

Západočeská univerzita v Plzni

Bakalářská práce

Klasifikace obrazů v rozsáhlých databázích

Autor:
Adam Fara

Vedoucí práce:
Ing. Pavel Campr

Prohlášení

Předkládám tímto k posouzení a obhajobě bakalářskou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

V Plzni dne 15. května 2013

Podpis: _____

Poděkování

Děkuji za cenné rady při vytváření práce a odborný dohled Ing. Pavlu Camprovi.

Děkuji MetaCentru za přístup k výpočetnímu prostředí, ve kterém proběhla většina výpočtů této práce. MetaCentrum je součástí programu "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005).

Abstrakt

Práce je zaměřena na rozpoznávání obrazu v rozsáhlých archivech, konkrétně na rozpoznávání typu scén, které se v obraze nacházejí. Budu se zabývat příznaky vhodnými pro popis obrazů a klasifikátory, které budou příznakové vektory vyhodnocovat a zařazovat obrazy do tříd.

Cílem práce je nalézt vhodné kombinace příznakových vektorů a klasifikátorů pro konkrétní datasety obrazů. Výstupem práce budou statistiky úspěšnosti různých kombinací použitých klasifikátorů a příznakových vektorů.

Využití programu by mohlo být pro automatické rozpoznávání obsahu ve videích a jejich indexaci. To by umožnilo snadné a rychlé vyhledávání požadovaných scén podle klíčových slov.

Klíčová slova: klasifikace scén, zpracování digitalizovaného obrazu, extrakce příznaků z obrazu

Abstract

This work is focused on image classification in large image archives, especially on classifying scene type. I deal with features appropriate for image description and classifiers, which evaluate features and classify images into classes.

The aim is to find best performing combinations of features and classifiers for specific datasets of images. Outcome of this work are statistics of success for various combinations of classifiers and features.

Program might be used for automatic content classification in videos and its indexing. It would facilitate fast and easy searching of required scenes by keywords.

Keywords: scene classification, digital image processing, feature extraction from image

Obsah

1	Úvod	1
2	Motivace	2
3	Analýza obrazu	4
3.1	Histogram of Oriented Gradients (HOG)	4
3.2	K-means HOG	5
3.3	GIST	6
3.4	Hue, Saturation, Value (HSV)	6
3.5	Local Binary Patterns (LBP)	7
4	Klasifikátory	8
4.1	Použité klasifikátory	8
4.1.1	Decision Tree Classification (DTC)	8
4.1.2	Gaussian Naive Bayes (GNB)	8
4.1.3	K-Nearest Nighbors (KNN)	8
4.1.4	Linear Discriminant Analysis (LDA)	8
4.1.5	Random Forest Classifier (RFC)	9
4.1.6	Support Vector Machines (SVM)	9
4.2	Způsob vyhodnocení	9
5	Experimentální data	11
5.1	Zprávy	11
5.2	SUN	12
6	Experimenty	14
6.1	Vyhodnocení	14
7	Závěr	31

1 Úvod

V práci se zabývám automatickým rozpoznáním obrazu a jejich rozdělením do tříd, konkrétně rozpoznáváním typu scén, které se v obraze nacházejí. Rozpoznání obrazu funguje tak, že se z obrazu nejprve spočtou příznaky, z těch jsou natrénovány klasifikátory a natrénovaný klasifikátor poté dokáže zařadit do třídy i obraz, který dříve neměl k dispozici. Tato problematika byla popsána například v publikacích [1][2][3].

Používám výhradně systémy učení s učitelem. To jsou typy klasifikátorů, kde pro vstupní množinu trénovacích dat znám jejich příslušnost ke třídě, do které mají být zařazeny. Klasifikátor se naučí rozpoznávat příznaky na základě příslušnosti obrazů ke konkrétním třídám.

To je hlavní rozdíl oproti systémům učení bez učitele, kde nejprve proběhne proces nalezení počtu clusterů (shluků) a jejich umístění v prostoru, a teprve na základě umístění shluků proběhne natrénování klasifikátoru.

V oblasti strojového učení se běžně používá výraz "obraz" pro označení vektorů popisujících nějakou věc. V případě této práce jsou popisovanou věcí také obrazy (snímky). Pro účely této práce je tedy nutné vymezit přesněji pojmy:

- Obraz - Původní snímek, který vstupuje do programu a jsou z něj vypočteny příznakové vektory.
- Příznakový vektor - Data vypočtená z jednotlivých obrazů, která jsou rozdělena na testovací a trénovací množinu a vstupují do klasifikátorů pro jejich natrénování a vyhodnocení úspěšnosti klasifikace.

Obrazy jsou rozdělené do velkých tříd (datasetů). Příkladem mohou být například obrazy získané z televizních zpráv. Každý dataset je pak navíc rozdělen do podtříd, do kterých se snažím obrazy zaklasifikovat. V případě datasetu "zprávy" mohou být jeho podtřídy "grafika", "indoor", "outdoor", "studio".

V práci se budu zabývat nejprve postupy, které jsou běžně používány pro rozpoznávání. V kapitolách "experimenty" a "vyhodnocení" se soustředím na svou vlastní práci a její výsledky.

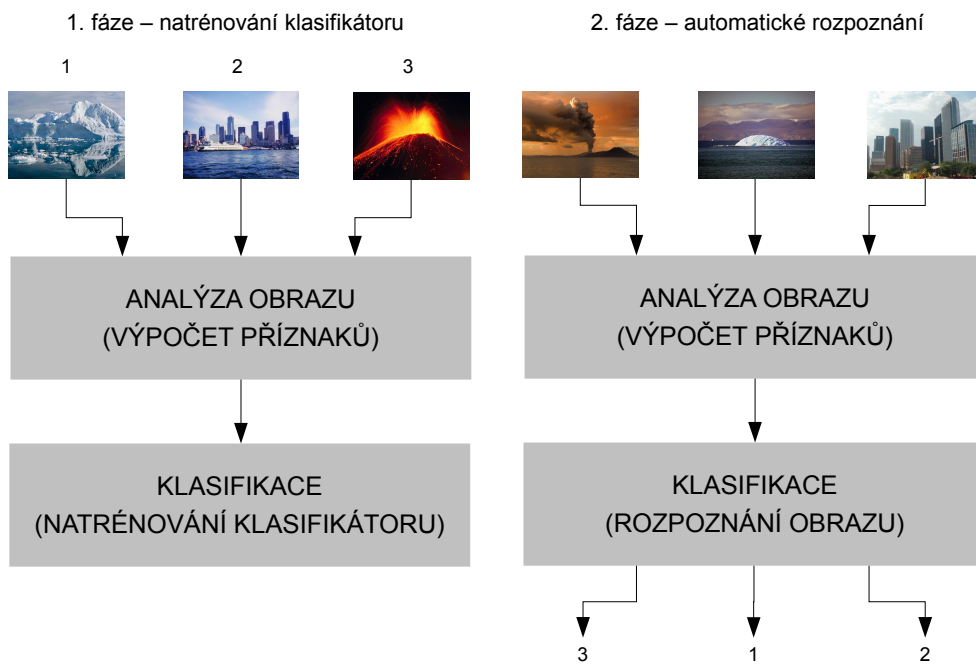
2 Motivace

Počítačové vidění (computer vision) úzce souvisí se strojovým učením (machine learning). Rozpoznání obrazů se může zaměřit na rozpoznání scény jako celku. To je také předmětem této práce. Klasifikace může jít do většího detailu, vyhledávat ve scéně konkrétní předměty nebo rozeznávat obličeje.

Rozpoznání scén jako celku lze využít například pro indexování videí tak, aby bylo možné vyhledávat ve videoarchivech pomocí klíčových slov. Ve videích by bylo vhodné nejprve vyhledat střihy. Pokud by se v rámci jednoho střihu neměnila scéna, jako je tomu například v televizních zprávách, stačilo by klasifikovat v každém střihu několik snímků a podle většinového rozhodnutí klasifikátorů zaklasifikovat i celou pasáž. Pokud by se v rámci střihu scény měnily, bylo by vhodné každý střih ještě rozdělit na menší části a klasifikovat každou zvlášť.

Rozpoznání předmětů v obrazech může být použito například v autonomních automobilech pro rozpoznání okolních předmětů, dopravních značek a podobně.

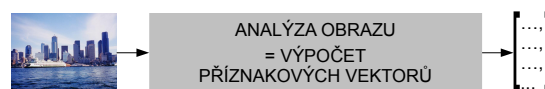
Princip automatického rozpoznávání obrazu spočívá ve dvou fázích. V první fázi jsou vybrány známé obrazy a je známá i jejich příslušnost ke třídě. Na těchto obrazech jsou natrénovány klasifikátory. Ve druhé fázi vstupují do natrénovaných klasifikátorů neznámé obrazy a probíhá automatické zařazení do třídy obrazů.



Obrázek 2.1: Princip automatického rozpoznání obrazu. Výstupem klasifikace je číslo třídy příslušného obrázku.

3 Analýza obrazu

Analýzou obrazu se rozumí aplikace algoritmů, které vypočtou příznakové vektory obrazu.



Příznakový vektor je uspořádaný vektor čísel charakterizujících obraz na základě jeho jasů, rozložení barev, rozložení hran a jejich směrů a podobně.

Předpokladem pro kvalitní klasifikaci je, že obrazy stejných tříd budou vykazovat podobnost v příznakových vektorech stejného typu. Vhodným nastavením parametrů pro výpočet příznakových vektorů se zvyšuje přesnost klasifikace. Dalším způsobem, jak zvýšit přesnost klasifikace, je kombinace nejúspěšnějších vektorů.

Například jeden typ výpočtu poskytne vektor dimenze 2 s hodnotami [1,2]. Druhý typ výpočtu poskytne vektor dimenze 3 s hodnotami [3,4,5]. Sloučením vektorů získám jeden vektor dimenze 5 s hodnotami [1,2,3,4,5].

Pro výpočet příznakových vektorů jsem použil pět metod. Každý typ příznaků vykazuje lepší výsledky při klasifikaci pro různé typy scén. Každý z příznaků má navíc volné parametry, které mění jeho způsob výpočtu, a tím i vlastnosti.

3.1 Histogram of Oriented Gradients (HOG)

Metoda [4] vychází z obrazu převedeného do stupňů šedi. Obraz je rozdělen do buněk. Velikost buněk je volitelným parametrem metody. Nad každou buňkou je vytvořena maska obrazu a posunuta o jeden pixel vpravo a o jeden pixel dolů. Odečtením hodnot jasů překrývajících se pixelů je spočten gradient jasů.

Tímto způsobem jsou určeny směry hran v každé buňce. Směry jsou rozděleny do histogramu. Počet intervalů histogramu je dalším volitelným parametrem metody.

V případě devíti intervalů je histogram rozdělen po dvaceti stupních. Metoda uvažuje směry hran v rozsahu 0 - 180 stupňů.

V metodě lze volitelně zapnout částečné překrytí buněk, čímž se zvětší popis obrazu. V tom případě jsou buňky sloučeny do větších bloků (například 2 x 2) a překryv bloků se provede posunutím bloku o šířku buňky sousedícího bloku.

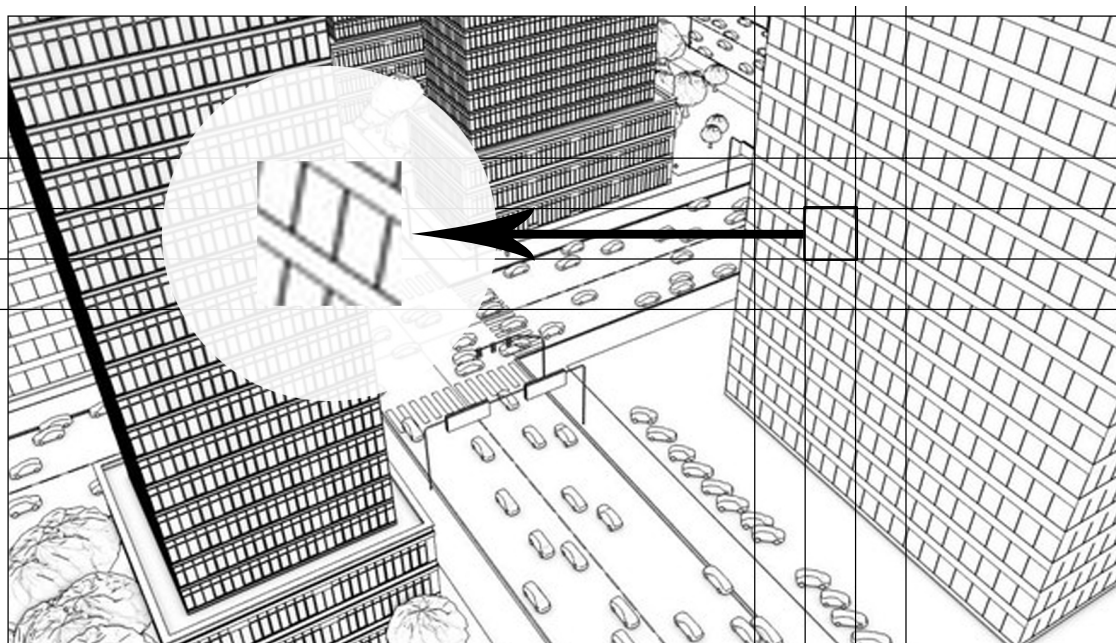
V práci je tato metoda označena jako *hog-a-b-c*, kde "a" označuje počet intervalů histogramu, "b" označuje šířku a výšku buňky, "c" označuje počet buněk v sloupcích a řádcích bloku. Pro výpočet jsem použil implementaci z knihovny Scikit - image [5].

3.2 K-means HOG

Jde o speciální typ výpočtu příznakových vektorů, protože vstupem jsou již příznakové vektory a nikoliv samotné obrazy.

Vstupem pro metodu [2] jsou již vypočtené příznakové vektory typu HOG. Jejím smyslem je vytvořit příznakový vektor, který bude robustní vzhledem k posunutí obrazu a bude popisovat strukturu obrazu pomocí histogramu.

V případě vstupních vektorů typu HOG s velikostí bloku 16x16 pixelů získám pro každý obraz vektor dimenze 36. Každý vektor popisuje konkrétní typ struktury. Budou v něm převyšovat například svislé čáry.

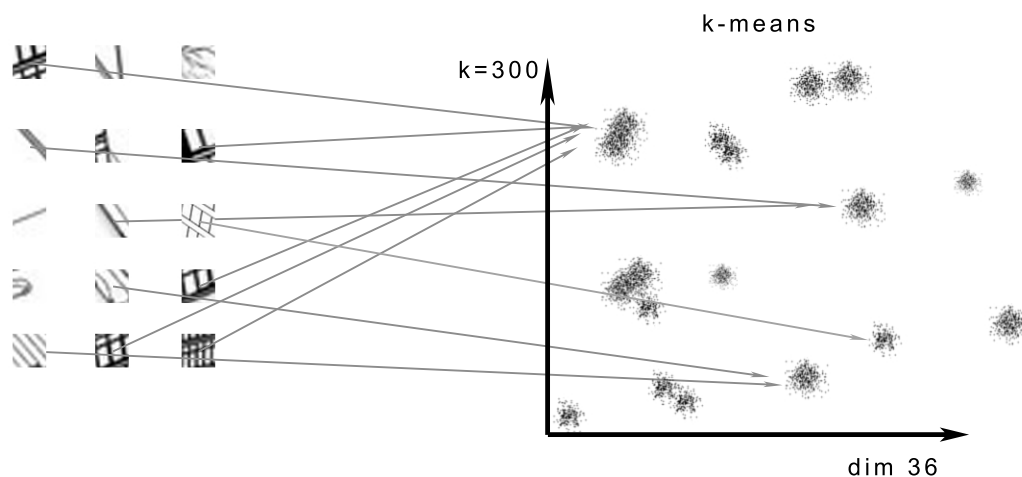


Obrázek 3.1: Ukázka struktury jedné buňky v obraze. (<http://www.turbosquid.com/3d-models/towers-collections-3d-model/587469>)

Z těchto vektorů celého jednoho datasetu je metodou k-means nalezeno 300 shluků. V prostoru shluků jsou tedy zastoupeny všechny charakteristické typy struktur. Výstupem je vektor dimenze 300.

Po vytvoření shluků, kdy jsou známy jejich středy, lze pro původní i nová data

použít klasifikátor "nearest neighbour" - "klasifikace podle nejbližšího souseda", kdy je příznakový vektor zařazen do třídy shluků, kterému je nejbliže.



Obrázek 3.2: Princip rozřídění struktur v obraze do k-means klasifikátoru.

3.3 GIST

Podobně jako v metodě HOG vypočítává i metoda GIST [6] gradienty jasu v obraze. Navíc používá při analýze obrazu až 24 filtrů a čtyři zvětšení. Tím dosahuje informativnějšího popisu obrazu. Pro výpočet příznaků GIST jsem použil implementaci programu Matlab.

3.4 Hue, Saturation, Value (HSV)

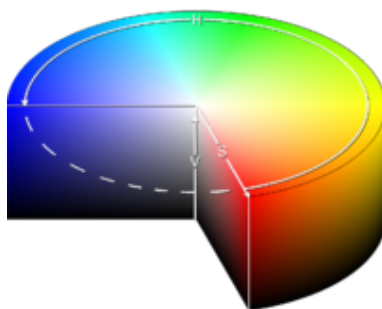
Metoda vychází z barevného obrazu převedeného do barevného spektra HSV. Obraz se rozdělí na M sloupců a N řádků a vznikne tak $M \times N$ buněk. Nejlepších výsledků dosahuje HSV vektor při rozdělení obrazu na tři sloupce a dva řádky, tedy šest buněk. Výstupem je v tomto případě vektor dimenze 96.

Hodnota H je odstín barvy v rozsahu 0 - 360°. Hodnota S je sytost barvy v rozsahu 0 - 255. Hodnota V je intenzita barvy, tedy namíchání s černou barvou, v rozsahu 0 - 255.

Pro každou buňku je spočten histogram o šestnácti intervalech hodnoty H. Pro každý pixel se určí hodnota H a do histogramu je přičtena hodnota V.

Tento typ příznakových vektorů je původně určen pro rozdělení scén typu indoor a outdoor. Ve venkovních scénách se předpokládá výrazně jiné rozložení barev a jasu

díky obloze. Jeho výsledky se ale ukázaly jako dobré i pro jiné datasey.



Obrázek 3.3: Princip barevného spektra HSV. (<http://processing.org/learning/color/>)

3.5 Local Binary Patterns (LBP)

Tato metoda [7] se původně používala pro popis textur, ale úspěšně se používá i pro klasifikaci obrazů. Její výhodou je, že výstupem je histogram spočtený přes celý obraz a díky tomu je příznakový vektor nezávislý na posunutí obrazu, a je tak více obecný než například HOG. Metoda pracuje s obrazem převedeným na stupně šedi. Projde postupně všechny pixely vyjma krajních a nalezne 8, resp. 16 bodů rovnoměrně rozložených na kružnici se zvoleným poloměrem. Každému z bodů je přiřazena hodnota 1, pokud je světlejší než středový pixel. V opačném případě je přiřazena hodnota 0. Pro každý pixel tak vznikne osmibitový vektor.

Volitelně se navíc může zapnout parametr "uniform", který umožní v osmiokolí maximálně dva přechody mezi 0 a 1. Tím se zmenší velikost výsledného histogramu. Obrázek se rozdělí do buněk defaultně o rozměru 16x16 pixelů a pro každou pozici v buňce je spočten histogram četnosti osmibitových vektorů z celého obrazu. Výstupem je tedy vektor dimenze 256. Pro výpočet jsem použil implementaci z knihovny Bob [8]. V práci je tato metoda označena jako *lbp_a_b*, kde "a" označuje poloměr kružnice a "b" aktivuje parametr uniform.

4 Klasifikátory

Klasifikátor je algoritmus, který se na zadaných datech nejprve natrénuje a poté dokáže klasifikovat neznámá data. Vstupními daty pro klasifikátory byly příznakové vektory vypočtené z obrazů.

V celé práci jsem používal výhradně klasifikátory s učitelem, tedy takové, kterým jsem spolu se vstupními daty dal i příslušnost obrazu ke správné třídě. Klasifikátory jsou zpravidla náročné algoritmy, uvádím proto jejich princip pouze zjednodušeně. Použil jsem klasifikátory z knihovny "scikit - learn" [9]. V dokumentaci této knihovny lze nalézt i detailní popis těchto klasifikátorů.

4.1 Použité klasifikátory

4.1.1 Decision Tree Classification (DTC)

Rozhodovací stromy porovnávají hodnoty prvků a vytvářejí tak v rozhodovacím stromu větve. Listy na koncích větví reprezentují třídu, do které má být obraz zařazen.

4.1.2 Gaussian Naive Bayes (GNB)

Algoritmus vytvoří pravděpodobnostní funkce, které charakterizují rozložení tříd v prostoru. Při klasifikaci používá bayesovu podmíněnou pravděpodobnost pro rozhodnutí, ke které třídě prvek náleží.

4.1.3 K-Nearest Nighbors (KNN)

Algoritmus nalezne K nejbližších sousedů v prostoru všech prvků. Poté vypočte průměrnou vzdálenost k jednotlivým třídám a jako správnou zvolí tu s nejnižší vzdáleností. Předpokládá se, že prvky stejné třídy budou v prostoru blízko u sebe a do jiných tříd se dostane jen několik jednotlivců.

4.1.4 Linear Discriminant Analysis (LDA)

Klasifikátor prokládá prostorem roviny rozdělující shluky prvků. Příslušnost nového prvku je určena podle toho, kam se zařadí v prostoru prvků.

4.1.5 Random Forest Classifier (RFC)

RFC je množina několika klasických rozhodovacích stromů, kdy každý ze stromů byl natrénován na náhodné podmnožině vstupních dat. Jako finální rozhodnutí klasifikace je vybráno to, pro které hlasovalo největší množství stromů.

4.1.6 Support Vector Machines (SVM)

Klasifikátor vytváří v prostoru dat nadroviny oddělující jednotlivé třídy od sebe. V práci jsem použil nelineární typ klasifikátoru SVM. Příslušnost nového prvku je určena podle toho, kam se zařadí v prostoru prvků. Na rozdíl od LDA mohou být rozdělovací nadroviny zakřivené.

4.2 Způsob vyhodnocení

Pro vyhodnocení výsledků jsem použil standardní statistické metody používané při predikci. Pro všechny záměny tříd uvažuji stejnou cenu chyby.

Confusion matrix (CM) - matice záměn

Matice záměn představuje kompletní přehled zařazení prvků do tříd. Je z ní dobře patrné, které třídy se při klasifikaci často zaměňují a které jsou naopak dobře oddělitelné. Je to proto silný nástroj, kterým se můžeme řídit například pro sloučení zaměňovaných tříd (pokud to experiment dovoluje) nebo pro ruční rozdělení často zaměňovaných tříd.

V sloupcích jsou běžně umístěny skutečné třídy a v řádcích pak predikované třídy. Predikce je tedy tím lepší, čím větší čísla (počet zařazení) dostaneme na diagonále. Ostatní hodnoty by se ideálně měly blížit nule. Matici záměn je vhodné zobrazovat také graficky.

skutečná třída	0	0	1	1
predikovaná třída	0	1	0	1
	tp true positive skutečně pozitivní	fn false negative falešně negativní	tn true negative skutečně negativní	fp false positive falešně pozitivní

Accuracy - přesnost

Přesnost je základním ukazatelem, který vypovídá o kvalitě klasifikace. Neposkytuje ale žádné bližší informace o úspěšnosti klasifikace jednotlivých tříd.

$$accuracy = \frac{\{tp\} + \{tn\}}{\{tp\} + \{tn\} + \{fp\} + \{fn\}}$$

Precision, recall

Ukazatelé precision a recall poskytují lepší představu o kvalitě klasifikace než pouhá úspěšnost. Precision (přesnost) stanovuje pravděpodobnost, že predikovaná třída bude správná. Recall (úplnost) stanovuje pravděpodobnost, že správná třída bude zvolena klasifikátorem.

Při špatném způsobu klasifikace může jeden ukazatel výrazně převyšovat druhý. Je proto dobré sledovat, zda jsou hodnoty ukazatelů podobné. Zároveň by při dobré klasifikaci měly být co nejvyšší.

$$precision = \frac{\{tp\}}{\{tp\} + \{fp\}}$$

$$recall = \frac{\{tp\}}{\{tp\} + \{fn\}}$$

F₁ score

F₁ score se používá pro porovnávání úspěšností různých klasifikátorů na stejných datech. Výpočet se provádí z čísel *precision* a *recall* a díky tomu poskytuje kvalitnější přehled než *accuracy*.

$$F_1 score = \frac{precision \cdot recall}{precision + recall}$$

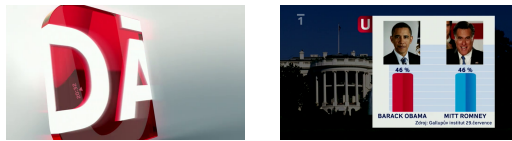
5 Experimentální data

Pro vlastní experimenty jsem použil dva datasety, které jsem dále rozdělil do tříd. Většina experimentů byla provedena na obrazech z datasetu SUN, který je velmi obsáhlý. To umožnilo otestovat klasifikaci do různých počtů tříd.

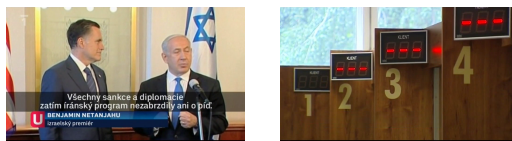
5.1 Zprávy

Jde o dataset, který jsem si vytvořil pro účely této práce. Pro experimenty jsem použil videa z televizních zpráv. Každé čtyři vteřiny jsem uložil obraz. Pro zvětšení datasetu a zároveň pro větší robustnost klasifikátorů jsem všechny obrazy uložil navíc horizontálně převrácené. Obrazy jsem rozdělil do čtyř tříd:

- grafika (202 obrazů)



- indoor (953 obrazů)



- outdoor (1 684 obrazů)



- studio (970 obrazů)



5.2 SUN

Databáze SUN [2] je databáze obrazů, která vzniká pro účely výzkumu počítačového vidění a rozpoznání obrazů. Od roku 2005 je databáze doplňována snímky dostupnými na internetu, které jsou roztrženy podle scén a podle konkrétních objektů. V době vzniku této práce obsahovala databáze 142 tisíc snímků rozdělených do více než 900 kategorií. Některé scény jsou v databázi zařazené do více podtříd. Takové snímky jsem při vytváření datasetů vynechal a použil jsem pouze snímky zařazené právě do jedné podtřídy.

indoor - outdoor (inout)

Dataset inout obsahuje 382 obrazů náhodně vybraných z vnitřních a venkovních scén.

indoor - outdoor - outdoor man made (in_out_outMM)

Dataset obsahuje 3 060 obrazů rozdělených do třech tříd. V každé podtřídě jsou zastoupeny obrazy z 51 scén.

- indoor - vnitřní scény
- outdoor - vnější přírodní scény
- outdoor man made - vnější scény vytvořené člověkem

15 classes

Dataset 15 vybraných scén. Scény pro tento dataset jsem vybíral tak, aby byly co nejvíce obsahově rozmanité. Dataset obsahuje 4 855 obrazů z těchto scén:

- indoor
 - airplane cabin
 - discotheque
 - power plant - indoor
 - sauna
 - subway station - platform



- outdoor

- badlands
- beach
- fjord
- glacier
- volcano



- outdoor man made

- crosswalk
- highway
- oilrig
- parking lot
- planetarium - outdoor



multiclass in

Dataset obsahuje všechny scény z databáze zařazené pouze v typu indoor. To je celkem 48 473 obrazů rozdělených do 346 scén.

multiclass out

Dataset obsahuje všechny scény z databáze zařazené pouze v typu outdoor. To je celkem 16 595 obrazů rozdělených do 88 scén.

multiclass out man made

Dataset obsahuje všechny scény z databáze zařazené pouze v typu outdoor man made. To je celkem 46 924 obrazů rozdělených do 362 scén.

6 Experimenty

Při přípravě experimentů jsem se soustředil na to, abych zaručil nezávislost obrazů. Všechny obrazy v datasetu jsou v průběhu výpočtu promíchány, abych vyloučil závislost jejich pořadí kvůli názvu. Dále je striktně oddělena trénovací a testovací množina obrazů.

Při trénování a vyhodnocování klasifikátorů jsem pro každou kombinaci *příznakové vektory - klasifikátor* použil pět různých namíchání trénovací a testovací množiny (cross validation). Výsledky jsou pak zprůměrované hodnoty ze všech pěti experimentů.

Pro všechny datasety jsem kombinoval všechny vypočtené typy příznaků se všemi použitými klasifikátory. Nakonec jsem zkombinoval spojením za sebe nejúspěšnější typy příznaků a znovu natrénoval klasifikátory. Tento způsob ještě zvyšuje úspěšnost klasifikace.

6.1 Vyhodnocení

V této části uvádím přehled úspěšností pro různé kombinace příznakových vektorů a klasifikátorů. Tabulka 6.1 ukazuje nejúspěšnější klasifikátory pro všechny datasety a použité příznaky. V tabulce je uvedena úspěšnost parametrem accuracy a příslušný klasifikátor, který poskytl nejlepší výsledek.

V tabulce 6.2 je přehled nejlepších klasifikací kombinovaných příznaků. Lze předpokládat, že sloučením nejúspěšnějších příznakových vektorů lze dosáhnout i nejlepší úspěšnosti kombinovaných příznaků. Toto pravidlo však neplatí vždy a příznakový vektor, který je sám o sobě velmi slabý, dokáže při sloučení s jinými příznaky významně vylepšit výsledek.

typ příznaků	dataset (počet tříd)						
	zprávy (4)	inout (2)	in_out _outMM (3)	15 classes (15)	multiclass in (408)	multiclass out (88)	multiclass outMM (362)
gist	81% KNN	77% SVM		67% SVM			
hog_9_8_1	73% GNB	66% LDA		51% LDA	10% GNB	15% GNB	5% LDA
hog_9_8_2	74% GNB	65% GNB	71% GNB	60% LDA		16% GNB	3% DTC
hog_9_16_2	73% GNB	71% GNB	71% GNB	60% LDA			11% GNB
hsv_3_2	73% KNN	76% KNN	56% RFC	46% LDA	5% RFC	16% LDA	6% LDA
hsv_6_4	74% KNN	77% GNB	56% RFC	45% LDA	5% RFC	16% LDA	6% LDA
k-means hog_9_8_2	81% KNN	74% RFC	71% GNB	60% LDA			
k-means hog_9_16_2	79% KNN	71% GNB	71% GNB	60% LDA	11% LDA	22% LDA	12% LDA
lbp_2_0	81% LDA	75% KNN	65% LDA	59% LDA		16% KNN	
lbp_2_1	78% LDA	75% GNB	66% LDA	56% LDA	11% LDA	23% LDA	11% LDA
lbp_4_0	80% LDA	78% GNB	64% LDA	57% LDA		25% LDA	
lbp_4_1	78% KNN	78% GNB	66% LDA	51% LDA		22% LDA	

Tabulka 6.1: Nejúspěšnější klasifikátory pro všechny datasety a použité příznaky. Prázdné kolonky nebyly spočteny kvůli paměťové a časové náročnosti nebo proto, že jsem od nich očekával horší výsledek.

kombinované příznaky	dataset (počet tříd)						
	zprávy (4)	inout (2)	in_out _outMM (3)	15 classes (15)	multiclass in (408)	multiclass out (88)	multiclass outMM (362)
k-means_hog_9_8_2 + lbp_2_0	86% LDA	78% LDA	71% GNB	67% LDA			
k-means_hog_9_8_2 + lbp_2_1	86% LDA	78% LDA	71% GNB	66% LDA			
k-means_hog_9_8_2 + lbp_4_0	86% LDA	75% LDA	71% GNB	67% LDA			
k-means_hog_9_8_2 + lbp_4_1	85% LDA	75% GNB	72% LDA	65% LDA			
k-means_hog_9_16_2 + lbp_2_0		77% LDA	70% GNB	68% LDA	18% LDA	31% LDA	18% LDA
k-means_hog_9_16_2 + lbp_2_1		77% LDA	73% LDA	67% LDA	16% LDA	28% LDA	15% LDA
k-means_hog_9_16_2 + lbp_4_0		73% GNB	72% GNB	66% LDA	19% LDA	31% LDA	18% LDA
k-means_hog_9_16_2 + lbp_4_1		74% LDA	72% LDA	65% LDA	15% LDA	27% LDA	15% LDA
k-means_hog_9_16_2 + lbp_2_1 + hsv_6_4		71% DTC	66% DTC	68% LDA		34% LDA	19% LDA
lbp_4_0 + hsv_6_4	81% LDA	79% GNB	62% RFC	60% LDA			

Tabulka 6.2: Kombinace příznakových vektorů a neúspěšnější klasifikátory.

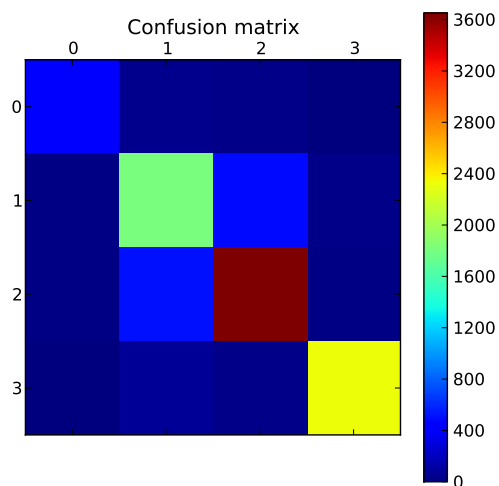
zprávy

typ příznaků	klasifikátor					
	DTC	GNB	KNN	LDA	RFC	SVM
gist	69%	76%	81%	79%	65%	44%
hog_9_8_1	73%	73%	68%	60%	55%	44%
hog_9_8_2	60%	74%	61%	60%	54%	44%
hog_9_16_2	61%	73%	67%	58%	56%	45%
hsv_3_2	65%	54%	73%	69%	64%	44%
hsv_6_4	64%	55%	74%	71%	62%	44%
hsv_6_6	64%	56%	74%	70%	62%	44%
k-means_hog_9_8_2	73%	69%	81%	60%	64%	44%
k-means_hog_9_16_2	64%	69%	79%	58%	56%	44%
lbp_2_0	69%	63%	78%	81%	68%	64%
lbp_2_1	71%	69%	77%	78%	72%	64%
lbp_4_0	71%	66%	80%	80%	68%	67%
lbp_4_1	70%	67%	78%	78%	71%	65%
k-means_hog_9_8_2 + lbp_2_0	72%	77%	81%	86%	69%	44%
k-means_hog_9_8_2 + lbp_2_1	72%	76%	81%	86%	67%	44%
k-means_hog_9_8_2 + lbp_4_0	73%	78%	81%	86%	69%	44%
k-means_hog_9_8_2 + lbp_4_1	72%	76%	81%	85%	67%	44%
lbp_4_0 + hsv_6_4	70%	69%	74%	81%	69%	44%

Tabulka 6.3: Přehled úspěšnosti všech použitých kombinací příznaků a klasifikátorů pro dataset "zprávy".

Na obrázku 6.1 je zobrazena matice záměn nejúspěšnější klasifikace datasetu "zprávy". Je z ní patrné, že největší záměna byla mezi třídami "indoor" a "outdoor".

V tabulce 6.4 je kompletní přehled úspěšnosti klasifikace jednotlivých tříd.



Obrázek 6.1: Matice záměn - zprávy, LDA, k-means_hog_9.8.2 + lbp_2.0

id	název	precision	recall	f1-score
0	grafika	0.89	0.84	0.86
1	indoor	0.74	0.77	0.75
2	outdoor	0.87	0.87	0.87
3	studio	0.98	0.96	0.97
celkem		0.86	0.86	0.86

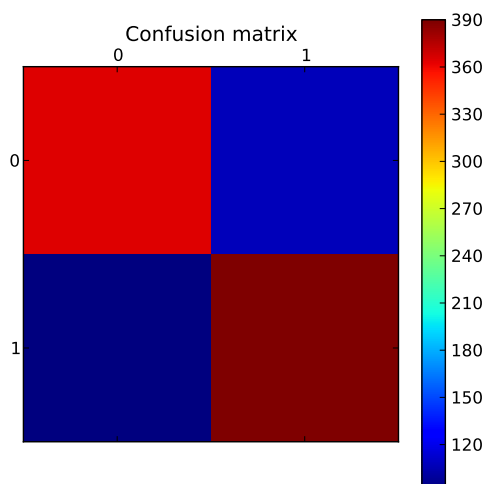
Tabulka 6.4: Klasifikace - zprávy, LDA, k-means_hog_9.8.2 + lbp_2.0

inout

typ příznaků	klasifikátor					
	DTC	GNB	KNN	LDA	RFC	SVM
gist	66%	75%	72%	66%	67%	77%
hog_9_8_1	56%	63%	51%	66%	57%	49%
hog_9_8_2	53%	65%	50%	65%	59%	49%
hog_9_16_2	55%	66%	52%	67%	61%	50%
hsv_3_2	69%	75%	76%	74%	73%	49%
hsv_6_4	70%	77%	75%	62%	72%	49%
hsv_6_6	67%	76%	75%	67%	72%	49%
k-means_hog_9_8_2	53%	73%	74%	68%	74%	49%
k-means_hog_9_16_2	55%	71%	70%	68%	68%	49%
lbp_2_0	67%	73%	75%	68%	72%	54%
lbp_2_1	70%	75%	71%	69%	74%	58%
lbp_4_0	70%	78%	76%	67%	75%	50%
lbp_4_1	70%	78%	75%	65%	71%	49%
k-means_hog_9_8_2 + lbp_2_0	68%	72%	74%	78%	71%	49%
k-means_hog_9_8_2 + lbp_2_1	68%	75%	74%	78%	74%	49%
k-means_hog_9_8_2 + lbp_4_0	70%	74%	74%	75%	72%	49%
k-means_hog_9_8_2 + lbp_4_1	68%	75%	74%	72%	74%	49%
k-means_hog_9_16_2 + lbp_2_0	67%	72%	70%	77%	68%	49%
k-means_hog_9_16_2 + lbp_2_1	69%	73%	70%	77%	73%	49%
k-means_hog_9_16_2 + lbp_4_0	69%	73%	70%	78%	70	49%
k-means_hog_9_16_2 + lbp_4_1	68%	72%	70%	74%	73%	49%
lbp_4_0 + hsv_6_4	71%	79%	75%	78%	74%	49%

Tabulka 6.5: Přehled úspěšnosti všech použitých kombinací příznaků a klasifikátorů pro dataset "inout".

Na obrázku 6.2 je zobrazena matice záměn nejúspěšnější klasifikace datasetu "inout". V tabulce 6.6 je kompletní přehled úspěšnosti klasifikace jednotlivých tříd.



Obrázek 6.2: Matice záměn - inout, GNB, lbp_4.0 + hsv_6.4

id	název	precision	recall	f1-score
0	indoor	0.8	0.77	0.78
1	outdoor	0.78	0.81	0.8
celkem		0.79	0.79	0.79

Tabulka 6.6: Klasifikace - inout, GNB, lbp_4.0 + hsv_6.4

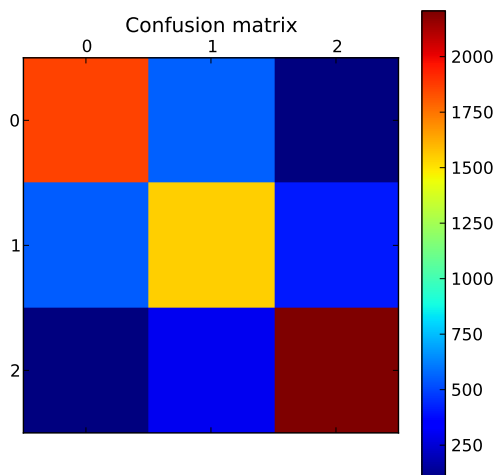
in_out_outMM

typ příznaků	klasifikátor					
	DTC	GNB	KNN	LDA	RFC	SVM
hog_9_8_2	46%	63%	35%	64%	51%	32%
hsv_3_2	54%	54%	53%	55%	55%	32%
hsv_6_4	55%	54%	53%	50%	56%	32%
k-means_hog_9_8_2	66%	63%	64%	70%	66%	32%
k-means_hog_9_16_2	62%	71%	60%	69%	64%	32%
lbp_2_0	60%	56%	58%	65%	57%	53%
lbp_2_1	59%	61%	55%	66%	60%	53%
lbp_4_0	60%		61%	64%	59%	55%
lbp_4_1	60%	62%	59%	66%	62%	59%
k-means_hog_9_8_2 + lbp_2_0	68%		64%	69%	68%	32%
k-means_hog_9_8_2 + lbp_2_1	67%	71%	64%	71%	67%	32%
k-means_hog_9_8_2 + lbp_4_0	69%		64%	68%	69%	32%
k-means_hog_9_8_2 + lbp_4_1	68%		64%	72%	68%	32%
k-means_hog_9_16_2 + lbp_2_0	64%		60%	69%	64%	32%
k-means_hog_9_16_2 + lbp_2_1	65%		60%	73%	65%	32%
k-means_hog_9_16_2 + lbp_4_0	65%	72%	60%	70%	62%	53%
k-means_hog_9_16_2 + lbp_4_1	65%		60%	72%	65%	32%
lbp_4_0 + hsv_6_4	62%	63%	54%	62%	63%	32%

Tabulka 6.7: Přehled úspěšnosti všech použitých kombinací příznaků a klasifikátorů pro dataset "in_out_outMM".

Na obrázku 6.3 je zobrazena matice záměn nejúspěšnější klasifikace datasetu "in_out_outMM". Je zřejmé, že nejlépe se podařilo oddělit třídy "indoor" a "outdoor man made".

V tabulce 6.8 je kompletní přehled úspěšnosti klasifikace jednotlivých tříd.



Obrázek 6.3: Matice záměn - in_out_outMM, LDA, k-means_hog_9_16_2 + lbp_2_1

id	název	precision	recall	f1-score
0	indoor	0.74	0.74	0.74
1	outdoor	0.64	0.62	0.63
2	outdoor man made	0.81	0.85	0.83
celkem		0.73	0.74	0.73

Tabulka 6.8: Klasifikace - in_out_outMM, LDA, k-means_hog_9_16_2 + lbp_2_1

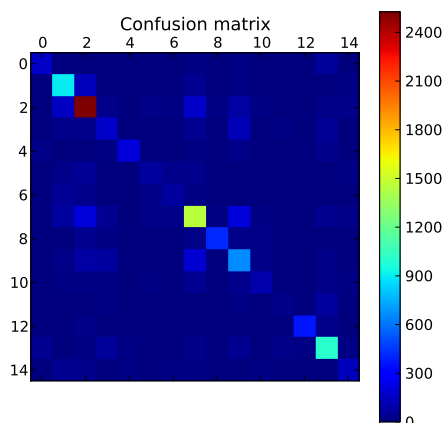
15 classes

typ příznaků	klasifikátor					
	DTC	GNB	KNN	LDA	RFC	SVM
gist	39%	41%	54%	55%	42%	67%
hog_9_8_1	30%	51%	33%	51%	29%	25%
hog_9_8_2	28%	52%	26%	51%	29%	25%
hsv_3_2	37%	20%	44%	46%	39%	26%
hsv_6_4	38%	22%	44%	45%	38%	26%
k-means_hog_9_8_2	41%	41%	48%	60%	40%	26%
k-means_hog_9_16_2	40%	40%	48%	60%	35%	26%
lbp_2_0	41%	35%	45%	59%	40%	33%
lbp_2_1	40%	37%	42%	56%	42%	34%
lbp_4_0	40%	34%	43%	57%	39%	32%
lbp_4_1	40%	33%	41%	51%	42%	32%
k-means_hog_9_8_2 + lbp_2_0	44%	43%	48%	67%	41%	26%
k-means_hog_9_8_2 + lbp_2_1	43%	43%	48%	66%	41%	26%
k-means_hog_9_8_2 + lbp_4_0	43%	45%	48%	67%	42%	26%
k-means_hog_9_8_2 + lbp_4_1	43%	43%	48%	65%	41%	26%
k-means_hog_9_16_2 + lbp_2_0	44%	43%	48%	68%	39%	26%
k-means_hog_9_16_2 + lbp_2_1	44%	41%	48%	67%	40%	26%
k-means_hog_9_16_2 + lbp_4_0	43%	44%	48%	66%	39%	26%
k-means_hog_9_16_2 + lbp_4_1	43%	41%	48%	65%	36%	26%
k-means_hog_9_16_2 + lbp_2_1 + hsv_6_4	44%	49%	44%	68%	40%	26%
lbp_4_0 + hsv_6_4	42%	40%	44%	60%	41%	26%

Tabulka 6.9: Přehled úspěšnosti všech použitých kombinací příznaků a klasifikátorů pro dataset "15 classes".

Na obrázku 6.4 je zobrazena matice záměn nejúspěšnější klasifikace datasetu "15classes".

V tabulce 6.10 je kompletní přehled úspěšnosti klasifikace jednotlivých tříd.



Obrázek 6.4: Matice záměn - 15 classes, LDA, k-means_hog_9_16_2 + lbp_2_1 + hsv_6_4

id	název	precision	recall	f1-score
0	airplane cabin	0.62	0.56	0.59
1	badlands	0.67	0.79	0.73
2	beach	0.79	0.82	0.8
3	crosswalk	0.34	0.33	0.34
4	discotheque	0.76	0.6	0.67
5	fjord	0.38	0.3	0.34
6	glacier	0.39	0.35	0.37
7	highway	0.7	0.69	0.69
8	oilrig	0.85	0.73	0.78
9	parking lot	0.53	0.59	0.56
10	planetarium - outdoor	0.44	0.34	0.38
11	power plant - indoor	0.26	0.17	0.2
12	sauna	0.95	0.79	0.86
13	subway station - platform	0.72	0.79	0.75
14	volcano	0.5	0.43	0.46
celkem		0.68	0.68	0.68

Tabulka 6.10: Klasifikace - 15 classes, LDA, k-means_hog_9_16_2 + lbp_2_1 + hsv_6_4

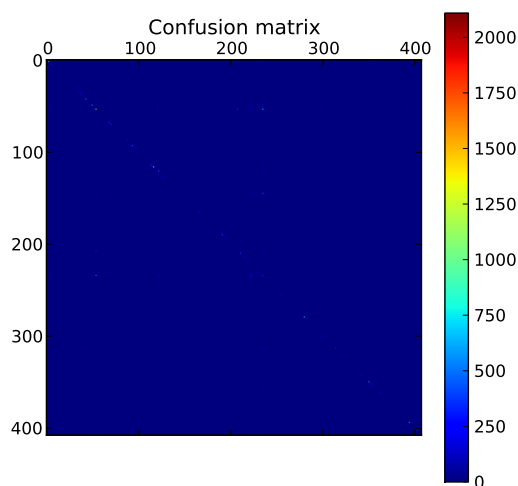
multiclass in

typ příznaků	klasifikátor					
	DTC	GNB	KNN	LDA	RFC	SVM
hog_9_8_1	4,5%	10%			4,4%	5,3%
hog_8_16_1	4,8%	9,2%	3,7%	8,7%	4,8%	5,3%
hsv_3_2	5,2%	0,7%	3,8%	5,3%	5,3%	5,3%
hsv_6_4	4,9%	1,2%	3,6%	5,0%	5,3%	5,3%
k-means_hog_9_16_2	5,3%	2,5%	7,0%	11%	5,3%	5,2%
lbp_2_1	6,0%	3,5%	5,1	11%	6,1%	
k-means_hog_9_16_2 + lbp_2_0	6,3%	3,4%	7,0%	18%	5,5%	5,3%
k-means_hog_9_16_2 + lbp_2_1	5,9%	3,1%	7,0%	16%	5,3%	5,2%
k-means_hog_9_16_2 + lbp_4_0	6,3%	3,8%	7,0%	19%	5,7%	
k-means_hog_9_16_2 + lbp_4_1	6,2%	3,1%	7,1%	15%		
k-means_hog_9_16_2 + lbp_2_1 + hog_9_8_1	5,9%	11%	7,5%	9,7%	4,5%	5,2%

Tabulka 6.11: Přehled úspěšnosti všech použitých kombinací příznaků a klasifikátorů pro dataset "multiclass in".

Na obrázku 6.5 je zobrazena matice záměn nejúspěšnější klasifikace datasetu "multiclass in". Klasifikace proběhla s úspěšností 19 %. Na diagonále je vidět několik lépe zaklasifikovaných tříd.

V tabulce 6.12 je kompletní přehled úspěšnosti klasifikace jednotlivých tříd.



Obrázek 6.5: Matice záměn - multiclass in, LDA, k-means_hog_9_16_2 + lbp_4_0

id	název	precision	recall	f1-score
30	auto mechanics - indoor	0	0	0
22	artists loft	0	0	0
294	pizzeria	0	0	0
147	elevator lobby	0	0	0
63	bistro - indoor	0.01	0	0
210	ice skating rink - indoor	0.5	0.53	0.51
88	car interior - frontseat	0.49	0.56	0.52
6	anechoic chamber	0.6	0.48	0.53
116	corridor	0.51	0.55	0.53
37	ball pit	0.89	0.5	0.64
celkem		0.18	0.19	0.18

Tabulka 6.12: Klasifikace - multiclass in, LDA, k-means_hog_9_16_2 + lbp_4_0. V tabulce je uvedeno pět nejúspěšnějších a pět nejméně úspěšných tříd.

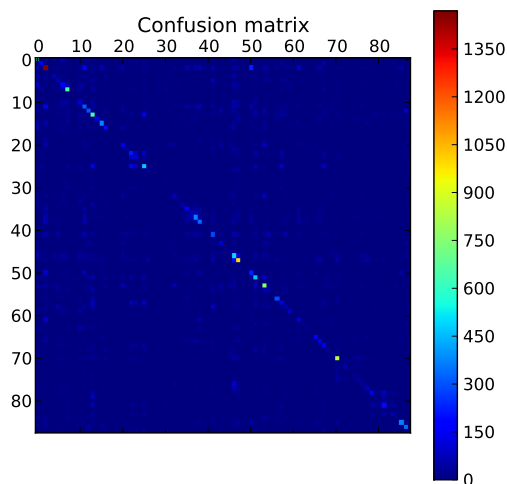
multiclass out

typ příznaků	klasifikátor					
	DTC	GNB	KNN	LDA	RFC	SVM
hog_9_8_1	9,7%	15%	9,9%	5,0%	9,2%	9,9%
hog_8_16_1	11%	13%	12%	12%	9,1%	8,3%
hsv_3_2	13%	3,7%	14%	16%	14%	8,3%
hsv_6_4	13%	4,7%	14%	16%	13%	8,3%
k-means_hog_9_16_2	22%	6,2%	15%	11%	12%	8,3%
lbp_2_1	13%	8,3%	14%	23%	15%	11%
k-means_hog_9_16_2 + lbp_2_0	14%	7,8%	15%	31%	13%	8,3%
k-means_hog_9_16_2 + lbp_2_1	14%	6,7%	7,0%	28%	12%	8,3%
k-means_hog_9_16_2 + lbp_4_0	14%	8,1%	15%	31%	14%	8,3%
k-means_hog_9_16_2 + lbp_4_1	14%	6,9%	15%	27%	13%	8,3%
k-means_hog_9_16_2 + lbp_2_1 + hsv_6_4	15%	11%	14%	34%	14%	8,3

Tabulka 6.13: Přehled úspěšnosti všech použitých kombinací příznaků a klasifikátorů pro dataset "multiclass out".

Na obrázku 6.6 je zobrazena matice záměn nejúspěšnější klasifikace datasetu "multiclass out". Na obrázku je dobře vidět diagonála správně zařazených tříd.

V tabulce 6.14 je kompletní přehled úspěšnosti klasifikace jednotlivých tříd.



Obrázek 6.6: Matice záměn - multiclass out, LDA, lbp_2.1 + k-means_hog_9.16.2 + hsv_6.4

id	název	precision	recall	f1-score
55	quicksand	0	0	0
8	cavern-outdoor	0	0	0
27	gorge	0	0	0
18	estuary	0	0	0
52	outcropping	0	0	0
47	mountain_snowy	0.54	0.51	0.53
86	wheat_field	0.5	0.53	0.52
56	raft	0.62	0.54	0.58
7	cavern-indoor	0.52	0.6	0.56
70	underwater-coral_reef	0.63	0.62	0.63
celkem		0.35	0.35	0.34

Tabulka 6.14: Klasifikace - multiclass out, LDA, lbp_2.1 + k-means_hog_9.16.2 + hsv_6.4. V tabulce je uvedeno pět nejúspěšnějších a pět nejméně úspěšných tříd.

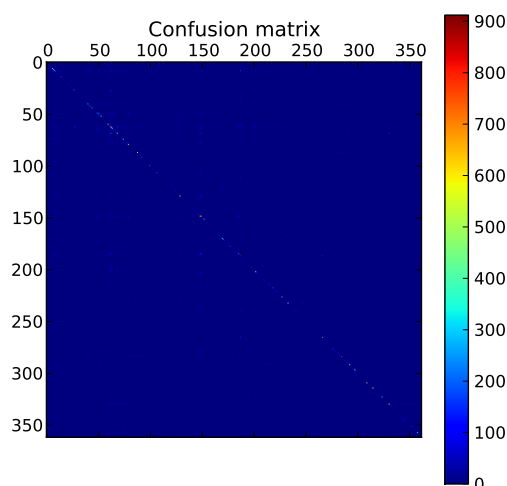
multiclass out man made

typ příznaků	klasifikátor					
	DTC	GNB	KNN	LDA	RFC	SVM
hog_9_8_1	3,7%		3,6%	5,4%	3,8%	
hog_8_16_1	3,9%	10%	5,3%	8,8%	4,7%	3,6%
hsv_3_2	4,6%	1,0%	5,0%	6,4%	5,4%	3,5%
hsv_6_4	4,5%	1,6%	5,5%	6,9%	5,2%	3,5%
k-means_hog_9_16_2	4,7%	2,3%	6,7%	12%	5,4%	3,5%
lbp_2_1	5,7%	4,6%	5,5%	4,3%	6,9%	11%
k-means_hog_9_16_2 + lbp_2_0	6,3%			18%		
k-means_hog_9_16_2 + lbp_2_1	5,8%			15%		
k-means_hog_9_16_2 + lbp_4_0	6,0%			18%		
k-means_hog_9_16_2 + lbp_4_1	5,9%			15%		
k-means_hog_9_16_2 + lbp_2_1 + hsv_6_4	6,1%			19%		
hog_8_16_1 + lbp_2_1	5,7%	11%	5,2%	11%	4,8%	3,6

Tabulka 6.15: Přehled úspěšnosti všech použitých kombinací příznaků a klasifikátorů pro dataset "multiclass out man made".

Na obrázku 6.7 je zobrazena matice záměn nejúspěšnější klasifikace datasetu "multiclass out man made". Klasifikace proběhla s úspěšností 19 %. Na diagonále je vidět několik lépe zaklasifikovaných tříd.

V tabulce 6.16 je kompletní přehled úspěšnosti klasifikace jednotlivých tříd.



Obrázek 6.7: Matice záměn - multiclass outMM, LDA, lbp_2.1 + k-means_hog_9.16.2 + hsv_6.4

id	název	precision	recall	f1-score
97	fence	0	0	0
181	lookout station - outdoor	0	0	0
47	bus station - outdoor	0	0	0
33	beach house	0	0	0
247	priory	0	0	0
44	bullring	0.38	0.48	0.43
151	ice skating rink - outdoor	0.42	0.44	0.43
296	stadium - baseball	0.41	0.45	0.43
232	phone booth	0.4	0.51	0.45
87	driving range - outdoor	0.51	0.44	0.47
celkem		0.19	0.19	0.19

Tabulka 6.16: Klasifikace - multiclass outMM, LDA, lbp_2.1 + k-means_hog_9.16.2 + hsv_6.4. V tabulce je uvedeno pět nejúspěšnějších a pět nejméně úspěšných tříd.

7 Závěr

V práci jsem vyzkoušel kombinace několika příznaků s klasifikátory. Nejlepšího výsledku (86 %) jsem dosáhl v klasifikaci datasetu "zpravy" s použitím klasifikátoru LDA a kombinací příznaků typu k-means_hog_9_8_2 + lbp_2_0. Ostatní výsledky jsou také uspokojivé. Například je velmi dobrá klasifikace s úspěšností 34 % v datasetu "multiclass in", který má 88 tříd.

V datasetech s velkým počtem tříd úspěšnost klasifikace klesá, ale kombinací dalších příznaků by se dala ještě vylepšit. Vhodná kombinace příznaků pro datasety s velkým počtem tříd je k-means_hog_9_16_2 + lbp_2_1 + hsv_6_4.

Velikého zlepšení klasických příznaků typu hog lze dosáhnout použitím clusterování, které zajistí nezávislost na posunu obrazu. Vektor příznaků je díky tomu více informativní a celá klasifikace je robustnější.

Výsledky práce by bylo možné použít pro automatické indexování scén v rozsáhlých databázích nebo ve videích. To by umožnilo rychlé a snadné vyhledávání pomocí klíčových slov.

Výpočty některých příznaků jsou datově i časově velmi náročné. Ještě náročnější je pak samotné natrénování klasifikátorů. U rozsáhlých datasetů může proces natrénování klasifikátoru trvat i několik dní. Z toho důvodu je výhodné použít vzdálené výpočetní prostředí. V této práci jsem využil MetaCentrum.

Budoucí práce

Pro zvýšení úspěšnosti klasifikace by bylo dobré vyzkoušet klasifikaci zkombinováním více typů vypočtených příznaků. Tento proces bude však výpočetně velmi náročný. Pro snížení výpočetních nároků (paměť i čas) lze použít metody redukce dimenze, například PCA (principal component analysis) - "Analýza hlavních komponent". Nejúspěšnější klasifikace by se poté provedly na všech obrazech.

Literatura

- [1] Martin Szummer and RW Picard. Indoor-outdoor image classification. *Content-Based Access of Image and ...*, 1998. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=646032.
- [2] Jianxiong Xiao, James Hays, Krista a. Ehinger, Aude Oliva, and Antonio Torralba. SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3485–3492, June 2010. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5539970>, doi:10.1109/CVPR.2010.5539970.
- [3] Aditya Vailaya, A Jain, and HJ Zhang. On image classification: City images vs. landscapes. *Pattern Recognition*, 31(12):1921–1935, 1998. URL: <http://www.sciencedirect.com/science/article/pii/S003132039800079X>.
- [4] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:886–893, 2005. URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1467360>, doi:10.1109/CVPR.2005.177.
- [5] Scikit - image. URL: <http://scikit-image.org/>.
- [6] Aude Oliva. Gist of the Scene. pages 251–257, 2005. URL: <http://cvcl.mit.edu/papers/oliva04.pdf>.
- [7] Tim Ojala, Matti Pietikhenl, David Harwood, and Laws Texture Measures. m i. pages 582–585, 1994. URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=576366>.
- [8] André Anjos, Laurent El Shafey, Roy Wallace, Manuel Günther, and Chris Mccool. Bob : A Free Signal Processing and Machine Learning Toolbox for Researchers Categories and Subject Descriptors. 1920.
- [9] Fabian Pedregosa, Ron Weiss, and Matthieu Brucher. Scikit-learn : Machine Learning in Python. 12:2825–2830, 2011.