Extrakce a zpracování příznaků pro automatickou klasifikaci bitmapových obrázků

Pavel Surynek[1] a Ivana Lukšová[2]

[1]Katedra teoretické informatiky a matematické logiky,
[2]Matematicko-fyzikální fakulta, Univerzita Karlova v Praze,
Malostranské náměstí 25, 118 00, Praha 1
pavel.surynek@mff.cuni.cz, ivana.luksova@gmail.com

Abstrakt. Článek pojednává o návrhu metody pro automatickou klasifikaci bitmapových obrázků do tříd definovaných na základě obsahu popsaného přirozeným jazykem. Příkladem takových přirozeně definovaných tříd mohou být obrázky zobrazující budovu, krajinu, umělecký obraz atd. Využívaná klasifikační metoda je založena na extrakci vhodných příznaků z bitmapového obrázku, jako jsou kontrast, barevný histogram, výskyt přímých linií atd. Příznaky jsou následně zpracovány pomocí předem natrénovaného rozhodovacího stromu. Provedené experimenty s 5 klasifikačními třídami ukázaly, že navržená metoda dosahuje 75%-85% úspěšnosti. Návrh metody je obecný v tom smyslu, že umožňuje jednoduše rozšiřovat sadu klasifikačních tříd i sadu extrahovaných atributů pro další zvyšování přesnosti klasifikace.

Klíčová slova: rozpoznávání obrazu, počítačové vidění, extrakce příznaků, rozhodovací strom

1 Úvod a motivace

Stále snadnější dostupnost zařízení pro záznam obrazových dat, ať už se jedná o digitální fotoaparáty či průmyslové kamery, postupně vyvolává poptávku po softwarových nástrojích pro automatické zpracování takto pořízených dat. Velké množství obrazových dat je přitom hlavním důvodem pro požadavek na automatizaci, neboť jejich ruční zpracování je často nad možnosti běžného lidského operátora či uživatele. Mnoho uživatelů digitálních fotoaparátů se s tímto problémem jistě setkalo při třídění několika tisíc fotografií z poznávací dovolené, což se zpravidla ukazuje jako časově velmi náročný úkol.

V představované práci jsme se zaměřili na analýzu statických obrazových dat, tedy klasických bitmapových obrázků. Základní východisko je přitom takové, že nijak neomezujeme původ bitmapových obrázků, může se tedy jednat o fotografie pořízené digitálním fotoaparátem či obrázky uměle vytvořené na počítači. Konkrétní úlohu, kterou v této práci řešíme, je *klasifikace bitmapových obrázků* do tříd definovaných popisem v přirozeném jazyce. Jedná se například o třídy obrázků zobrazující *krajiny*, *umělecké obrazy* či obrázky zobrazující *budovy*.

Tato práce vznikla za podpory Grantové agentury České republiky - GAČR v rámci projektu číslo 201/09/P318 a za podpory Ministerstva školství, mládeže a tělovýchovy ČR – MŠMT ČR v rámci výzkumného záměru číslo MSM 0021620838.

Cílem bylo navrhnout zcela **automatickou** metodu, která pro předem danou pevnou sadu klasifikačních tříd (nejsou součástí vstupu) rozhodne o vstupním bitmapovém obrázku, do které třídy (či více tříd) patří. Sekundárním cílem bylo navrhnout takovou metodu, která by byla snadno **rozšiřitelná** o další klasifikační třídy a měla potenciál k postupnému **zvyšování úspěšnosti** klasifikace. Vzhledem k náročnosti klasifikační úlohy lze očekávat, že prototypová implementace metody nebude mít příliš vysokou úspěšnost a je tedy třeba si hned na počátku návrhu vymezit prostor pro další zpřesňování a zvyšování kvality.

Na základě počáteční analýzy úlohy klasifikace byl jako základní stavební prvek klasifikační metody zvolen *rozhodovací strom* [4, 7]. Rozhodovací strom je struktura, která popisuje přiřazení určité hodnoty vstupnímu vektoru příznaků. V našem případě je hodnotou indikátor příslušnosti do klasifikační třídy a příznaky jsou zvolené vlastnosti bitmapového obrázku, jako jsou *jas*, barevný *histogram* [8] nebo *výskyt přímých linií* vyjádřené číselnou hodnotou.

Dostatečně přesná klasifikace pomocí rozhodovacího stromu v daném kontextu je podmíněna nalezením vhodných příznaků, které třídy bitmapových obrázků dostatečně dobře charakterizují. Je evidentní, že prostor pro hledání výstižných příznaků je velký a tím pádem potenciál pro zvyšování úspěšnosti klasifikace pomocí rozhodovacího stromu je rovněž značný. Další hlavní požadavky rozhodovací strom plní sám o sobě, neboť rozhodovací fáze je zcela automatická a rozšíření množiny atributů a klasifikačních tříd je v podstatě rutinní záležitost.

Článek vychází z bakalářské práce spoluautorky [2]. Text článku je členěn následujícím způsobem. Nejprve je představena klasifikační úloha a klasifikační třídy, které požadujeme, aby navrhovaná metoda byla schopná rozlišovat. Dále jsou představeny použité techniky z umělé inteligence a počítačové grafiky, jež byly použity jako stavební prvky pro klasifikační metodu, a samotná klasifikační metoda. Poslední část je věnována experimentálnímu vyhodnocení implementace navržené metody v jazyce Java na několika testovacích sadách bitmapových obrázků.

2 Vymezení klasifikační úlohy s bitmapovými obrázky

Nechť K je konečná množina klasifikačních tříd a nechť $\mathcal{I}$ je množina bitmapových obrázků. Každá klasifikační třída $t \in \mathcal{K}$ má přiřazen popis v přirozeném jazyce $d(t)$. Dále nechť zobrazení $c: \mathcal{I} \longrightarrow 2^{\mathcal{K}}$ definuje přiřazení klasifikačních tříd obrázkům tak, že $\forall I \in \mathcal{I}\ \forall t \in c(I)\ d(t)$ dobře popisuje obrázek I z pohledu uživatele. Uvědomme si, že zobrazení c není známo v explicitní formě; implicitně zná zobrazení c uživatel, když dokáže říci, jaký je výsledek $c(I)$. Následující definice umožní přesněji zformulovat řešený problém.

Definice 1 (*klasifikační úloha, klasifikační metoda*). *Klasifikační metoda* je algoritmus realizující zobrazení $c': \mathcal{I} \longrightarrow 2^{\mathcal{K}}$. *Klasifikační úloha* s bitmapovými obrázky odpovídá výpočtu $c'(I)$ pro daný vstupní obrázek $I \in \mathcal{I}$. □

Definice 2 (*úspěšnost klasifikační metody*). Klasifikační metoda realizující zobrazení $c': \mathcal{I} \longrightarrow 2^{\mathcal{K}}$ je *úspěšná* pro klasifikační třídu $t \in \mathcal{K}$ a vstupní obrázek $I \in \mathcal{I}$, jestliže $t \in c'(I) \Leftrightarrow t \in c(I)$. Nechť $\mathcal{S} \subset \mathcal{I}$ je konečná množina bitmapových obrázků a

nechť $\mathcal{S}' \subseteq \mathcal{S}$ je množina obrázků, kde je klasifikační metoda úspěšná pro třídu $t \in \mathcal{K}$, potom $|\mathcal{S}'|/|\mathcal{S}|$ je *úspěšnost* klasifikační metody pro třídu t a množinu obrázků $\mathcal{S}$. □

Cílem pochopitelně je, aby klasifikační metoda měla co nejvyšší úspěšnost vzhledem k daným klasifikačním třídám pro co největší množinu vstupních obrázků (nejlépe pro všechny, které uživatel může metodě zadat).

2.1 Vybrané klasifikační třídy

Při počátečním návrhu klasifikační metody jsme požadovali zvládnutí klasifikace obrázků vzhledem k množině pěti klasifikačních tříd $\mathcal{K} = \{P, A, B, L, M\}$. Tyto klasifikační třídy byly inspirovány požadavky uživatele digitálního fotoaparátu. V následujícím textu je u každé třídy $t \in \mathcal{K}$ definován popis $d(t)$ a jsou zmíněny příklady obrázků, které do třídy t spadají či nikoli.

Obr. 1. Bitmapový obrázek klasifikovaný uživatelem do třídy ***fotografie*** (**P**) [levá část] a příklad obrázku, který uživatel do této třídy nezařadil [pravá část].

P (photography): fotografie

$d(P)$ =“fotografie pořízená analogovým či digitálním fotoaparátem“

Pozitivní a negativní příklady bitmapových obrázků vzhledem ke klasifikační třídě P jsou ukázány na Obr. 1. Do třídy fotografie nepatří například uměle vytvořené obrazy.

Obr. 2. Bitmapový obrázek uživatelem klasifikovaný jako ***umělecký obraz*** (**A**).

A (artistic): umělecké obrazy

$d(I)$ =“obrázek vytvořený kreslířskou, malířskou technikou“

Příklad uměleckého bitmapového obrázku je ukázán na Obr. 2. Do třídy umělecké obrazy nepatří například fotografie, ovšem rovněž do této třídy nezařazujeme obrázky různých diagramů či renderované obrazy.

B (buildings): budovy

$d(B)$ =“obrázek zobrazující budovu či budovy“

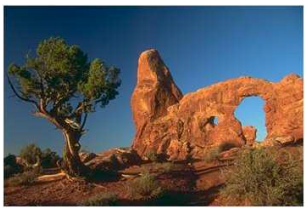
Příklad obrázku spadajícího do třídy P je ukázán na Obr. 3. Naopak do této třídy nezařazujeme například obrázek krajiny. Ovšem na druhou stranu jistě zařadíme obrázek domu obklopeného krajinou.

Obr. 3. Příklad bitmapového obrázku uživatelem zařazeného do klasifikační kategorie ***budovy*** (**B**).

L (landscapes): krajiny

$d(L)$ =“obrázek zobrazující krajinu“

Příklad bitmapového obrázku spadajícího do třídy krajiny je ukázán na Obr. 4. Do třídy krajiny nezařazujeme obrázky umělých objektů, ovšem umělý objekt umístěný v krajině může celkově představovat krajinu.



Obr. 4. Bitmapový obrázek řazený uživatelem do klasifikační kategorie ***krajiny*** (**L**).

M (macro objects): makro objekty

$d(M)$ ="fotografie objektu snímaného z malé vzdálenosti s neostrým pozadím"

Příklad makro objektu je ukázán na Obr. 5. Mezi makro objekty ovšem například nezařazujeme kreslené obrázky či malby (umělecké obrazy).

Všimněme si, že při podávání slovního popisu klasifikační třídy $t \in \mathcal{K}$ se v $d(t)$ zaobíráme pozitivním vymezením dané třídy (tj. popisem, jak vypadá pozitivní příklad). Negativní vymezení třídy t je dané tím, že do třídy nezahrnujeme obrázek, který není dobře popsán pomocí $d(t)$. Navržené klasifikační třídy mimo jiné dovolují existenci vstupního obrázku $I_1 \in \mathcal{I}$, že $c(I_1) = \{A, B, L\}$, naopak existence obrázku $I_2 \in \mathcal{I}$, že $c(I_2) = \{P, A\}$ je vyloučená.

Obr. 5. Příklad bitmapového obrázku klasifikovaného uživatelem jako ***makro objekt*** (**M**).

3 Počáteční analýza úlohy klasifikace

Při vývoji klasifikační metody bylo třeba zodpovědět několik klíčových otázek. Především nebylo jasné, jak se formálně vypořádat s podmínkou, že $\forall I \in \mathcal{I}\ \forall t \in c(I)$ $d(t)$ dobře popisuje obrázek I z pohledu uživatele. Zde je jednak nutné zohlednit, že $c: \mathcal{I} \longrightarrow 2^{\mathcal{K}}$ je v podstatě definováno nazíráním konkrétního uživatele a může tedy být silně subjektivní. To znamená, že návrh klasifikační metody by měl počítat s různými uživateli a umožnit příslušnou konfiguraci. A dále je nutné zohlednit již zmíněný aspekt, že zobrazení c neznáme explicitně.

Jako nejvhodnější se nakonec ukázalo zachytit zobrazení $c: \mathcal{I} \longrightarrow 2^{\mathcal{K}}$ částečně pomocí anotované (konečné) trénovací množiny vybraných obrázků. Toto řešení zohledňuje jak individuální uživatele (různí uživatelé mohou použít různé trénovací množiny obrázků), tak fakt, že znalost c není explicitní (k vytvoření anotované trénovací množiny stačí konečně mnoho dotazů na uživatele, aby spočetl c pro zadaný vstup). Samotná klasifikační metoda realizující zobrazení $c': \mathcal{I} \longrightarrow 2^{\mathcal{K}}$ by se měla shodovat s c na všech (nebo téměř všech) vstupních obrázcích z trénovací množiny a měla by být co nejúspěšnější pro co největší množinu vstupních obrázků vzhledem k zobrazení c. Ke splnění tohoto požadavku je třeba, aby navrhovaná klasifikační metoda byla schopna dobře **generalizovat** [7].

Uvedené požadavky vedly k použití *rozhodovacího stromu* [4, 7] jako hlavního klasifikačního nástroje. Pro rozhodovací stromy existují účinné učící algoritmy [6], které přesně zohledňují trénovací množinu (pokud ta splňuje jisté základní předpoklady – tj. neobsahuje rozpory) a výrazně podporují schopnost generalizace výsledného rozhodovacího stromu. Jak již bylo zmíněno v úvodní sekci, volbu rozhodovacích stromů podporuje také jejich snadná adaptace pro větší počet klasifikačních tříd.

Další důležitou otázkou byla identifikace a získávání příznaků, které dobře charakterizují bitmapové obrázky a které by mohl používat příslušný rozhodovací strom.

Vzhledem k předpokládaným klasifikačním třídám bylo intuitivně navrženo používat tři kategorie charakteristik: *základní barevné charakteristiky* (např. počet barev, relativní zastoupení barev), charakteristiky založené na *histogramu* (např. kontrast, lokální kontrast) a charakteristiky založené na *hranové informaci* [3] (např. výskyt přímých linií či četnost pravých úhlů). Předpoklad přitom byl, že základní barevné charakteristiky pomohou odlišit fotografie a umělecké obrazy. Dále bylo očekáváno, že charakteristiky založené na histogramu budou významné mimo jiné u fotografií lišících se kontrastem (poslouží například k rozlišení fotografie makro objektu s rozostřeným pozadím od ostré fotografie krajiny). A konečně charakteristiky založené na hranové informaci pomohou odlišit obrázky umělých a přírodních objektů (tedy například fotografii krajiny od fotografie budovy). Přitom se předpokládá, že ke kompletní klasifikaci bude nutné zkombinovat různou měrou příznaky všech zvolených charakteristik (důležitost příznaků pro jednotlivé klasifikační třídy je rovněž zajímavá otázka; shodou okolností nám tuto otázku zodpoví vytvořený rozhodovací strom, neboť ten prioritně provádí rozhodování podle důležitějších příznaků).

4 Použité techniky a proces klasifikace

Hlavním **přínosem** práce je návrh příznaků využitelných pro klasifikaci a návrh procesu jejich extrakce ze vstupního bitmapového obrázku. Proces extrakce příznaků popíšeme podrobněji, zatímco samotné využití příznaků při klasifikaci pomocí rozhodovacího stromu pouze stručně, neboť se převážně jedná o aplikaci existující techniky. Zvláštní pozornost bude věnována složitějším typům příznaků založených na hranové informaci.

4.1 Rozhodovací strom

Výběr charakteristických příznaků využitelných při klasifikaci je podstatně určen použitými technikami. Hlavní použitou technikou je přitom rozhodovací strom [4, 7]. Předpokládejme, že máme množinu příznaků extrahovaných z obrázku $\mathcal{P}$, z nichž každý může nabývat jisté množiny hodnot (celá čísla nebo čísla v plovoucí desetinné čárce) a klasifikační třídu $t \in \mathcal{K}$. *Binární rozhodovací strom* pro třídu t je orientovaný strom, kde každý uzel kromě listů má dva následníky. Kořen a vnitřní uzly jsou ohodnoceny příznakem $p \in \mathcal{P}$ a listy jsou ohodnoceny booleovskou hodnotou $False$ nebo $True$. Hrany vedoucí z uzlu ohodnoceného příznakem p jsou ohodnoceny disjunktními podmnožinami hodnot příznaku p.

Rozhodovací strom přiřazuje vektoru ohodnocení příznaků indikátor náležení do dané třídy t. Máme-li vektor hodnot jednotlivých příznaků, je jim přiřazený indikátor náležení do klasifikační třídy určený řízeným průchodem rozhodovacího stromu od kořene k listu, jehož ohodnocení představuje výsledný indikátor náležení do třídy t. Řízený průchod začíná v kořeni a je definován induktivně. Předpokládejme, že aktuální uzel je ohodnocen příznakem p, potom průchod pokračuje do koncového uzlu hrany, jejíž ohodnocující podmnožina hodnot příznaku p obsahuje hodnotu tohoto příznaku ve vstupním vektoru ohodnocení příznaků.

Říkáme, že rozhodovací strom je *úplný*, jestliže je řízený průchod stromem od kořene k listům definován pro každý vstupní vektor ohodnocení příznaků. Pro reprezentaci klasifikačního zobrazení $c: \mathcal{I} \longrightarrow 2^{\mathcal{K}}$ budeme potřebovat $|\mathcal{K}|$ rozhodovacích stromů. Všimněme si, že je nyní dobře definována úspěšnost úplného rozhodovacího stromu podle Definice 2. Učení úplného rozhodovacího stromu, které bylo využito v navržené klasifikační metodě, probíhá pomocí algoritmu ID3, který je popsán v [6].

4.2 Extrakce příznaků z barev

Nadále budeme pracovat s bitmapovými obrázky s 8-bitovou hloubkou pro každou barevnou složku [8], případná úprava na jinou barevnou hloubku je přímočará. Základní charakteristiku, kterou budeme z bitmapového obrázku extrahovat, představuje barevná informace. Následující definice barevnou informaci přesně formalizují.

Definice 3 (*bitmapový obrázek – barevný, monochromatický*). *Barevný bitmapový obrázek* je pětice $I_{rgb} = (r, g, b, x_{max}, y_{max})$, kde $r: \{0,1, \dots, x_{max}\} \times \{0,1, \dots, y_{max}\} \longrightarrow \{0,1, \dots,255\}$, $g: \{0,1, \dots, x_{max}\} \times \{0,1, \dots, y_{max}\} \longrightarrow \{0,1, \dots,255\}$ a $b: \{0,1, \dots, x_{max}\} \times \{0,1, \dots, y_{max}\} \longrightarrow \{0,1, \dots,255\}$ jsou funkce, které po řadě jednotlivým pixelům obrázku přiřazují hodnotu červené, zelené a modré barevné složky; x_{max} vyjadřuje horizontální rozměr obrázku a y_{max} vertikální rozměr obrázku (rozměry obrázku tedy jsou $(x_{max} + 1) \times (y_{max} + 1)$ pixelů). *Monochromatický bitmapový obrázek* je trojice $I_{mono} = (i, x_{max}, y_{max})$, kde $i: \{0,1, \dots, x_{max}\} \times \{0,1, \dots, y_{max}\} \longrightarrow \{0,1, \dots,255\}$ je funkce, která jednotlivým pixelům přiřazuje hodnotu jasu. Pro barevný bitmapový obrázek zkonstruujeme odpovídající monochromatický položením $i(x, y) = 0.299\, r(x, y) + 0.587\, g(x, y) + 0.114\, b(x, y)\ \ \forall x \in \{0,1, \dots, x_{max}\};\ \forall y \in \{0,1, \dots, y_{max}\}$ (vztah byl určen empiricky [8]). Velikost obrázku je dána jako počet jeho pixelů, tedy $s_{xy} = (x_{max} + 1) * (y_{max} + 1)$. □

Užitečnou charakteristikou je *celkový počet barev* v obrázku. Máme-li například umělecký obraz, který není zašuměný, lze očekávat, že celkový počet barev bude nízký. Pro daný barevný obrázek je celkový počet barev určen následujícím vztahem: $c_{rgb} = |\{(r(x, y), g(x, y), b(x, y)) | x \in \{0,1, \dots, x_{max}\} \wedge\ y \in \{0,1, \dots, y_{max}\}\}|$.

Další zajímavou charakteristikou je *zastoupení jednotlivých barev*. Jelikož je ovšem počet všech možných barev příliš vysoký, je třeba se omezit jen na jistou *barevnou paletu* [8] a počítat pouze zastoupení barev blízkých barvám ze zvolené palety. Nejjednodušší volbu barevné palety představuje reprezentace barev (r, g, b), kde $\frac{\alpha}{\eta}256 \leq r < \frac{\alpha+1}{\eta}256 \wedge \frac{\beta}{\eta}256 \leq g < \frac{\beta+1}{\eta}256 \wedge \frac{\gamma}{\eta}256 \leq b < \frac{\gamma+1}{\eta}256$ pro $\eta \in \mathbb{N}$ a $\alpha, \beta, \gamma \in \{0,1, \dots, \eta - 1\}$ pomocí jedné barvy. Jelikož požadujeme, aby počet barev palety, nebyl příliš vysoký, je třeba, aby hodnota parametru η byla nízká. V implementaci klasifikační metody bylo použito $\eta = 4$. Zastoupení barev palety v obrázku je definováno funkcí: $rgb(\alpha, \beta, \gamma) = |\{(x, y) | x \in \{0,1, \dots, x_{max}\} \wedge\ y \in \{0,1, \dots, y_{max}\} \wedge\ \frac{\alpha}{\eta}256 \leq r(x, y) < \frac{\alpha+1}{\eta}256 \wedge\ \frac{\beta}{\eta}256 \leq g(x, y) < \frac{\beta+1}{\eta}256 \wedge\ \frac{\gamma}{\eta}256 \leq b(x, y) < \frac{\gamma+1}{\eta}256\}|/s_{xy}$, kde $\eta \in \mathbb{N}$ a $\alpha, \beta, \gamma \in \{0,1, \dots, \eta - 1\}$. Užitím funkce rgb byla pro každou klasifikační třídu určena paleta nejčastější barev, přičemž jako příznak pak byla použita shoda s touto paletou.

4.3 Příznaky extrahované z histogramu a z kontrastu

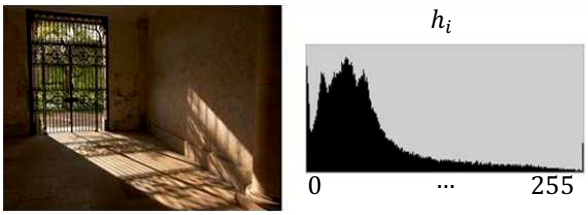
Histogram rovněž představuje důležitou charakteristiku bitmapových obrázků. Ve fotografické technice jsou často používány pojmy *tmavá* (*low-key*), *středně-tónová* (*mid-key*) či *světlá* (*high-key*) fotografie, které po řadě vyjadřují převahu tmavých, středně-tónových či světlých barev ve fotografii. Budeme se tedy těmito pojmy inspirovat a provedeme jejich přesnou kvantifikaci pomocí histogramu.

Definice 4 (*histogram*). *Histogram* pro monochromatický obrázek $I_{mono} = (i, x_{max}, y_{max})$ je zobrazení $h_i: \{0,1,\dots,255\} \longrightarrow \mathbb{N}_0$ (přirozená čísla včetně nuly), kde $h_i(j) = |\{(x,y)|x \in \{0,1,\dots,x_{max}\} \wedge y \in \{0,1,\dots,y_{max}\} \wedge i(x,y) = j\}|$. Pro barevný obrázek $I_{rgb} = (r, g, b, x_{max}, y_{max})$ definujeme histogram analogickým způsobem pomocí trojice funkcí $h_r: \{0,1,\dots,255\} \longrightarrow \mathbb{N}_0$, $h_g: \{0,1,\dots,255\} \longrightarrow \mathbb{N}_0$ a $h_b: \{0,1,\dots,255\} \longrightarrow \mathbb{N}_0$ pro jednotlivé barevné složky. □

Důležitou charakteristikou, která souvisí s histogramem, je *kontrast* obrázku. Intuitivně označujeme za kontrastní takový obrázek, který má velký rozdíl mezi nejsvětlejšími a nejtmavšími odstíny, jejichž podíl je zároveň relativně vysoký. Histogram takového obrázku tedy není soustředěný v úzké části spektra jasu nebo spektra barevných složek.

Tabulka 1. Některé charakteristiky založené na histogramu.

Charakteristika	Výpočet
Podíl tmavých barev $r_{low}^{h_i}$	$r_{low}^{h_i} = \frac{1}{s_{xy}} \sum_{j=0}^{85} h_i(j)$
Podíl středně-tónových barev $r_{mid}^{h_i}$	$r_{mid}^{h_i} = \frac{1}{s_{xy}} \sum_{j=86}^{170} h_i(j)$
Podíl světlých barev $r_{high}^{h_i}$	$r_{high}^{h_i} = \frac{1}{s_{xy}} \sum_{j=170}^{255} h_i(j)$
Kontrast podle **Michelsona** $c_M^{h_i}$	$c_M^{h_i} = \frac{i_{max} - i_{min}}{i_{max} + i_{min}}$
Kontrast podle **Webera** $c_{rms}^{h_i}$	$c_{rms}^{h_i} = \sqrt{\frac{1}{s_{xy}} \sum_{x=0}^{x_{max}} \sum_{y=0}^{y_{max}} (i(x,y) - i_\mu)^2}$



Obr. 6. Ukázka tmavého obrázku (*low-key*) a histogramu.

Kontrast lze formálně definovat různými způsoby, při návrhu klasifikační metody jsme použili definice z [5], tedy definice podle Michelsona a Webera. K zavedení těchto pojmů je třeba znát hodnotu minimálního, maximálního a průměrného jasu, které jsou dány po řadě vztahy $i_{min} = \min\{j|j \in \{0,1,\dots 255\} \wedge h_i(j) > 0\}$, $i_{max} = \max\{j|j \in \{0,1,\dots 255\} \wedge h_i(j) > 0\}$ a $i_\mu = \frac{1}{s_{xy}} \sum_{j=0}^{255} (j * h_i(j))$. Některé charakteristiky založené na histogramu podle jasové složky ukazuje Tabulka 1. Uvědomme si, že analogicky lze všechny uvedené charakteristiky definovat podle jednotlivých barevných složek (ty však experimentální implementace navržené klasifikační metoda nepoužívá).

U některých obrázků, jako jsou například fotografie objektů snímaných z malé vzdálenosti (makro objekty), je vysoký kontrast soustředěn pouze v jisté plošně ome-

zené oblasti (samotný objekt), zatímco zbylé oblasti mají kontrast nízký (pozadí). V takovém případě se vyplatí uvažovat o tzv. *lokálním kontrastu* tedy kontrastu určitého výřezu obrázku. V klasifikační metodě bylo použito rozdělení obrázku pravidelnou mřížkou, přičemž kontrast v každém poli mřížky (jak podle Michelsona $c_M^{h_i}$ tak podle Webera $c_{rms}^{h_i}$) byl uvažován jako zvláštní příznak (experimentální implementace klasifikační metody používá mřížku, kde rozměr každé buňky je 25×25 pixelů).

4.4 Získávání příznaků pomocí hranové detekce

Stěžejní technika, která umožní odlišit klasifikační třídu *budovy* (**B**) od ostatních tříd, využívá *hranovou informaci*. Jelikož extrakce a využití hranové informace představuje nejkomplexnější techniku vyvinutou v rámci práce, budeme se jí v této sekci věnovat podrobněji a klíčové operace popíšeme v pseudo-kódu.

Algoritmus 1. *Algoritmus **hranové detekce** vyjádřený v pseudo-kódu.* Vstupem algoritmu je monochromatický obrázek I_{mono} a konvoluční h matice velikosti $(2k_{max}+1) \times (2k_{max}+1)$. Výstupem je monochromatický obrázek zobrazující detekované hrany.

function *Detect-Edges* $(I_{mono} = (i, x_{max}, y_{max}), h, k_{max})$: **grayscale image**
1: **for** $x = 0,1,\dots,x_{max}$ **do**
2: | **for** $y = 0,1,\dots,y_{max}$ **do**
3: | | $s \leftarrow 0$
4: | | **for** $k_x = -k_{max},\dots,-1,0,1,\dots,k_{max}$ **do**
5: | | | **for** $k_y = -k_{max},\dots,-1,0,1,\dots,k_{max}$ **do**
6: | | | | $s \leftarrow s + h(k_{max}+k_x+1, k_{max}+k_y+1) * i(x+k_x, y+k_y)$
7: | | $i_E(x,y) \leftarrow s$
8: **return** (i_E, x_{max}, y_{max})

Cílem je získat explicitní znalost o hranách v obrázku ve formě seznamu úseček, aby bylo možné zkoumat charakteristiky jako je četnost *výskytu pravých úhlů*. Budeme postupovat v několika fázích. Nejprve použijeme standardní postup pro **detekci hran** založený na konvoluci, jehož výsledkem jsou hrany v implicitní podobě, přesněji **zvýraznění** hran. Konkrétně je využita konvoluce pomocí *Laplaceova operátoru* [8] vyjádřeného maticí $h = ((1,1,1);(1,-8,1);(1,1,1))$. Formálně je postup hranové detekce vyjádřen v pseudo-kódu jako Algoritmus 1. Výsledek procesu hranové detekce je ukázán na Obr. 7.

Obr. 7. Výsledek provedení *hranové detekce*. Výsledný obrázek se zvýrazněnými hranami je ukázán vpravo.

Další fází je převod implicitní informace o hranách do **explicitní** formy. K tomu slouží tzv. *Houghova transformace* [1]. Transformační algoritmus pro každý výrazný pixel obrázku (jehož jas je vyšší než práh θ_H) po hranové detekci určí všechny přímky, které

tímto pixelem procházejí a jsou vyjádřené předpisem $\rho = x * \cos\vartheta + y * \sin\vartheta$. Přesněji jsou tedy pro daný výrazný bod (x, y) spočítány parametry ϑ a ρ podle uvedeného předpisu (řešení existuje nekonečně mnoho - použije se diskrétní vzorkování).

Algoritmus 2. *Upravená **Houghova transformace** vyjádřená v pseudo-kódu.* Vstupem algoritmu je monochromatický obrázek I^e_{mono}, který vznikl hranovou detekcí. Výstupem algoritmu je množina úseček v explicitní formě.

function *Hough-Transformation* $(I^e_{mono} = (i^e, x^e_{max}, y^e_{max}), \varphi_{max}, \rho_{max}, \theta_H, \theta_S)$: **set**

$a_H \leftarrow \vec{0}$
for $x = 0,1,\dots,x^e_{max}$ **do**
  for $y = 0,1,\dots,y^e_{max}$ **do**
    if $i^e(x,y) \geq \theta_H$ **then**
      for $\varphi = 0,1,\dots,\varphi_{max}$ **do**
        $\rho \leftarrow x * \cos((\varphi/\varphi_{max})\pi) + y * \sin((\varphi/\varphi_{max})\pi)$
        $a_H(\varphi,\rho) \leftarrow a(\varphi,\rho) + 1$
$L \leftarrow \emptyset$
for $\varphi = 0,1,\dots,\varphi_{max}$ **do**
  for $\rho = 0,1,\dots,\rho_{max}$ **do**
    if a_H nabývá lokálního svého maxima v bodě (φ,ρ) **then**
      $L \leftarrow L \cup \{(\varphi,\rho)\}$
$L_S \leftarrow \emptyset$
for each $(\varphi,\rho) \in L_S$ **do**
  let $[(x_1,y_1),(x_2,y_2),\dots,(x_m,y_m)]$ je posloupnost (celočíselných) bodů přímky (φ,ρ) v obrázku I^e_{mono} uspořádaných podél přímky
  $(a_x, a_y) \leftarrow \perp$
  for $k = 1,2,\dots,m$ **do**
    if $(a_x,a_y) = \perp$ **then**
      if $i_E(x_k,y_k) \geq \theta_S$ **then**
        $(a_x,a_y) \leftarrow (x_k,y_k)$
    else
      if $i_E(x_k,y_k) < \theta_S$ **then**
        $(a_x,a_y) \leftarrow \perp$
        $L_S \leftarrow L_S \cup \{((a_x,a_y),(x_k,y_k),(\varphi,\rho))\}$
return L_S

Tabulka 2. Některé charakteristiky založené na detekovaných úsečkách.

Charakteristika	Výpočet
Délka nejdelší úsečky $d^{L_S}_{max}$	$d^{L_S}_{max} = \max\{d_k \mid k = 1,2,\dots,n\}$, kde $d_k = \sqrt{(a^k_x - b^k_x)^2 + \left(a^k_y - b^k_y\right)^2}$
Průměrná délka úseček $d^{L_S}_{\mu}$	$d^{L_S}_{\mu} = \frac{1}{n}\sum_{k=1}^{n} d_k$
Rozptyl délek úseček $d^{L_S}_{\sigma}$	$d^{L_S}_{\sigma} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(d_k - d^{L_S}_{\mu})^2}$
Podíl krátkých úseček $r^{L_S}_{short}$	$r^{L_S}_{short} = \frac{1}{n}\left\lvert\{d_k \mid 0 < d_k \leq \frac{1}{3}d^{L_S}_{max}\}\right\rvert$
Podíl středních úseček $r^{L_S}_{mid}$	$r^{L_S}_{mid} = \frac{1}{n}\left\lvert\{d_k \mid \frac{1}{3}d^{L_S}_{max} < d_k \leq \frac{2}{3}d^{L_S}_{max}\}\right\rvert$
Podíl dlouhých úseček $r^{L_S}_{long}$	$r^{L_S}_{long} = \frac{1}{n}\left\lvert\{d_k \mid \frac{2}{3}d^{L_S}_{max} < d_k \leq d^{L_S}_{max}\}\right\rvert$

Algoritmus 3. ***Detekce pravých úhlů***. Vstupem algoritmu je množina úseček L_S reprezentovaných explicitně. Dále pak interval úhlů φ^H_{min} až φ^H_{max}, které jsou považovány za horizontální a interval úhlů φ^V_{min} až φ^V_{max}, které jsou považovány za vertikální. Výstupem algoritmu množina dvojic úseček, které spolu svírají pravý úhel.

function *Detect-Right-Angles* $(L_S, \varphi^H_{min}, \varphi^H_{max}, \varphi^V_{min}, \varphi^V_{max})$: **set**
1: $A_R \leftarrow \emptyset$
2: **let** $L_S = \{((a_x^1, a_y^1), (b_x^1, b_y^1), (\varphi^1, \rho^1)), \dots, ((a_x^n, a_y^n), (b_x^n, b_y^n), (\varphi^n, \rho^n))\}$
3: **for** $k_1 = 1,2, \dots, n$ **do**
4: | **for** $k_2 = 1,2, \dots, n$ **do**
5: | | **if** $k_1 \neq k_2$ **then**
6: | | | **if** $\varphi^H_{min} \leq \varphi^{k_1} \leq \varphi^H_{max}$ **and** $\varphi^V_{min} \leq \varphi^{k_2} \leq \varphi^V_{max}$ **then**
7: | | | | **if** úsečka s krajními body $(a_x^{k_1}, a_y^{k_1}), (b_x^{k_1}, b_y^{k_1})$ protíná úsečku s krajními body $(a_x^{k_1}, a_y^{k_1}), (b_x^{k_1}, b_y^{k_1})$ **then**
8: | | | | | $A_R \leftarrow A_R \cup \{(k_1, k_2)\}$
9: **return** A_R

Výskyt přímky je pak indikován lokálním maximem funkce závislé na parametrech ϑ a ρ, která vyjadřuje počet výrazných bodů na dané přímce. Detekované přímky jsou následně rozdělené na úsečky podle toho, zda je jas pixelů na přímce dostatečný (pokles jasu pod práh θ_S indikuje nutnost rozdělení přímky na úsečky). Výsledkem je pak množina úseček $L_S = \{((a_x^1, a_y^1), (b_x^1, b_y^1), (\varphi^1, \rho^1)), \dots, ((a_x^n, a_y^n), (b_x^n, b_y^n), (b_x^n, b_y^n), (\varphi^n, \rho^n))\}$, kde (a_x^k, a_y^k) je počáteční bod, (b_x^k, b_y^k) je koncový bod k-té úsečky a (φ^k, ρ^k) jsou parametry přímky, na které k-tá úsečka leží.

Na základě této explicitní znalosti úseček lze definovat různé charakteristiky. Některé z nich ukazuje Tabulka 2. Z explicitní znalosti úseček lze dále extrahovat znalost o pravých úhlech, což je charakteristika zřejmě důležitá pro klasifikaci třídy budovy (B). Postup na zjištění pravých úhlů je popsán jako Algoritmus 3. Rozhodující charakteristika založená na detekci pravých úhlů použitá v navrhované klasifikační metodě je jejich počet, tedy hodnota $c^{A_R} = |A_R|$.

5 Experimentální vyhodnocení

Klasifikační metoda založená na popsaných příznacích a rozhodovacím stromu byla implementována v jazyce Java. Trénovací sady obrázků obsahovaly 350 ručně anotovaných obrázků celkem, přičemž pro každou klasifikační třídu bylo použito alespoň 40 trénovacích obrázků. Trénovací sady jsou dostupné na webu: http://ktiml.mff.cuni.cz/~surynek/research/znalosti2011. Na stejném webu jsou dostupné i dvě testovací sady obrázků nazývané *Sada A* a *Sada B*, se kterými byly provedeny experimenty. Sada A obsahovala 297 obrázků a Sada B 405 obrázků.

Tabulka 3. Úspěšnost klasifikace na třídě ***fotografie*** (**P**).

	Počet obrázků	**Správně klasifikované**	**Úspěšnost**
Trénovací sada	155	154	99.35%
Sada A	297	243	81.82%
Sada B	405	300	74.07%

Tabulka 4. Úspěšnost klasifikace na třídě ***umělecké*** obrázy (**A**).

	Počet obrázků	Správně klasifikované	Úspěšnost
Trénovací sada	104	104	100.00%
Sada A	297	251	84.51%
Sada B	405	331	81.73%

Tabulka 5. Úspěšnost klasifikace na třídě ***budovy*** (**B**).

	Počet obrázků	Správně klasifikované	Úspěšnost
Trénovací sada	104	104	100.00%
Sada A	297	232	78.11%
Sada B	405	350	86.42%

Vyhodnocována byla úspěšnost klasifikace testovacích obrázků vzhledem ke klasifikačním třídám P, A, B, L, M (jak je definována v Definici 2). Zároveň jsme zkoumali, které příznaky jsou pro danou třídu nejvýznamnější. Míra významnosti příznaků je přímo dána jejich výskytem a umístěním v rozhodovacím stromu – významnější příznaky se nacházejí blíže kořeni.

Tabulka 6. Úspěšnost klasifikace na třídě ***krajiny*** (**L**).

	Počet obrázků	Správně klasifikované	Úspěšnost
Trénovací sada	90	89	98,89%
Sada A	297	232	84.51%
Sada B	405	350	81.73%

Výsledky pro třídu **fotografie (P)** jsou ukázány v Tabulce 3. Jako nejvýznamnější charakteristiky se ukázaly: podíl barev z barevné palety, počet barev, počet lokálních maxim v histogramu a rozptyl v okolí lokálních maxim v histogramu. Pro fotografie je charakteristické že počet barev nedosahuje extrémních hodnot.

Výsledky pro třídu **umělecké obrazy (A)** jsou ukázány v Tabulce 4. Nejvýznamnějšími charakteristikami byly: počet lokálních maxim v histogramu, rozptyl v okolí lokálních maxim v histogramu a počet barev. Tato třída se vyznačuje omezeným počtem barev, které ale zabírají velkou plochu obrázku.

Výsledky klasifikace pro třídu **budovy (B)** jsou ukázány v Tabulce 5. Podle očekávání jsou nejvýznamnějšími příznaky: počet pravých úhlů, délka nejdelší úsečky a podíl dlouhých úseček. Tyto výsledky jsou v souladu s očekáváním.

Klasifikační výsledky pro třídu **krajiny (L)** ukazuje Tabulka 6. Zde jsou nejvýznamnějšími charakteristikami: podíl modrých odstínů, podíl světlých odstínů (high-key), kontrast obrázku a podíl tmavých odstínů (low-key). Lze říci, že podle očekávání hraje velkou roli kontrast. Důležitost modrých odstínů lze vysvětlit faktem, že s krajinou je typicky zobrazována i obloha.

Tabulka 7. Úspěšnost klasifikace na třídě ***makro objekty*** (**M**).

	Počet obrázků	Správně klasifikované	Úspěšnost
Trénovací sada	118	118	98,89%
Sada A	297	259	87.20%
Sada B	405	295	72.84%

Konečně výsledky pro třídu **makro objekty (M)** jsou ukázány v Tabulce 7. Nejdůležitějšími příznaky jsou: podíl nekontrastní plochy, podíl nekontrastní plochy

v okrajích obrázku, podíl světlých, středních a tmavých odstínů. Opět lze konstatovat, že podle očekávání hrál velkou roli lokální kontrast.

6 Závěr a budoucí práce

Navrhli jsme relativně úspěšnou metodu pro klasifikaci bitmapových obrázků do několika vybraných tříd popsaných přirozeným jazykem založenou na extrakci příznaků a rozhodovacím stromu. Úspěšnost současné implementace metody se pohybuje mezi 75% a 85%. Podali jsme formální analýzu úlohy a popsali důležité technologie pro návrh metody. Jako klíčový se ukázal návrh extrakce důležitých příznaků. Zejména netriviální příznaky, jako jsou hranové charakteristiky, se ukázaly být zvláště důležité. Do budoucna by bylo možné klasifikační metodu rozšířit o další příznaky za účelem zvýšení úspěšnosti klasifikace.

Reference

1. R. O. Duda, P. E. Hart. *Use of the Hough Transformation to Detect Lines and Curves in Picutres*. Communications of the ACM, ACM Press, 1972.
2. I. Lukšová. *Klasifikace bitmapových obrázků.* Bakalářská práce, Matematicko-fyzikální fakulta, Univerzita Karlova v Praze, 2010.
3. D. Marr, E. C. Hildreth. *Theory of edge detection.* Proceedings of the Royal Society of London, Series B, Volume 207 (1167), pp. 187–217, The Royal Society, 1980.
4. T. M. Mitchell. *Machine Learning.* McGraw Hill, 1997, ISBN: 978-0070428072.
5. E. Peli. *Contrast in Complex Images.* Journal of the Optical Society of America A: Optics, Image Science, and Vision, Volume 7, Issue 10 , pp.2032-2040, OSA The Optical Society, 1990.
6. J. R. Quinlan. *Induction of Decision Trees.* Machine Learning, Volume 1, pp. 81-106, Springer, 1986.
7. S. Russel P. Norvig. *Artificial Intelligence – A modern approach*. Prentice Hall, 2003, ISBN: 978-0137903955.
8. J. Žára, B. Beneš, J. Sochor, P. Felkel. *Moderní počítačová grafika (druhé vydání).* Computer Press, 2005, ISBN: 978-8025104545.

Annotation:

Attribute Extraction and Processing for Automated Classification of Bitmap Images

The paper describes a method for automated classification of bitmap images into classes such as images displaying building, landscape or illustration. The method is based on attribute extraction from the input image such as contrast, occurrence of lines and right angles. Extracted attributes are then processed by a decision tree. Approximately 75% to 85% of tested images were correctly classified in experiments.