



KATEDRA APLIKOVANEJ INFORMATIKY
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITA KOMENSKÉHO, BRATISLAVA
9.2.9 APLIKOVANÁ INFORMATIKA

ROZPOZNÁVANIE A KLASIFIKÁCIA MALIARSKYCH PLÁTIEN NA FOTOGRAFIÁCH

(Diplomová práca)

ZUZANA HALADOVÁ

Vedúci: Elena Šikudová, PhD.

Bratislava, 2010

ZADANIE DIPLOMOVEJ PRÁCE

Študent(ka) *Zuzana Haladova*

vypracuje za účelom obhajoby v rámci záverečného konania vo vysokoškolskom magisterskom štúdiu na UK FMFI diplomovú prácu s (predbežným) názvom:

Rozpoznávanie a klasifikácia maliarskych plátien na fotografiách

pod vedením pracovníka *RNDr. Elena Šikudová, PhD.* z pracoviska *KAI*

Cieľ diplomovej práce a ďalšie poznámky:

1. *Naštudovanie problematiky a špecifikácia úlohy*.....

2. *Implementácia navrhnutej metódy riešenia*

3. *Analýza výsledkov*

podpis študenta *Haladova* podpis ved. dipl. práce

Šikudová

Diplomová práca je tematicky priradená k špecializácii *Aplikovaná informatika*.....

V Bratislave dňa *17.12.2008* podpis gestora špec.: *Šikudová*

Poznámka: Vyplňujte v 3 exemplároch: 1 pre študenta, 1 pre vedúceho práce, 1 pre študijné oddelenie (odovzdá študent pri zápise do 4. ročníka).

Čestne prehlasujem, že som túto diplomovú prácu vypracovala samostatne s použitím citovaných zdrojov.

.....

Pod'akovanie

Ďakujem Elene Šikudovej za nedocniteľnú pomoc pri písaní tejto práce. Ďakujem svojej rodine, priateľovi, kamarátkam a spolužiakom za psychickú (a fyzickú) podporu a ďakujem tiež všetkým čo mi poskytli svoje fotografie obrazov.

Abstrakt

Od vzniku prvých internetových vyhľadávačov, sa oblasť vyhľadávania obrázkov posunula až k dnešným sofistikovaným systémom vyhľadávajúcim na základe vizuálneho obsahu. Prekážka vo forme sémantickej priepasti však spôsobila, že dodnes nie je možné správne identifikovať všetky objekty na obrázku. Napriek tomu, táto diplomová práca ponúka riešenie na klasifikáciu konkrétneho typu objektov - maliarskych plátnien. Navrhnuté riešenie zahŕňa segmentáciu maľby z obrázku, vytvorenie deskriptora maľby a jej klasifikáciu pomocou porovnávania jej deskriptora s vytvorenou databázou deskriptorov originálnych obrázkov malieb. Segmentácia maľby je zabezpečená pomocou štyroch rôznych metód založených na metódach Anisotropic diffusion, Gauss gradient, Kuwahara filter a segmentácia s využitím povodí. Na nájdenie kľúčových bodov a vytvorenie deskriptora boli použité techniky SIFT a SURF. Všetky použité metódy boli validované pomocou databázy 100 fotografií a 15 originálnych obrázkov malieb Rembrandta Harmenszoon van Rijna.

Kľúčové slová: Maliarske plátna, Segmentácia, Klasifikácia, SIFT, SURF, Rembrandt

Abstract

Since the development of the first text-based image search on the internet, the area of image retrieval has come a long way to sophisticated content based image retrieval systems. On the other hand, the semantic gap causes that it is still not possible to create a system which can correctly identify any object in the image. However, this thesis proposes a solution for classifying the one sort of objects - paintings. This approach includes segmentation of the painting from the image, creation of the descriptor file from the segmented painting, and classification of the painting by matching its descriptor file to the created database of descriptor files of original paintings. The segmentation of the painting is achieved with 4 different methods based on Anisotropic diffusion, Gauss gradient, Kuwahara filter and Watershed transform. For the estimation of key points and creation of the descriptor file, the SIFT (Scalable Invariant Feature Transform) or the SURF (Speeded Up Robust Features) technique is used. The performance of all techniques are validated within the thesis. The solution proposed in this paper was tested on the database of 100 photographs and 15 original images of the Rembrandt Harmenszoon van Rijn's paintings.

Key words: Fine art paintings, Segmentation, Classification, SIFT, SURF, Rembrandt

Obsah

1	Úvod do problematiky	2
1.1	Vyhľadávanie obrazov	2
1.1.1	Vyhľadávanie pomocou textovej informácie	2
1.1.2	Vyhľadávanie na základe vizuálneho obsahu	3
1.1.3	Budúci vývoj	10
1.2	Klasifikácia malieb	12
1.2.1	Prehľad prác	12
1.2.2	Prístup počítačovej grafiky a videnia k umeleckým štýlom	14
1.3	Prehľad analogických prác	17
2	Špecifikácia	18
2.1	Cieľ	18
2.2	Požiadavky	18
2.3	Databáza obrázkov	19
2.3.1	Originály obrazov	19
2.3.2	Fotografie turistov	19
2.4	Softvér	20
2.5	Funkcie	22
2.6	Algoritmy	23
2.6.1	Anisotropic diffusion	23
2.6.2	Kuwahara filter	23
2.6.3	Segmentácia s využitím povodí (Watershed)	23
2.6.4	Houghova transformácia	24

2.6.5	Sobelov operátor	25
2.6.6	Ekvalizácia histogramu	26
2.6.7	Top hat	26
2.6.8	Bottom hat	27
2.6.9	Extended minima transformácia	27
2.6.10	SIFT	27
2.6.11	SURF	29
3	Implementácia	31
3.1	Predspracovanie	31
3.2	Segmentácia	32
3.2.1	Gauss gradient	32
3.2.2	Anisotropic diffusion	34
3.2.3	Watershed	37
3.2.4	Kuwahara filter	39
3.3	Klasifikácia	39
3.3.1	SIFT	40
3.3.2	SURF	42
3.3.3	Pomer strán	43
3.4	MATLAB aplikácia	44
4	Validácia	47
4.1	Segmentácia	47
4.2	Klasifikácia	49
4.2.1	Pomer strán	52
4.2.2	Ďalšie metódy (ColorSIFT a HueSIFT)	53

Úvod

Ľudí, ktorí zbožňujú výtvarné umenie tak ako ja, určite teší nepreberné množstvo umeleckých galérií nachádzajúcich sa na internete. Tieto galérie ponúkajú možnosť si pri akejkoľvek príležitosti (napríklad aj vo vani) naštartovať zmysly a vychutnávať prostredníctvom internetu, zadarmo a neobmedzene potešenie, ktoré umelecké diela ponúkajú. Napriek týmto zjavným výhodám sa však (aspoň pre mňa) nič nevyrovná pocitu, keď človek vstúpi do Rijksmusea, Musée d'Orsay, Ermitáže, Albertiny alebo úchvatnej Sixtínskej kaplnky a môže obdivovať, oňuchávať a donekonečna skúmať na vlastné oči. Pre väčšinu obdivovateľov umenia preto zostávajú umelecké internetové galérie len akýmsi doplnkom, kde si po (alebo pred) návšteve galérie môžu dodatočne pozrieť (alebo stiahnuť ako wallpaper) ten svoj obraz. Tu však niekedy nastáva problém, ktorý súvisí s neskutočným množstvom obrazov nachádzajúcich sa v internetových galériách, a tým je vyhľadávanie. Väčšina týchto galérií neponúka žiadne, alebo len textové vyhľadávanie podľa názvu obrazu. Čo teda má človek urobiť v prípade, že bol na úžasnej výstave, odfoťil si nádherný obraz (a okamžite zabudol jeho názov) a chce si ho vyhľadať v kvalitnejšej podobe. Keďže sama som sa v takej situácii ocitla už milión krát, rozhodla som sa formulovať vo svojej diplomovej práci metódy, ktoré by umožnili vyhľadávanie názvu obrazu podľa fotografie a ich neskoršiu implementáciu do internetových galérií. Cieľom tejto diplomovej práce je teda navrhnúť a popísať metódu na riešenie problému klasifikácie obrazu na fotografii.

Kapitola 1

Úvod do problematiky

Téma tejto diplomovej práce zasahuje do dvoch oblastí rozpoznávania obrazcov. V prvom rade je to vyhľadávanie obrazov (Image retrieval), ale vzhľadom na databázu s ktorou pracujem sa dotýka aj klasifikácie umeleckých obrazov (Classification of fine art paintings). Nasledujúca časť je venovaná vývojom týchto odborov a najvýznamnejším prácam v každom z nich.

1.1 Vyhľadávanie obrazov

V 90-tych rokoch sa vedecká komunita začala zaoberať technológiami vyhľadávania obrázkov v súkromných databázach a na internete [18]. Dôvodom bol najmä rastúci počet obrázkov na internete a teda aj záujem užívateľov o ich triedenie a vyhľadávanie. Táto oblasť je podľa dát nad ktorými pracuje diferenciovaná na 2 prístupy: vyhľadávanie pomocou textovej informácie a vyhľadávanie na základe vizuálneho obsahu.

1.1.1 Vyhľadávanie pomocou textovej informácie

Prístup spočíva v triedení obrázkov podľa krátkej textovej informácie pripojenej k obrázku, prípadne informácií, ktoré sa vyskytujú v jeho tesnej blízkosti na web stránke. Obrázky sú kvôli jednoduchosti vyhľadávania tiež označované kľúčovými slovami a zaradované do kategórií. Pozitívom tohoto prístupu je možnosť rýchleho

spracovania veľkého množstva textu, a teda rýchle a jednoduché vyhľadávanie. Negatívnym faktorom je skutočnosť, že množstvo obrázkov je nedostatočne alebo nesprávne označených a tiež tzv. Thesaurus problém [62], ktorý vzniká ak obrázok môže byť pomenovaný pomocou viacerých synonym. Zlepšenie tohoto prístupu bolo dosiahnuté technikami zahŕňajúcimi odozvu od používateľa (feedback). Najefektívnejšou (a najzábavnejšou) technikou je momentálne ESP game [56]. ESP game je hra, ktorej dlhodobým cieľom je označiť (všetky) obrázky na internete krátkou textovou informáciou. Hra spočíva v tom, že dvom online užívateľom sa postupne ukazujú neoznačené obrázky z databázy (ak hráte ESP game cez Google Image Labeler [23] označujete obrázky vyhľadané z celého internetu) a každý z nich má za úlohu popísať obrázok identickým slovom ako jeho spoluhráč. Ak obrázok obaja označia rovnako, hra ponúkne ďalší obrázok, až dokiaľ nevyprší časový limit. Za každú hru dostávajú hráči body a zaraďujú sa do tabuľky úspešnosti. Táto hra výborne skĺbila zábavu a zmysluplný cieľ a získala takú popularitu, že na stránke gwap.com [27] sú hráči, ktorí hrajú ESP game 40 hodín do týždňa.

1.1.2 Vyhľadávanie na základe vizuálneho obsahu

CBIR - Content Base Image Retrieval je akákoľvek technológia, ktorá pomáha organizovať digitálne archívy obrázkov na základe ich vizuálneho obsahu. Do kategórie CBIR teda zahrňame čokoľvek od funkcie počítajúcej podobnosť obrázkov až po robustný nástroj na označovanie obrázkov. CBIR je odvetvie, v ktorom kooperujú odborníci z rôznych odborov ako rozpoznávanie obrazcov, počítačové videnie, strojové učenie, vyhľadávanie informácií, z oblasti Web and data mining, informačnej teórie a štatistiky [18]. V oblasti CBIR sa pracuje najmä na riešení problémov súvisiacich s dvomi fenoménmi:

1. Senzorická priepasť (Sensory gap) je rozdiel medzi objektom v reálnej scéne a informáciou o objekte zo zaznamenatej scény.
2. Sémantická priepasť (Semantic gap) vyjadruje rozdiel medzi informáciou, ktorá sa dá automaticky extrahovať z danej scény a informáciou, ktorú môže užívateľ získať zo scény v nejakej konkrétnej situácii.

Dôležitejším je problém sémantickej priepasti, ktorý predstavuje rozdiel medzi obsahom na nízkej úrovni a vyššou úrovňou konceptu [49]. Problém je stále otvorený a je kľúčový aj v oblasti rozpoznávania obrázkov. Obsahom na nízkej úrovni sú mienené informácie extrahované z obrázkov pomocou tzv. príznakov. Príznak (feature) je nositeľom vizuálnej vlastnosti obrázku, buď pre malú skupinu pixelov alebo pre celý obrázok. Existujú dva druhy týchto príznakov [18]:

1. Lokálne príznaky (Local features) - reprezentujú detaily v obrázku. Pri extrakcii lokálnych príznakov je skupina príznakov vypočítaná pre každý pixel z obrázku s použitím jeho okolia. Tieto výpočty môžu byť časovo náročné, najmä pre veľké obrázky, na ich zjednodušenie môžeme rozdeliť obrázok na malé neprekrývajúce sa bloky a vypočítať príznaky pre každý z nich. Lokálne príznaky sú napr. body záujmu (interest points), významné body (salient points), príznaky oblasti (region-based features).
2. Globálne príznaky (Global features) - vyjadrujú súhrnnú charakteristiku obrázka. Výhodou globálnych príznakov je rýchla extrakcia príznakov a určovanie podobnosti obrázkov. Nevýhodou je fakt, že príznaky môžu byť príliš senzitívne na polohu a zlyhávajú pri identifikácii kľúčových vizuálnych charakteristík. Globálne príznaky sú napr. histogram farieb (color histogram), globálne popisy tvaru (global shape descriptors).

Najpoužívanejšie príznaky sú:

1. Farba - pri práci s farebnými príznakmi je dôležitý správny výber farebného priestoru, v ktorom sa bude pracovať. Farebné priestory sú triviálne rozdelené na hardvérovo orientované (RGB, CMYK), užívateľsky orientované (HSV, HSL, HSI) a vnemovo orientované (CIE Lab). Pri extrakcii príznakov sa najčastejšie pracuje vo vnemovo orientovaných, ktoré sa najviac zhodujú s ľudským vnímaním farieb, alebo v užívateľsky orientovaných priestoroch. Príkladom farebných príznakov je histogram, alebo correlogram.
2. Textúra - Cieľom textúrnych príznakov je zachytenie zrnitosti a opakujúcich sa vzorov na obrázku. Rôzne objekty od trávy cez kvety až k plyšovým medvedíkom

sa líšia svojou textúrou, čo je dôležité pri segmentácii obrázkov. Textúrne príznaky sa extrahujú napr. pomocou granulometrie alebo Gaborových filtrov.

3. Tvar je kľúčovým atribútom segmentovaných oblastí. Je popisovaný na základe oblasti, ktorú objekt pokrýva, alebo na základe jeho kontúry - hrán. Pri hľadaní hrán sa aj vo farebných obrázkoch dá 90% hrán zistiť zo šedotónového obrazu, používajú sa preto klasické metódy ako Houghova transformácia [29], alebo techniky hľadania hranice (Canny [10], Kirsch [33], Sobel [50]). Pri hľadaní oblasti sa používajú techniky ako Watershed (technika hľadania povodí) [54].
4. Príznaky založené na lokálnych invariantoch - ako napríklad body záujmu alebo rohové body sú príznaky nezávislé na škálovaní, afinných transformáciách a osvetlení. Jednou z metód na extrakciu tohoto druhu príznakov je transformácia príznakov nezávislá na škálovaní (Scalable Invariant Feature Transform alebo SIFT) [37].

V ďalšej časti práce, kde budú spomínané príznaky z pohľadu klasifikácie maliarskych plátien, budú rozlišované príznaky textúry a farby a okrem toho príznaky línií a svetla. Pri práci s vizuálnym obsahom v CBIR systémoch v prvom kroku prichádza k extrakcii informácie z obrázka pomocou spracovania obrazu (napr. segmentácia) alebo extrakcie príznakov (lokálnych, globálnych...). V druhom kroku potom prebieha spájanie, vytváranie matematického popisu obrázka - takzvanej signatúry a indexovanie extrahovaných informácií. V treťom kroku prichádza usporiadanie výsledkov a výstup z vyhľadávania. Kroky sú schematické a v priebehu vývoja CBIR systémov do jednotlivých krokov vstupuje strojové učenie, odozva od užívateľa, predspracovanie obrazu atď. Pre vytvorenie komplexného CBIR systému je dôležitá správna špecifikácia dát nad ktorými má systém pracovať, ako aj vstupu, ktorý systém od užívateľa dostane, preto v tejto časti uvádzam aj niektoré základné rozdelenia CBIR systémov.

Pre špecifikáciu systémov na vyhľadávanie obrázkov sú definované dva druhy domén:

1. Úzka doména (Narrow domain) má jasne definované vizuálne charakteristiky a malú variabilitu (napr. fotografie kvetov). Na úzkej doméne sa dá jednoduchšie formulovať CBIR.

2. Široká doména (Broad domain) má veľkú variabilitu a nemá definované spoločné vizuálne charakteristiky (napr. obrázky na internete).

Nad širokou doménou sú formulované 3 typy vyhľadávania:

1. Asociačné - hľadaný obrázok sa dosiahne iteratívnym doladovaním vyhľadávania (na začiatku nie je jasne definovaný hľadaný obrázok).
2. Cieľovo orientované - hľadanie špecifického obrázka
3. V kategóriách - hľadanie reprezentanta sémantickej triedy (kategórie).

Pre CBIR systémy sa tiež upresňuje, s akou skupinou dát budú pracovať, z čoho potom vyplývajú konkrétne úlohy a prekážky pre daný systém:

1. Osobné kolekcie - vyznačujú sa homogénnym obsahom, malým rozsahom a primárne sú určené pre potreby vlastníka.
2. Doménovo špecifické kolekcie - homogénne kolekcie, určené na špecifický zámer, kontrolovane prístupné skupine užívateľov.
3. Podnikové kolekcie - heterogénne kolekcie, prístupné zamestnancom cez firemný intranet.
4. Archívy - kolekcie veľkého rozsahu, semi-štruktúrované homogénne dáta, kontrolovane prístupné používateľom internetu.
5. Web - semi-štruktúrovaná heterogénna kolekcia obrovského rozsahu, prístupná všetkým užívateľom internetu pomocou vyhľadávačov ako Google alebo Yahoo!.

Pre lepšiu interakciu medzi CBIR systémom a užívateľom je dôležitá variabilita užívateľských vstupov pre vyhľadávanie. Vstupy môžu byť rôznych typov:

1. Kľúčové slová
2. Neobmedzený text
3. Obrázok

4. Grafika – ručne kreslený, alebo počítačom generovaný obrázok
5. Kombinácia - kombinácia predchádzajúcich typov

Rôzne typy užívateľského vstupu do CBIR systému vyžadujú rôzne spôsoby spracovania:

1. Textový
2. Obsahový
3. Zložený
4. Interaktívny - Jednoduchý
5. Interaktívny - Zložený

Vývoj technológií CBIR systémov sa delí na dve etapy:

Prvá etapa (1995-2000) Komplexné zhrnutie tohoto obdobia sa nachádza v Smeulders et al. [49]. V tejto etape sa prínosmi pre fázu extrakcie príznakov stali: použitie farebných histogramov viacfarebných objektov na indexovanie obrázkov Swain a Ballard 1990 [51], využitie spekulárneho odrazu a tvaru na zlepšenie konštantnosti farebného vnemu (color constancy) Finlayson 1996 [21], využitie Gaborových filtrov na extrakciu lokálneho tvaru pre zrýchlenie porovnávania a vyhľadávania Manjunath a Ma 1996 [38]. Lokálne príznaky nezávislé od pozorovateľa a uzavretia objektu boli využité v Schmid a Mohr 1997 [47]. V oblasti segmentácie bolo významným prínosom napr. využitie delenia spektrálnym grafom Shi a Malik 2000 [48] a nakoniec metóda na automatický výber príznakov bola vyvinutá v Swets a Weng 2007 [52]. V Smeulders et al. [49] boli publikácie pojednávajúce o indexovaní a párovaní (matching) extrahovaných príznakov rozdelené do 6 skupín: Párovanie podľa príznakov (napr. technika Histogram incremental, ktorá zlučuje histogramy obrázku a modelu Swain a Ballard 1990 [51]), Párovanie podľa siluet objektu (napr. elastické priradovanie obrázkov vo vyhľadávaní podľa užívateľovho nákresu tvaru objektu Del Bimbo

a Pala 1997 [19]), Štruktúrované párovanie príznakov (napr. štruktúrované priradovanie príznakov na základe diskkrétnej relaxácie Wilson a Hancock 1997 [58]), Párovanie významných príznakov (salient) (napr. geometrické hashovanie Wolfson a Rigoutsos 1997 [59]), Párovanie na sémantickej úrovni (napr. využitie stupňovaných (graded) množín na sémantické zlučovanie výsledkov vyhľadávania Fagin 1997 [20]), Prístupy zlučovania podľa podobnosti, založené na učení (learning-based approaches for similarity matching) - vyhľadávanie obrázkov ako transdukčný problém učenia Wu et al. 2000 [60].

Druhá etapa (2000-2010) Hnacím motorom tejto etapy bol fakt, že vďaka expanzii systémov ako Picasa od spoločnosti Google, sociálnym sieťam ako Facebook, MySpace alebo Twitter, a lacnému a prístupnému webhostingu sa z bežných konzumentov obrázkov na internete stali ich producenti. Zvyšujúci sa počet užívateľov aktívne tvoriacich vizuálny obsah dramaticky zvýšil počet obrázkov na internete, čo ešte viac motivovalo k ďalšiemu rozvoju CBIR systémov. Datta et al. 2008 [18] vyjadril cieľ tejto etapy: „Navrhnutie všemocného nástroja na vyhľadávanie obrázkov schopného vykonať všetky skupiny užívateľských požiadaviek a chápaného a predstavujúceho interakciu medzi užívateľom a systémom na vyhľadávanie obrázkov z pohľadu užívateľa a systému.“ V porovnaní s predchádzajúcou etapou sa po roku 2000 objavuje veľká diverzita obrazových signatúr. Toto bolo spôsobené odvodením nových príznakov (napr. tvaru) a pokrokom v konštrukcii signatúr podľa týchto príznakov. V oblasti príznakov sa prešlo od globálnych (ako napr. farebný histogram) k lokálnym príznakom, (napr. charakteristické body, príznaky oblasti alebo robustná charakteristika lokálneho tvaru). Signatúry sa dajú z obrázku získať nielen statickou extrakciou, ale aj pomocou metód strojového učenia, buď na základe učenia sa na obrázkoch, alebo podľa užívateľskej odozvy.

Podľa ich adaptivity sa signatúry delia na: statické, adaptované na užívateľa, a adaptované na obrázok. V skupine signatúr adaptovaných na obrázok bol vo Wang et al. 2001 [57] zavedený pojem sémantickej senzitivity. Princípom tohto mechanizmu je rozdelenie obrázkov do malého množstva predpripravených

tried a podľa príslušnej triedy, potom výber signatúry vhodnej na popis obrázka. Triedami sú napríklad graf vs. obrázok, s textúrou vs. bez textúry. Z pohľadu matematickej reprezentácie môžu byť signatúry vyjadrené ako:

1. jediný vektor
2. množina vektorov
3. zoznamy (codebooks)
4. funkcie hustoty a pravdepodobnosti (probability density function).

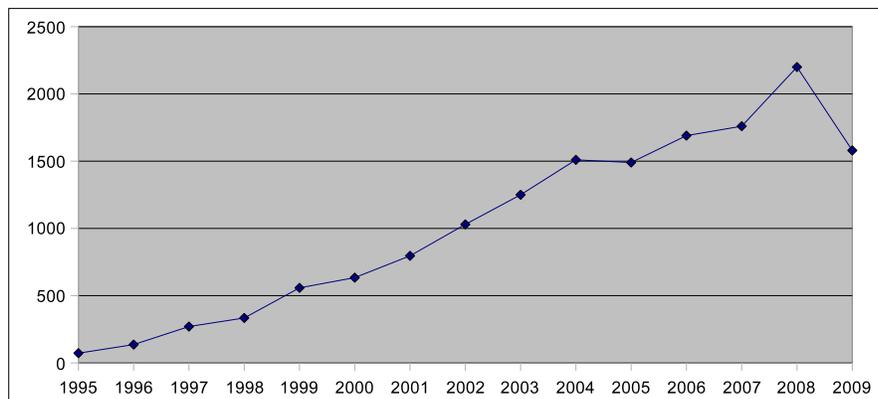
Zoznamy (codebooks) a funkcie hustoty a pravdepodobnosti (probability density function) slúžia ako prehľad vektorov lokálnych príznakov, codebooks sú generované vektorovou kvantizáciou, s kódovými slovami sa pracuje symbolicky metódami textového vyhľadávania. Keď je určená konkrétna matematická reprezentácia signatúr dvoch obrázkov, je dôležité vybrať vhodnú metódu ako počítať ich vzdialenosť. Pri príznakoch reprezentovaných jediným vektorom je riešením Euklidovská alebo geodetická vzdialenosť.

Pre množinu vektorov $[(z_1, p_1), (z_2, p_2), \dots, (z_n, p_n)]$ kde \mathbf{z} je vektor príznaku a \mathbf{p} je jemu priradená váha vyzerá triviálne riešenie nasledovne. Sú dané dve signatúry $I_m = (z_1^{(m)}, p_1^{(m)}), (z_2^{(m)}, p_2^{(m)}), \dots, (z_n^{(m)}, p_n^{(m)})$, $m = 1, 2$ vypočítajú sa vzdialenosti medzi $z_i^{(1)}$ a $z_j^{(2)}$ a spoja vzdialenosti medzi týmito vektormi do vzdialenosti množín vektorov. Ďalším prístupom je [12] priradenie váhy s_{ij} pre každú dvojicu $z_i^{(1)}, z_j^{(2)}$, $1 \leq i \leq n_1, 1 \leq j \leq n_2$ pričom váha s_{ij} vyjadruje dôležitosť spojenia $z_i^{(1)}$ a $z_j^{(2)}$. Na váhy sa vzťahujú obmedzenia, napríklad $\sum_i s_{i,j} = p_j^{(2)}, \sum_j s_{i,j} = p_i^{(1)}$. Po určení váh je vzdialenosť medzi obrázkami súčtom vzdialeností jednotlivých vektorových párov: $D(I_1, I_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} s_{i,j} d(z_i^{(1)}, z_j^{(2)})$. Vektor $d(\cdot, \cdot)$ je definovaný v závislosti od konkrétnej aplikácie. Na určenie vzdialenosti medzi množinami vektorov existujú aj ďalšie metódy ako Hausdorffova, Mallowsova vzdialenosť, alebo tzv. EMD (Earth Mover's Distance)

Vývojom prešli aj metódy zadávania dopytov (queries). Jednou z nových metód bolo vyhľadávanie farebných obrázkov pomocou náčrtku v Chalechale et al. 2005 [14]. Ďalšími prístupmi je vyhľadávanie pomocou 3D modelov v Assfalg et al. 2002 [5], alebo vyhľadávanie pomocou podobnosti s iným obrázkom, kombináciou podobnosti vo viacerých príznakoch ako farba, tvar alebo textúra. Metóda vyhľadávania podľa podobnosti s iným obrázkom bola vylepšená v Carson et al. 2002 [12] keď bola do metódy pridaná možnosť užívateľa špecifikovať segment obrázku (blob), ktorý je pre vyhľadávanie kľúčový. Užívateľ dostane svoj vzorový obrázok po segmentácii a má možnosť vyznačiť jeden alebo viac segmentov. Ak teda užívateľ chce vyhľadať obrázky červených lodičiek, tak na obrázku skrinky na topánky, ktorý ponúkol ako vzorový, môže vyznačiť segment obsahujúci červené lodičky ako cieľ vyhľadávania.

1.1.3 Budúci vývoj

CBIR je napriek 15 rokov trvajúcej záujmu stále otvorená téma, čo potvrdzuje aj neklesajúci počet vedeckých článkov a publikácií, každoročne vydávaných najznámejšími vydavateľstvami ako IEEE a Springer. Na ilustráciu je uvedený pokus navrhnutý v Datta et al. (2005) [18]. Pokus predstavuje vizualizáciu počtu publikovaných článkov nájdených internetovým vyhľadávačom Google Scholar [25] za roky 1995-2009 pre kľúčové slová „content based image retrieval“. Na obrázku 1.1 je zobrazený počet článkov, ktorý od roku 1995 rovnomerne rástol a v posledných rokoch sa drží na relatívne vysokej hodnote. Ďalším dôkazom dôležitosti CBIR techník je fakt, že Yushi Jing a Shumeet Baluja z Google, Inc. v roku 2008 vo svojom článku [32] predstavili novú metódu pre vyhľadávanie obrázkov cez Google Image Search, ktorá je spojením textového vyhľadávania a vyhľadávania na základe vizuálneho obsahu. Na extrakciu príznakov z obrázku je v tejto metóde použitý SIFT [37] s detektorom bodov záujmu pomocou rozdielu Gaussiánov (rovnako ako v tejto diplomovej práci). Po extrakcii príznakov je z obrázkov databázy vytvorený ohodnotený graf podobnosti, podľa ktorého sa vyhodnotí dôležitosť (rank) jednotlivých vyhľadaných obrázkov.



Obr. 1.1: Vývoj počtu článkov vyhľadaných pre kľúčové slová „content based image retrieval“ za roky 1995-2009. Vytvorené dňa 30.3.2010.

Datta et al. [18] v svojej práci naznačuje aj budúcnosť CBIR systémov. Novými problémami a aplikáciami, ktorými by sa v budúcnosti mala vedecká komunita zaoberať sú napríklad:

Automatická notácia - popis obrázkov - keby sme vedeli z obrázkov extrahovať sémanticky správnu informáciu a sformulovať ju do kľúčových slov, mohli by sme pre množstvo aplikácií využívať textové vyhľadávanie oproti CBIR.

Automatizované ilustrovanie textov - takýto systém by musel správne extrahovať kľúčové slová z textu, a tiež vybrať z databázy obrázkov vhodne reprezentujúci dané kľúčové slovo.

Estetika fotografií - možnosť zoraďovania obrázkov podľa ich estetickej hodnoty, je zaujímavá pre väčšinu tvorcov CBIR systémov. Jednou z možností ako vytvoriť model pre estetiku fotografií je sledovanie hodnotenia na portáloch, kde ľudia zdieľajú a hodnotia svoje umelecké fotografie (napr. ephoto.sk [43]).

Informačná bezpečnosť – CBIR môže byť pre oblasť informačnej bezpečnosti potrebná v oblasti autorských práv pri odhaľovaní plagiátorstva, napr. pri vyhľadávaní duplikátov lôg firiem.

1.2 Klasifikácia malieb

30 000 rokov pred naším letopočtom sa datuje nález prvých jaskynných malieb v jaskyni Chauvet vo Francúzsku [16]. Od tých čias až po súčasnosť vzniklo množstvo malieb, obrazov a iných umeleckých predmetov, ktoré sú teraz súčasťou nášho kultúrneho dedičstva. Od roku 1761, keď bolo pre verejnosť otvorené prvé múzeum (moderného typu) v Bazileji [2] sa v zbierkach múzeí a galérií po celom svete zhromažďuje nepreberné množstvo (len v depozitároch slovenských múzeí sa v súčasnosti nachádza 9 miliónov [55] artefaktov) umeleckých predmetov.

V osemdesiatych rokoch dvadsiateho storočia [13] sa vedecká komunita z oblasti počítačovej grafiky a videnia začala týmito umeleckými predmetmi zaoberať a začala pracovať na zachovávaní kultúrneho dedičstva (preservation of cultural heritage). V rámci zachovávaní a spracovávaní kultúrneho dedičstva sa pracuje na mnohých úrovniach, počínajúc vytváraním 3D modelov zničených sôch (rekonštrukcia malého Budhu z oblasti Bamiyan v Afganistane, ktorá bola zničená v roku 2001 Talibanom [1]), vytváranie 3D rekonštrukcií historických podôb miest (Rome Reborn projekt, vytváranie 3D modelu Antického Ríma [26]), Digitalizácia historických dokumentov, Skenovanie umeleckých predmetov (napr. Skenovanie Michelangelovho Davida týmom zo Stanfordskej univerzity [35]) a v neposlednom rade rozpoznávanie, klasifikácia a uchovávanie maliarskych plátien [17, 31, 46, 36].

1.2.1 Prehľad prác

O maliarske plátna, ako o ďalšiu oblasť aplikácie spracovania obrazu sa začali vedci zaujímať začiatkom 80-tych rokov. Jedným z prvých počínov v tejto oblasti bolo využitie kvantitatívnej analýzy a techník spracovania obrazu na virtuálne reštaurovanie (doplnenie chýbajúcich častí) obrazu Giottovej (Giotto di Bondone 1267 – 1337 Florencia) Madony v Casazza et al. [13] a práca Johna F. Asmusa o digitálnom archivovaní maliarskych plátien [4].

Po roku 1990 sa na medzinárodných konferenciách objavilo množstvo prác z tejto oblasti. V nasledujúcej časti sú spomenuté niektoré z najvýznamnejších. Rozpozná-

vaním umeleckých diel (maliarskych plátien, fresiek a nástenných malieb) od skutočných scén sa zaoberá komunita okolo A. Leykina a F. Cutzu z roku 2005. V svojich prácach zistili, že najvhodnejšou metódou na rozdelenie do dvoch spomínaných množín najlepšie slúži kombinácia 3 druhov príznakov: hrany, farba a textúra, ktorá pri použití neurónových sietí dosahuje až 92% úspešnosť. Pri detekcii hrán je rozdiel medzi reálnou scénou a umeleckým dielom v type hrán, ktoré obsahujú. Reálna scéna obsahuje veľké množstvo hrán vytvorených rôznou intenzitou odrazeného svetla od susedných povrchov, zatiaľ čo umelecké dielo má prevládajúci počet tzv. farebných hrán. Rozdiel vo farebných príznakoch je v počte jedinečných farieb, vo veľkosti farebnej zmeny medzi okolitými pixelami a v počte vysoko saturovaných pixelov, pričom každá z týchto hodnôt je výrazne vyššia pri umeleckých dielach, než pri fotografiách reálnych scén. Textúrne príznaky sa extrahovali s využitím Gaborovych filtrov (16 masiek, pre 4 smery a 4 orientácie). Autori zistili, že reálne scény mali väčšie hodnoty pre horizontálne a vertikálne smery a umelecké diela pre diagonálne smery.

Z roku 2005 pochádza práca Shuqiang Jiang et al. zaoberajúca sa detekciou a kategorizáciou tradičných čínskych malieb [31]. Tradičné čínske maľby sa od západných líšia štýlom, ale aj použitými materiálmi. Maľuje sa na papier alebo hodváb s použitím atramentu a/alebo minerálnych pigmentov. Tradičné maľby sa rozdeľujú do dvoch skupín Xieyi (jednoduché náčrtky s použitím malého množstva atramentu) a Gongbi (prepracované vizuálne komplexné maľby). Autori sa v práci zamerali na odlíšenie tradičných čínskych malieb od ostatných malieb a potom ich rozdelenie do príslušných skupín. Jeden z príznakov, ktoré použili na klasifikáciu, bol vektor farebnej koherencie, ktorý obsahuje pre každú farbu počet koherentných pixelov, teda pixelov, ktoré boli súčasťou nejakej spojitkej oblasti tej konkrétnej farby (spojité oblasti boli hľadané pomocou 8-susednosti). Ďalším využitým príznakom bola autokorelácia, ktorá vyjadruje „hrubosť“ textúry pomocou lineárnej vzdialenosti textúrnych primitívov. Posledným použitým príznakom v tejto práci bol 1-rozmerný histogram dĺžky hrán, ktorý reprezentuje počty hrán jednotlivých dĺžok nájdených pomocou Sobelovho filtra.

Rozpoznávaníu ťahov ťetca sa venovali Sablatnik et al. [46] a Melzer et al. [41] v PRIP skupine (rozpoznávanie obrazcov a spracovanie obrazu). Melzer et al. extrahovali ťahy ťetca pomocou techník MBO - Model Based Stroke Operator (operátor založený na modeli ťahov ťetca) a MLFN - Multi Layered Feed Forward (Viacvrstvá dopredná neurónová sieť). MBO pomocou modelu ťahov ťetca, ktorý bol vytvorený v spolupráci s umeleckým historikom, detekuje lokálne ťahy z obrázka, ktoré potom spája do jediného modelu ťahov ťetca. MLFN je sieť s jednou skrytou vrstvou a 225 bunkami na vstupnej vrstve, ktorá bola trénovaná na 40 000 vzorkách ťahov ťetca. Technika MBO dosiahla úspešnosť od 72 do 77 percent správne klasifikovaných autorov a MLFN od 80 do 84 percent pri testovaní na troch rôznych množinách.

1.2.2 Prístup počítačovej grafiky a videnia k umeleckým štýlom

Práce zaoberajúce sa maliarskymi plátnami z pohľadu počítačovej grafiky a videnia sa dajú rozdeliť do troch prúdov: Analýza štýlu, Klasifikácia štýlu a autora a Vyhľadávanie umeleckých obrázkov. Analýza štýlu sa zaoberá analýzou príznakov, zachovaním umeleckých plátien a verifikáciou hypotéz umeleckých historikov. Klasifikácia štýlu pracuje s extrakciou lokálnych a globálnych príznakov, segmentáciou a extrakciou ťahov ťetca. Tieto metódy sa vykonávajú v 4 krokoch: 1. Predspracovanie - filtrovanie obrázka pomocou mediánového, priemerovacieho alebo Gausiánového filtra, korekcia veľkosti obrázka. 2. Extrakcia príznakov. 3. Porovnávanie príznakov. 4. Klasifikácia.

Príznyky sú dôležité pre vyjadrenie vizuálnej informácie každého obrázka a špeciálne maliarskych plátien. Ako bolo spomenuté v časti 1.1.2, pre klasifikáciu štýlu a vyjadrenie informácií o umeleckom obraze sú dôležité štyri druhy príznakov.

Línie: Príznyky línie zahŕňajú rozpoznávanie čiar, tvarov, perspektívy a hĺbky. Na detekciu čiar (hrán) v obrázku sa využívajú rôzne metódy, ako Sobelov, Cannyho, Robertsov, Kirshov detektor. Po identifikácii čiar a rohov v obrázku, je možné určiť orientáciu a dĺžku čiar, ale aj sofistikovanejšie informácie, ako napr. tvar, úbežníky alebo hĺbku. Tieto príznaky boli využité napr. pri oddeľovaní prasklín od pôvodnej maľby na infračervených skenoch [28] (praskliny na

obrazoch sú spôsobené starnutím a dajú sa extrahovať na základe apriornej informácie o ich hrúbke a pravdepodobných miestach výskytu) a takisto vo vyššie spomínanej práci zaoberajúcej sa odlišovaním reálnych scén od maliarskych plátien [17]. Pri klasifikácii štýlov sa využívajú informácie z analýzy hrán, triviálne - počet hrán nájdených Sobelovým detektorom, gradientné mapy (koeficienty v mape reprezentujú súhrn intenzity hrán v obrázku). Výsledky výskumu v oblasti klasifikácie štýlov na základe analýzy línií ukázali, že príznaky nesúce vyššiu sémantickú informáciu klasifikujú s rovnakou presnosťou ako príznaky s nízkou sémantickou relevanciou.

Svetlo: Využitie osvetlenia v obraze charakterizuje mnohé umelecké štýly, medzi nimi napríklad barok. Príznaky vyjadrujúce aspekty svetla sú teda efektívne pri klasifikácii umeleckých štýlov. Využívajú sa príznaky ako priemerná hodnota intenzity, štandardná odchýlka priemernej intenzity, zošíkmenie intenzity (tretí centrálny moment) - tento príznak je invariantný zmenám v jase, početnosť tmavých pixelov v obrázku, alebo priemerná odchýlka priemernej intenzity, ktorá je nezávislá od ostatných príznakov z tejto kategórie a je odolná voči šumu v obrázku.

Textúra: Vedci považujú textúru za najefektívnejší typ príznaku na klasifikáciu umeleckého štýlu, pretože najvhodnejšie vystihuje ťahy štetca. Dôležitosť ťahov štetca pre klasifikáciu štýlu je podľa Carr a Leonard [11] v tom, že "je to v podstate priamy odraz tlaku a pohybu autorovej ruky naprieč povrchom maľby. Ťahy štetca sú jedným z najintímnejších spojení, ktoré máme ako pozorovatelia s autorovou myslou a prácou." Problém s textúrnymi príznakmi spočíva v absencii presnej definície pojmu textúra. Ďalším problémom je, že textúra ako ju vnímame v rámci počítačového videnia, nie je totožná s pojmom z oblasti dejín umenia. Umeleckí historici rozlišujú na rozdiel od počítačových grafikov dva druhy textúry, skutočnú a vnemovú. Skutočná textúra obrazu je 3D maľba navrstvená na plátne, zatiaľ čo vnemová textúra je 2D ilúzia textúry vytvorená umelcovou zručnosťou. Napriek týmto nezrovnalostiam sa textúrne príznaky preukázali ako veľmi efektívne pri klasifikácii v tejto doméne. Medzi

textúrne príznaky patria napríklad Gaborove filtre [30], rýchla Furierova transformácia FFT, diskretná kosínusová transformácia DCT, matica opakovaných výskytov, alebo waveletová transformácia [50]. Waveletová transformácia má schopnosť sumarizovať obrazy v rozličných rozlíšeniach a orientáciách a produkuje 4 skupiny koeficientov. Nízko-frekvenčné, horizontálne, vertikálne a diagonálne. Táto transformácia bola využitá pri určení „počtu rúk“, teda počtu pomocníkov, ktorí maľovali časti obrazu Madona s dieťaťom, ktorého autorstvo je pripísané Perugiovi. Jedno z ďalších využití textúrnych príznakov ponúka práca Richarda Taylora [53], ktorá analyzovala maľby Jacksona Pollocka z hľadiska textúry, a ukázala že jeho maľby majú fraktálny základ (Pollockov štýl označil Taylor ako fraktálny expresionizmus).

Farba: Farebné príznaky poskytujú asi najdôležitejšiu informáciu o štýle maľby. Napríklad dostupnosť farebných pigmentov závisí od historického obdobia, v ktorom umelec žil. Využitie farby je takisto význačnou charakteristikou autorovho individuálneho štýlu. Jedným z príznakov je rozsah farebnej palety, ktorý bol využitý aj v Cutzu et al. [17]. Tento sa dá získať spočítaním každej jedinečnej farby z obázku (jedinečnej RGB trojice), alebo namapovaním jedinečných farieb na zmyslupnú paletu (napr. pomocou MVQ - kvantizácie podľa minimálneho rozdielu) a spočítaním výskytu týchto farebných zhlukov v obrázku. Ďalším dôležitým príznakom z tejto skupiny je histogram, ktorý podáva informáciu o frekvencii výskytu farieb. Najdôležitejšími typmi histogramu v tomto druhu aplikácií sú dynamický histogram, ktorý zobrazuje výskyt farieb určených pomocou MVQ v obrázku, statický histogram a statický histogram aplikovaný na farebnú mapu. Histogram bol použitý napr. v práci študentov [15] Stanfordskej univerzity, ktorý rozpoznávali obrazy v Cantorovom centre umenia z fotografií. Jedným z najambicióznejších počinov v oblasti farebných príznakov bola digitálna analýza Van Goghových olejových malieb, ktorá mala verifikovať hypotézu, že Van Gogh používal komplementárne farby na zdôraznenie kontúr [7]. Obrazy boli spracovávané v priestore CIE Lab a hypotéza bola verifikovaná. Extrakcia farebných príznakov sa využíva aj v oblasti reštaurátorstva na zistenie pôvodnej distribúcie farieb z vyblednutého obrazu [44]. Táto infor-

mácia potom napomôže reštaurátorom správne zrekonštruovať maľbu bez jej poškodzovania pri manuálnom zisťovaní pôvodnej farby.

V dizertačnej práci Lombardi [36] boli popísané niektoré trendy v oblasti klasifikácie štýlu. Niektoré z nich sú: klasifikácia zameraná len na určité typy malieb (napr. portréty [41]), zapojenie len určitého typu príznakov (fraktály [53] a ťahy štetca [41]), priamy prístup - vytváranie matematických modelov na popis autora či hnutia, alebo prístupy zamerané na autentifikáciu a identifikáciu autora a nie klasifikáciu štýlu.

1.3 Prehľad analogických prác

Pri vytváraní prehľadu problematiky bola objavená len jedna práca zaoberajúca sa totožným problémom a teda segmentáciou a klasifikáciou maliarskych plátien z fotografií. Touto prácou bola seminárna práca študentov zo Stanfordu pod názvom A day at the museum [15]. Na rozdiel od našej práce, Chang C. et al. vytvorili svoju databázu fotografií turistov len zo svojich vlastných fotografií odfotoграфovaných v jednej galérii jedným fotoaparátom so známymi parametrami a pod konštantným osvetlením, čo umožnilo použiť študentom na segmentáciu jednoduché prahovanie a na klasifikáciu príznak histogramu. V našej práci sa používa databáza získaná od rôznych autorov a obsahuje fotografie z rôznych galérií odfotoграфované nešpecifikovanými fotoaparátmi.

Kapitola 2

Špecifikácia

V tejto kapitole je podrobne popísaný cieľ diplomovej práce, požiadavky na navrhnuté riešenie ako aj algoritmy použité pri riešení.

2.1 Cieľ

Cieľom diplomovej práce je navrhnúť a popísať metódu na riešenie problému klasifikácie obrazu na fotografii a implementovať ju v prostredí MATLAB. V rámci problému klasifikácie fotografie práca rieši dva samostatné problémy: segmentácia obrazu z fotografie a klasifikácia obrazu (porovnanie s vytvorenou databázou malieb). Podobná aplikácia implementovaná vo vhodnom jazyku by mohla byť vhodným doplnkom funkčnosti webovských galérií, ako napr. Olga's gallery [39].

2.2 Požiadavky

Navrhnutá metóda by mala byť **rýchla** pri segmentácii a klasifikácii, **nemala by nadsegmentovať obraz**, mala by byť **robustná** pri klasifikácii a mala by umožňovať použitie viacerých algoritmov ako pri segmentácii tak aj pri klasifikácii. **Implementácia** do ďalších jazykov by mala byť priamočiara a **jednoduchá**.

MATLAB aplikácia vytvorená v rámci tejto diplomovej práce musí spĺňať niektoré

základné požiadavky. Aplikácia musí na vstupe prijímať čiernobiele alebo farebné fotografie typu JPG. Vstupom do aplikácie sú bežné fotografie obrazov vyfotografovaných turistami v galériách a múzeách. Správna segmentácia by mala fungovať na všetkých fotografiách, ktoré obsahujú aspoň polovicu plochy rámu obrazu, pričom obraz aj rám môžu byť čiastočne prekryté postavami turistov. Obraz na fotografii by mal byť správne klasifikovaný, ak je uložený v databáze Originálov. V tom prípade funkcia vráti názov obrazu. V prípade, že sa obraz v databáze nenachádza, aplikácia oznámi, že obraz nemá obsahnutý v databáze a nemôže ho klasifikovať. Ďalšou z požiadaviek je jednoduchosť rozširovania databázy Originálov.

2.3 Databáza obrázkov

Na overenie funkčnosti metód vytvorených v tejto diplomovej práci bolo potrebné zhromaždiť dve množiny obrázkov: Originály obrazov a Fotografie turistov. V týchto dvoch množinách boli použité fotografie obrazov od Rembrandta Harmenszoon van Rijna, ktorý je považovaný za najvýznamnejšieho holandského maliara¹. Jeho zbierky sú súčasťou stálych expozícií múzeí ako Louvre v Paríži, Ermitáž v Petrohrade, Rijksmuseum v Amsterdame, Metropolitné múzeum umenia v New Yorku alebo Prado v Madride.

2.3.1 Originály obrazov

Množina originálov obsahuje 15 fotografií Rembrandtových obrazov z internetovej galérie Olga's gallery [39]. Tieto fotografie obsahujú len samotné plátno bez rámu a sú zosnímané pod rovnomerným osvetlením v rozlíšení 600 x 600 dpi.

2.3.2 Fotografie turistov

Množina fotografií turistov obsahuje 100 fotografií odfotografovaných v múzeách alebo galériách bežnými návštevníkmi nešpecifikovanými digitálnymi fotoaparátmi. Obsahuje fotografie z autorkinej zbierky z múzeí Ermitáž a Rijksmuseum, fotografie,

¹Rembrandt Harmenszoon van Rijn (* 15. júl 1606, Leiden - † 4. október 1669, Amsterdam)

	Názov diela	Počet fotografií turistov
1	Danae	12
2	Nočná hliadka	9
3	Artemis	4
4	Židovská nevesta	4
5	Jeremiáš žiali nad zničením Jeruzalema	7
6	Predstavenstvo súkenníckeho cechu	5
7	Hodina anatómie doktora Tulupa	3
8	Návrat márnokratného syna	9
9	Portrét Marie Trip	3
10	Portrét Saskie	1
11	Flora	3
12	Kúpeľ	3
13	Svätá rodina	3
14	Autoportrét ako Sv. Pavol	8
15	Portrét muža	2
16	Neklasifikovaných	24

Tabuľka 2.1: Tabuľka počtu fotografií turistov v použitej databáze.

ktoré boli získané vďaka iniciatíve na stránke² pochádzajúce z múzeí Louvre a Metropolitaného múzea umenia a z portálu travel.webshots [3], odkiaľ boli získané fotografie z rôznych múzeí a galérií po celom svete. Tieto fotografie sú v rôznom rozlíšení, majú rozličné veľkosti a sú zachytené pod variabilným osvetlením závislým od danej galérie a umiestnenia obrazu. Na fotografiách je obraz často prekrytý inými návštevníkmi galérie, niekedy je zachytená iba časť rámu a obraz je pravidelne fotografovaný pod uhlom iným ako 90 stupňov, čo spôsobuje deformáciu rovnobežnosti jednotlivých strán rámu. V tabuľke 2.1 je zoznam obrazov (a ich počet) obsiahnutých v databáze. Obrázok 2.1 obsahuje ukážky fotografií z jednotlivých databáz.

2.4 Softvér

V tejto diplomovej práci bol použitý softvér MATLAB s doplnením Image processing toolbox. MATLAB je prostredie a zároveň programovací jazyk štvrtej generácie. Skratka MATLAB znamená MATrix LABoratory (Maticové laboratórium), čo nazna-

²<http://members.chello.sk/halada-j/diplomovka.html>



Obr. 2.1: Príklad obrázkov z databázy Originálov obrazov (vľavo) a databázy Fotografií turistov (vpravo).

čuje, že matica je v tomto jazyku najzákladnejšou dátovou štruktúrou. V prostredí MATLABu je možné používať matematické funkcie pre lineárnu algebru, štatistiku, Fourierovu analýzu, filtrovanie, optimalizáciu, numerickú integráciu, funkcie na zobrazenie 3D a 2D dát, nástroje na vytvorenie užívateľského grafického rozhrania a množstvo rozšírení v podobe toolboxov. MATLAB na konci sedemdesiatych rokov vyvinul Cleve Moler s použitím programovacieho jazyka Fortran a jeho balíčkov LINPACK a EISPACK (na riešenie vlastných hodnôt matíc a lineárnych rovníc). Pôvodná verzia MATLABu obsahovala 80 funkcií a pre pridanie novej funkcie bolo potrebné modifikovať zdrojový kód vo Fortrane a skompilovať celý MATLAB. MATLAB sa odvtedy stále zdokonaľoval, v roku 1984 bola založená spoločnosť Mathworks a v roku 1991 bol MATLAB preprogramovaný z Fortranu do jazyka C, neskôr do C++ [42]. V súčasnosti je MATLAB využívaný najmä v oblasti lineárnej algebry, numerickej analýzy a tiež spracovania signálu a obrazu.

Ďalšou možnosťou bolo vytvorenie funkcie pomocou knižnice OpenCV. OpenCV je open source multiplatformová knižnica pre počítačové videnie, vyvinutá v jazyku C.

Alfa verzia tejto knižnice bola zverejnená na konferencii IEEE Conference on Computer Vision and Pattern Recognition v roku 2000. Knižnicu OpenCV je možné používať v jazyku C, C++ a Python. Aplikačné oblasti tejto knižnice sú rozpoznávanie tváří, gest, segmentácia, sledovanie pohybu, stereo videnie alebo pohyblivá robotika. OpenCV tiež zahŕňa knižnicu pre strojové učenie, ktorá obsahuje: rozhodovacie stromy, neurónové siete, naivné Bayesovské klasifikátory, algoritmus k-najbližších susedov a iné.

MATLAB bol zvolený najmä preto, že mnohé nové práce z konferencií v oblasti spracovania obrazu obsahujú implementáciu svojho algoritmu v MATLABe.

2.5 Funkcie

Funkčnosť MATLAB aplikácie implementujúcej navrhnutú metódu je obsiahnutá v 3 základných funkciách: `PreprocessImg`, `SegmentImg`, `ClassifyImg`.

PreprocessImg: Funkcia zabezpečujúca správne načítanie fotografie, jej konverziu na šedotónový obrázok (v prípade farebnej fotografie), upravenie jeho veľkosti na výšku 300px a ďalšie prípadné predspracovanie (ekvalizácia histogramu, zväčšenie rozlíšenia...).

SegmentImg: Na vstupe funkcie `SegmentImg` je načítaný predspracovaný obrázok a číslo/názov jedného z možných algoritmov a na výstupe je obrázok obsahujúci plátno vysegmentované pomocou zadaného algoritmu.

ClassifyImg: Cieľom funkcie je vytvoriť popis obrázka vytvoreného funkciou `SegmentImg` a vyhľadať v databáze popisov originálov ten najviac korešpondujúci. Podľa hodnoty korešpondencie potom klasifikuje vstupný obraz ako originál z databázy, alebo v databáze sa nenachádzajúci.

2.6 Algoritmy

Nasledujúca sekcia popisuje algoritmy použité v tejto diplomovej práci. Texty k jednotlivým algoritmom vychádzajú z prednášok doc. M. Ftáčnika [22] a z niektorých ďalších prác [45, 34, 37, 6].

2.6.1 Anisotropic diffusion

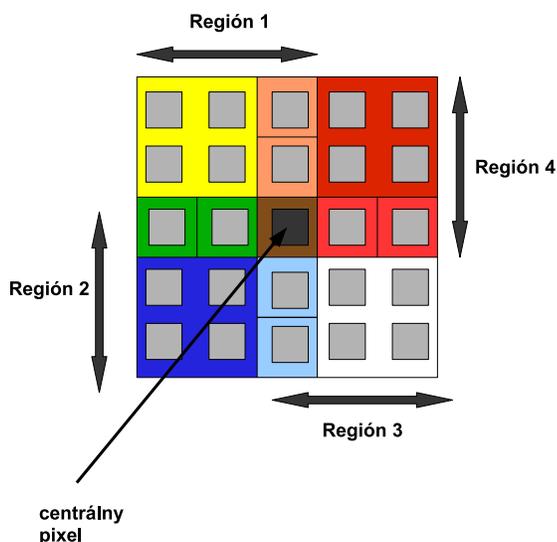
Anisotropic diffusion alebo Perona–Malik diffusion [45] je metóda filtrovania obrazu ktorá vyhladí obraz, ale zachováva hrany. Metóda využíva „scale-space“, teda vytvára rodinu po sebe idúcich obrázkov, kde nasledujúci obrázok vznikne z predchádzajúceho aplikovaním difúznej rovnice. Tento proces pokračuje až kým nedosiahne požadovanú hodnotu vyhladenia. Rovnica Anisotropic diffusion pre nejaký obrázok z rodiny (dolný index t označuje o ktorý z obrázkov ide, v prípade $t=0$ ide o pôvodný obrázok) je $I_t = \text{div}(c(x, y, t)\nabla I) = c(x, y, t)\Delta I + \nabla c \cdot \nabla I$, kde div je operátor divergencie, ∇ je gradient, Δ Laplacián a $c(x,y,t)$ je koeficient difúzie. Hlavnou výhodou metódy je, že vyhladzuje obrázok vo vnútri jednotlivých oblastí a nie medzi oblasťami.

2.6.2 Kuwahara filter

Kuwahara filter [34] patrí do skupiny filtrov zachovávajúcich hrany. Môže byť implementovaný pre rôzne veľkosti okna, v tejto práci bol použitý pre veľkosť štvorcového okna $J = K = 4L + 1$ pre $L = 1$. Okno je rozdelené do štyroch regiónov o veľkosti $\lfloor (J+1)/2 \rfloor \times \lfloor (K+1)/2 \rfloor$ (pozri obrázok 2.2). Pre každý región je vypočítaná priemerná intenzita (pri šedotónovom obrázku) a rozptyl. Výsledná intenzita centrálného obrazového bodu je potom rovná priemernej intenzite regiónu s najmenším rozptylom.

2.6.3 Segmentácia s využitím povodí (Watershed)

V tejto metóde predstavujú oblasti segmentovaného obrazu zberné nádrže a reprezentujú lokálne minimum šedej úrovne. Prvý prístup cez watershed segmentáciu spočíva

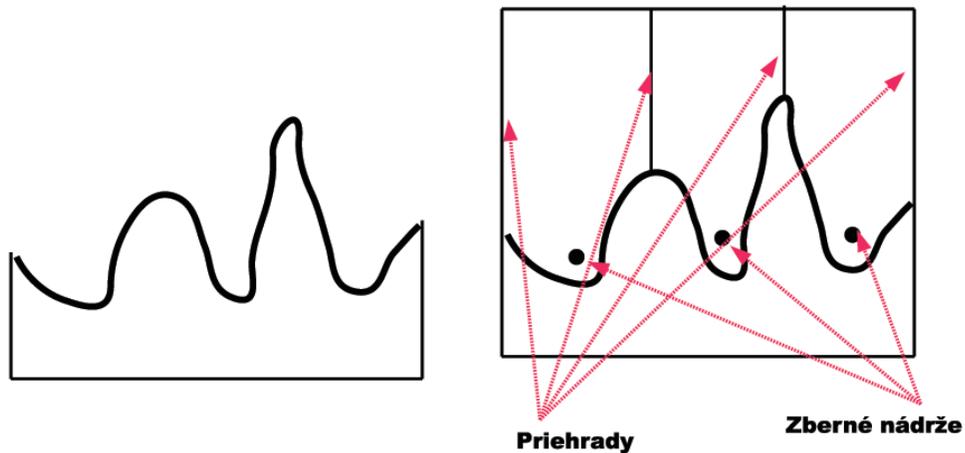


Obr. 2.2: Rozdelenie okna pri Kuwahara filtri.

v nájdení cesty vedúcej od daného pixla smerom dolu k lokálnemu minimu v zbernej nádrži (obrázok 2.3). Tu sa pixle usporiadajú podľa toho, v ktorom minime sa skončí ich cesta. Pri druhom prístupe každé minimum šedej úrovne reprezentuje jednu zbernú nádrž a stratégia spočíva v plnení týchto nádrží smerom nahor.

2.6.4 Houghova transformácia

Segmentačná technika použiteľná vtedy, keď je potreba detekovať objekty so známym tvarom hranice. Houghova transformácia môže detekovať rovné čiary aj krivky (hranice objektu), ak sú známe ich analytické vyjadrenia. Je robustná pri identifikácii zakrytých a zašumených objektov. Najjednoduchšia forma Houghovej transformácie detekuje rovné čiary. Každá priamka môže byť vyjadrená ako $y = mx + b$ a graficky zobrazená pomocou bodov (x,y) . Houghova transformácia reprezentuje priamku pomocou dvojice b,m , konkrétne pomocou r a Θ kde r je vzdialenosť najbližšieho bodu priamky od počiatku a Θ je uhol vektora od počiatku do najbližšieho bodu priamky.



(a) Profil šedých úrovní obrázku

(b) Prvý prístup k segmentácii s využitím povodí - lokálne minimá početnosti šedých úrovní (výšky) vytvoria zberné nádrže, v lokálnych maximách sa vytvoria priehrady

Obr. 2.3: Watershed segmentácia.

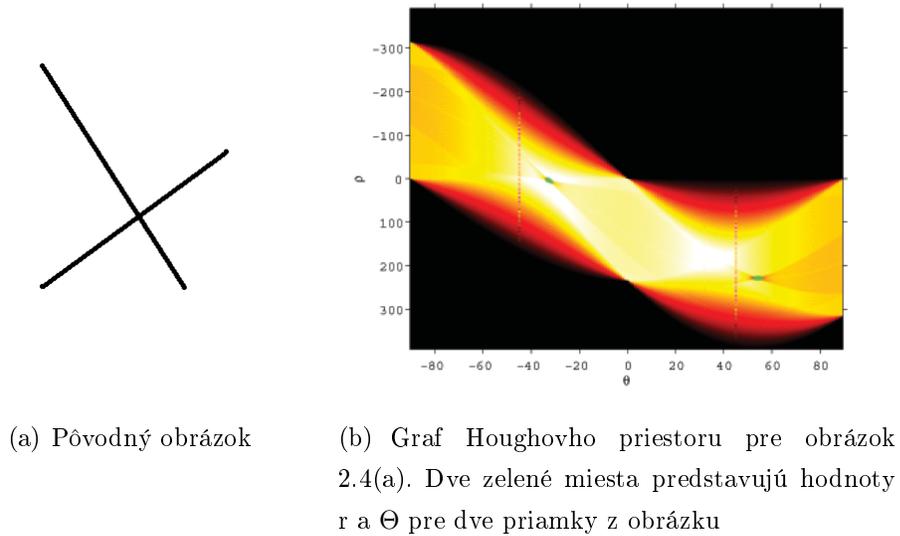
Parametrické vyjadrenie priamky pomocou r a Θ je:

$$y = \left(-\frac{\cos \Theta}{\sin \Theta} \right) x + \left(\frac{r}{\sin \Theta} \right)$$

Houghova transformácia pre každý bod obrázka pre všetky priamky ním prechádzajúce pod rôznym uhlom vypočíta r a Θ týchto priamok. V grafe závislosti r a Θ potom zobrazí krivky pre všetky body, pričom ich riadiace vrcholy sú hodnoty r a Θ pre jednotlivé priamky. Miesto, kde sa pretne viacero kriviek (väčšinou je to výrazné svetlé miesto v grafe Houghovho priestoru) určuje r a Θ priamky na obrázku. Obrázok 2.4 znázorňuje graf Houghovho priestoru vytvorený v MATLABe.

2.6.5 Sobelov operátor

Patrí medzi diferenčné gradientné operátory a teda ho možno vyjadriť pomocou konvolučnej masky. Vrátí odozvu tam, kde je hrana a nízku alebo nulovú odozvu má



Obr. 2.4: Houghova transformácia.

tam, kde hrana nie je. Príklad Sobelových konvolučných masiek:

$$h_1 = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix} \quad h_2 = \begin{vmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{vmatrix} \quad h_3 = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix}$$

2.6.6 Ekvalizácia histogramu

Cieľom ekvalizácie histogramu je vytvoriť obraz s rovnako rozloženými jasovými úrovňami v celej jasovej škále. Na histogram sa môžeme pozeráť ako na diskretnú funkciu hustoty pravdepodobnosti. Môžeme napísať $\sum_{i=0}^k G(q_i) = \sum_{i=0}^k H(p_i)$. Ekvalizovaný histogram $G(q)$ zodpovedá uniformnej funkcii hustoty pravdepodobnosti f , ktorej funkčné hodnoty sú konštantné: $f = \frac{N^2}{q_k - q_0}$

2.6.7 Top hat

Top hat filtrovanie je operácia matematickej morfológie, ktorá odpočíta morfológické otvorenie obrázku od pôvodného obrázku. **Morfologické otvorenie** zodpovedá

filtrovaníu obrázku pomocou erózie a následne pomocou dilatácie. **Erózia** alebo Minkowského rozdiel je operácia, ktorá znižuje množinu, odstraňuje štruktúry určitej veľkosti a tvaru v závislosti od štruktúrného prvku (SE^3). Eróziu môžeme zapísať ako: $A \ominus B = \{x \mid B_x \subseteq A\}$ kde B je SE a x jeho referenčný bod. **Dilatácia** alebo Minkowského súčet zväčšuje množinu, vyplní diery a zálivy určitej veľkosti a tvaru v závislosti od SE. Dilatácia: $A \oplus B = \bigcup_{b \in B} A_b$ kde B je SE. Teraz môžeme zapísať otvorenie ako: $A \circ B = (A \ominus B) \oplus B$ Top hat je v tomto kontexte: $topHat(A) = A - (A \circ B)$

2.6.8 Bottom hat

Bottom hat filtrovanie je operácia matematickej morfológie, ktorá odpočíta pôvodný obrázok od morfológického uzavretia obrázku. **Morfologické uzavretie** zodpovedá filtrovaníu obrázku pomocou dilatácie a následne pomocou erózie. Uzavretie môžeme zapísať ako: $A \bullet B = (A \oplus B) \ominus B$ Bottom hat je potom: $bottomHat(A) = (A \bullet B) - A$

2.6.9 Extended minima transformácia

Extended minima transformácia je operácia, ktorá najskôr filtruje obrázok tak, že odstráni všetky minimá v obrázku, ktorých hodnota je menšia ako zadaný prah - h. Potom vráti regionálne minimá ⁴ z filtrovaného obrázku.

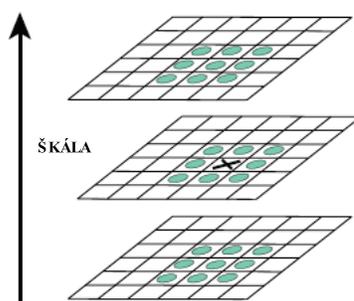
2.6.10 SIFT

SIFT (Scaleable Invariant Feature Transform) je technika vytvorená Lowem [37], ktorá predstavuje škálovo a rotačne invariantný detektor a deskriptor bodov záujmu (interest points IP). V prvom kroku techniky sú detektorom nájdené IP v obrázku a v druhom kroku je vytvorený deskriptor pre každý z týchto bodov. Konkrétne metódy detektora a deskriptora sú:

³Štruktúrný prvok má definovanú veľkosť, tvar, orientáciu a pozíciu X - teda referenčný bod, alebo počiatok súradnicovej sústavy

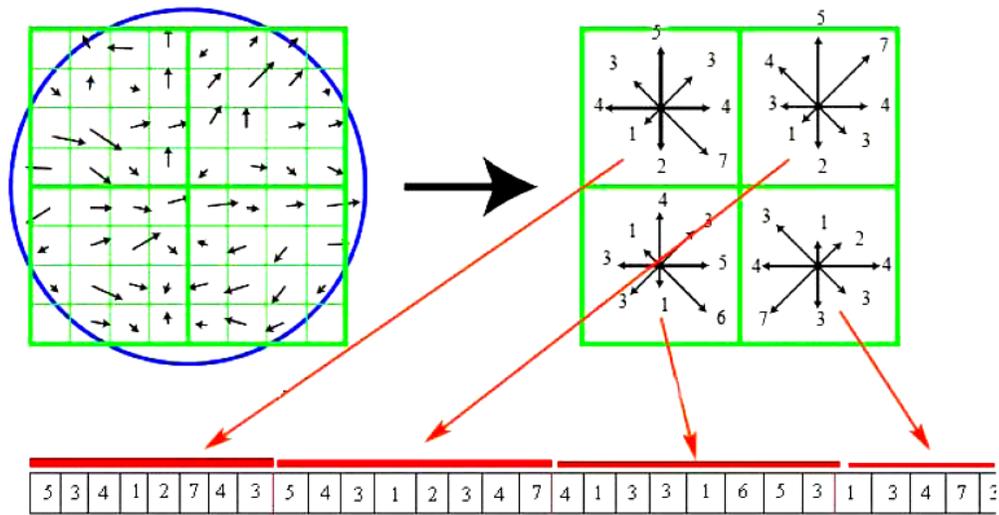
⁴Regionálne minimá sú spojitú oblasti pixelov s konštantnou hodnotou intenzity, pre ktoré platí, že všetky pixely na ich vonkajšej hranici majú väčšie hodnoty intenzity

detektor: SIFT využíva takzvaný „scale space“ s využitím rozdielu Gaussiánov. Obrázok je podvzorkovaný do viacerých veľkostí - „oktáv“. V každej oktáve je obrázok opätovne filtrovaný Gaussiánom, DoG je rozdiel dvoch nasledovných filtrovaných obrázkov v jednej oktáve. V DoG sa potom hľadajú lokálne extrémery v okolí $3 \times 3 \times 3$. Okolie $3 \times 3 \times 3$ znamená, že sa v rámci jedného obrázku hľadá v okolí 3×3 , ale taktiež sa hľadá aj v okolí 3×3 na nasledovnom a predchádzajúcom obrázku v rámci „scale-space“ (pozri obrázok 2.5). Tieto sú identifikované ako IP, sú to kandidáti na kľúčové body (keypoints KP). Do KP sa dostanú len dostatočne kontrastné IP.



Obr. 2.5: Lokálne extrémery boli nájdené porovnávaním pixelu s okolím $3 \times 3 \times 3$, teda 3×3 okolím v aktuálnom, predchádzajúcom a nasledujúcom obrázku. Obrázok bol prebratý z [37]

deskriptor: Pre každý KP bod je pomocou jeho okolia vypočítaný deskriptor z histogramov orientácií o veľkosti 4×4 . V každom bode okolia je vypočítaná veľkosť a orientácia gradientu, ktorým sú potom priradené váhy pomocou Gaussovho kruhového okna (pozri obrázok 2.6). Z týchto hodnôt sú vytvorené histogramy orientácií, každý pokrývajúci jeden zo šestnástich (4×4) regiónov z okolia KP. Dĺžka šípky v histograme predstavuje sumu veľkostí gradientov v blízkosti daného smeru v oblasti. V každom histograme je 8 smerov, čo pri veľkosti 4×4 znamená 128 hodnôt deskriptora pre jeden KP.



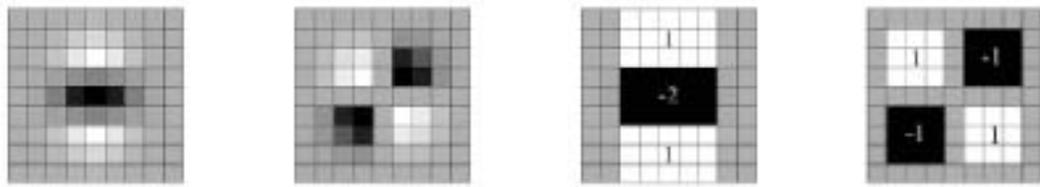
Obr. 2.6: Príklad vytvorenia deskriptora z okolia kľúčového bodu pri SIFT metóde. Na obrázku sú použité štyri (2x2) regióny namiesto šestnásť(4x4). Obrázok prebratý z [37].

2.6.11 SURF

SURF (Speeded Up Robust Features) je technika, ktorú autori Bay et al. [6] považujú za vylepšenie SIFTu a iných podobných techník. SURF obsahuje podobne ako SIFT detektor na nájdenie a deskriptor na popísanie záujmových bodov.

detektor: SURF, podobne ako SIFT detekuje IP v „scale space“, namiesto Gausiánov však využíva aproximáciu Gaussiánov druhej parciálnej derivácie v smere y a xy pomocou box filtrov (obrázok 2.7). Na vytvorenie „scale space“ sa pri tejto metóde nevyužíva zmenšovanie obrázka a následná konvolúcia, ale naopak, konvolúcia pôvodného obrázka pomocou postupne sa zväčšujúcich filtrov. Na nájdenie IP je na obrázok aplikovaná supresia ne-maximálnych hodnôt v okolí 3x3x3 (Okolie 3x3x3 je okolie spomínané pri technike SIFT).

deskriptor: Pre vytvorenie deskriptora KP je potrebné v prvom kroku zistiť jeho orientáciu a v druhom kroku vytvoriť štvorcovú oblasť zarovnanú podľa tejto orientácie (z nej sa bude deskriptor extrahovať). Pre určenie orientácie KP



Obr. 2.7: Prvé dva: druhé parciálne derivácie Gaussiánov v smere y a xy Druhé dva: aproximácie druhej parciálnej derivácie Gaussiánov v smere y a xy pomocou box filtrov. Obrázok prebratý z [6].

sa najskôr vypočítajú odozvy Haarovho waveletu v smere x a y v kruhovom okolí KP s priemerom $6s$, pričom s je veľkosť aktuálnej škály. Vypočítaným odozvám sú potom priradené váhy pomocou Gaussovho kruhového okna a sú reprezentované ako vektor v priestore. Dominantná orientácia KP je najväčší z vektorov vypočítaných ako suma vertikálnych a horizontálnych odoziev, získaných pomocou otočného okna ako suma všetkých odoziev v danej orientácii. V ďalšom kroku je vytvorená v okolí KP štvorcová oblasť zarovnaná podľa orientácie. Oblasť je rozdelená na 4×4 subregióny, pričom v každom sú vypočítané 4 príznaky na 5×5 dobre rozmiestnených bodoch. Príznakmi sú: suma odoziev Haarovho waveletu v horizontálnom a vertikálnom smere a suma absolútnych hodnôt odoziev Haarovho waveletu v horizontálnom a vertikálnom smere. Deskriptor KP potom obsahuje 4 hodnoty pre každú zo 16 oblastí (spolu 64 hodnôt).

Kapitola 3

Implementácia

Metóda na identifikáciu obrazu nachádzajúceho sa na fotografii popísaná v tejto práci bola implementovaná v prostredí MATLAB. Navrhnutý prístup je rozdelený na 3 fázy: predspracovanie obrazu (`PreprocessImg`), segmentácia maľby z obrazu (`SegmentImg`) a klasifikácia maľby (`ClassifyImg`). V segmentačnej a klasifikačnej fáze boli implementované rôzne prístupy, ktoré sú validované v 4. kapitole tejto práce. V tejto kapitole je ďalej popísaná aplikácia s jednoduchým GUI ⁵ využívajúca tieto tri funkcie, ktorá bola v prostredí MATLAB vytvorená kvôli lepšej prezentácii jednotlivých metód.

3.1 Predspracovanie

Hlavnou úlohou fázy predspracovania je načítanie obrázka, jeho prevzorkovanie, konverzia na šedotónový (v prípade farebného obrázka) a ekvalizácia histogramu. Všetky tieto súčasti sú zahrnuté vo funkcii `PreprocessImg`. Na vstupe funkcie sa nachádza názov fotografie a absolútna cesta v tvare `'abs.cesta/názov.jpg'`. Tento obrázok je načítaný do premennej `I`:

```
I = imread('abs.cesta/názov.jpg');
```

V prvom kroku sa zistí, či je obrázok farebný, ak áno je prevedený na šedoúrovňový pomocou MATLAB funkcie `rgb2gray(I)`. Funkcia `rgb2gray` vytvorí šedoúrovňový obrá-

⁵Graphical User Interface - Grafické užívateľské rozhranie

zok tak, že jeho hodnoty prevedie z farebného modelu RGB na model HSL ⁶ a vytvorí šedoúrovňový obrázok s hodnotami intenzity rovnajúcej sa L - lightness [40]. V ďalšom kroku je obrázok prevzorkovaný na výšku 300 px so zachovaním pomeru strán pomocou MATLAB funkcie `resize`. Na zlepšenie kvality obrazu zvýšením kontrastu sa využíva ekvalizácia histogramu. V tejto práci bola na ekvalizáciu histogramu obrázku I využitá MATLAB funkcia

```
I = histeq(I);
```

Takto predspracovaný obraz je potom výstupom funkcie `PreprocessImg`.

3.2 Segmentácia

V časti segmentácia bolo hlavným cieľom vysegmentovať maliarske plátno z fotografie. Na tento účel boli v rámci tejto práce naimplementované 4 rôzne metódy. Prvá metóda využíva techniku Gauss gradient [61], vylepšené metódy využívajú Anisotropic diffusion [45] a Kuwahara filter [34] a posledná metóda predstavuje odlišný prístup s pomocou segmentácie s využitím povodí [22].

3.2.1 Gauss gradient

Prvým krokom je spracovanie vstupného obrázku 3.2(a) pomocou Gauss gradient filtra.

```
[Gx,Gy]=gaussgradient(I);
```

Výstupom filtra sú obrázky gradientu v smere x (G_x 3.2(b)) a y (G_y 3.2(c)), ktoré vznikli konvolúciou vstupného obrázku prvou deriváciou Gausiánu. V prvom kroku sú vytvorené prvé derivácie Gausiánu podľa x a y. V smere x (algoritmus 1 obr 3.1) kde `gauss` a `dgauss` sú funkcie ktoré vytvoria Gausián a jeho prvú deriváciu. Matica v smere y je totožná s transponovanou maticou v smere x.

```
hy=hx'
```

Výstupné obrázky G_x , G_y potom vzniknú konvolúciou vstupného obrázku maticami h_x a h_y :

```
Gx=imfilter(IM,hx,'replicate','conv');
```

⁶HSL - farebný model pozostávajúci z 3 zložiek: odtieň, saturácia a jas

```

for i=1:size
    for j=1:size
        u=[i-halfsize-1 j-halfsize-1];

        hx(i,j)=gauss(u(1),sigma)*dgauss(u(2),sigma);
    end
end

```

Obr. 3.1: algoritmus 1.

```
Gy=imfilter(IM,hy,'replicate','conv');
```

Kde 'replicate' znamená, že hodnotám obrázku, ktoré sú mimo rozsah, sú priradené najbližšie okrajové hodnoty obrázku. Obrázky Gx a Gy sú následne prevedené zo šedotónových na binárne (Bx, By), pomocou prahu s hodnotou 80 (pri hodnotách 0-100). Najvhodnejší prah bol zvolený pomocou testovania rôznych prahov na 5 obrázkoch.

Konečná segmentácia s pomocou štandardnej Houghovej transformácie

Na nájdenie úsečiek ohraničujúcich rám v obrázku je na binárne obrázky použitá MATLAB implementácia Houghovej transformácie (pozri 2.6). V prvom kroku je použitá štandardná Houghova transformácia (SHT):

```
[Hx,tx,rx]=hough(Bx);
```

kde Hx^7 je Houghova transformačná matica (HTM), tx - theta (v stupňoch) a rx - rho sú polia rho a theta hodnôt nad ktorými bola vygenerovaná transformačná matica [40]. V ďalšom kroku sú lokalizované vrcholy v HTM pomocou funkcie:

```
peaksx = houghpeaks(Hx,5);
```

kde hodnota 5 špecifikuje maximálny počet vrcholov ktoré sa majú v HTC vyhľadať a $peaksx$ je $Q \times 2$ matica, kde Q sa pohybuje od 0 do 5 a obsahuje číslo riadku a stĺpca, v ktorom sa vrchol v HTC nachádza. Následne je použitá funkcia `houghlines`, ktorá extrahuje úsečky z obrázka Bx, ktoré sú prepojené s poliami vo vytvorenej SHT. Na vstupe funkcie sú 4 parametre získané pomocou predchádzajúcich funkcií Bx, tx, rx, peaksx a 2 voliteľné parametre 'FillGap' a 'MinLen'. FillGap je kladné reálne číslo, ktoré špecifikuje maximálnu vzdialenosť medzi dvoma úsečkami

⁷Pre Hy platí: $[Hy,ty,ry]=hough(By)$, v ostatných príkladoch podobne

(prepojenými s poliami v SHT), pri ktorej sa spoja do jednej úsečky. Hodnota 80 bola získaná pomocou testovania rôznych hodnôt na databáze 20 fotografií z databázy Fotografií turistov. `MinLen` je kladné reálne číslo, ktoré určuje najmenšiu veľkosť úsečky ktorá má byť zachovaná (menšie úsečky sú odstránené).

```
linesx=houghlines(Bx,tx,rx,peaksx,'FillGap',80,'MinLen',50)
```

kde `linesx` je štruktúrované pole dĺžky rovnajúcej sa počtu nájdených úsečiek, ktoré obsahuje 4 hodnoty pre každú úsečku:

1. `point1`: 2 zložkový vektor [X Y] špecifikujúci začiatkový bod úsečky
2. `point2`: 2 zložkový vektor [X Y] špecifikujúci koncový bod úsečky
3. `theta`: uhol (v stupňoch) poľa v SHT
4. `rho`: pozícia poľa v SHT na osi rho.

Následne je pre každý zo smerov hore, dole (`linesy`), vpravo, vľavo (`linesx`) určená úsečka najlepšie zodpovedajúca podmienkam na segmentačnú priamku. Pre segmentačnú priamku (vertikálnu) musí platiť, že bod, ktorý má x-ovú súradnicu totožnú so stredom obrázku musí mať y-ovú súradnicu čo najbližšie k stredu obrázka, nie však bližšie ako jednu osminu šírky obrázka⁸. Úsečka zvolená ako segmentačná priamka pre jeden zo smerov je potom predĺžená na segmentačnú priamku tak, aby sa jej koncové body nachádzali na okraji obrázku 3.2(d). 4 segmentačné priamky potom jednoznačne identifikujú vysegmentovanú oblasť ako vnútro štvoruholníka, ktorý vytvorili 3.2(e).

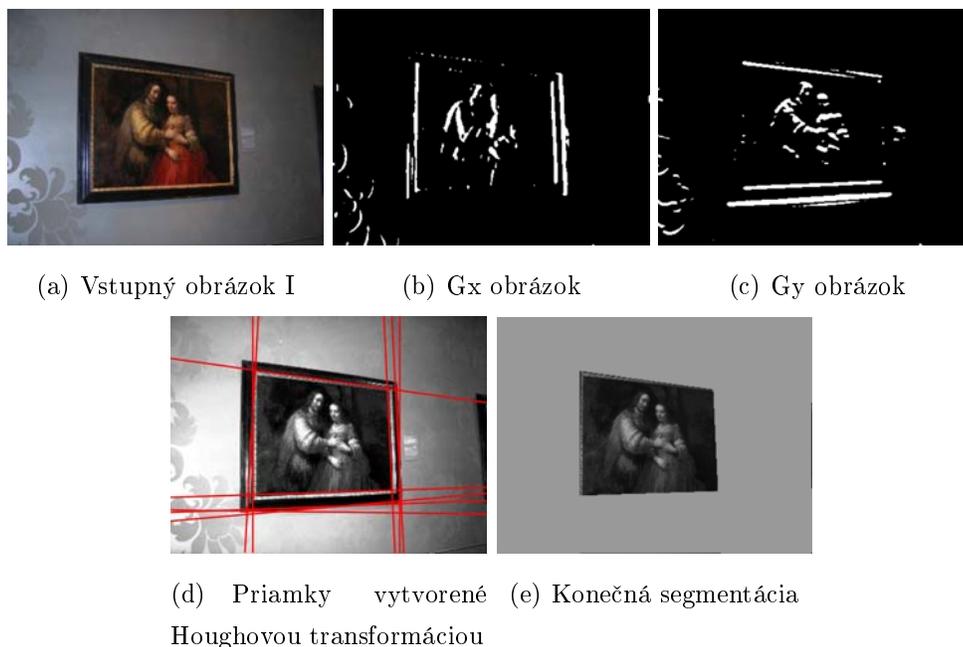
3.2.2 Anisotropic diffusion

Pri druhej metóde bolo prvým krokom filtrovanie vstupného obrázka I 3.3(a) metódou Anisotropic diffusion (pozri 2.6), ktorá obrázok vyhladí pri zachovaní hrán.

```
diff = anisodiff(I, 10, 30, .2, 1);
```

kde na vstupe do funkcie sú argumenty (v poradí) vstupný obrázok, počet iterácií,

⁸Pre horizontálnu priamku platí, že bod, ktorý má y-ovú súradnicu totožnú so stredom obrázka musí mať x-ovú súradnicu čo najbližšie k stredu obrázka, nie však bližšie ako jednu osminu výšky obrázka



Obr. 3.2: Proces segmentácie pomocou metódy Gauss gradient.

kappa - koeficient kondukcie (20-100), lambda (maximálna hodnota stability - 0.25), a typ Perona - Malik difúznej rovnice (1 alebo 2). Na výstupe funkcie je vyhladený obrázok 3.3(b), ktorý je následne konvolvovaný vertikálnym a horizontálnym Sobelovým operátorom (pozri 2.6):

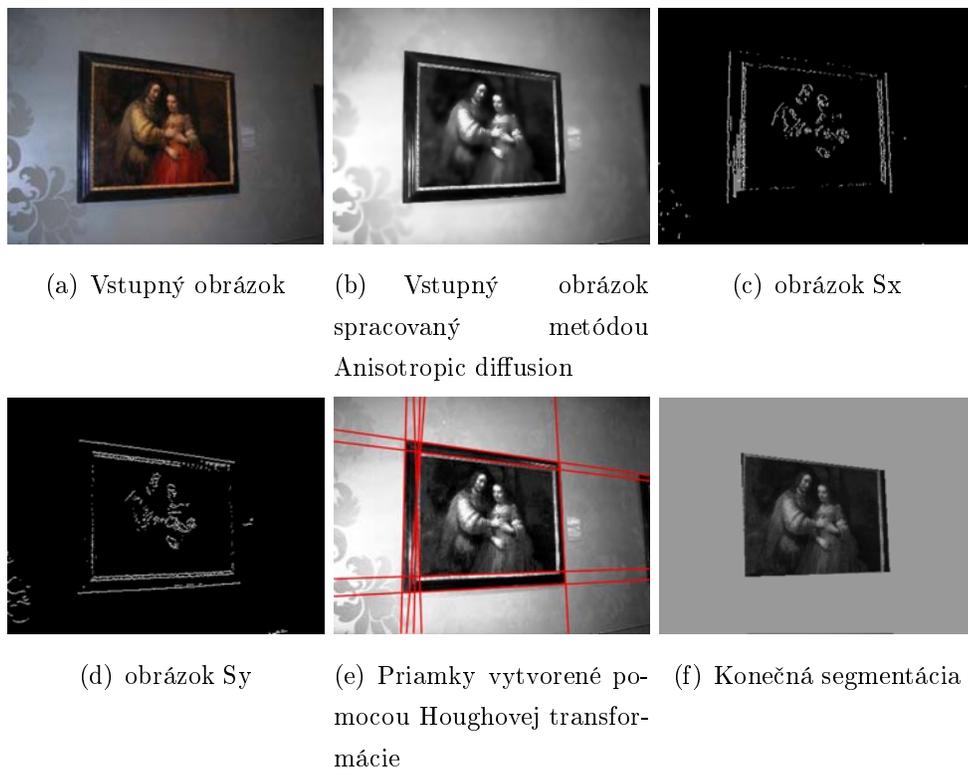
```
Sy=edge(diff/255,'sobel',[], 'horizontal');
```

```
Sx=edge(diff/255,'sobel',[], 'vertical');
```

kde funkcia edge na vstupe požaduje (v poradí): vstupný obrázok, typ operátora, prah⁹ a smer hrán. Na výstupe funkcie sú obrázky S_x 3.3(c) a S_y 3.3(d) zobrazujúce hrany.

Ďalšie kroky tejto metódy sú totožné s postupom v časti Konečná segmentácia s pomocou štandardnej Houghovej transformácie v metóde Gauss gradient.

⁹V prípade, že prah nieje špecifikovaný, je zvolený adaptívne, čo sa ukázalo ako najlepšie riešenie v tomto prípade



Obr. 3.3: Proces segmentácie pomocou metódy Anisotropic diffusion.

3.2.3 Watershed

Metóda založená na segmentácii s využitím povodí bola v tejto práci naimplementovaná, aby prezentovala odlišný prístup k segmentácii. Iniciálnym krokom metódy je prevedenie vstupného obrázka I 3.4(a) na dvojité presnosť. Cieľom nasledujúcich krokov je zintenzívniť kontrast, zvýrazniť hrany a predspracovať obrázok pred použitím segmentácie s využitím povodí. V prvom kroku je vytvorený Top hat (I_{top} 3.4(b)) a Bottom hat (I_{bot} 3.4(c)) z obrázka I s využitím štruktúrneho elementu se (v tvare disku, o veľkosti 5 px):

```
 $I_{top} = \text{imtophat}(I, se);$ 
```

```
 $I_{bot} = \text{imbothat}(I, se);$ 
```

Obrázok so zvýšeným kontrastom ($I_{enhance}$ 3.4(d)) je docieľený vzťahom ($I + \text{Top Hat} - \text{Bottom hat}$):

```
 $I_{enhance} = \text{imsubtract}(\text{imadd}(I_{top}, I), I_{bot});$ 
```

V ďalšom kroku je vytvorený komplement (I_{ec}) obrázka $I_{enhance}$, na ktorý je potom využitá Extended minima transformácia (I_{emin} 3.4(e)), ktorá vytvorí binárny obrázok I_{emin} :

```
 $I_{ec} = \text{imcomplement}(I_{enhance});$ 
```

```
 $I_{emin} = \text{imextendedmin}(I_{ec}, .49);$ 
```

kde .49 v druhom priradení je hodnota pre h-minima transformáciu. Posledným krokom pred použitím povodí je vytvorenie obrázku I_{impose} 3.4(f) pomocou I_{impose} minima¹⁰ transformácie obrázku I_{ec} so značkovačom I_{emin} :

```
 $I_{impose} = \text{imimposemin}(I_{ec}, I_{emin});$ 
```

Na obrázok I_{impose} je aplikovaná segmentácia a je vytvorený obrázok wat 3.4(g):

```
 $wat = \text{watershed}(I_{impose});$ 
```

Konečná segmentácia 3.4(h) je stanovená tak, že za pozadie určíme tie body, ktoré majú rovnaké RGB hodnoty ako jeden z rohov obrázku.

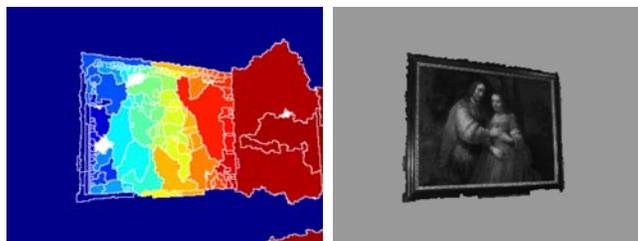
¹⁰Impose minima transformuje pomocou morfolologickej rekonštrukcie intenzitu obrázku tak, aby mal regionálne minimum tam, kde má značkovač nenulovú hodnotu



(a) Vstupný obrázok

(b) Top hat

(c) Bottom hat

(d) $(I + \text{tophat}) - \text{bottomhat}$ (e) Extended minima z (d) (f) Minima imposition (d) z markrom (e)

(g) Oblasti vytvorené pomocou Houghovej transformácie

(h) Konečná segmentácia

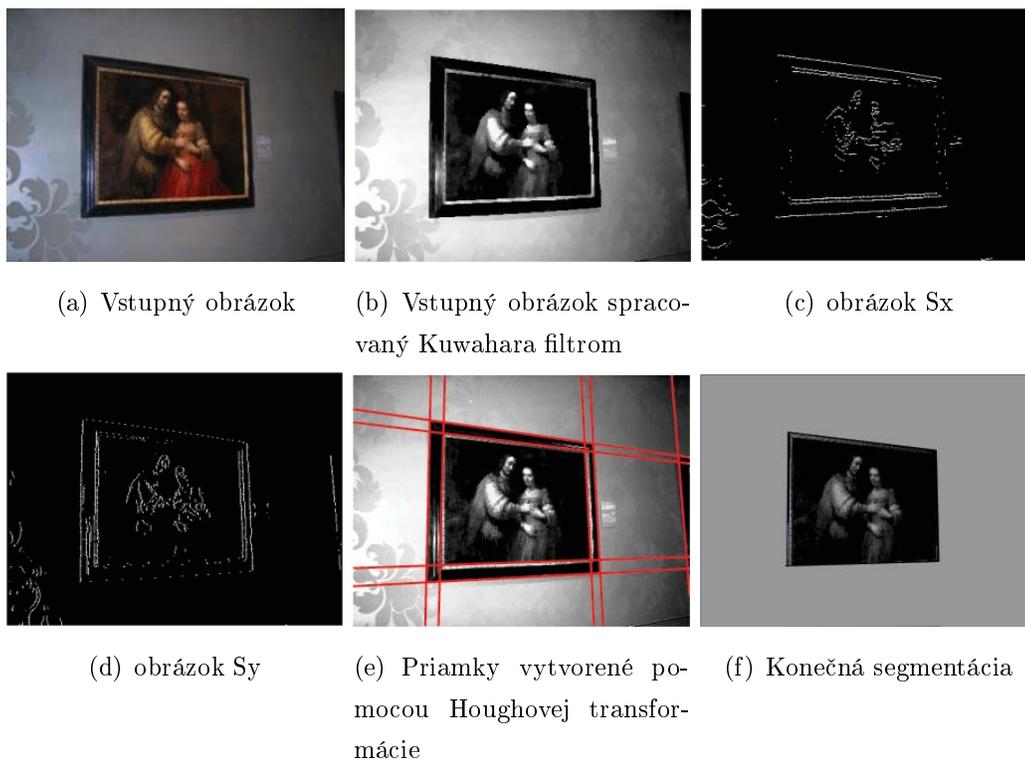
Obr. 3.4: Proces segmentácie pomocou metódy Watershed.

3.2.4 Kuwahara filter

Posledná metóda je takmer totožná s metódou Anisotropic diffusion. Jediným rozdielom je filter, ktorým je vstupný obrázok 3.5(a) na začiatku spracovaný. Ako už názov napovedá, pri tejto metóde bol použitý Kuwahara filter (pozri odsek 2.6):

```
kuw = Kuwahara(I,5);
```

kde prvý argument funkcie je vstupný obrázok a druhý veľkosť Kuwaharovho filtrovacieho okna. Výstupný obrázok `kuw` 3.5(b) je potom spracovaný rovnakými technikami ako obrázok `diff` 3.5(b) v metóde Anisotropic diffusion (Sobel, Hough..).



Obr. 3.5: Proces segmentácie pomocou metódy Kuwahara filter.

3.3 Klasifikácia

Výstupom zo segmentačnej fázy, teda funkcie `SegmentImg`, je obrázok obsahujúci vysegmentovanú fotografiu s čiernym pozadím, ktorý je vstupom do klasifikačnej fázy,

do funkcie `ClassifyImg`. Cieľom tejto fázy je vytvoriť deskriptor príznakov (prízna-ky založené na lokálnych invariantoch) popisujúci daný obrázok a potom porovnávaním deskriptora s deskriptormi obrázkov z databázy nájsť korešpondujúci obrázok. Klasifikácii obrázkov z databázy Fotografií turistov teda musí predchádzať vytvorenie databázy deskriptorov Originálov obrazov. Dôležitými úlohami tejto fázy sú: vytvorenie deskriptora obrázka, vytvorenie databázy deskriptorov a nájdenie najviac korešpondujúceho deskriptora z databázy. V tejto práci sú tieto úlohy riešené dvomi prístupmi: prvý je založený na metóde SIFT a druhý na metóde SURF.

3.3.1 SIFT

V SIFT prístupe je použitá C++ implementácia funkcie na extrakciu SIFT deskriptorov kľúčových bodov (KP) z obrázku vytvorená tvorcom metódy Lowem¹¹ [37]. Vstupom do tejto funkcie je šedoúrovňový obrázok a výstupom je súbor '`názov.m`' obsahujúci deskriptory všetkých kľúčových bodov identifikovaných v obrázku (ďalej DF). Štruktúra DF súboru je nasledovná:

2 x celé číslo Prvé číslo **num** je počet nájdených KP a druhé číslo **len** je počet hodnôt deskriptora pre jeden KP.

num x 4 x racionálne číslo Hodnoty určujú polohu KP. V poradí: riadok, stĺpec, škála a orientácia.

len x celé číslo 128 - hodnotový deskriptor pre každý KP (pozri 2.6).

V prípravnej fáze metódy bolo potrebné vytvoriť databázu DF súborov pre 15 obrazov z databázy Originálov. Každý nový DF súbor vytvorený zo segmentovaného vstupného obrazu bude potom porovnávaný s touto databázou. Databáza DF súborov bola vytvorená nasledovne: V prvom kroku je každý obraz načítaný a prevedený na šedoúrovňový pomocou techník spomínaných v sekcii 3.1. Následne je zavolaný vykonateľný súbor SIFT [37], ktorý vytvorí DF súbor '`názov.m`'. DF súbor je zanalyzovaný kvôli prípadným chybám. Prvým krokom je pokus o otvorenie súboru pomocou príkazu (algoritmus 2 obr 3.6):

¹¹Keďže táto funkcia je naimplementovaná autorom metódy a je rýchla a efektívna, bolo zbytočné implementovať túto funkciu znova

```

g = fopen(name '.m', 'r');
if g == -1
    error('Nemôžem otvoriť súbor.');
```

Obr. 3.6: algoritmus 2.

Potom sú načítané prvé dve hodnoty v súbore, teda `len` a `num`, a je otestované, či je číslo `len` rovné 128, teda či majú deskriptory správnu dĺžku (algoritmus 3 obr 3.7):

```

[header, count] = fscanf(g, '%d %d', [1 2]);
if count ~= 2
    error('Zlá hlavička súboru.');
```

Obr. 3.7: algoritmus 3.

Potom je v cykle od 1 do `num` otestované či pre každý KP sú v DF súbore 4 hodnoty pre polohu a 128 hodnôt deskriptora. Ak kontrola prebehne úspešne, môže byť 'názov.m' zaradený do databázy.

Po vytvorení databázy nasleduje vlastná klasifikácia neznámych obrazov vysegmentovaných z fotografií turistov. Prvým krokom je, podobne ako pri vytváraní databázy, vytvorenie DF súboru a kontrola jeho obsahu. Okrem kontroly súboru sú deskriptory KP zapísané do poľa `deskriptorsx` o veľkosti `num` x `len`. V ďalšom kroku je postupne spracovaná databáza DF súborov, z ktorých sú vytvorené polia `deskriptors1-15`. Nasleduje postupné porovnávanie (získovanie počtu zhôd) polí `deskriptorsx` a `deskriptors1-15` pomocou techniky najbližšieho suseda (nearest neighbour). Pre jednu dvojicu (napr. `deskriptorsx` a `deskriptors1`) technika najbližšieho suseda nájde pre každý deskriptor KP (teda každý riadok v poli `deskriptorsx`) **najbližšieho suseda** (NS) a **druhého najbližšieho suseda** (DNS) v poli `deskriptors1`. NS je taký riadok v poli `deskriptors1`, ktorého euklidovská vzdialenosť je najmenšia.

Dôležitou hodnotou pri porovnávaní technikou najbližšieho suseda je počet zhôd. Za zhodu považujeme prípad, keď pre riadok v `descriptorx` platí, že euklidovská vzdialenosť NS je menšia ako euklidovská vzdialenosť DNS vynásobená `distanceRatio` (premenná `distanceRatio` bola autorom SIFTu zadaná ako 0.6). Vstupný obraz reprezentovaný poľom `descriptorx` je teda klasifikovaný ako originál (1-15) reprezentovaný poľom `descriptor1-15` s najvyšším počtom zhôd s poľom `descriptorx`.

Problém môže nastať pri klasifikácii fotografií obrazov, ktoré nie sú prezentované v databáze Originálov obrazov. SIFT metóda nie je 100% rozdeľujúca a následkami šumu a/alebo rozmazania sa stáva, že dva nekorešpondujúce obrazy majú počet zhôd väčší ako 0. Kvôli týmto prípadom bol v tejto diplomovej práci zavedený **prah** (viac v 4.2), na rozlíšenie správnej klasifikácie. Ak je najvyšší počet zhôd poľa `descriptorx` s poľom z `descriptor1-15` menší ako **prah**, vstupný obrázok je klasifikovaný ako nenachádzajúci sa v databáze. Ak je tento počet väčší ako **prah**, obrázok je klasifikovaný ako jeden z obrazov 1-15.

Ďalším vylepšením tejto metódy je orezanie prehľadávania. Kvôli zrýchleniu metódy bol ustanovený tzv. **najlepší prah**. Ak pri porovnávaní poľa `descriptorx` a `descriptor1-15` nastane počet zhôd väčší alebo rovný najlepšiemu prahu, je vstupný obrázok klasifikovaný ako obraz s číslom daného poľa a ďalej sa už neprehľadáva. Hodnota najlepšieho prahu bola určená ako 30, čo zvýšilo rýchlosť v 60% výpočtov.

3.3.2 SURF

Pri implementácii SURF metódy bol opäť použitý spustiteľný súbor napísaný v jazyku C++ autorom metódy [6]. Na vstupe funkcie SURF ktorú súbor obsahuje, je šedotónový obrázok a na výstupe súbor 'názov.surf'. Štruktúra súboru je nasledovná:

2 x celé číslo Prvé číslo **len** je počet hodnôt deskriptora pre jeden KP + 1 (zvyčajne 65) a druhé číslo **num** je počet nájdených KP.

num x 2 x racionálne číslo Hodnoty určujú polohu KP.

3 x racionálne číslo 1, 2, 3 = [1 2; 2 3] vstup do matice druhého momentu. SURF používa kruhové oblasti.

1 x 1/-1 Znamienko Laplaciánu



(a) Body detekované pomocou SIFTu (b) Body detekované pomocou SURFu

Obr. 3.8: Porovnanie kľúčových bodov detekovaných metódami SIFT (354 KP) a SURF (140 KP) na obraze Danae z databázy Originálov obrazov.

len-1 x racionálne číslo 64 - hodnotový deskriptor pre každý KP (pozri 2.6).

Okrem rozdielov v štruktúre súborov 'názov.m' a 'názov.surf' sú medzi implementáciami SURF a SIFT metód v tejto práci len drobné odlišnosti. Prvou odlišnosťou je dĺžka deskriptora. Pri SIFT metóde je to 128 hodnôt: 8 hodnôt histogramu orientácií pre každú zo 16 oblastí a pri SURF metóde 64: 4 hodnoty odzviav Haarovho waveletu zo 16 oblastí. Ďalším rozdielom je určenie premennej distanceRatio pri technike najbližšieho suseda. Zatiaľ čo pre SIFT bola určená ako 0.6 pre SURF je to 0.7. Metódy majú taktiež rozlične určené prahové hodnoty správnej klasifikácie (viac v 4.2). Hodnota **najlepšieho prahu** bola pri metóde SURF určená ako 20, keďže počet detekovaných KP, teda aj počet zhôd je pri metóde SURF nižší ako pri metóde SIFT.

3.3.3 Pomer strán

Klasifikačná fáza bola kvôli komplexnosti rozšírená o porovnávanie na základe pomeru strán obrazu. Keďže vo výtvarnom umení nie je veľkosť obrazu konštantná, ale naopak veľmi rôznorodá v závislosti od autora a typu obrazu, hodnota pomeru strán obrazu môže byť pomocným príznakom pri klasifikácii obrazov. Obrázky vstupujúce do segmentačnej fázy obsahujú vysegmentované obrazy, z ktorých je možné vypočítať výšku

a šírku obrazu a vyjadriť ich pomer. Tento pomer môže byť potom porovnaný s predpočítanými pomermi strán obrazov z databázy Originálov. So znižujúcim sa rozdielom pomerov sa zvyšuje pravdepodobnosť správnej klasifikácie.

Okrem využitia pri klasifikácií sa dá príznak pomeru strán využiť aj na zrýchlenie porovnávania DF súboru s databázou v technikách SIFT a SURF. So znalosťou pomeru strán môžeme DF súbor daného obrazu porovnávať len s DF súbormi obrazov, ktorých pomer strán je od neho vzdialený najviac o hodnotu **vzdialenosť pomerov**. Určenie tejto hodnoty a validácia techniky je popísaná v časti 4.2

3.4 MATLAB aplikácia

Kvôli vizualizácii a lepšiemu porovnaniu jednotlivých metód bola v prostredí MATLAB vytvorená aplikácia ProjectRembrandt, ktorá obsahuje funkcie PreprocessImg, SegmentImg a ClassifyImg popísané v predchádzajúcej časti. Aplikácia bola vytvorená pomocou sprievodcu GUIDE - GUI Creator. Nasleduje krátky popis použitých komponentov.

1 x obrázok - vymedzuje priestor, do ktorého sa načíta vstupný obrázok.

1 x textové okno - textové okno sprostredkúva priebežné informácie a výsledky používateľovi. Po klasifikácii obrazu, nachádzajúceho sa v obrázku, sa v okne zobrazí názov obrazu. V prípade, že sa obraz v databáze nenachádza, v okne sa o tom objaví správa.

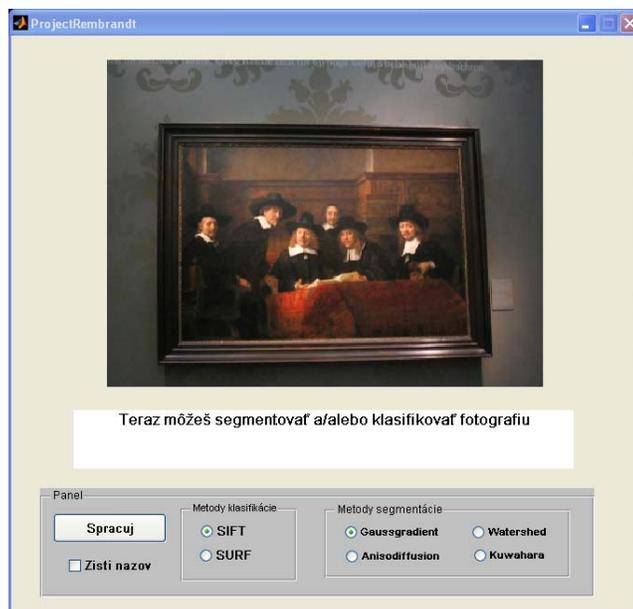
1 x panel obsahujúci:

2 x tlačidlo - 2 prekrývajúce sa tlačidlá umožňujú užívateľovi najskôr otvoriť adresárovú štruktúru pre načítanie obrázku a potom odoslať obrázok na klasifikáciu. Po kliknutí na prvé tlačidlo sa toto stane neviditeľným a užívateľovi je sprístupnená funkčnosť druhého tlačidla.

1 x check box - pri zaškrtnutom check boxe je obrázok následne po segmentácií klasifikovaný. V prípade, že check box zaškrtnutý nie je, vykoná sa len fáza segmentácie.

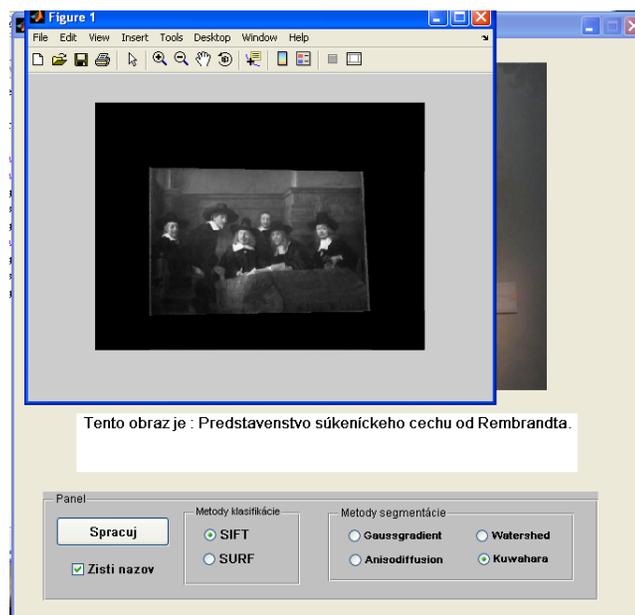
1 x skupina tlačidiel - 2 x radio button - radio buttony umožňujú užívateľovi výber klasifikačnej metódy. Na začiatku je aplikácia nastavená na metódu SIFT.

1 x skupina tlačidiel - 4 x radio button - radio buttony umožňujú užívateľovi výber segmentačnej metódy. Pôvodne je aplikácia nastavená na metódu Gauss gradient.



Obr. 3.9: Okno s načítanou fotografiou.

Po spustení aplikácie sa objaví okno s textom obsahujúcim výzvu na zvolenie fotografie. Po načítaní fotografie sa objaví okno ako na obrázku 3.9. Teraz si môže užívateľ zvoliť metódu na segmentáciu fotografie z obrazu a aj metódu na klasifikáciu obrazu. Po úspešnej segmentácii a klasifikácii sa objaví okno ako na obrázku 3.10 s vysegmentovanou fotografiou a názvom obrazu, ktorý sa na nej nachádza (ak je tento obsiahnutý v databáze).



Obr. 3.10: Okná po segmentácii a klasifikácii.

Kapitola 4

Validácia

Štvrtá kapitola tejto práce je venovaná overeniu výsledkov použitých algoritmov, ich porovnaniu a vyhodnoteniu najlepších metód. Metódy boli testované na jednej alebo oboch databázach z dvojice databáza Originálov obrazov a databáza Fotografií turistov.

4.1 Segmentácia

V časti Segmentácia boli metódy testované na obrázkoch z databázy Fotografií turistov (databáza Originálov obrazov obsahuje už vysegmentované obrázky). Obrázky z tejto databázy boli odfotografované pomocou rôznych fotoaparátov, pod rôznym osvetlením a s rôznym rozlíšením, čo ovplyvnilo ich výslednú kvalitu. Najviac problémov v segmentačnej fáze bolo spôsobených nízkym kontrastom a neostrosťou fotografií. Problémy s nízkym kontrastom boli čiastočne vyriešené použitím ekvalizácie histogramu obrázku vo fáze predspracovania.

Pri segmentácií môžu vznikáť dva druhy problémov. Sú to nadsegmentácia a podsegmentácia. V tejto práci spôsobuje väčší problém nadsegmentácia, keďže sa môže stať, že do výslednej segmentácie nezaradí charakteristické (pre klasifikáciu dôležité) časti obrazu. Nadsegmentácia je najmä pri metóde Kuwahara filter a Gauss gradient spôsobená výraznými hranami v obraze.



Obr. 4.1: Obraz nočná hliadka.

Dobrým príkladom je obraz Nočná hliadka¹², kde sú na veľmi tmavom podklade namaľované dlhé, rovné kopije, muškety a vlajky v svetlých farbách, ktoré vytvárajú výrazné hrany na obrázkoch. Ďalším problémom spojeným so segmentáciou tohoto obrazu je skutočnosť, že je vložený do tmavého rámu a umiestený na tmavo zelenej stene. To spôsobuje, že hrany rámu nie sú na niektorých fotografiách dostatočne výrazné, Houghova transformácia ich nezarádi medzi výsledných kandidátov na segmentačné priamky a uprednostní výraznejšie hrany, napríklad práve dlhé oštepky na obraze.

Podsegmentácia bola vo väčšine prípadov spôsobená slabým kontrastom medzi rámom a obrazom a rámom a stenou spojeným s kontrastným rohom, alebo ozdobnou lištou obsiahnutou na fotografii.

Cieľom tejto časti práce bolo nájsť metódu, ktorá dosiahne na testovaných obrázkoch najväčší počet správne klasifikovaných obrázkov a zároveň sa čo najviac vyhne nadsegmentovaniu. Sto obrázkov z databázy Fotografií turistov bolo segmentovaných pomocou štyroch rôznych metód a podľa výslednej segmentácie boli za-

¹²Nočná hliadka (De Nachtwacht) je obraz - olej na plátne autora Rembrandta Harmenszoon van Rijna z roku 1642. Veľkosť obrazu je 363 na 438 cm. Obraz patrí do stálej expozície Rijksmusea v Amsterdame

Metóda	Gauss gradient	Anisotr. diffusion	Watershed	Kuwahara filter
Správna segmentácia	73%	89%	49%	89%
Nadsegmentované	6%	3%	1%	6%
Podsegmentované	21%	8%	50%	5%

Tabuľka 4.1: Percentuálne zastúpenie správne segmentovaných, nadsegmentovaných a podsegmentovaných obrázkov v segmentácii vytvorenej metódami Gauss gradient, Anisotropic diffusion, Watershed a Kuwahara filter.

radené do je/-dnej z troch skupín: správna segmentácia, nadsegmentácia a podsegmentácia. Výsledky tejto validácie môžeme vidieť v tabuľke 4.1. Pri výbere najlepšej metódy boli teda najskôr zvolené metódy s navyšším počtom správne klasifikovaných obrázkov, teda Anisotropic diffusion (3.2.2) a Kuwahara filter (3.2.4). Keďže metóda Anisotropic diffusion spôsobila menej nadsegmentovaných obrázkov, bola zvolená ako najvhodnejšia metóda na tento druh segmentácie.

4.2 Klasifikácia

Metódy použité pri klasifikácii obrázkov boli validované na sto obrázkoch z databázy Fotografií turistov spracovaných v segmentačnej fáze a pätnástich z databázy Originálov obrazov. Pre každý zo sto obrázkov bol určený jeden z originálov s najväčším počtom zhôd medzi DF súbormi. Keďže sa v databáze Fotografií nachádzajú aj obrazy, ktoré nie sú zastúpené v databáze Originálov, bol po rozbere výsledkov určený **prah** na oddelenie týchto obrazov do triedy číslo šestnásť (nenachádzajúci sa v databáze). Prah bol určený z výsledkov tak, aby maximalizoval počet správne klasifikovaných obrazov. Tabuľka 4.2 obsahuje percentuálne vyjadrený počet správne klasifikovaných obrázkov, pri nastavení rôznych hodnôt prahu. X% správne klasifikovaných obrazov znamená, že na x obrázkoch zo sto bola správne identifikovaná maľba a v textovom okne aplikácie sa objavil správny názov obrazu. Na 100-x obrázkoch teda bola buď identifikovaná nesprávna maľba, alebo žiadna maľba z databázy.

hodnota prahu =	0	6	8	10	12	14
SIFT	75%	88%	89%	89%	90%	90%
SURF	73%	90%	88%	83%	82%	79%

Tabuľka 4.2: Percentuálne zastúpenie správne klasifikovaných obrazov metódami SIFT a SURF pri rôznych prahoch.

metóda	SIFT	SURF
nesprávna maľba	5%	7%
žiadna maľba	5%	3%

Tabuľka 4.3: Percentuálne zastúpenie zle identifikovaných a neidentifikovaných obrazov na fotografiách metódami SIFT a SURF pri thresholde 12, 13, 14 pre SIFT a 6 pre SURF.

Z tabuľky 4.2 je zrejmé že SIFT dosahuje najlepšie výsledky pri hodnote prahu 12,13,14 a SURF pri hodnote 6. Tieto techniky teda v najlepšom prípade správne identifikovali maľby na deväťdesiatich fotografiách. Tabuľka 4.3 ukazuje koľko z desiatich zle klasifikovaných obrazov bolo identifikovaných ako nesprávne obrazy a koľko ako obrazy neobsiahnuté v databáze.

Hoci pri technike SIFT bolo viac malieb identifikovaných ako nenachádzajúcich sa v databáze, čo je kvôli užívateľom prijateľnejšie, bolo použité aj ďalšie kritérium na určenie najvhodnejšej metódy. Je ním rýchlosť metódy - v tomto prípade čas potrebný na vytvorenie jedného DF súboru. Hodnoty v tabuľke 4.4 ukazujú, že metóda SURF je približne dva a pol krát rýchlejšia ako metóda SIFT. Pri rovnakých výsledkoch dosiahnutých pri správnej klasifikácii tento údaj naznačuje, že metóda SURF je napriek vyššiemu počtu nesprávne klasifikovaných obrazov identifikovaných ako odlišné obrazy výhodnejšia na použitie pri tomto druhu klasifikácie.

Okrem voľby vhodnej klasifikačnej metódy nám validačné techniky môžu ponúknuť informácie o klasifikácii vzhľadom na jednotlivé triedy. V práci bolo využitých šesťnásť tried, jedna pre každý z obrazov z databázy Originálov (pre počet obrazov prislúchajúcich do jednotlivých tried pozri tabuľku 2.1) a jedna trieda pre obrazy mimo

Metóda	SIFT	SURF
čas	0,8125 s	0,32025 s

Tabuľka 4.4: Čas spotrebovaný metódami SIFT a SURF na vytvorenie jedného DF súboru z obrázku.

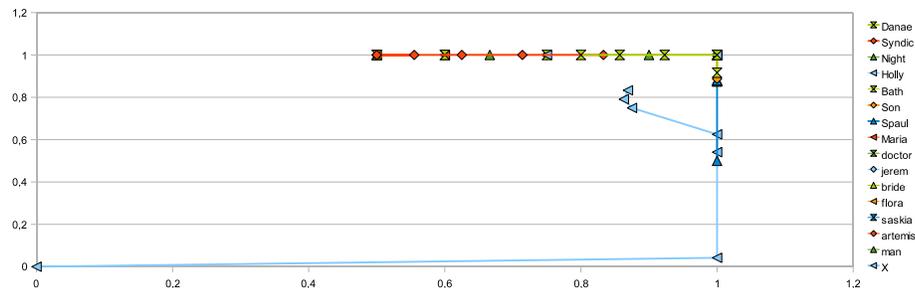
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
14	0	1	0	0	0	0	0	0	0	0	0	0	0	7	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
16	3	0	1	0	0	6	2	0	2	1	3	3	1	0	2	0

Tabuľka 4.5: Matica zámen pre metódu SIFT a prah rovný nule.

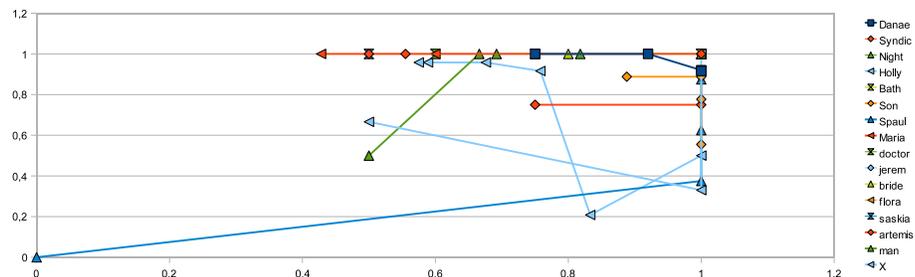
databázy. Prvým krokom v tomto procese je vytvorenie matice zámen (confusion matrix). Matica zámen má veľkosť $n \times n$, kde n je počet tried a je vytvorená pre každú metódu a každý prah (2 metódy \times 6 prahov = 12 matíc). Riadky reprezentujú triedy, do ktorých obrazy skutočne patria a stĺpce reprezentujú triedy, kam boli obrazy klasifikáciou zaradené. Ak sa teda obraz nachádza v prvom riadku a prvom stĺpci, znamená to, že je to obraz Danae a bol danou metódou s daným prahom správne klasifikovaný. Ak sa obraz nachádza napríklad v druhom riadku a treťom stĺpci znamená to, že ide o obraz Nočná hliadka nesprávne klasifikovaný ako Artemis. Tabuľka 4.5 reprezentuje maticu zámen pre metódu SIFT a prah rovný nule.

Po skonštruovaní matíc zámen z nich môžeme odčítať pre každú triedu hodnoty recall a precision, kde recall je počet správne klasifikovaných obrazov vydelený počtom všetkých obrazov v danej triede a precision je počet správne klasifikovaných

obrazov vydelený počtom všetkých obrazov klasifikovaných ako daná trieda. Graf vzťahu precision a recall (obr. 4.2 pre SIFT a obr. 4.3 pre SURF) nám ukazuje úspešnosť jednotlivých prahov vzhľadom na dané triedy.



Obr. 4.2: Graf vzťahu precision (x-ova os) a recall (y-ova os) pre metódu SIFT. Značky na krivkách tried sú hodnoty pri jednotlivých prahoch.



Obr. 4.3: Graf vzťahu precision (x-ova os) a recall (y-ova os) pre metódu SURF. Značky na krivkách tried sú hodnoty pri jednotlivých prahoch.

4.2.1 Pomer strán

V časti 3.3.3 bola spomínaná doplnková technika klasifikácie - porovnávanie na základe pomeru strán obrazu. Ak je pomer strán jediným príznakom pri klasifikácii obrazu, potom je segmentovaný obraz klasifikovaný ako originál s najbližšou hodnotou pomeru strán. Táto klasifikačná technika bola validovaná na päťdesiatich vysegmentovaných obrazoch (boli použité len obrazy, ktoré zodpovedajú jednému z pätnástich

obrazov databázy Originálov) a dosiahla úspešnosť dvadsaťdva percent. Tento výsledok vypovedá o tom, že pomer strán nie je sám o sebe dostatočne diskriminatívny na správnu klasifikáciu obrazov. Problémom tejto techniky sú obrazy, ktoré nie sú reprezentované žiadnym originálom v databáze, z čoho vyplýva, že ich táto technika vždy priradí k nesprávnemu originálu.

Využitie pomeru strán na zrýchlenie porovnávania DF súborov navrhnuté v časti 3.3.3 vyžaduje určenie hodnoty **vzdialenosť pomerov**, ktorá určí, ktoré DF súbory sa budú porovnávať a ktoré nie. Hodnoty pomerov sa pre originály pohybovali medzi 0,715 a 1,476 a pre vysegmentované obrazy medzi 0,557 a 1,703. Podľa hodnôt rozdielov pomerov päťdesiatich validovaných obrázkov od korešpondujúcich pomerov originálov bola určená hodnota **vzdialenosti pomerov** na 0,5. Ohraničenie touto hodnotou zrýchli porovnávanie pri osemdesiatich percentách obrázkov (bez zapojenia techniky orezávania prehľadávania). Ďalší problém pri využívaní príznaku pomeru strán môže nastať, ak užívateľ odfotografuje obraz na výšku a neskôr nezmení orientáciu fotografie, čo spôsobí že výška a šírka budú mať vymenené hodnoty a pomer strán bude mať zlú hodnotu. Systém navrhnutý v tejto práci nevyžaduje na vstupe správne orientovanú fotografiu, keďže metódy SIFT a SURF sú invariantné k zmene orientácie.

4.2.2 Ďalšie metódy (ColorSIFT a HueSIFT)

Okrem metód SIFT a SURF, spomínaných v predchádzajúcej časti, boli počas hľadania vhodných klasifikačných metód otestované aj dve ďalšie metódy, menovite HueSIFT a ColorSIFT.

Metóda ColorSIFT bola vytvorená G. J. Burghoutsom a J. M. Geusebroekom [9]. ColorSIFT je modifikáciou metódy SIFT, ktorá pracuje nad farebnými obrázkami a ich hodnotami v RGB a pri tvorbe histogramu orientácií využíva hodnotu farebného gradientu. Na validáciu tejto metódy bolo použitých 10 obrázkov z databázy Fotografií turistov, patriacich do tried Danae, Predstavenstvo súkenníckeho cechu a Nočná hliadka a 3 obrázky z databázy Originálov, reprezentujúce rovnaké triedy. Na otestovanie metódy bola použitá spustiteľná implementácia autorov. Validácia ukázala, že v porovnaní s metódou SIFT, nie je ColorSIFT vhodný na klasifikáciu

tohto druhu dát. Pri 10 obrázkoch našiel ColorSIFT v priemere len desať percent z počtu zhôd nájdených pomocou metódy SIFT a v jednom prípade nenašiel žiadne zhody medzi obrázkom a korešpondujúcim originálom.

Metóda HueSIFT bola navrhnutá a testovaná kvôli domnienke, že odtieň je jedným z najdôležitejších aspektov obrazu a hodnoty odtieňa (Hue) by mali teda reprezentovať maľbu robustnejšie ako hodnoty intenzity. Metóda HueSift spočíva vo využití šedotónového obrázku reprezentujúceho hodnoty Hue (namiesto intenzity) ako vstupu do metódy SIFT. Na získanie hodnoty Hue bolo treba previesť obrázok z farebného modelu RGB do HSL. Táto metóda sa ukázala ako nevhodná, čo bolo spôsobené najmä použitím rôznych fotoaparátov s rôznymi farebnými snímačmi, ktoré zapríčinili rôzne reprezentácie RGB a teda aj HSL rovnakých farieb.

Záver

Táto diplomová práca splnila svoj cieľ. V práci sa podarilo vytvoriť systém na klasifikáciu obrazov na fotografiách. Boli formulované metódy na segmentáciu obrazov z fotografií a na klasifikáciu vysegmentovaných obrazov. Vo validačnej fáze bola ako najlepšia segmentačná metóda zvolená Anisotropic diffusion s 89 percentnou úspešnosťou pri segmentácii. Validácia ukázala, že obe klasifikačné techniky dosahujú 90 percentnú úspešnosť pri klasifikácii. Ako najvhodnejšia bola zvolená metóda SURF, ktorá pracuje 2,5 krát rýchlejšie ako SIFT.

Systém navrhnutý v tejto práci by implementovaný vo vhodnom programovacom jazyku predstavoval skvelé rozšírenie umeleckých galérií na internete. Pri tvorbe takéhoto nástroja sa spracuje najskôr databáza obrázkov z galérie a vytvorí sa databáza deskriptorov. Pre každý dotaz na klasifikáciu obrazu na fotografii, by bol jej deskriptor vypočítaný a následne porovnaný s databázou.

Literatúra

- [1] T. H. A. Gruen. Image- and model-based reconstruction of the small buddha of bamiyan, afghanistan. In *Proc. of the The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 37, pages 963–968, 2008.
- [2] H. Ackermann. The basle cabinets of art and curiosities in the sixteenth and seventeenth centuries. In *The Origins of Museums: The cabinet of curiosities in sixteenth- and seventeenth-century Europe*, 2002.
- [3] AG.com. Travel webshots., 2009. <http://travel.webshots.com>.
- [4] J. F. Asmus. Digital image processing in art conservation. *BYTE*, 12(3):151–165, 1987.
- [5] J. Assfalg, A. Del Bimbo, and P. Pala. Three-dimensional interfaces for querying by example in content-based image retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 8(4):305–318, 2002.
- [6] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359, 2008.
- [7] I. Bereznoy, E. Postma, and J. van den Herik. Computer analysis of van gogh’s complementary colours. *Pattern Recogn. Lett.*, 28(6):703–709, 2007.
- [8] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRAPH ’00: Proceedings of the 27th annual conference on Computer graph-*

- ics and interactive techniques*, pages 417–424, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
- [9] G. J. Burghouts and J.-M. Geusebroek. Performance evaluation of local colour invariants. *Comput. Vis. Image Underst.*, 113(1):48–62, 2009.
- [10] J. F. Canny. A computational approach to edge detection. *Readings in computer vision: issues, problems, principles, and paradigms*, pages 184–203, 1987.
- [11] D. W. Carr and M. Leonard. *Looking at paintings : a guide to technical terms*. J. Paul Getty Museum in association with British Museum Press, Malibu, Calif., USA, 1992.
- [12] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8):1026–1038, 2002.
- [13] O. Casazza and P. Franchi. Digital image processing techniques for preservation and restoration of artworks. In *Scientific methodologies applied to works of art. Proceedings of the symposium*, pages 116–118, 1998.
- [14] A. Chalechale, G. Naghdy, and A. Mertins. Sketch-based image matching using angular partitioning. *IEEE Transactions on Systems, Man and Cybernetics*, 35(1):28–41, 2005.
- [15] C. Chang, M. Etezadi-Amoli, and M. Hewlett. A day at the museum., 2009. <http://www.stanford.edu/class/ee368/Project07/reports/ee368group06.pdf>.
- [16] J. Clottes. Heilbrunn timeline of art history: Chauvet cave (ca. 30,000 b.c.), október 2002. http://www.metmuseum.org/toah/hd/chav/hd_chav.htm.
- [17] F. Cutzu, R. Hammoud, and A. Leykin. Distinguishing paintings from photographs. *Comput. Vis. Image Underst.*, 100(3):249–273, 2005.
- [18] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):1–60, 2008.

- [19] A. Del Bimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(2):121–132, 1997.
- [20] R. Fagin. Combining fuzzy information from multiple systems (extended abstract). In *PODS '96: Proceedings of the fifteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 216–226, New York, NY, USA, 1996. ACM.
- [21] G. D. Finalyson. Color in perspective. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):1034–1038, 1996.
- [22] M. Ftacnik. Prednasky., 2009. <http://www.sccg.sk/~ftacnik/imageproc.htm>.
- [23] Google. Google image labeler., november 2009. <http://www.google.com/imagelabeler/>.
- [24] Google. Google milestones., november 2009. <http://www.google.com/corporate/history.html>.
- [25] Google. Google scholar., november 2009. <http://scholar.google.com/>.
- [26] G. Guidi, B. Frischer, M. D. Simone, A. Cioci, A. Spinetti, L. Carosso, L. L. Micoli, M. Russo, and T. Grasso. Virtualizing ancient rome: 3d acquisition and modeling of a large plaster-of-paris model of imperial rome. volume 5665, pages 119–133. SPIE, 2005.
- [27] gwap.com. Esp game., november 2009. <http://www.espgame.org/gwap/gamesPreview/espgame/>.
- [28] A. Hanbury, P. Kammerer, and E. Zolda. Painting crack elimination using viscous morphological reconstruction. In *ICIAP '03: Proceedings of the 12th International Conference on Image Analysis and Processing*, page 226, Washington, DC, USA, 2003. IEEE Computer Society.
- [29] P. Hough. Method and means for recognizing complex patterns, 1962. U.S. Patent 3,069,654.

- [30] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recogn.*, 24(12):1167–1186, 1991.
- [31] S. Jiang, Q. Huang, Q. Ye, and W. Gao. An effective method to detect and categorize digitized traditional chinese paintings. *Pattern Recogn. Lett.*, 27(7):734–746, 2006.
- [32] Y. Jing and S. Baluja. Pagerank for product image search. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 307–316, New York, NY, USA, 2008. ACM.
- [33] R. Kirsh. Computer determination of the constituent structure of biomedical images. *Computers and Biomedical Research*, 4(3):315–328, 1971.
- [34] M. Kuwahara. Processing of ri-angiocardigraphic images. in *Digital Processing of Biomedical Images*, pages 187–203, 1976.
- [35] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project: 3d scanning of large statues. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 131–144, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
- [36] T. E. Lombardi. *The classification of style in fine-art painting*. PhD thesis, New York, NY, USA, 2005. Pace University.
- [37] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2002.
- [38] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):837–842, 1996.
- [39] O. Mataev. Olga's gallery., 2009. <http://www.abcgallery.com/>.
- [40] MathWorks. Documentation - image processing toolbox, 2010. <http://www.mathworks.com/access/helpdesk/help/toolbox/images>.

- [41] T. Melzer, P. Kammerer, and E. Zolda. Stroke detection of brush strokes in portrait miniatures using a semi-parametric and a model based approach. In *ICPR '98: Proceedings of the 14th International Conference on Pattern Recognition-Volume 1*, 1998.
- [42] C. Moler. Origins of matlab., 2004. http://www.mathworks.com/company/newsletters/news_notes/clevescorner/dec04.html.
- [43] netropolis. ephoto.sk – portál o fotografování a o všem, čo s fotografiou súvisí, november 2009. <http://www.ephoto.sk/>.
- [44] M. Pappas and I. Pitas. Digital color restoration of old paintings. *IEEE Transactions on Image Processing*, 9:291–294, Feb. 2000.
- [45] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. Technical report, Berkeley, CA, USA, 1988.
- [46] R. Sablatnig, P. Kammerer, and E. Zolda. Hierarchical classification of paintings using face - and brush stroke models. In *in Proc. of 14th Int. Conf. on Pattern Recognition*, pages 172–174, 1998.
- [47] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(5):530–535, 1997.
- [48] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.
- [49] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.
- [50] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.
- [51] M. J. Swain. *Color indexing*. PhD thesis, 1990. Supervisor-Ballard, Dana H.

- [52] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):831–836, 1996.
- [53] R. P. Taylor, R. Guzman, T. P. Martin, G. D. R. Hall, A. P. Micolich, D. Jonas, B. C. Scannell, M. S. Fairbanks, and C. A. Marlow. Authenticating pollock paintings using fractal geometry. *Pattern Recogn. Lett.*, 28(6):695–702, 2007.
- [54] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(6):583–598, 1991.
- [55] vláda SR. Analýza súčasného stavu a postavenia múzeí a galérií na slovensku, 20 december 2006. <http://www.rokovania.sk/appl/material.nsf/0/6975D036FCF212ADC1257244003A9525?OpenDocument>.
- [56] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, NY, USA, 2004. ACM.
- [57] J. Z. Wang, J. Li, R. M. Gray, and G. Wiederhold. Unsupervised multiresolution segmentation for images with low depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(1):85–90, 2001.
- [58] R. C. Wilson and E. R. Hancock. Structural matching by discrete relaxation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(6):634–648, 1997.
- [59] H. J. Wolfson and I. Rigoutsos. Geometric hashing: An overview. *IEEE Comput. Sci. Eng.*, 4(4):10–21, 1997.
- [60] Y. Wu, Q. Tian, and T. S. Huang. Discriminant-em algorithm with application to image retrieval. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:1222, 2000.

- [61] G. Xiong. Gradient using first order derivative of gaussian., 2009. <http://www.mathworks.com/matlabcentral/fileexchange/-8060-gradient-using-first-order-derivative-of-gaussian>.
- [62] X. S. Zhou and T. S. Huang. Cbir: from low-level features to high-level semantics. volume 3974, pages 426–431. SPIE, 2000.