová. Keďže obloha sa vyskytuje na obidvoch fotografiách, budú s vysokou pravdepodobnosťou bloky s oblohou z obidvoch fotografií v jednom centroide (príp. centroidoch, závisí od vyhodnotenia podobnosti deskriptorov). V takomto centroide sa po „príspevku“ druhej fotografie budú nachádzať kľúčové slová: dvakrát „obloha“, jedenkrát „hory“ a jedenkrát „rieka“. Touto cestou sa v centroidoch pri učení potláčajú nevhodné slová a ich skóre sa znižuje s narastajúcim počtom vhodných korešpondencií.

Procedúra metódy pre spracovanie trénovacej množiny je takáto:

1. Každá fotografia F je rozdelená na rovnako veľké bloky. Pre každú sa vypočíta 96-rozmerný deskriptor na základe farieb a textúry.
2. Každý blok „zdedí“ všetky kľúčové slová, ktoré boli asociované s pôvodnou fotografiou.
3. Jednotlivé bloky sú následne zoskupené (*clustered*) do tzv. centroidov. Centroid reprezentuje akýsi vzor bloku (opísaný jeho deskriptorom), v ktorom môžu byť obsiahnuté iné bloky na základe podobnosti k danému vzoru (podobnosť je limitovaná prahom P). Zoskupenie do centroidov sa realizuje na základe nasledujúcej podmienky:

Pre každý deskriptor R , ktorý opisuje blok B fotografie F :

ak

existuje centroid C taký, že euklidovská vzdialenosť medzi R a C je menšia ako prah P centroidu C , potom zarad' blok B do C .

inak

vytvor nový centroid C' a zarad' B do tohto centroidu.

4. Pre každý centroid c_j , pre každé slovo w_i , ktoré sa v ňom vyskytuje je spočítané tzv. skóre, ktoré reprezentuje početnosť výskytu slova w_i v centroide c_j . Vypočítané hodnoty skóre pre centroid c_j následne normalizujeme tak, že ich súčet je rovný 1:

$$P(w_i, c_j) = \frac{m_{ji}}{M_j},$$

kde m_{ji} je počet všetkých slov w_i v centroide c_j a M_j je počet všetkých slov w , ktoré sú asociované s centroidom c_j .

Procedúra metódy pre spracovanie novej fotografie, ktorá nedisponuje opisom (kľúčovými slovami) je nasledovná:

1. Každá fotografia sa rozdelí na bloky, pre ktoré sú následne vypočítané deskripty.
2. Na základe deskriptorov je pre každý blok nájdený centroid, do ktorého patrí – vyhladá sa tzv. „najbližší príbuzný“ (porovnanie deskriptora centroidu s deskriptorom bloku).
3. S nájdenými centroidmi sú asociované slová, ktoré majú vypočítané svoje skóre (pozri procedúru spracovania trénovacej množiny). Jedno a to isté slovo sa môže vyskytovať vo viacerých centroidoch s rôznymi hodnotami skóre. Z týchto hodnôt sa vypočíta priemer. Daný výpočet sa prevedie pre každé slovo vo všetkých

nájdených centroidoch. Následne sa priemery zoradia vzostupne a vyberie sa n slov, ktoré sa priradia k novej fotografii.

3.1.2 Translation model

Jedná sa o populárny model automatického anotovania fotografií, ktorý používa obrazovú segmentáciu na rozdelenie vstupnej fotografie. Dané rozdelenie tvoria nepravidelné oblasti, ktoré sa označujú ako „blob“. Na opis blobov sa používajú nízkoúrovňové vlastnosti (angl. *low-level features*). Model publikoval Daygulu a kol. v roku 2002 [13]. Vytvorili diskretný slovník blobov (angl. *vocabulary*) naprieč kolekcii fotografií a aplikovali svoj model, v ktorom sa inšpirovali strojovým prekladom (angl. *machine translation*). Pri preklade blobov (obsiahnutých vo fotografii) na slová, odhadujú pravdepodobnosť prekladu $p(a_{nj} = i)$ - vo fotografii n je nejaký blob b_i asociovaný so špecifickým slovom w_j . Táto pravdepodobnosť je maximalizovaná trénovacou množinou. Pravdepodobnostná funkcia je definovaná nasledovne:

$$p(w|b) = \prod_{n=1}^N \prod_{j=1}^{M_n} \sum_{i=1}^{L_n} (p_{nj} = i) t(w = w_{nj} | b = b_{ni}),$$

kde N je počet fotografií; M_n a L_n sú počty slov a blobov, ktoré sú asociované s fotografiou n ; a_n je priradenie $a_n = \{a_{n1}, \dots, a_{nM_n}\}$ také, že $a_{nj} = i$, ak sa b_{ni} preloží na w_{nj} ; $t(w|b)$ je pravdepodobnosť „získania“ inštancie slova w pri poskytnutí inštancie blobu b . Túto pravdepodobnosť postupne maximalizujú štatistickým učením sa pravdepodobnosťami prekladov (angl. *translation probabilities*), pri ktorom používajú EM algoritmus (angl. *Expectation-Maximisation*). Po vytvorení tabuľky odhadov je cieľová fotografia (bez opisov) anotovaná vybratím najpravdepodobnejších slov pre každý z jej blobov. Model bol vyhodnotený na množine s 5000 fotografiami (*Corel5k dataset*). Každéj fotografii bolo priradených 1-5 slov zo slovníka, ktorý tvorilo celkovo 371 slov. Na obr. 10 sú znázornené ukážky výsledkov v podobe označovaných segmentov. Opísaný model je pomerne úspešný v predikcii všeobecných slov ako obloha, strom, tráva.



Obrázok 10 – Translation model – ukážka výsledku v podobe označovania segmentov.

4 Zhodnotenie súčasného stavu a ciele práce

Automatické anotovanie fotografií je úloha a súčasne výzva, ktorá je veľmi náročná, a to hneď z niekoľkých dôvodov.

Prvým, a to asi najväčším je, že vytvoriť systém, pre všeobecné rozpoznávanie objektov je v súčasnosti prakticky nemožné (otázne, či niekedy vôbec). Jeden a ten istý objekt môže byť zachytený z rôznych uhlov, vzdialeností, príp. za rôznych svetelných podmienok. Pri zovšeobecnení, objekt s rovnakým pomenovaním môže mať rôznu podobu (meniaci sa tvar, farba). Vhodným ilustrujúcim príkladom sú metódy pre lokalizáciu a rozpoznávanie tvárí na fotografiách. Na Internete je k dispozícii na zhladnutie viacero videí, keď autentifikačný systém založený na detekcii tváre zlyhá pri ľuďoch s tmavou pokožkou. Metód pre detekciu tvárí existuje viacero, napr. metódy založené na porovnávaní šablón, ktoré vyžadujú robustnú databázu tvárí. Tvár vyhľadávajú na základe korelácie medzi vstupom a šablónami. Iné metódy sa sústreďujú na hľadanie morfológických rysov ako oči, ústa, farba pokožky atď. Jedným z variantov sú zložité znalostné metódy, ktoré sú založené na definovaných pravidlách skutočnej podoby ľudskej tváre.

Metódami pre rozpoznávanie ľudskej tváre sme načrtli dva základné prístupy pre rozpoznávanie objektov. Prvým z nich je, že nemáme definované exaktné pravidlá o podobe objektov, ale disponujeme robustnou databázou s ich podobou (vzorky) a objekty rozpoznávame porovnávaním. Správne rozpoznanie je pri určitej invariancii možné do tej miery, ako kvalitná je naša databáza podobných objektov, a aké algoritmy použijeme pri určení korelácie medzi objektmi. Disponovať databázou, v ktorej budeme mať uložené vzorky pre rozpoznanie objektov zachytených na akejkoľvek fotografii je nemysliteľné. Druhým z prístupov je, že objekt máme zadaný pravidlami, opismi atď. Podobne ako pri nemožnosti disponovať obrazovou databázou sveta, nemôžeme zdefinovať pravidlá a opisy pre „všetky“ objekty sveta.

Ďalším z problémov je časová a priestorová (pamäťová) zložitosť. Časová zložitosť pri extrakcii vlastností z fotografií, ich zatriedení do tried podľa vzájomnej podobnosti (vytvorenie zhlukov) a vyhľadani podobných kandidátov. Pri priestorovej zložitosti zohráva významnú úlohu veľkosť a počet extrahovaných vlastností (rozmer vektora príznakov). Napríklad, pri extrahovaní lokálnych vlastností z fotografií vzniká problém uloženia veľkého množstva údajov tak, aby k nim bol prístup v reálnom čase. Pri extrahovaní 500 lokálnych vlastností z fotografie metódou SIFT pre 1000 fotografií vzniká potreba uložiť 500 000 vlasností. V mnohých prípadoch býva priestor, ktorý potrebujeme pre uloženie (pred)vypočítaných informácii akousi daňou za čas, za ktorý sme schopní k informáciám prístupit'. Inými slovami, rýchlejší prístup k informáciám znamená potrebu alokácie väčšieho priestoru.

Aj napriek uvedeným prekážkam existuje veľa prístupov automatického anotovania fotografií. Vo všeobecnosti je základom automatického anotovania fotografií kvalitná

trénovacia množina. Jednoduché prístupy sú založené na vyhľadani n podobných fotografií voči cieľovej fotografii a extrahovaním kľúčových slov na základe ich početnosti. Robustnejšie metódy využívajú segmentačné algoritmy na vytvorenie tzv. slovníkov vizuálnych termov, ku ktorým sú priradené opisy. Spracovanie cieľovej fotografie potom prebieha tak, že sa z nej extrahujú vizuálne termy, vyhľadajú sa podobné, a na základe miery podobnosti a ich početnosti sa priradia k cieľovej fotografii opisy. Ďalšími rozšíreniami sú rôzne pravdepodobnostné modely, ktoré počítajú pravdepodobnosť prepojenia medzi fotografiou a slovami. Napríklad, fotografii, na ktorej sú tigre sú častejšie priradené slová ako tráva, strom, obloha a menej často slová ako auto, počítač.

Kvalita väčšiny prístupov závisí primárne od veľkosti trénovacej množiny, anotácií, ktoré sú priradené k fotografiám a od metód pre extrakciu vlastností a vyhodnotenia podobnosti. Ak je naša metóda založená na extrakcii globálnych vlastností, potom „trpí“ všeobecnosťou a pri extrakcii kľúčových slov z podobných fotografií treba skutočne pristúpiť k pravdepodobnostným modelom. Výsledok, keď sú cieľovým fotografiám priradené občas správne slová, občas menej správne a inokedy úplne chybné, nie je určite uspokojivý. Pri použití lokálnych vlastností je naša metóda príliš diskriminujúca a v prípade, že je prah minimálnej podobnosti nastavený príliš vysoko, môže nastať situácia, že k cieľovej fotografii nenájdeme ani len podobné fotografie. Kde je správny kompromis? Našou motiváciou je skúsiť ho nájsť.

5 Návrh metódy pre automatické anotovanie fotografií

Naším cieľom je navrhnuť novú metódu, ktorá bude pri vyhľadaní podobných fotografií k cieľovej fotografii kombinovať lokálne a globálne vlastnosti. Pri hľadaní podobných kandidátov na extrakciu kľúčových slov sa zameriame na vizuálnu analýzu fotografií z pohľadu pravdepodobnosti, že obsahujú správne kľúčové slová pre cieľovú fotografiu. Väčšina existujúcich prístupov hľadá najpodobnejšie fotografie voči dopytu vo svojej trénovacej množine. V našej metóde pri vyhodnotení podobných kandidátov zvýhodníme tých, na ktorých sú extrahované objekty záujmu z cieľovej fotografie dominantnejšie, alebo je ich frekvencia výskytu vyššia. Pri takýchto fotografiách je vyššia pravdepodobnosť, že niektoré z kľúčových slov (slová), ktoré sú k nim priradené, opisujú práve objekty záujmu. Tiež sa zameriame na identifikáciu menej dominantných objektov. Pre priradené kľúčové slová cieľovej fotografii vypočítame ich kvalitu. Výsledkom metódy bude, že dokážeme nie len priradiť anotáciu k fotografii, ale dokážeme priamo objekty identifikovať, resp. označiť a pomenovať.

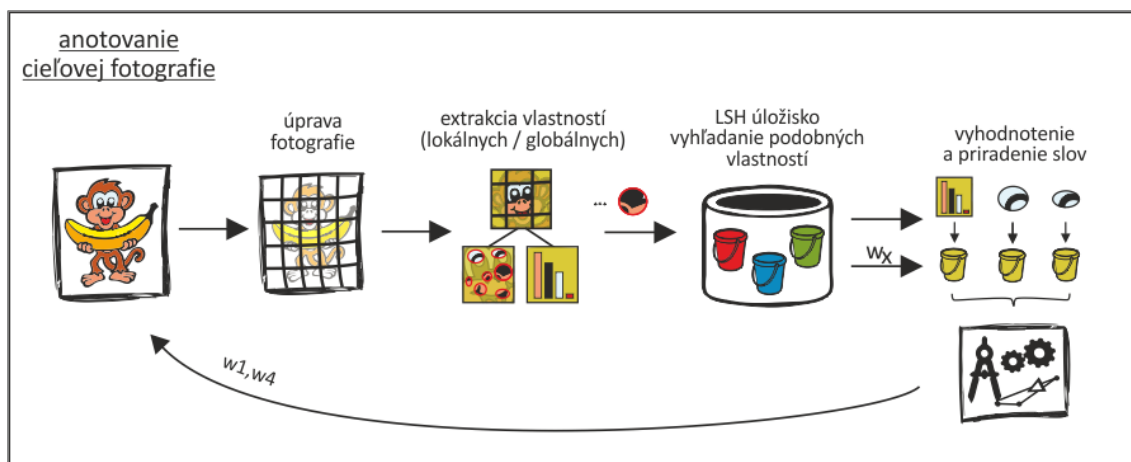
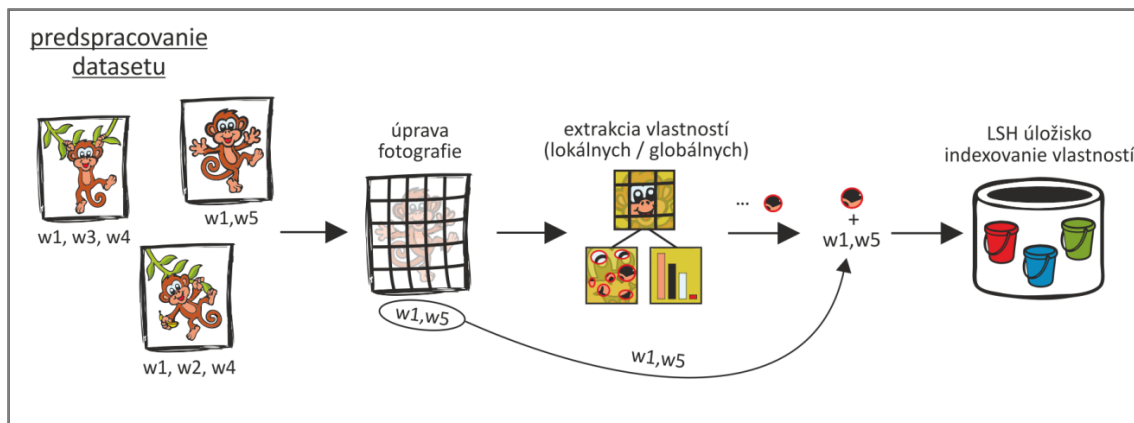
Nami navrhovaná metóda pozostáva z dvoch základných častí (pozri obr. 11), a to predspracovanie trénovacej množiny (datasetu) a anotovanie cieľovej fotografie (dopytu).

Predspracovanie množiny pozostáva z týchto krokov:

1. úprava cieľovej fotografie – interpolácia, rozdelenie do blokov,
2. extrakcia lokálnych (SIFT) a globálnych (JCD) vlastností,
3. priradenie slov z cieľovej fotografie extrahovaným vlastnostiam,
4. indexovanie extrahovaných vlastností do LSH úložiska.

Anotovanie cieľovej fotografie pozostáva z týchto krokov:

1. úprava cieľovej fotografie – interpolácia, rozdelenie do blokov,
2. extrakcia lokálnych (SIFT) a globálnych (JCD) vlastností,
3. vyhľadanie podobných vlastností k dopytu
4. vyhodnotenie a priradenie slov cieľovej fotografii – pravdepodobnostný model, overenie geometrickej konzistencie lokálnych vlastností.



Obrázok 11 - Všeobecný model navrhutej metódy pre automatické anotovanie fotografií.

5.1 Extrakcia vlastností

5.1.1 Výpočet lokálnych vlastností

Lokálne vlastnosti sú viac diskriminujúce ako globálne a vďaka tejto vlastnosti sú vhodné pre vyhľadanie konkrétnych (špecifických) objektov alebo tých, ktoré majú niektoré časti (v zmysle obsahu na fotografii) spoločné.

V prípade, že má fotografia horizontálne/vertikálne rozlíšenie viac ako 768 pixelov, tak je zmenšená v pôvodnom pomere na požadovaný rozmer. Fotografia s menším rozmerom je zachovaná. Pre fotografiu extrahujeme maximálne 500 vlastností metódou SIFT. Ich počet na fotografiu závisí od jej vizuálneho obsahu. Napríklad, pre fotografiu, na ktorej je odfočená jasná obloha (homogénne farby) s letiacim vtákom sa extrahuje výrazne menej vlastností ako pri fotografii, na ktorej je odfočená kontrastná pestrofarebná záhrada s rôznorodými kvetmi. Každý extrahovaná vlastnosť obsahuje informácie o jej pozícii (koordináty x , y), orientácii, veľkosti a 128-rozmerný vektor príznakov (deskriptor). Nevýhodou je, že SIFT vlastnosti sú extrahované z fotografie v stupnici šedej (*grayscale*), a preto chýbajú v deskriptore informácie týkajúce sa farieb. Tento nedostatok kompenzuje aj kombináciou s globálnymi vlastnosťami.

5.1.2 Výpočet globálnych vlastností

Globálne vlastnosti nám zabezpečia všeobecnosť. Ako bolo spomenuté, ak nedokážeme identifikovať objekty na cieľovej fotografii (trénovacia množina neobsahuje výrazne podobné fotografie), tak vďačným opisom je aj určenie komplexnej kategórie (napr. všeobecne auto bez určenia značky a typu), alebo priradenie frekventovaných kľúčových slov akými sú napr. obloha, more, slnko. Ďalším prínosom sú pri veľkých (relatívne) homogénnych oblastiach, ktoré spravidla neopisujeme lokálnymi vlastnosťami.

V prípade výpočtu globálnych vlastností, fotografiu interpolujeme na pomer 3:2 podľa toho, k akému rozlíšeniu (768x512, 384x256, 192x128, 96x64, 48x32) má pôvodné rozlíšenie bližšie. Napríklad, ak má pôvodná fotografia rozmery 768x576 pixelov, tak ju upravíme na rozmer 768x512 pixelov. Interpoláciou môže v určitých prípadoch dôjsť k deformácii, avšak z globálneho pohľadu na fotografiu nie príliš zásadnej. Ďalej fotografiu rozdelíme na fixný počet blokov. Ich počet závisí od upraveného rozlíšenia. V poradí uvedených rozlíšení, rozdelíme fotografiu na 16x16, 8x8, 4x4, resp. 2x2 bloky. Fotografia s rozmerom 48x32 pixelov predstavuje jeden blok. Ak má fotografia menej ako 24x16 pixelov, tak nie je zahrnutá do výpočtu globálnych vlastností. Pre každý blok vypočítame 168-rozmerný deskriptor JCD.

5.2 Indexovanie vlastností do LSH úložiska

Na ilustráciu, uvažujme fotografiu, z ktorej extrahujeme 500 lokálnych vlastností. Už pri 1000 fotografiách disponujeme s 500 000 vlastnosťami, ktoré sú opísané 128-rozmerným vektorom typu *double*. Extrahovaných údajov je veľa, sekvenčné vyhľadávanie je nemysliteľné, a preto sme použili na zhlukovanie podobných vlastností metódu Lokálneho senzitivného hašovania (angl. *LSH - Locality Sensitive Hashing* [11]). LSH patrí do rodiny metód prezývaných ako hľadanie najbližších susedov (angl. *NNS - Nearest Neighbour Search*).

Kľúčovou myšlienkou LSH je, že objekty, ktoré sú si podobné sú vložené do rovnakého koša (angl. *bucket*) s vysokou pravdepodobnosťou. Základnými parametrami pre LSH algoritmus je dvojica L a k . Majme množinu $G = \{g_1, \dots, g_L\}$, ktorú značíme ako rodina hašovacích funkcií g a ďalej majme množinu hašovacích funkcií H . Každá funkcia g_i ($i = 1, 2, \dots, L$) je tvorená zret'azením k hašovacích funkcií $h_1, \dots, h_k \in H$, tzn. $g(v) = [h_1(v), \dots, h_k(v)]$, kde v je d -rozmerný vektor príznakov (deskriptor, v našom prípade 128, alebo 168 rozmerný vektor). Inými slovami, náhodná hašovacia funkcia g je vytvorená zret'azením k náhodne vybratých hašovacích funkcií z množiny H . Funkcia h je definovaná takto:

$$h(p)_{(a,b)} = \left\lfloor \frac{a \cdot v + b}{w} \right\rfloor,$$

kde a je d -rozmerný náhodný vektor, ktorého elementy sa určujú z p -stabilného rozdelenia pravdepodobností (napr. *Gaussovo rozdelenie*), b je reálne číslo z intervalu $(0, w)$, parameter w je veľkosť koša, $a \cdot v$ je skalárny súčin náhodného vektora a a vektora príznakov v . Optimálna hodnota parametra w závisí od celkovej množiny údajov a od vektora príznakov v , ktorý tvorí dopyt. V článku [11] ukázali, že parameter

w nastavený na hodnotu 4.0 poskytuje všeobecne dobré výsledky, a preto sme experimentálne zvolili túto hodnotu.

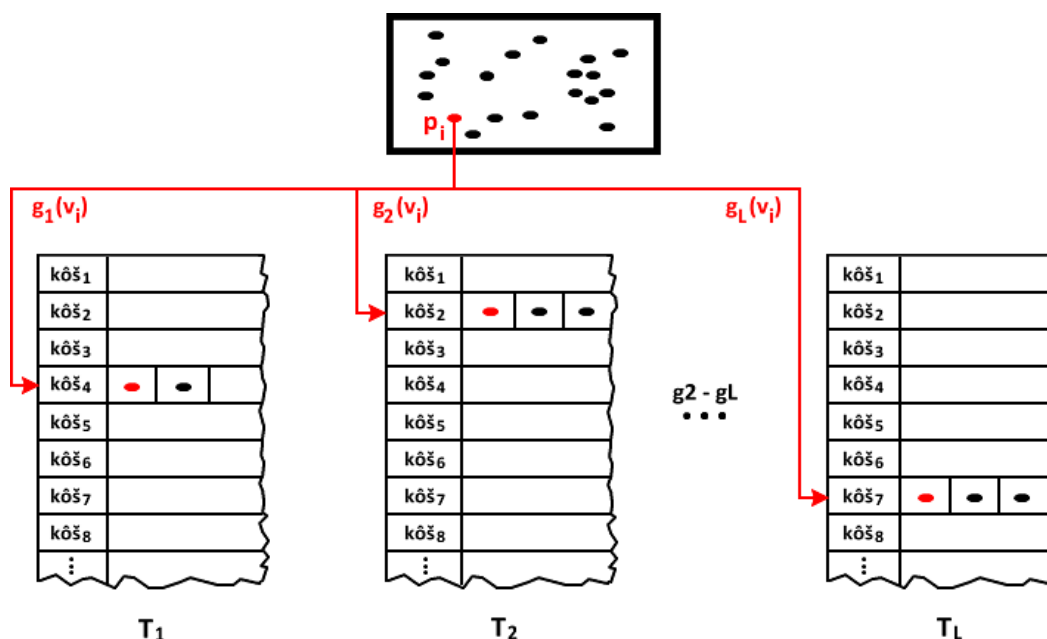
Keďže výsledkom každej funkcie g_i je kľúč (angl. *hash*) v tvare k -rozmerného vektora $x_i = (c_{i_1}, c_{i_2}, \dots, c_{i_k}) = g_i(h_1(v), h_2(v), \dots, h_k(v)) = g_i(v)$, je žiaduce vytvoriť z dôvodu efektivity celočíselný identifikátor koša [1]. Na jeho vytvorenie použijeme funkciu f , ktorá zabezpečuje vysokú pravdepodobnosť, že dva rozdielne koše budú mať aj rozdielne identifikátory. Funkcia f je definovaná takto:

$$f(x_i) = \left(\sum_{n=1}^k r_n c_{i_n} \right) \text{ mod } \textit{prime},$$

kde x_i je k -rozmerný kľúč vypočítaný pre vektor príznačov v hašovacou funkciou g_i , r_d sú náhodné celé kladné čísla, c_{i_n} sú elementy vektora x_i a konštanta $\textit{prime} = 2^{64} - 59$ je najväčšie 64-bitové prvočíslo. Pre každú hašovaciu funkciu g_i je vytvorená hašovacia tabuľka T_i .

Vloženie (zaindeoxvanie) lokálnej/globálnej vlastnosti p do zodpovedajúceho koša prebieha tak, že vektor príznačov v (deskriptor) je postupne zahašovaný funkciami g_1, \dots, g_L (pozri obr. 12). Výsledkom zloženej funkcie $f(g_i(v))$ je identifikátor koša (kľúč), do ktorého je vlastnosť p pridaná v tvare: súradnica X, súradnica Y a identifikátor fotografie, z ktorej bola vlastnosť extrahovaná. Vektor príznačov v , orientácia a veľkosť sú uložené separátne v tabuľke extrahovaných vlastností.

Pre parametre k a L platí, že so zvyšujúcim k sa redukuje šanca, že budú v jednom koši najbližší susedia, resp. so zvyšujúcim L sa zvyšuje šanca, že budú v jednom koši najbližší susedia. Tieto parametre určíme experimentálne pomocou metódy opísanej v [1]. Vstupom metódy pre určenie parametra L je parameter k a súčasne pravdepodobnosť Pr , s akou budú dva podobné si body v jednom koši. Pre parametre $k = 7$ a $Pr = 0.97$ má vypočítaný parameter L hodnotu 15 ($L = 15$).



Obrázok 12 – LSH, indexovanie vlastnosti p do zodpovedajúceho koša vo všetkých L tabuľkách.

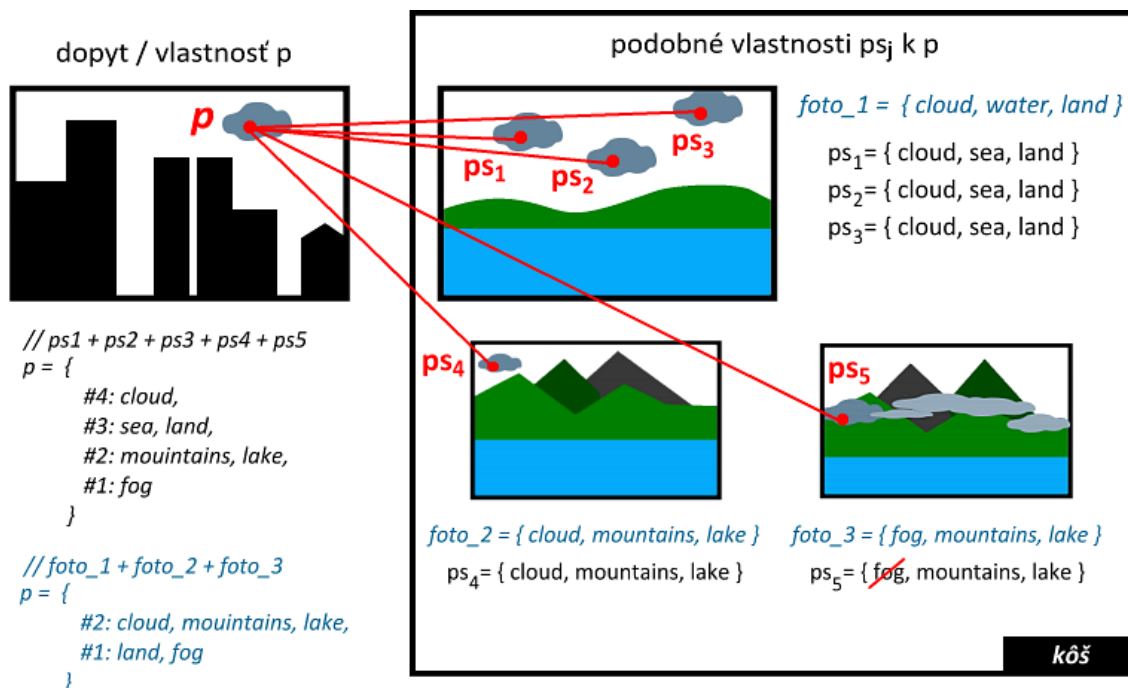
Metóda LSH sa spravidla používa pre ukladanie údajov v operačnej pamäti (angl. *memory-based*). Z dôvodu veľkého množstva extrahovaných vlastností sme museli navrhnúť variant, ktorý bude uchovávať údaje na disku (angl. *disk-based*) s tým, že bude prístup k údajom možný v reálnom čase. Keďže metóda používa náhodne generované parametre, je potrebné tieto vygenerovať „raz“ a uložiť pre opätovné použitie. V prípade potreby ukladania takého enormného množstva údajov je relačná databáza nevyhovujúca, pretože by s narastajúcim počtom extrahovaných vlastností (pribúdaním fotografií) nebolo možné zabezpečiť dopyty v reálnom čase. Z tohto dôvodu sme sa rozhodli použiť vysoko škálovateľný distribuovaný databázový systém *Cassandra*⁴, ktorý je špeciálne navrhnutý pre spracovanie veľkého objemu údajov. Jedná sa o hybrid medzi stĺpcovo orientovaným DBMS (angl. *Database Management System*) a riadkovo orientovaným DBMS. Model sme špeciálne navrhli s ohľadom na vlastnosti LSH a možnosti DBMS *Cassandra* (podrobnosti v technickej dokumentácii).

5.3 Vyhľadanie podobných vlastností k dopytu

Z dopytu, resp. cieľovej fotografie sa vytvorí kópia, ktorá je upravená pre potreby výpočtov lokálnych a globálnych vlastností. Po extrakcii vlastností je postupne deskriptor zahašovaný rovnakou metódou ako pri mapovaní (vkladaní) do LSH úložiska. Pre každú cieľovú vlastnosť p teda vypočítame L kľúčov (K_{p1}, \dots, K_{pL}), kde L závisí od počtu hašovacích tabuliek ($L=15$). Každý kľúč K_{pi} ($i = 1, \dots, L$) je v jemu prislúchajúcej tabuľke T_i odkazom do koša (angl. *bucket*), v ktorom sa nachádzajú podobné vlastnosti ps_{ij} k p ($T_i(K_{pi}) = \{ps_{ij}\}$). Vlastnosť ps_{ij} má priradené slovo (slová) a identifikátor fotografie, z ktorej bola extrahovaná. Všetky získané slová z podobných vlastností ps_{ij} zoskupíme a usporiadame podľa počtu ich výskytov. Usporiadany zoznam následne priradíme cieľovej vlastnosti p . Získané usporiadanie slov reprezentuje pravdepodobnosť, s akou slovo opisuje cieľovú vlastnosť. Ak je p lokálna vlastnosť, tak všetky podobné lokálne vlastnosti ps_{ij} zoskupíme podľa fotografií, z ktorých boli extrahované (pomocou identifikátora fotografie). Tieto zoskupenia tiež priradíme p .

Na obr. 13 je vyobrazená situácia, keď cieľová vlastnosť p (objektu záujmu) má viac podobných vlastností ps_j na fotografii *foto_1*. V takom prípade je vyššia pravdepodobnosť, že nejaké zo slov, ktoré sú priradené k *foto_1* opisuje práve objekt záujmu. Medzi podobné vlastnosti k p bola do rovnakého koša vložená aj vlastnosť ps_5 , z dôvodu, že sa časť „hmlového mraku“ výrazne podobala časti mraku na cieľovej fotografii. Keďže každá z vlastností ps_1 , ps_2 a ps_3 má priradené slová z *foto_1* osobitne, tak sme docielili, že cieľovej vlastnosti p vieme určiť najpravdepodobnejšie slovo, ktoré ju „opisuje“ (#4: *cloud*). Ak by sme uvažovali, že cieľovej vlastnosti p priradíme len slová fotografií, v ktorých sa nachádzajú nájdené podobné vlastnosti ps_j , tak v danej situácii nevieme určiť, ktoré zo slov „opisuje“ cieľovú vlastnosť, keďže početnosť slov *cloud*, *mountains* a *lake* je rovnaká (#2: *cloud*, *mountains*, *lake*).

⁴ <http://cassandra.apache.org/>

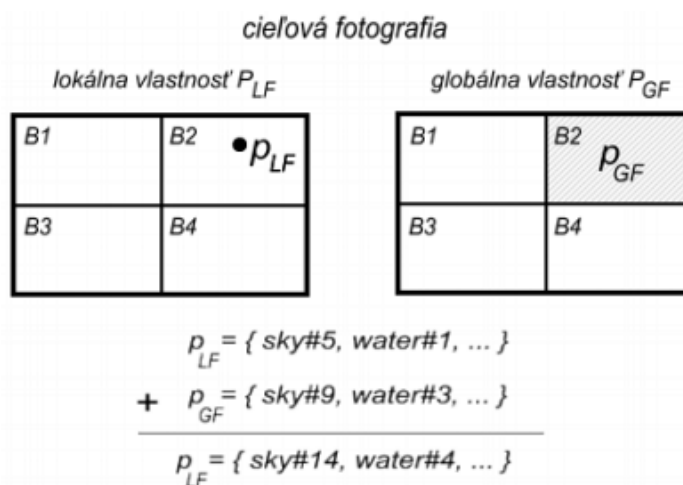


Obrázok 13 – Známenie výhodnosti riešenia, keď má každá vlastnosť priradené slová osobitne.

5.4 Vyhodnotenie a priradenie slov

Výstupom predchádzajúceho kroku je, že každá cieľová vlastnosť p (lokálna p_{LF} , globálna p_{GF}) extrahovaná z cieľovej fotografie (dopytu) „disponuje“ zoznamom usporiadaných slov podľa ich početnosti a v prípade p_{LF} aj podobnými lokálnymi vlastnosťami, ktoré sú zoskupené podľa fotografií, z ktorých boli extrahované. Naším cieľom je v tomto kroku priradiť cieľovej fotografii všetky slová, ktoré ju opisujú a súčasne na nej pomenovať objekty záujmu.

Pre cieľové lokálne vlastnosti p_{LF} vieme pomocou ich súradníc zistiť, do ktorých blokov B patria. Na obr. 14 je znázornená cieľová lokálna vlastnosť p_{LF} , ktorá patrí do bloku B_2 . Cieľová globálna vlastnosť p_{GF} bola vypočítaná z bloku B_2 . Ku každej p_{LF} preto pripočítame početnosti slov z prislúchajúcej p_{GF} (za predpokladu, že existujú).



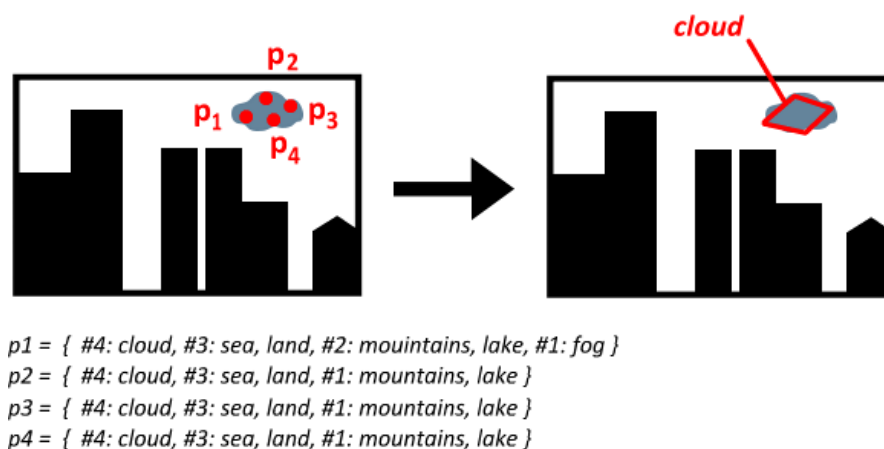
Obrázok 14 – Pripočítanie početnosti slov k cieľovej lokálnej vlastnosti P_{LF} z prislúchajúcej cieľovej globálnej vlastnosti P_{GF} .

Následne cieľové vlastnosti zoskupíme podľa prvých slov (s najvyššou početnosťou), vyberieme prvé slová v zoskupeniach, spočítame ich početnosť, usporiadame a priradíme cieľovej fotografii.

Z cieľových lokálnych vlastností p_{LF} , ktoré majú priradené rovnaké slovo s najvyššou početnosťou (aspoň 2) vytvoríme skupiny. Z nich vytvoríme pomenované polygóny na cieľovej fotografii, ktoré budú ohraničovať objekty záujmu. Na určenie ohraničení objektov je však potrebné overiť geometrickú konzistenciu zoskupených vlastností voči korešpondujúcim lokálnym vlastnostiam, ktoré k nim boli priradené. Každá z vlastností p_{LF} má priradené podobné vlastnosti, ktoré sú zoskupené podľa fotografií. Vyberieme tie vlastnosti, ktoré patria rovnakej fotografii a ich počet je najvyšší. Týmto vytvoríme korešpondujúce páry a hľadáme projektívnu transformáciu T (homografiu) použitím metódy RANSAC (angl. *RANdom SAMple Consensus* [15]). Výsledkom metódy je T , ktorá správne transformuje najvyšší počet korešpondujúcich párov. Opis metódy:

1. Vytvorenie vzorky: Označme cieľovú fotografiu f_1 a k nej podobnú fotografiu f_2 . Náhodne sa vyberie k korešpondujúcich párov $(p_1, q_1), (p_2, q_2), \dots (p_k, q_k)$, kde $p_i \in f_1$ a $q_i \in f_2$.
2. Zo vzorky sa vypočíta projektívna transformácia T .
3. Všetky korešpondujúce body $p_i \in f_1$ sa transformujú použitím T .
4. Ak je T vypočítaná zo správnych párov, potom by mali transformované body p_i ležať na miestach korešpondujúcich bodov q_i , tzn. $p_i = T(q_i)$ (tieto sa angl. označujú *inliers*). Kvalita transformácie je vyhodnotená z počtu bodov, ktoré spĺňajú podmienku: $d(p_i, T(q_i)) < \varepsilon$, tzn., euklidovská vzdialenosť medzi p_i a $T(q_i)$ je menšia ako zvolený prah ε (typicky 1 -3 pixely). Tie, ktoré podmienku nespĺňajú sa v angl. označujú *outliers*.
5. Body 1 – 4 sa iteratívne opakujú a hľadá sa taká T , ktorá správne transformuje najvyšší počet bodov. Počet iterácií je buď fixne daný (napr. 1000), alebo je vypočítaný z odhadu pravdepodobnosti nájdania správnej vzorky.

Z cieľových lokálnych vlastností, ktoré zostali v skupinách (*inliers*) vytvoríme polygóny ohraničujúce objekty záujmu a pomenujeme ich slovom s najvyššou početnosťou v skupine (pozri obr. 15).



Obrázok 15 – Z cieľových vlastností je vytvorený pomenovaný polygón ohraničujúci objekt záujmu.

6 Overenie metódy

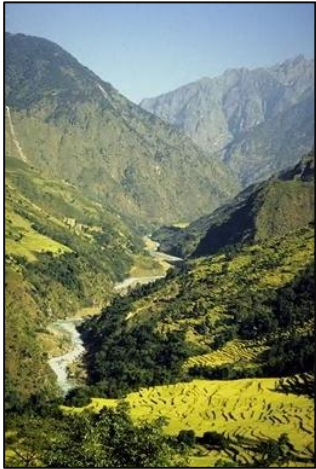
Na overenie metódy sme použili Corel5K korpus, ktorý obsahuje 5000 fotografií. Tento korpus sa všeobecne používa pri metódach automatického anotovania na vyhodnotenie a overenie výsledkov. Obsahuje rôzne témy počnúc mestskými scenériami, fotografie zachytávajúce rozmanitosť prírody, zvieratá, príp. umelecké objekty. Je rozdelený do dvoch množín, a to trénovacej a testovacej. Každá fotografia má priradené 1 – 5 slov, ktoré vytvoril človek. Jedná sa prevažne o opisy s charakterom všeobecnej kategórie ako obloha, voda, budovy, piesok, pláž, let vtáka, pohorie, krajina, skaly, oblak, strom atď. Pre trénováciu množinu je celkovo vytvorených 374 slov.

Metódu sme vyhodnotili so 40 fotografiami z testovacej množiny. V tab. 1 sú uvedené 4 ukážky výsledkov priradenia anotácií nami navrhnutou metódou fotografiám z testovacej množiny (v prílohe sú uvedené ďalšie ukážky). Naša metóda vráti maximálne 5 najlepších výsledkov, ktoré sú usporiadané zľava podľa relevancie (najvyššieho „skóre“), inými slovami, pravdepodobnosť, že sa prvé slovo nachádza na cieľovej fotografii je najvyššia.

Zo získaných výsledkov sme vypočítali presnosť (angl. *precision*) a pokrytie (angl. *recall*). Pre každú fotografiu sme vypočítali presnosť ako podiel počtu získaných relevantných slov k počtu všetkých získaných slov. Pokrytie sme vypočítali ako podiel počtu získaných relevantných slov k počtu relevantných slov (slov vložených človekom). Celkovú presnosť a pokrytie sme vypočítali ako podiel súčtu čiastkových výsledkov a počtu anotovaných fotografií (40). Výsledky pre jednotlivé fotografie sú taktiež uvedené v prílohe. Celková presnosť je 0,435 a pokrytie 0,612.

V niektorých prípadoch slová získané našou metódou boli relevantné, avšak anotácie vytvorené človekom tieto slová neobsahovali. Preto sme vyhodnotili presnosť osobitne pre získané slová tak, že sme určili relevantné slová zo získaných slov. Relevantné slová sú označené tučným písmom a nová presnosť je uvedená v hranatých zátvorkách. Celková presnosť je 0,565.

Tabuľka 1 – Ukážka výsledkov priradenia anotácií nami navrhnutou metódou fotografiám z testovacej množiny korpusu Corel5K.

<p>Názov súboru: 187072.jpg</p> <p>Anotácia človekom: mountain, river, sky, water</p> <p>Metóda: sky, tree, stone, mountain, clouds</p> <p>Presnosť, pokrytie: 2/5; 2/4</p>	
<p>Názov súboru: 276029.jpg</p> <p>Anotácia človekom: mountain, sky, tree, water</p> <p>Metóda: sky, tree, mountain, water, stone</p> <p>Presnosť, pokrytie: 4/5; 4/4</p>	
<p>Názov súboru: 384073.jpg</p> <p>Anotácia človekom: <i>people, sand, sky, water</i></p> <p>Metóda: <i>sky, water, people, sand, beach</i></p> <p>Presnosť, pokrytie: 4/5; 4/4 [5/5]</p>	

Názov súboru:

46096.jpg

Anotácia človekom:

kauai, people

Metóda:

people, kauai, hawaii, rocks, hats

Presnosť, pokrytie:

2/5; 2/2



7 Zhodnotenie

Navrhli sme metódu pre automatické anotovanie fotografií, ktorá kombinuje lokálne a globálne vlastnosti za účelom získania najlepších výsledkov pre cieľovú fotografiu. Touto kombináciou sme sa snažili docieľiť kompromis tak, aby bola naša metóda použiteľná pre identifikáciu objektov a súčasne bola vhodná aj pre určenie všeobecnej kategórie. Väčšina existujúcich metód je založená na vyhľadani najpodobnejších fotografií k cieľovej fotografii. V našej metóde sa pri hľadaní podobných kandidátov zameriavame na analýzu fotografií. Vychádza z hypotéz o početnosti a dominancii objektov na fotografii a pravdepodobnosti, že sa k nim opis vzťahuje.

Jedným z najväčších problémov bol spôsob ukladania extrahovaných vlastností tak, aby bolo možné vykonať dopyt pre ich výber s odozvou v reálnom čase. Za týmto účelom sme analyzovali rôzne možnosti a zvolili sme DBMS Cassandra, ktorý je špeciálne navrhnutý pre uloženie veľkého objemu údajov.

Sila lokálnych vlastností spočíva v rozpoznávaní objektov s určitou invarianciou. Corel5K korpus obsahuje však prevažne fotografie, ktoré sú tematicky zamerané na všeobecnú kategóriu, čomu zodpovedajú aj anotácie, ktoré sú k fotografiám priradené. Z tohto dôvodu zohrali pri experimente svoju úlohu prevažne globálne vlastnosti, ktoré sú schopné zabezpečiť všeobecnosť. Preto by bolo vhodné overiť metódu na korpuse, ktorý je orientovaný viac na anotovanie objektov. Na zlepšenie výsledkov sme preto implementovali nástroj ANNO^R, ktorý umožňuje vytváranie modelov. Modely sú vytvárané ručnou segmentáciou fotografie. Každý model je jednoznačne pomenovaný a reprezentuje nejaký objekt sveta. Do nástroja sme importovali celý Corel5K korpus a požiadali 30 zodpovedných študentov, aby z fotografií extrahovali modely. Používateľ dostal k dispozícii fotografiu a slovo, pre ktoré mal vytvoriť model. Na korektné porovnanie výsledkov so súčasnými by sme potrebovali vytvorenie modelov zo všetkých fotografií. V súčasnosti disponujeme modelmi z 1085 fotografií a celkovo bolo vytvorených 4590 modelov. Podrobnejšie informácie o nástroji ANNO^R sa nachádzajú v prílohe.

Ďalším významným nadobudnutým poznatkom je, že pri lokálnych vlastnostiach je kľúčové overenie ich geometrickej konzistencie. Bez overenia sú výsledky prakticky nepoužiteľné. V súčasnej implementácii sme extrahovali v priemere 1MB metadát na jednu fotografiu.

Literatúra

1. Andoni, A., Indyk, P.: LSH Algorithm and Implementation (E2LSH), manual June 21, 2005, [November 5, 2010]. <http://www.mit.edu/~andoni/LSH/manual.pdf>
2. Agarwal, A., Roth, D.: Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(11): pp. 1475–1490.
3. Blei, D., M., Jordan, M., I.: Modeling annotated data. In *Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in informaion Retrieval*. SIGIR '03. ACM, New York, NY, 2003, pp. 127-134.
4. Bin, W., et al: Automatic Refinement of Keyword Annotations for Web Image Search, *Advances in Multimedia Modeling*, 2006, pp. 259-268.
5. Cosmin, A., Philippe, B.: SIFT-CCH: Increasing the SIFT distinctness by Color Co-occurrence Histograms, *Proceedings of 5th IEEE International Symposium on Image and Signal Processing and Analysis*, ISPA 2007, Istanbul, 2007, pp.130-135.
6. Chatzichristofis, S., A., et al: SpCD - Spatial Color Distribution Descriptor - A fuzzy rule-based compact composite descriptor appropriate for hand drawn color sketches retrieval. In: *Proceedings ICAART*. 2010, pp. 58–63.
7. Chatzichristofis, S., A., Boutalis, Y., S.: Content based radiology image retrieval using a fuzzy rule based scalable composite descriptor, *Multimedia Tools and Applications, Special Issue on Data Semantics for Multimedia Systems*, Springer, 2009.
8. Chatzichristofis, S., A., et al: Selection of the proper compact composite descriptor for improving content based image retrieval. In: *Proceedings SPPRA*. 2009, pp. 134–140.
9. Chatzichristofis, S., A., Boutalis, Y., S.: CEDD: Color and edge directivity descriptor - A compact descriptor for image indexing and retrieval, *6th International Conference in advanced research on Computer Vision Systems ICVS 2008*, Santorini, Greece, 2008.
10. Chatzichristofis, S., A., Boutalis, Y., S.: FCTH: Fuzzy color and texture histogram - A low level feature for accurate image retrieval, *9th International Workshop on Image Analysis for Multimedia Interactive Services*, *Proceedings: IEEE Computer Society*, Klagenfurt, Austria, .
11. Datar, M., et al: Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the 20th annual symposium on Computational geometry (SCG '04)*. ACM, New York, NY, USA, 2004, pp. 253-262.
12. H. Tamura, et al: Textural Features Corresponding to Visual Perception. *IEEE Transaction on Systems, Man, and Cybernetcs*, Vol. SMC-8, No. 6, 1978, pp. 460–472.
13. Duygulu, P., et al: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the European Conference on Computer Vision*, 2002, pp. 97–112.

14. Eitz, M., et al: A descriptor for large scale image retrieval based on sketched feature lines. In *Proceedings of the 6th Eurographics Symposium on Sketch-Based interfaces and Modeling* (New Orleans, Louisiana, 2009). SBIM '09. ACM, New York, NY, 2009, pp. 29-36.
15. Fischler, M., A., Bolles, R., C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981, pp. 24(6):381–395.
16. Grabner, M., et al: Fast approximated SIFT. In: ACCV, 2006, pp. I:918-927.
17. Heesch, D., et al: NNk networks and automated annotation for browsing large image collections from the world wide web. In *Proceedings of the 14th Annual ACM international Conference on Multimedia* (Santa Barbara, CA, USA, October 23 - 27, 2006). MULTIMEDIA '06. ACM, New York, NY, 2006, pp. 493-494.
18. Kanade, T.: *Picture Processing System by Computer Complex and Recognition of Human Faces*. Ph.D dissertation, Kyoto University, 1973.
19. Ke, Y., et al: An efficient parts-based near-duplicate and sub-image retrieval system. In *Proc. of the 12th Annual ACM Int. Conf. on Multimedia* (New York, NY, USA, 2004). New York, NY, 2004, pp. 869-876.
20. Ke, Y. and Sukthankar, R., "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. 506-513.
21. K. Rodden: Evaluating similarity-based visualisations as interfaces for image browsing. PhD thesis, University of Cambridge, Computer Laboratory, September 2002. [May 10, 2009]. <http://www.cl.cam.ac.uk/Research/SRG/opera/publications/Papers/UCAM-CL-TR-543.pdf>.
22. Lavrenko, R., M., Jeon, J: A model for learning the semantics of pictures. In: *Proc of the 16th Conference on Advances in Neural Information Processing Systems NIPS*, MIT Press, 2003.
23. Li, Z., et al: Intuitive and effective interfaces for WWW image search engines. In: *Proc. of the 12th Annual ACM international Conference on Multimedia*, MULTIMEDIA '04. ACM, New York, NY, 2004, pp. 748-749.
24. Liu, J., Li, M., Ma, W., Liu, Q., and Lu, H. 2006. An adaptive graph model for automatic image annotation. In: *Proc. of the 8th ACM international Workshop on Multimedia information Retrieval* (Santa Barbara, California, USA, October 26 - 27, 2006). MIR '06. ACM, New York, NY, 61-70.
25. Lowe, D., G.: Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision* 60, 2004, pp. 91-110.
26. Mach, L.: Scale Invariant Feature Transform. Prezentácia. [May 10, 2009]. <http://artax.karlin.mff.cuni.cz/~machl5bm/sift/SIFT.pdf>
27. Mori, H., T., Oka, R.: Image-to-word transformation based on dividing and vector quantizing images with words. In: *Proc. of the International Workshop on Multimedia Intelligent Storage and Retrieval Management*, 1999.
28. MPEG-7, Multimedia content description interfaces. Part 3: Visual, ISO/IEC 15938-3:2002, 2002.
29. Smeulders, A., et al: Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, pp. 22(12):1349–1380.
30. Torralba, Oliva, A: Statistics of natural image categories. *Network: Computation in Neural Systems*, 2003, pp. 14:391–412.

31. Viola, Jones, M: Rapid object detection using a boosted cascade of simple features. In: *International Conference on Pattern Recognition*, 2001, pp. 511–518.
32. Wang, L., et al: Automatic image annotation and retrieval using subspace clustering algorithm. In: *Proc.of the 2nd ACM international Workshop on Multimedia Databases* (Washington, DC, USA, November 13 - 13, 2004). MMDB '04. ACM, New York, NY, 2004, pp. 100-108.
33. Wang, X., et al: AnnoSearch: Image Auto-Annotation by Search. In: *Proc. of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. CVPR*. IEEE Computer Society, Washington, DC, 2006, pp. 1483-1490.
34. Xu, Z., et.al: Robust image copy detection using multi-resolution histogram. In *Proceedings of the international Conference on Multimedia information Retrieval* (Philadelphia, Pennsylvania, USA, 2010). MIR '10. ACM, New York, NY, 2010, pp.129-136.
35. Zagoris, K., et al: IMG(ANAKTISI): A web content based image retrieval system, *2nd International Workshop on Similarity Search and Applications (SISAP)*, Proceedings: IEEE Computer Society, Prague, Czech Republic , 2009, pp.154-155.
36. Žára, J., et al: *Moderní počítačová grafika*. Computer Press, Brno, Czech Republic, 2004.

Príloha A – Technická dokumentácia

A.1 Dátový model LSH úložiska

Na obr. 16 je vyobrazený dátový model pre LSH úložisko. Tento model bol navrhnutý pre použitie v DBMS Cassandra (angl. *Database Management System*), ktorý predstavuje hybrid medzi stĺpcovo orientovaným DBMS a riadkovo orientovaným.

Keyspace v analógii s RDMS (angl. *Relational Database Management System*) reprezentuje databázu a *ColumnFamily* tabuľku. V našom modeli sú vytvorené 3, a to *Annor_Core*, *Annor_LF* a *Annor_GF*.

Annor_Core

V tabuľke *LSH_Data* sú uložené konštanty a náhodne vygenerované čísla, ktoré sa používajú pri výpočte kľúčov, resp. pri indexovaní vlastností do košov metódou LSH.

V tabuľke *photos_keypoints_list* sú uložené všetky lokálne vlastnosti, ktoré sú zoskupené podľa fotografií (*photoID*). Každá lokálna vlastnosť je identifikovateľná kľúčom v tvare *X_Y_photoID*, kde *X*, *Y* sú celočíselné súradnice vlastnosti a *photoID* je identifikátor fotografie, z ktorej bola vlastnosť extrahovaná. Identifikátor je fotografii pridelený pri jej spracovaní a je získaný z tabuľky *Sequences* (*photo_sequence_id*). Každá lokálna vlastnosť má uložené tieto údaje: vektor príznakov (128-rozmerný deskriptor vypočítaný metódou SIFT), kde každý príznak je typu *double*; súradnice *X*, *Y*; orientáciu a veľkosť.

V tabuľke *photos_gfeatures_list* sú uložené globálne vlastnosti, ktoré sú rovnako zoskupené podľa fotografií, z ktorých boli extrahované. Každá globálna vlastnosť je identifikovateľná kľúčom v tvare *#block_photoID*, kde *#block* je poradové číslo bloku (počítané z ľavého horného rohu fotografie po riadkoch) a *photoID* je identifikátor fotografie. Každá globálna vlastnosť má uložený vektor príznakov (168-rozmerný JCD deskriptor).

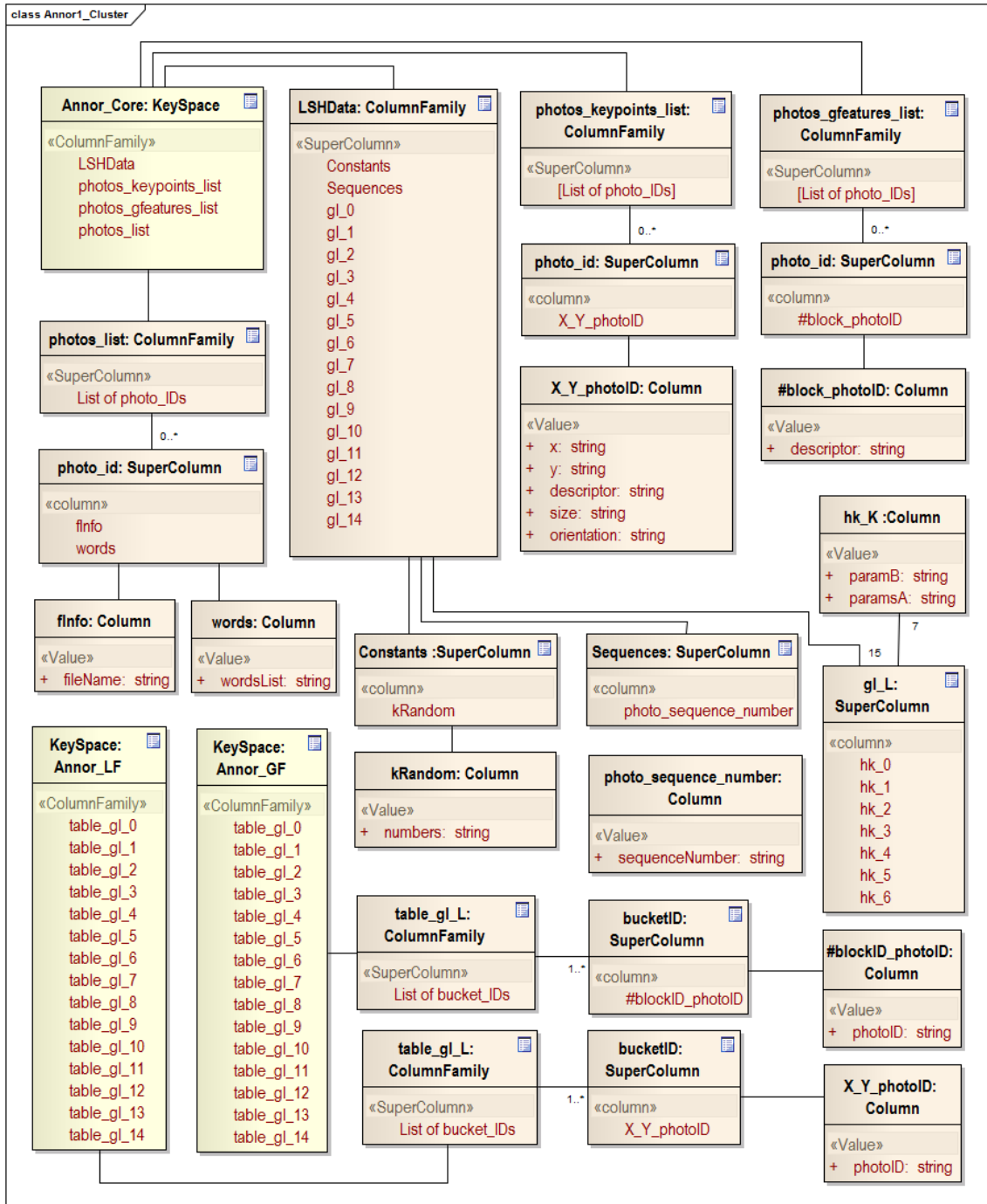
V tabuľke *photos_list* sú uložené všetky spracované fotografie, ktoré sú identifikovateľné pomocou kľúča *photoID*, ktorý im bol priradený pri ich spracovaní. Každá fotografia obsahuje informácie o názve súboru s cestou v súborovom systéme (*fInfo*) a slová, ktoré sú k fotografii priradené (*words*).

Annor_LF

V *Annor_LF* je vytvorených 15 tabuliek (*table_gl_0*, ..., *table_gl_14*), kde každá tabuľka reprezentuje jednu hašovaciu tabuľku, v ktorej sú zoskupené podobné lokálne vlastnosti do košov. V každej tabuľke sa nachádza zoznam košov, v ktorých sú uložené identifikátory lokálnych vlastností v tvare *X_Y_photoID*.

Annor_GF

V *Annor_GF* sú uložené (zoskupované) globálne vlastnosti. Štruktúra je rovnaká ako v *Annor_LF*, ale identifikátory globálnych vlastností sú v tvare #block_photoID.



Obrázok 16 - Model LSH úložiska navrhnutý pre použitie v DBMS Cassandra.

A.2 Ukážka zdrojového kódu

```
// vypocet LSH kluca pre i-tu tabulku (i = 0, ..., L-1)
public List<UInt32> ComputeHash(double[] featureVector, int gL)
{
    ValidateFeatureVector(featureVector);
    List<UInt32> vectorHashes = new List<UInt32>(featureVectorDimension);

    double value;
    for (int hK = 0; hK < LSHInitParams.ParameterK; hK++)
    {
        value = 0d;
        for (int d = 0; d < featureVectorDimension; d++)
        {
            value += featureVector[d] * gL_hK_Functions[gL][hK].ParamsA[d];
        }

        // p-stable distribution formula
        Int32 hashValInt = Convert.ToInt32((value +
            gL_hK_Functions[gL][hK].ParamB) / LSHInitParams.ParameterW);
        UInt32 hashValUInt = (hashValInt < 0)
            ? Convert.ToUInt32(hashValInt - Int32.MinValue)
            : Convert.ToUInt32(hashValInt);
        vectorHashes.Add(hashValUInt);
    }

    return vectorHashes;
}

// mapovanie LSH kluca na bucketID
public UInt64 ComputeBucketId64b(List<UInt32> fpHash,
    List<UInt64> kRandomConstants)
{
    CheckFeaturePointHash(fpHash);
    CheckKRandomConstants(kRandomConstants);

    BigInteger sum = new BigInteger(0);
    for (int i = 0; i < LSHInitParams.ParameterK; i++)
    {
        sum += kRandomConstants[i] * fpHash[i];
    }
    // MAX_64b_PRIME - maximlane 64 bitove prvcislo
    sum = (sum % SpecialFunction.MAX_64b_PRIME)

    return (UInt64) sum;
}
```

```

// pridanie lokalnej vlastnosti do LSH uloziska
public String AddLocalFeature(Keypoint keypoint, UInt64 photoId)
{
    ...

    // identifikator vlastnosti v kosi
    String keypointId = String.Format("{0}_{1}_{2}",
        Convert.ToUInt32(keypoint.X),
        Convert.ToUInt32(keypoint.Y),
        photoId);

    for (int gl = 0; gl < LSHInitParams.ParameterL; gl++)
    {
        // vlozenie lokalnej vlastnosti do zodpovedajuceho kosa v gl
        using (var dbLF = new CassandraContext(
            Properties.Settings.Default.keyspaceLF, connection))
        {
            List<UInt32> kpFeatureVector = lsh.ComputeHash(keypoint.FV, gl);
            UInt64 bucketId = lsh.ComputeBucketId64b(kpFeatureVector,
                KRandomConstants);

            var cfGlTable = dbLF.GetColumnFamily<BytesType, BytesType>
                (String.Format("table_gl_{0}", gl));

            dynamic bucket = cfGlTable.Get(String.Format("{0}", bucketId)).
                Take(1).FirstOrDefault();
            if (bucket.Columns.Count == 0)
            {
                bucket = cfGlTable.CreateRecord(key:
                    String.Format("{0}", bucketId));
                dbLF.Attach(bucket);
            }

            dynamic keypointGlTableData = bucket.CreateSuperColumn();
            keypointGlTableData.p = String.Format("{0}", photoId);
            bucket[String.Format("{0}", keypointId)] = keypointGlTableData;

            dbLF.SaveChanges();
        }

        ...
    }

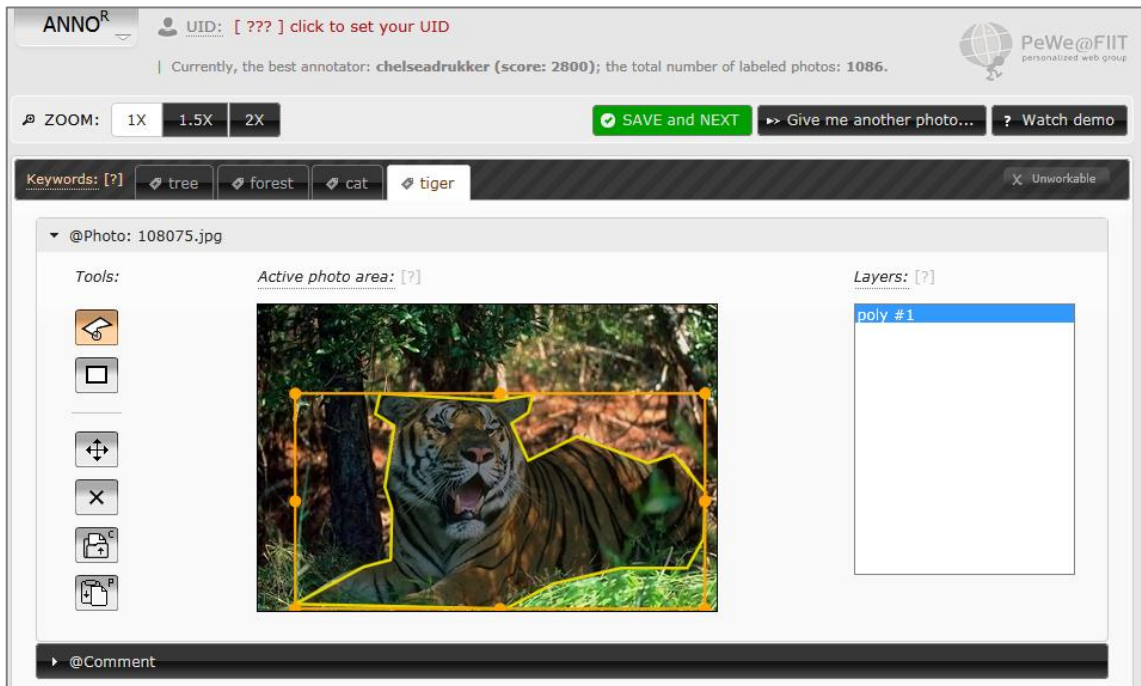
    return keypointId;
}

```


A.3 Nástroj ANNO^R

V súčasnosti závisí kvalita našej metódy predovšetkým od trénovacej množiny a množiny fotografií, ktorým chceme priradiť anotácie (cieľové fotografie). Identifikácia objektov záujmu na cieľovej fotografii je možná vtedy, ak náš trénovací dataset disponuje fotografiami, na ktorých sa vyskytujú objekty záujmu z cieľovej fotografie v „prípustnej invariancii“.

Naším cieľom je vybudovať robustný dataset, ktorý bude obsahovať „vizuálne modely“ objektov sveta, a za týmto účelom sme vytvorili online nástroj⁵ ANNO^R (pozri obr. 17) na ich vytváranie.



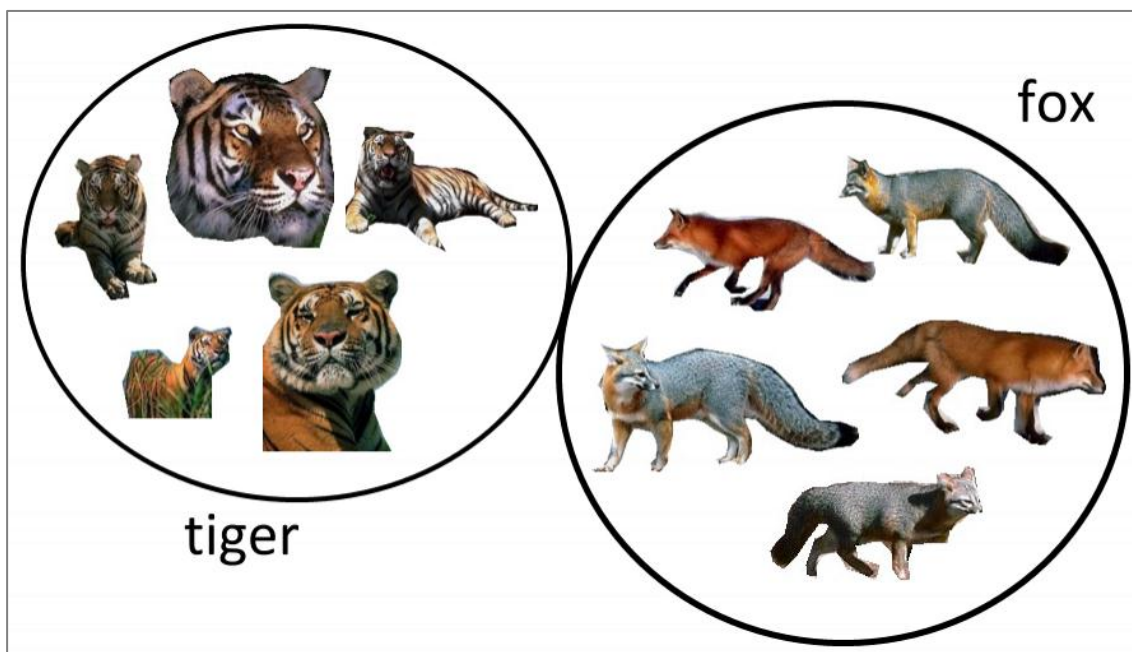
Obrázok 17 – Snímok z nástroja ANNO^R s vytvoreným segmentom ohraničujúcim tigra.

Používateľovi je vygenerovaná fotografia a slová, ktorá sú k nej priradené. Úlohou používateľa (anotátora) je pomocou nástroja obdĺžnik a polygón vytvoriť ohraničenia (segmenty) pre každé zo slov. Používateľ má možnosť vytvárať modely pod svojim pseudonymom a každé vytvorenie modelu je ohodnotené ziskom bodov. Čím je segment podrobnejší, tým získa viac bodov. Na obr. 18 sú vyobrazené vizuálne modely vytvorené používateľmi pomocou nástroja z fotografií Corel5K. Každý z modelov je jednoznačne pomenovaný, a tak je možné vytvárať ich zhľuky podľa typu objektu (pozri obr. 19).

⁵ <http://annor.laude.sk>



Obrázok 18 – Vizuálne modely zvierat, ktoré boli vytvorené používateľmi pomocou nástroja ANNO^R.






Obrázok 19 – Zhľuky vizuálnych modelov vytvorené podľa „typu“ objektov reálneho sveta.

Príloha B – Ukážka výsledkov anotovania

V tejto prílohe uvádzame výsledky anotovania fotografií z testovacej množiny datasetu Corel5K.

Tabuľka 2 - Testovacia množina fotografií z datasetu Corel5K s výsledkami metódy.

<p>Názov súboru: 103034.jpg</p> <p>Anotácia človekom: albatross, flight, sky</p> <p>Metóda: flight, wings, sky, snow, deer</p>	
<p>Názov súboru: 143076.jpg</p> <p>Anotácia človekom: desert, mountain, plants, valley</p> <p>Metóda: sky, sand, water, mountain, clouds</p>	
<p>Názov súboru: 103060.jpg</p> <p>Anotácia človekom: birds, booby, water</p> <p>Metóda: water, rocks, birds, tree, people</p>	

Názov súboru:
103088.jpg

Anotácia človekom:
iguana, lizard, marine, rocks

Metóda:
iguana, marine, lizard, rocks,
water



Názov súboru:
113007.jpg

Anotácia človekom:
field, foals, horses, mare

Metóda:
foals, horses, field, fence, mare



Názov súboru:
118021.jpg

Anotácia človekom:
buildings, harbor, shore, water

Metóda:
sky, water, people, sand, rocks



Názov súboru:
118041.jpg

Anotácia človekom:
buildings, cafe, shore, water

Metóda:
water, sky, buildings, stone,
mountain



<p>Názov súboru: 120003.jpg</p> <p>Anotácia človekom: sky, landscape, monastery, water</p> <p>Metóda: tree, sky, stone, water, mountain</p>	
<p>Názov súboru: 122062.jpg</p> <p>Anotácia človekom: buildings sky water waves</p> <p>Metóda: sky, water, people, sand, boats</p>	
<p>Názov súboru: 131004.jpg</p> <p>Anotácia človekom: land, grass, sky, tree</p> <p>Metóda: tree, mountain , sky, stone, ruins</p>	
<p>Názov súboru: 131027.jpg</p> <p>Anotácia človekom: horizon, house, tree, wall</p> <p>Metóda: tree, stone, sky, ruins, water</p>	

Názov súboru:
140089.jpg

Anotácia človekom:
farms, house, reflection, water

Metóda:
sky, water, horizon, chapel,
buildings



Názov súboru:
144014.jpg

Anotácia človekom:
coast, hills, lighthouse, water

Metóda:
water, sky, boats, people, **stone**



Názov súboru:
147064.jpg

Anotácia človekom:
flowers, hills, people, water

Metóda:
light, **people,** frozen, sculpture,
ice











Názov súboru:
144023.jpg





Anotácia človekom:
horizon, grass, hills, sky

Metóda:
desert, shrubs, **sky,** field, frost



<p>Názov súboru: 171074.jpg</p> <p>Anotácia človekom: frozen, water, snow</p> <p>Metóda: sky, snow, water, mosque, pillar</p>	
<p>Názov súboru: 276048.jpg</p> <p>Anotácia človekom: mountain, sky, tree, water</p> <p>Metóda: sky, stone, grass, watter, pyramid</p>	
<p>Názov súboru: 46007.jpg</p> <p>Anotácia človekom: oahu, people, waves</p> <p>Metóda: people, waves, oahu, water, buildings</p>	
<p>Názov súboru: 46093.jpg</p> <p>Anotácia človekom: kauai, people</p> <p>Metóda: kauai, tree, people, water, hawaii</p>	

<p>Názov súboru: 143078.jpg</p> <p>Anotácia človekom: mountain, tree, valley</p> <p>Metóda: sky, mountain, tree, water, sand</p>	
<p>Názov súboru: 384068.jpg</p> <p>Anotácia človekom: beach, rocks, sand, sky</p> <p>Metóda: sky, water, beach, sand, tree</p>	
<p>Názov súboru: 33065.jpg</p> <p>Anotácia človekom: clouds, jet, ruins, stairs</p> <p>Metóda: ruins, sky, stone, water, tree</p>	
<p>Názov súboru: 276087.jpg</p> <p>Anotácia človekom: mountain, sky, snow, tree</p> <p>Metóda: sky, mountain, tree, sand, water</p>	

<p>Názov súboru: 276071.jpg</p> <p>Anotácia človekom: clouds, mountain, sky, water</p> <p>Metóda: water, sky, mountain, grass, tree</p>	
<p>Názov súboru: 276019.jpg</p> <p>Anotácia človekom: mountain, sky, snow, water</p> <p>Metóda: sky, water, mountain, tree, stone</p>	
<p>Názov súboru: 201096.jpg</p> <p>Anotácia človekom: buildings, fence, sky</p> <p>Metóda: sky, water, buildings, stone, tree</p>	
<p>Názov súboru: 187091.jpg</p> <p>Anotácia človekom: mountain, people</p> <p>Metóda: people, mountain, water, sky, tree</p>	

Názov súboru:
171003.jpg

Anotácia človekom:
frozen, water, snow

Metóda:
water, snow, sky, birds, tree



Názov súboru:
17017.jpg

Anotácia človekom:
house, sky, water

Metóda:
sky, tree, water, stone, mountain



Názov súboru:
163070.jpg

Anotácia človekom:
birds, tree, flag

Metóda:
birds, tree, people, girl, branch







Názov súboru:
163062.jpg

Anotácia človekom:
birds, branch, bridge

Metóda:
birds, branch, tree, fly, people



<p>Názov súboru: 144029.jpg</p> <p>Anotácia človekom: coast, sky, water, waves</p> <p>Metóda: sky, water, stone, coast, tree</p>	
<p>Názov súboru: 142067.jpg</p> <p>Anotácia človekom: reflection, sky, tree, water</p> <p>Metóda: sky, tree, mountain, water, stone</p>	
<p>Názov súboru: 118008.jpg</p> <p>Anotácia človekom: boats, water, waves</p> <p>Metóda: water, waves, sky, people, shore</p>	
<p>Názov súboru: 103057.jpg</p> <p>Anotácia človekom: birds, booby, water</p> <p>Metóda: water, birds, sky, tree, waves</p>	

Názov súboru:
103057.jpg

Anotácia človekom:
beach, boats, sand, water

Metóda:
sky, stone, water, sand, grass,



V tab. 3 sú uvedené vypočítané presnosti a pokrytia pre testovaciu množinu. RI – počet relevantných získaných slov, R – počet relevantných slov, I – počet získaných slov.

Tabuľka 3 - Výpočet presnosti a pokrytia pre testovaciu množinu fotografií.

Názov fotografie	RI	I	R	Presnosť (RI/I)	Pokrytie (RI/R)
103034.jpg	2 [3]	5	3	0,4 [0,6]	2/3
143076.jpg	1 [3]	5	4	0,2 [0,6]	0,25
103060.jpg	2 [3]	5	3	0,4 [0,6]	2/3
103088.jpg	4 [5]	5	4	0,8 [1]	1
113007.jpg	4	5	4	0,8	1
118021.jpg	1 [3]	5	4	0,2 [0,6]	0,25
118041.jpg	2 [3]	5	4	0,4 [0,6]	0,5
120003.jpg	2 [3]	5	4	0,4 [0,6]	0,5
122062.jpg	2	5	4	0,4	0,5
131004.jpg	2	5	4	0,4	0,5
131027.jpg	1 [2]	5	4	0,2 [0,4]	0,25
140089.jpg	1 [2]	5	4	0,2 [0,4]	0,25
144014.jpg	1 [3]	5	4	0,2 [0,6]	0,25
147064.jpg	1	5	4	0,2	0,25
144023.jpg	1	5	4	0,2	0,25
171074.jpg	2 [3]	5	3	0,4 [0,6]	2/3
187072.jpg	2	5	4	0,4	0,5
276029.jpg	4	5	4	0,8	1
276048.jpg	2	5	4	0,4	0,5
46007.jpg	3 [4]	5	3	0,6 [0,8]	1
46093.jpg	2	5	2	0,4	1
46096.jpg	2	5	2	0,4	1
143078.jpg	2	5	3	0,4	2/3
384073.jpg	4 [5]	5	4	0,8 [1]	1
384068.jpg	3	5	4	0,6	0,75
33065.jpg	1 [3]	5	4	0,2 [0,6]	0,25
276087.jpg	3	5	4	0,6	0,75
276071.jpg	3 [4]	5	4	0,6 [0,8]	0,75
276019.jpg	3	5	4	0,6	0,75
201096.jpg	2	5	3	0,4	2/3
187091.jpg	2	5	2	0,4	1
171003.jpg	2	5	3	0,4	2/3
17017.jpg	2 [3]	5	3	0,4 [0,6]	2/3
163070.jpg	2	5	3	0,4	2/3
163062.jpg	2 [4]	5	3	0,4 [0,8]	2/3
144029.jpg	3 [4]	5	4	0,6 [0,8]	0,75
142067.jpg	3	5	4	0,6	0,75
118008.jpg	2	5	3	0,4	2/3
103057.jpg	2 [3]	5	3	0,4 [0,6]	2/3
103057.jpg	2 [4]	5	4	0,4 [0,8]	0,5
				0,435 [0,565]	0,612

Príloha C – Obsah dátového nosiča

- **obsah.txt** - popis súborov na nosiči
- **/Data** - dáta použité pri experimentoch
- **/Src** – zdrojové kódy
- **/Libs** – nástroje použité pri realizácii metódy
- **/Thesis** - text diplomovej práce v elektronickej forme, anotácie

Príloha D – Návrh článku na konferenciu SMAP 2011