

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE**  
**FAKULTA CHEMICKÉJ A POTRAVINÁRSKEJ TECHNOLOGIE**

**Evidenčné číslo: FCHPT-5414-68704**

**Analýza drevných vlákien z farebnej fotografie**

**Diplomová práca**

**Bratislava, 2016**

**Bc. Timea Némethová**

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE**  
**FAKULTA CHEMICKEJ A POTRAVINÁRSKEJ TECHNOLOGIE**

**Evidenčné číslo: FCHPT-5414-68704**

**Analýza drevných vlákien z farebnej fotografie**

**Diplomová práca**

<b>Študijný program:</b>	Automatizácia a informatizácia v chémii a potravinárstve
<b>Číslo študijného odboru:</b>	2621
<b>Názov študijného odboru:</b>	5.2.14 automatizácia
<b>Školiace pracovisko:</b>	Ústav informatizácie, automatizácie a matematiky
<b>Vedúci práce:</b>	Ing. Juraj Števek, PhD.

**Bratislava, 2016**

**Bc. Timea Némethová**



## ZADANIE DIPLOMOVEJ PRÁCE

Študentka: **Bc. Timea Némethová**  
ID študenta: 68704  
Študijný program: automatizácia a informatizácia v chémii a potravinárstve  
Študijný odbor: 5.2.14. automatizácia  
Vedúci práce: Ing. Juraj Števek, PhD.  
Miesto vypracovania: Oddelenie informatizácie a riadenia procesov (ÚIAM FCHPT)

Názov práce: **Analýza drevných vlákien z farebnej fotografie**

Špecifikácia zadania:

Cieľom tejto práce je automaticky rozlíšiť rôzne druhy drevných vlákien na základe farebnej fotografie z mikroskopu. Táto práca stavia na výsledkoch z publikácie (Jablonský a kol. 2015). Cieľom práce je vytvoriť balík funkcií, ktorá z databázy farebných fotografií vlákien vykoná požadované výpočty farebných súradníc a štatistické analýzy.

Zoznam odbornej literatúry:

1. Katuščák, S. – Kučera, L. – Varga, Š. New method of recognition of wood species. Increasing of the effectiveness of colorimetric recognition of picea excelsa and abies alba. *Wood Research*, s. 1–12.
2. Dubinyová, L. – Jablonský, M. – Varga, Š. – Fikar, M. – Katuščák, S.: Cellulose Materials Identification? The Effect of Dimensionality of Colour Photography Data. *BioResources*, č. 3, zv. 10, str. 71–86, 2016.
3. Michal Jablonský, Lenka Dubinyová, Štefan Varga, Katarína Vizárová, Jozef Šima a Svetozár Katuščák. Cellulose Fibre Identification through color vectors of stained fibre. *Bioresources* 3(10), 5845-5862, 2015

Riešenie zadania práce od: 15. 02. 2016  
Dátum odovzdania práce: 22. 05. 2016

L. S.

**Bc. Timea Némethová**  
študentka

**prof. Ing. Miroslav Fikar, DrSc.**  
vedúci pracoviska

**prof. Ing. Miroslav Fikar, DrSc.**  
garant študijného programu

# Podakovanie

Chcela by som poďakovať svojmu vedúcemu, Ing. Jurajovi Števekovi, PhD.,  
za cenné rady a odborné vedenie pri vypracovaní tejto práce.

Timea Némethová  
Bratislava, 2016

# Abstrakt

Táto práca sa zaoberá problematikou objektívnej klasifikácie drevných vlákien na základe vizuálnej informácie z mikroskopických fotografií týchto vlákien. Práca uvádza diskriminačnú analýzu ako jednu z možností ako klasifikovať. Analyzuje rôzne spôsoby prípravy (predspracovania) vzoriek pre túto analýzu. Prínosmi práce sú (1) dve metódy automatického výberu vzoriek z mikrofotografií drevných vlákien a ich experimentálne overenie, (2) experimentálne preskúmanie úspešnosti klasifikácie pri použití dvoch farebných schém a ďalších parametrov klasifikácie a (3) praktický vizuálny nástroj pre prostredie Matlab, pomocou ktorého je možné klasifikáciu drevných vlákien ľahko vykonávať.

**Kľúčové slová:** Identifikácia celulózných vlákien, Analýza materiálov, RGB, Helmholtzova farebná schéma, Diskriminačná analýza

# Abstract

This work deals with objective classification of wood fibers based on visual information from microscopic photographs of these fibers. The work presents the discriminative analysis as one of the ways of classification. It also analyzes various ways of preprocessing of samples for this analysis. The main contributions of this work are (1) two methods for automated selection of samples from microscopic photographs and their experimental evaluation, (2) the experimental exploration of classification success rate when using two different color schemes and other classification parameters, (3) a practical visual tool for Matlab environment, which can be easily used for classification of wood fibers.

**Keywords:** Fibre material identification, Materials analysis, RGB, Helmholtz color scheme, Discriminatory analysis

# Obsah

<b>1</b>	<b>Úvod</b>	<b>1</b>
1.1	Ciele a zameranie práce . . . . .	2
<b>2</b>	<b>Teoretická časť</b>	<b>4</b>
2.1	Teória farieb . . . . .	4
2.1.1	RGB systém . . . . .	5
2.1.2	Helmholtzove súradnice . . . . .	5
2.2	Optimalizácia . . . . .	8
2.2.1	Optimalizačná metóda Luus-Jaakola . . . . .	8
2.2.2	Sekvenčné kvadratické programovanie . . . . .	9
2.3	Základná teória pravdepodobnosti . . . . .	9
2.4	Diskriminačná analýza . . . . .	13
<b>3</b>	<b>Praktická časť</b>	<b>19</b>
3.1	Automatický výber regiónov pixelov . . . . .	20
3.1.1	Metóda krátkej spojnice . . . . .	22
3.1.2	Metóda vpísanej elipsy . . . . .	25
3.1.3	Porovnanie metód výberu regiónov . . . . .	26
3.2	Vyhodnotenie úspešnosti klasifikácie pri rôznych nastaveniach . .	28
3.2.1	Scenár a nastavenie experimentov . . . . .	29
3.2.2	Presnosť pre rôzne vstupné údaje . . . . .	30
3.2.3	Úspešnosť lineárnej a kvadratickej diskriminačnej analýzy	32
3.2.4	Zvyšné hľadiská (metódy výberu vzoriek a farebné systémy)	33
3.2.5	Zhrnutie experimentov a diskusia . . . . .	35
3.3	Nástroj pre identifikáciu drevných vlákien pre prostredie Matlab .	37
<b>4</b>	<b>Zhodnotenie</b>	<b>42</b>

<b>Literatúra</b>	<b>43</b>
<b>A Podrobné výsledky experimentov</b>	<b>44</b>
<b>B Odvodenie vzťahov pre LDA a KDA</b>	<b>50</b>
<b>C Fotografie použité v experimentoch</b>	<b>52</b>



# Zoznam obrázkov

2.1	Zmiešavanie základných farieb RGB. . . . .	5
2.2	Schéma na určovanie atribútov farieb podľa Helmholtza. . . . .	7
2.3	Gaussova krivka normálneho rozdelenia . . . . .	12
2.4	Na obrázku vidíme dva spôsoby akými pomyselná priamka rozdeľuje dve triedy. Prvá priamka je nevyhovujúca, pretože body z jednej triedy sa nachádzajú v priestore druhej triedy a naopak. Druhá priamka separuje triedy oveľa lepšie. . . . .	14
2.5	Na obrázku je znázornená separácia tried pre Kvadratickú diskriminačnú analýzu. . . . .	15
2.6	Na obrázku sú dve triedy (trieda 1 a trieda 2) a obe sú zložené z objektov reprezentovaných dvojzložkovými farebnými vektormi. Na osi $x$ je červená farba ( $R = \text{red}$ ), na osi $y$ je modrá farba ( $B = \text{blue}$ ). . . . .	17
3.1	Mikrofotografia dreveného vlákna. Modrou farbou je vyznačený možný reprezentatívny región. Cieľom našich metód je takýto región vybrať automaticky. . . . .	21
3.2	Vľavo: pôvodná fotografia, vpravo: vytvorená maska . . . . .	22
3.3	Vľavo: vytvorená maska, v strede: upravená maska, vpravo: vybrané najväčšie vlákno . . . . .	23
3.4	Hranice regiónov pozadia . . . . .	23
3.5	Fotka s vybraným regiónom . . . . .	24
3.6	Elipsa vpísaná do polygónu. Červená úsečka predsavuje úzky polomer elipsy a zároveň miesto výberu vzorky pixelov. . . . .	26
3.7	Fotografia s vybraným regiónom . . . . .	31

3.8	Zhluky v priestore (RGB) zodpovedajúce jednotlivým triedam. Možno vidieť, že vzorky pripravené priemerovaním skupín pixelov (vľavo) sú kompaktnějšíe ako vzorky pozostávajúce zo surových pixelov (vpravo). . . . .	37
3.9	Nástroj pre identifikáciu drevných vlákien pre Matlab. Rozhranie pre tréovanie klasifikátora. . . . .	38
3.10	Nástroj pre identifikáciu drevných vlákien pre Matlab. Rozhranie pre vykonanie a testovanie klasifikácie. . . . .	40

# Zoznam tabuliek

3.1	Výsledky spustenia metód pre 18 rozličných fotografií drevných vláken. . . . .	28
3.2	Porovnanie presnosti klasifikácie pre tri rôzne vstupné dátové sady a dva rôzne klasifikátory . . . . .	33
3.3	Prehľad výsledných presností klasifikácie pre rôzne alternatívy v troch rôznych hľadiskách. . . . .	34
3.4	Indexy rozlíšiteľnosti pre všetky dvojice tried, vypočítané s použitím RGB a Helmholtzovho systému farieb. . . . .	35
A.1	Porovnanie presnosti klasifikátorov s použitím pôvodnej sady údajov z článku Jablonského. Tabuľka uvádza počty pokusov klasifikácie v 20 experimentoch. V jednom experimente bolo pre každú triedu vykonaných 100 pokusov. . . . .	45
A.2	Porovnanie presnosti klasifikátorov s použitím pôvodných fotografií ale rozličných vzorkových regiónov. Tabuľka uvádza počty pokusov klasifikácie v 20 experimentoch. V jednom experimente bolo pre každú fotografiu (celkovo 17) vykonaných 10 pokusov. . . . .	46
A.3	Porovnanie presnosti klasifikátorov s použitím rozšírenej množiny fotografií rozdelenej na trénovaciu a testovaciu podmnožinu. Tabuľka uvádza počty pokusov klasifikácie v 20 experimentoch. V jednom experimente bolo pre každú testovaciu fotografiu (celkovo 17) vykonaných 10 pokusov. . . . .	47
A.4	Podrobné výsledky klasifikácie pre alternatívne farebné systémy a metódy výberu regiónov za použitia prímerov skupín pixelov ako vzorkovacieho prístupu pre trénovaciu množinu údajov. 20 experimentov po 170 pokusoch (17 testovacích fotografií po 10 pokusov). . . . .	48

A.5	Podrobné výsledky klasifikácie pre alternatívne farebné systémy a metódy výberu regiónov za použitia surových pixelov ako vzorkovacieho prístupu pre trénovaciu množinu údajov. 20 experimentov po 170 pokusoch (17 testovacích fotografií po 10 pokusov). . . . .	49
-----	--	----

# 1. Úvod

Bežnou úlohou v priemyselnej praxi je identifikácia neznámych materiálov. Vo všeobecnosti sa takáto úloha vyskytuje v mnohých profesiách: lekárstve, potravinárstve, poľnohospodárstve, drevárstve, forenznnej analýze a mnohých ďalších. Existuje veľký počet metód, ktorými sa identifikácia vykonáva. Využíva sa chemická analýza, spektrometria ale aj jednoduchšie makroskopické pozorovanie a následná identifikácia človekom. Existujúce metódy sa navzájom kvalitatívne líšia v aspektoch ako dostupnosť (cena), náročnosť (na uskutočnenie identifikácie) a presnosť. Keďže tieto vlastnosti sú často v protiklade, existuje zrejma motivácia hľadať nové metódy, ktoré by poskytovali výhodný pomer “ceny a výkonu”.

Máloktorá z existujúcich metód je univerzálna. Pri niektorých typoch materiálov už existujú praxou overené prístupy identifikácie. V našej práci sa zameriavame na drevné vlákna. Presnejšie, zaoberáme sa problémom klasifikácie vzorky drevného vlákna do niektorej z dopredu definovaných tried. Identifikácia drevných vlákien má využitie v mnohých oblastiach ako produkcia textílií, ochrana kultúrneho dedičstva, archeológia, poľnohospodárstvo a pod. Podobne ako materiály vo všeobecnosti, aj drevné vlákna možno identifikovať viacerými prístupmi: skúmaním farebných vlastností, skúmaním mechanických vlastností, chromatografiou plynov, rôznymi druhmi spektroskopie, termomechanickou či mikroskopickou analýzou. Aj tu platí, že sa metódy identifikácie líšia cenou, náročnosťou aj presnosťou.

Existujúce metódy identifikácie drevných vlákien možno rozdeliť aj podľa ďalšieho hľadiska: objektivity. Za objektívne považujeme metódy, ktoré identifikujú látku bez rozhodovania človeka, kým subjektívne toto rozhodovanie vyžadujú (samozrejme, možno rozlišovať rôzne miery objektivity). Subjektívne metódy majú nevýhodu práve v ľudskom faktore. Človek sa môže pomýliť, jeho výkon nie je stabilný a často krát ho musíme na danú úlohu školiť. Objektívne metódy majú výhodu konzistentnosti, fungujú okamžite. Zároveň sú však objektívne me-

tódy často drahé (vyžadujú drahé zariadenia), zatiaľ čo subjektívne metódy sú lacnejšie, pretože vyžadujú len jednoduché zariadenia a opierajú sa o expertné schopnosti človeka.

Jednou z tendencií výskumu je subjektívne metódy objektivizovať, teda nahradiť kroky vykonávané človekom strojovým spracovaním (s motiváciou odstrániť nevýhody subjektivity a ponechať výhodu ceny). Keďže sa subjektívne metódy v drvivej väčšine opierajú o spracovanie vizuálnej informácie človekom, jednou z príležitostí je využitie metód počítačového videnia a umelej inteligencie (najmä strojového učenia) na vykonávanie vizuálnej analýzy. Často používanou subjektívnou metódou identifikácie drevných vlákien je farbenie vzoriek rozlišovacím farbivom (takým, ktoré pozmení farbu vlákien tak, aby sa jednotlivé druhy človeku ľahšie identifikovali) a následná mikroskopická vizuálna analýza. Práve v tomto postupe možno spracovanie obrazu počítačom využiť – na mieste vizuálnej analýzy.

## 1.1 Ciele a zameranie práce

Cieľom našej práce je objektivizovať hodnotenie drevných vlákien, čiže vylúčiť ľudské subjektívne hodnotenie v kvalitatívnom posudzovaní drevných produktov. Praktickým výstupom práce je program, ktorý vie automaticky rozoznávať rôzne typy drevných vlákien z mikrofotografií.

V tejto práci sa zameriavame na určovanie typov vlákien na základe farby, kde vzorky vyhodnocujeme na mikroskopickej úrovni. Existujú pritom dve základné vizuálne vlastnosti drevnej hmoty, farba a morfológia, pričom v tejto práci sa zameriavame len na farbu a morfológiu nevyužívame. V tomto prístupe prebieha klasifikácia na základe farebných súradníc, ktoré získame z farebných fotografií vlákien. Identifikácia iba na základe farby je zjednodušený prístup, je však dostatočne reprezentatívny. V rámci práce sme najskôr zreprodukovali experiment založený na tomto prístupe, ktorý bol opísaný v článku [3]. Potom sme uskutočnili ďalšie experimenty, v ktorých sme experimentovali s inými nastaveniami a vstupnými hodnotami prístupu, predovšetkým s farebnými vektormi, ktoré sú založené na dvoch farebných schémach (RGB a Helmholtz). Zamerali sme sa tiež na predspracovanie fotografií, ktorého úlohou je výber vzoriek do klasifikácie. V práci sme navrhli automatický postup na výber reprezentatívnych pixelov z fotografie do analýzy, ktorý odstraňuje subjektívny výber experta.

Za najdôležitejšie prínosy práce považujeme:

- Vytvorenie dvoch automatických metód extrakcie reprezentatívnych pixelov z mikrofotografií drevných vlákien (pre zvýšenie automatizácie a teda

aj objektivity identifikácie) a ich experimentálne overenie.

- Experimentálne preskúmanie rozlišovacej schopnosti dvoch farebných schém pri klasifikácii mikrofotografií drevných vlákien.
- Vytvorenie praktického nástroja na vykonávanie klasifikácie mikrofotografií drevných vlákien pre prostredie Matlab.

Naše metódy identifikácie drevných vlákien sa opierajú o viaceré modely, princípy a algoritmy. V rámci teoretickej časti práce ich preto postupne uvádzame. Najskôr sa venujeme teórii farieb, keďže identifikácia má prebiehať na základe získanej farebnej informácie z mikrofotografií drevných vlákien. Predovšetkým sa zameriavame na dve farebné schémy (RGB, Helmholtz), ktoré definujú farbu pomocou rôznych atribútov. Ich rozlišovaciu schopnosť budeme v práci tiež analyzovať. Samotnú identifikáciu budeme vykonávať využitím diskriminačnej analýzy [2]. Jej princíp uvádzame v teoretickej časti práce.

Jedným z prínosov práce je aj automatická extrakcia reprezentatívnych pixelov z mikrofotografií drevných vlákien. Dve metódy, ktoré sme na extrakciu vytvorili využívajú rozličné metódy spracovania obrazu (napr. prahová identifikácia regiónov, morfológické úpravy) a optimalizácie (Luus-Jaakola, nekonvexná nelineárna optimalizácia), ktorým sa časť analýzy tak isto venuje. Samotný opis metód a ich experimentálne overenie sa nachádza v praktickej časti práce.

## 2. Teoretická časť

### 2.1 Teória farieb

Každý objekt sa dá vidieť ako farba alebo kombinácia farieb. Fyziologický proces videnia je reakciou na kontrastné pôsobenie farieb, farby ako vlastnosti odrazeného svetla pôsobia na sieťnicu oka. Prostredníctvom farebného videnia získavame 70 - 90% informácií. Tiež môžeme povedať, že farba je súčasťou tzv. Ekologickej kvality (EQ) materiálov, technológií architektonických diel a tiež výrobkov a výrobových komplexov. S farebnosťou a stavom optických vlastností systému súvisí vnútorný obsah a celková kvalita, napr. charakteristický priebeh zmien farebnosti pri stárnutí a zániku materiálov.

Na základe tohto je zrejmé, že analýzou farieb sa vieme mnohé dozvedieť o povahe materiálov, ktoré pozorujeme (a to jednak priamo ľudským pozorovaním, alebo “imitáciou” tohto pozorovania počítačom). Vieme napríklad určovať druhy materiálov, čím sa zaoberá aj táto práca.

Vedecké skúmanie farieb sa datuje od 16. - 17. storočia a je založené na exaktných výskumoch. Dnes už vieme vytvárať rôzne knižnice materiálov alebo katalógy, ktoré sa využívajú napr. na testovanie a predvídanie starnutia a životnosti. Na margo objektívnosti skúmania však treba povedať, že farba nie je fyzikálna veličina, ale predstavuje subjektívny ľudský vnem. Farbe možno priradiť fyzikálnu charakteristiku (napr. vlnovú dĺžku svetla) ale jej vnímanie bude stále závisieť od pozorovateľa.

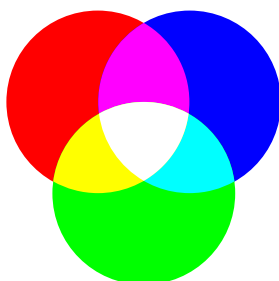
Podľa [4] je farba definovaná nasledovne: “Farba je taká charakteristika viditeľného žiarenia, pomocou ktorej môže pozorovateľ rozoznať rozdiely medzi dvoma pozorovanými poliami bez štruktúry povrchu rovnakého tvaru a veľkosti, ktoré môžu byť zapríčinené rozdielmi v spektrálnom zložení radiačnej energie zúčastnenej v pozorovacom procese. Psychofyzikálna farba je špecifikovaná trisťmiulárnymi hodnotami radiačnej energie, ktorá dopadá do oka.”



Farba sa dá určovať viacerými spôsobmi. V tejto práci sa zameriavame na špecifikáciu farby prostredníctvom RGB systému a Helmholtzových súradníc.

### 2.1.1 RGB systém

RGB je aditívny farebný model, pri ktorom svetlo želanej farby vzniká zmiešaním červeného (R-red), zeleného (G-green) a modrého (B-blue) svetla vhodnej intenzity. Všetky viditeľné farby sú kombináciou týchto troch základných farebných zložiek. Miešanie farieb je založené na aditívnom spôsobe, čiže sčítaním (pridávaním farby k farbe). Pre uchovanie v pamäti počítačov sa používa 256 (0-255) odtieňov každej farby. To zodpovedá 8 bitom, čiže jednému bajtu. Takto vie vzniknúť až 16 777 216 farieb. Z toho vyplýva, že každá farba je tvorená tromi bajtmi. Prvý bajt udáva odtieň červenej farby, druhý zelenej a tretí modrej. Príklad : najsýtejšia červená je daná kombináciou 255 0 0, najsýtejšia zelená je 0 255 0 a najsýtejšia modrá je 0 0 255.



Obr. 2.1: Zmiešavanie základných farieb RGB.

Model RGB sa pri práci s digitálnymi fotografiami používa najčastejšie, pretože zodpovedá spôsobu, akým sa informácia o farbe uchováva. Jeho nevýhodou však je, že je neprírodný pre človeka. Ľudia spravidla nie sú schopní pre ľubovoľnú farbu určiť číselne jej RGB zložky. Ak sa teda zaoberáme problémom, v ktorom chceme od človeka, aby farby rozlišoval a vedel ich opísať, prípadne pomenovať, musíme využiť farebnú schému, ktorá je spôsobu ľudského vnímania farieb bližšia, napríklad Helmholtzovu schému.

### 2.1.2 Helmholtzove súradnice

V Helmholtzových súradniciach je farba opísaná tromi atribútmi, a síce dominantnou vlnovou dĺžkou  $\lambda_D$ , excitačnou čistotou  $p_E$  a luminiscenciou  $L$  [4].

Na určovanie farieb v Helmholtzovej schéme sa používa grafická schéma na obrázku 2.2. V nej môžeme vidieť farebné spektrum v tvare podkovy. V jeho priestriedku sa nachádza achromatický (bezfarebný) bod  $W$  a okraj schémy zodpovedá najsýtejším farbám a slúži na určovanie dominantnej vlnovej dĺžky.

Farba určená podľa Helmholtzovej schémy má nasledovné atribúty:

- **Dominantná vlnová dĺžka**  $\lambda_D$  reprezentuje vlnovú dĺžku monochromatického svetla, ktoré má rovnakú farbu ako naša vzorka. Zodpovedá farbe takej vlnovej dĺžky, ktorá keď sa primieša k určitému množstvu achromatickej farby (achromatické farby sú čierna, biela, sivá; čiže sú to farby, ktoré nemajú farebný odtieň), poskytne požadovanú farbu (farbu s uvažovanou  $\lambda_D$ ). Vlnová dĺžka farby prepája svet ľudského vnímania farieb so svetom fyzikálnym, ktorý je všeobecne zrozumiteľný vo všetkých oblastiach (napr. biológia, medicína).
- **Excitačná čistota**  $p_E$  farby je pomer vzdialenosti danej farby od achromatického bodu k celkovej vzdialenosti medzi achromatickým bodom a spektrálne čistou farbou danej dominantnej vlnovej dĺžky  $\lambda_D$ .
- **Žiarivosť/luminiscencia**  $L$  je atribútom farby určujúcim jej vnímanú svietivosť resp. svetlosť. Presnejšie, každej farbe možno priradiť odtieň sivej, ktorý by vznikol, ak by sme z farby odstránili všetku "farebnosť". Tieto odtiene sivej možno zoradiť od čiernej po bielu, pričom čierna má nulovú a biela najvyššiu žiarivosť. Žiarivosť farby je potom určená žiarivosťou sivej, podľa toho, kde sa táto sivá medzi bielou a čiernou nachádza.

Hodnoty Helmholtzových atribútov farby možno vypočítať z jej RGB hodnôt (postup uplatňujeme aj my v praktickej časti práce). Postup spočíva v prepočte cez ďalšie farebné schémy. Prvým krokom je preklad RGB súradníc do tzv. CIE schémy (pomenovanej podľa Medzinárodnej komisie osvetľovania, organizácie, ktorá skúma farby a okrem iného definuje štandardy ich reprezentácie). CIE schéma sa označuje ako XYZ, podľa zložiek, ktoré tieto tri písmená označujú:

- X predstavuje lineárnu kombináciu kriviek reakcie čapíkov na svetlo danej farby (buniek ľudského oka, citlivých na farbu).
- Y predstavuje svetlosť. Tejto hodnote v Helmholtzovej schéme zodpovedá hodnota  $L$ .
- Z približne zodpovedá modrej stimulácii (čapíka reagujúceho na modrú).

Vypočítat koordináty XYZ z RGB možno podľa nasledujúceho vzorca:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.769 & 1.752 & 1.130 \\ 1 & 4.590 & 0.060 \\ 0 & 0.566 & 5.594 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.1)$$

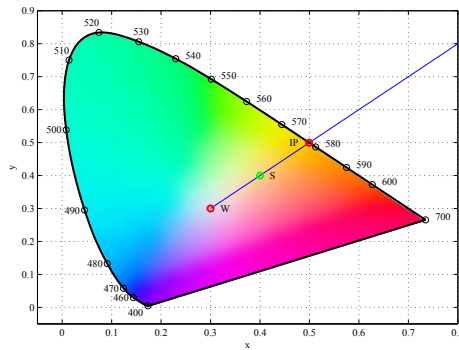
Farebnosť (chromaticita) farby je následne určená hodnotami  $x$  a  $y$ , ktoré dostaneme normalizáciou:

$$x = \frac{X}{X + Y + Z} \quad (2.2)$$

$$y = \frac{Y}{X + Y + Z} \quad (2.3)$$

Vypočítané  $x$  a  $y$  následne vieme použiť pri výpočte Helmholtzových koordinátov  $\lambda_D$  a  $p_E$ .  $x$  a  $y$  určujú bod  $S$ , cez ktorý prekladáme priamku smerujúcu z achromatického bodu  $W$  k okraju diagramu (pozri obrázok 2.2). Tam, kde táto priamka pretína okraj “podkovy” sa nachádza bod  $IP$ , ktorý určuje dominantnú vlnovú dĺžku  $\lambda_D$ . Zároveň pomer vzdialeností  $W-S$  a  $W-IP$  určuje hodnotu excitačnej čistoty  $p_E$ .

$$p_E = \frac{W - S}{W - IP} \quad (2.4)$$



Obr. 2.2: Schéma na určovanie atribútov farieb podľa Helmholtza.

Charakter Helmholtzových atribútov, na rozdiel od RGB, má bližšie k fyzikálnej podstate svetla (atribút dominantnej vlnovej dĺžky) a okrem toho oveľa

lepšie vystihuje spôsob, akým by človek opisoval farbu. Táto schéma preto môže mať význam pri úlohách, pri ktorých človek farebne rozlišuje nejaké objekty. V našej práci sa zaoberáme takouto úlohou – rozlišovaním typov drevných vlákien.

## 2.2 Optimalizácia

V rámci tejto práce sme vytvorili dve automatické metódy, ktorých cieľom je extrakcia reprezentatívnych pixelov z mikrofotografií. Táto extrakcia má nahradiť manuálny výber vzoriek z fotografie človekom. V rámci týchto metód teda musíme automaticky spracúvať mikrofotografie. Predovšetkým v nich musíme identifikovať regióny, v ktorých sa samotné vlákna nachádzajú (odlíšiť ich od “pozadia”). Následne určujeme, ktorá časť vlákien je najvhodnejšia na vytvorenie vzorky (všeobecne nám ide o hľadanie priečnych rezov na miestach, kde vlákna nie sú nevhodne ohnuté, prekryté či rozmazané).

Okrem iného pritom využívame dve všeobecné optimalizačné metódy: (1) metódu Luus-Jaakola a (2) metódu Sekvenčného kvadratického programovania.

### 2.2.1 Optimalizačná metóda Luus-Jaakola

Optimalizačná metóda Luus-Jaakola [8] patrí medzi tzv. bezgradientové metódy, to znamená, že nepotrebujeme poznať gradient účelovej funkcie, čo umožňuje jej použitie pre nediferencovateľné a nekonvexné úlohy. Táto metóda je založená na náhodnom hľadaní optima. Tzn., že záruku nájdenia globálneho optima nemáme. Avšak, algoritmus v praktickom použití poskytuje veľmi dobré výsledky.

Princíp metódy spočíva v tom, že v okolí  $d$ , ktoré si zvolíme náhodne vyberieme bod  $x_0$  a uložíme si jeho funkčnú hodnotu. Potom okolie  $d$  prispôbujeme bodu  $x_0$  tak, že  $x_0 + d$  a  $x_0 - d$ . V tomto okolí sa náhodne zvolí ďalší bod  $x_1$ . Ak je jeho funkčná hodnota menšia ako funkčná hodnota predchádzajúceho bodu, presunieme sa doň a veľkosť okolia bude rovnaká ako v predchádzajúcom kroku. Ak je však funkčná hodnota väčšia, zostaneme v bode v ktorom sa práve nachádzame a okolie zmenšíme (miera zmenšenia je parametrom algoritmu). Takto postupujeme až kým sa neuplatní zastavovacie kritérium (minimálna veľkosť okolia  $d$ , ktorú si tiež určíme na začiatku).

## 2.2.2 Sekvenčné kvadratické programovanie

Sekvenčné kvadratické programovanie (SQP) predstavuje súbor metód, ktoré efektívne riešia konvexné optimalizačné problémy [5]. Pri SQP optimalizujeme funkciu  $f$  pri obmedzeniach  $g$ .

$$\begin{array}{ll}\min_{x \in R^n} & f(x) \\ \text{s.t.} & g(x) \leq 0_m\end{array}$$

SQP metódy sú iteratívne, teda začnú na nejakom iniciálnom riešení problému (nemusí byť nutne prípustné) a postupne hľadajú zlepšujúce riešenia (také, ktoré sú prípustné vzhľadom na obmedzenia a zároveň lepšie ako predchádzajúce nájdené). V každej iterácii sa testuje prípustnosť aktuálneho riešenia a hľadá sa vektor smerom k ďalšiemu riešeniu, ktorého výpočet berie do úvahy zlepšenie hodnoty účelovej funkcie a zníženie hodnoty tzv. penalizačnej funkcie (ktorá pre aktuálne riešenie vypočíta “ako veľmi” porušuje dané obmedzenia). Vďaka použitiu penalizačnej funkcie nie sú obmedzenia ostré a v priebehu riešenia sa môže algoritmus pomerne dlho pohybovať v neprípustných riešeniach a nakoniec dospieť k riešeniu prípustnému [5].

Metódy SQP sú schopné riešiť konvexné problémy. Konvexný problém je taký, v ktorom lineárne obmedzenia (nerovnosti)  $g$  v rámci ktorých optimalizujeme funkciu  $f$  musia tvoriť konvexný priestor. Uvažujme napríklad problém hľadania najväčšej možnej elipsy vpísanej do polygónu. V tejto úlohe maximalizujeme funkciu obsahu elipsy. Obmedzeniami sú pre nás nerovnosti “strážiacie” či polohy jednotlivých bodov z obvodu elipsy neprekročili hranicu polygónu. Pomocou SQP možno riešiť aj nekonvexné problémy, ale bez záruky, že bude nájdené globálne optimum.

## 2.3 Základná teória pravdepodobnosti

V tejto sekcii uvádzame základné vzťahy z teórie pravdepodobnosti, pretože ich budeme neskôr používať pri lineárnej diskriminačnej analýze (LDA) a kvadratickej diskriminačnej analýze (KDA).

K tomu aby sme mohli realizovať diskriminačnú analýzu potrebujeme vedieť, čo znamenajú nasledujúce pojmy:

- (náhodný) elementárny jav
- náhodná udalosť

- pravdepodobnosť
- pravdepodobnostný priestor
- náhodná premenná
- pravdepodobnostné rozdelenie (Gaussovské)
- stredná hodnota náhodnej premennej
- smerodajná odchýlka
- variancia náhodnej premennej

*Elementárny jav* je nedeliteľný výsledok nejakého deja. Ak dopredu nevieme predpovedať, či tento jav nastane, môžeme povedať, že je náhodný. Napríklad výsledkom hodu kockou môže byť, že padne trojka.

*Náhodná udalosť* je množina elementárnych javov ktoré môžu nastať. Náhodné udalosti označujeme veľkými písmenami latinskej abecedy. Náhodnou udalosťou je tak napríklad to, že pri hode kockou padne trojka (vtedy je v množine jeden elementárny jav - padnutie trojky) ale aj to, že padne trojka alebo štvorka (vtedy sú v množine dva elementárne javy, padnutie trojky a padnutie štvorky). Náhodná udalosť pritom nemusí obsahovať žiaden elementárny jav (vtedy ide o tzv. nemožnú udalosť) ale aj všetky elementárne javy, vtedy ide o istú udalosť).

Náhodnej udalosti možno priradiť jej *pravdepodobnosť*. Pravdepodobnosť je číslo od 0 po 1 vyjadrujúce do akej miery je pravdepodobné (očakávateľné), že daná náhodná udalosť nastane. Štatisticky ju definujeme ako podiel počtu pokusov, kedy sa vyskytla náhodná udalosť a počtu všetkých pokusov.

Ďalej definujeme *výberový priestor*  $\Omega$ . Je to množina elementárnych javov, ktoré môžu nastať (resp. ktoré uvažujeme). Napríklad v prípade hodu kockou ide o šesť javov (hody 1 až 6). Vo výberovom priestore sa musia nachádzať všetky javy, ktoré pri pokuse môžu nastať. Zároveň sa tieto javy musia navzájom vylučovať (nemôžu žiadne dva nastať súčasne). Súčet pravdepodobností elementárnych javov je 1.

S uvedenými pojmi súvisí pojem *pravdepodobnostný priestor*, ktorý je definovaný ako trojica  $(\Omega, \mathcal{F}, \mathbb{P})$ , pričom

- $\Omega$  je výberový priestor.
- $\mathcal{F}$  je množina náhodných udalostí, ktoré môžu nastať (ktoré uvažujeme).

- $\mathbb{P}$  je funkcia pravdepodobnosti, ktorá každej náhodnej udalosti vráti jej pravdepodobnosť.

*Náhodná premenná* je zobrazenie  $X : \Omega \rightarrow \mathbb{R}$ , ktoré elementárnym javom priradzuje ich pravdepodobnosť. Náhodnou premennou je také zobrazenie  $X : \Omega \rightarrow \mathbb{R}$  ak pre každé  $x \in \mathbb{R}$  platí

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F} \quad (2.5)$$

Náhodné premenné môžu byť diskkrétne alebo spojité, podľa toho, či sú javy diskkrétne (vymenovatelné) alebo spojité. Pri diskkrétnych náhodných premenných možno pre každý jav určiť jeho pravdepodobnosť. Pri spojitých náhodných premenných určujeme pravdepodobnosť či sa náhodná premenná dostane do zadaného intervalu.

Pre spojitú náhodnú premennú môžeme definovať zákon rozdelenia pomocou funkcie hustoty pravdepodobnosti.

Pravdepodobnostné (Gaussovske, normálne) rozdelenie: je rozdelenie pravdepodobnosti spojitej náhodnej premennej. Graficky ho môžeme znázorniť pomocou tzv. Gaussovskej krivky (obr. 2.3). Gaussovská krivka zobrazuje rozdelenie pomocou funkcie hustoty pravdepodobnosti. Toto rozdelenie je definované v tvare:

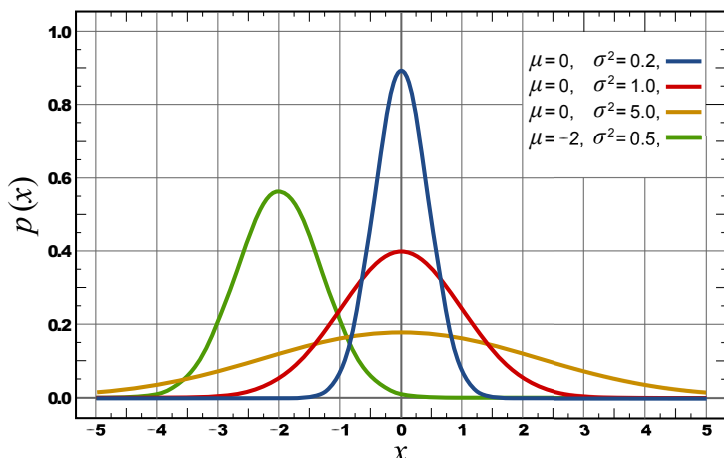
$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.6)$$

kde:

- $\mu$  - stredná hodnota
- $\sigma$  - odchýlka od strednej hodnoty - určuje ako veľmi sú hodnoty jednotiek odchýlené od priemernej hodnoty (je to priemer vzdialeností jednotiek súboru od strednej hodnoty).
- $\sigma^2$  - variancia - je to kvadratická odchýlka nameraných hodnôt od priemeru; je definovaná vzťahom:

$$\sigma^2 = \int_{-\infty}^{\infty} p(x)(x - \mu)^2 dx \quad (2.7)$$

- $x$  - náhodná premenná - je číselné vyjadrenie náhodnej udalosti
- $p(x)$  - funkcia hustoty pravdepodobnosti - túto funkciu znázorňuje Gaussova krivka



Obr. 2.3: Gaussova krivka normálneho rozdelenia

Ak pracujeme s viacerými náhodnými premennými, môžeme ich združiť do *náhodného vektora*.

Dôležitým pojmom pre našu prácu je *kovariancia*, ktorú definujeme ako súčin odchýliek dvoch náhodných premenných. Pracovať s ňou potrebujeme predovšetkým v tzv. *kovariančnej matici*, čo je matica, ktorá má na priesečníku  $i$ -teho riadku a  $j$ -teho stĺpca kovarianciu medzi  $i$ -tým a  $j$ -tým prvkom náhodného vektora. Dôsledkom toho okrem iného je, že na diagonále sa nachádzajú variancia jednotlivých náhodných premenných. Vzťah pre kovarianciu vyzerá nasledovne:

$$\text{cov}(x, y) = E[(x - E[x])(y - E[y])] \quad (2.8)$$

kde:

- $x$  a  $y$  - náhodné premenné
- $E[\ ]$  - stredná hodnota (náhodnej premennej)

Ďalej si vysvetlíme *podmienujú pravdepodobnosť* (podľa [6]). Pri podmienenej pravdepodobnosti sa zaoberáme problémom určenia pravdepodobnosti nejakej náhodnej udalosti, pričom je známa doplňujúca informácia, týkajúca sa výskytu inej náhodnej udalosti, ktorá môže túto pravdepodobnosť ovplyvniť.



Definícia hovorí: Nech  $(\Omega, \mathcal{F}, \mathbb{P})$  je pravdepodobnostný priestor a  $A$  a  $B$  náhodné udalosti a  $\mathbb{P}(B) > 0$ . Potom pravdepodobnosťou udalosti  $A$  podmienenou udalosťou  $B$  nazývame podiel.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (2.9)$$

Ďalej si zadefinujeme pojem úplný systém udalostí (podľa [6]). Nech  $(\Omega, \mathcal{F}, \mathbb{P})$  je pravdepodobnostný priestor a  $H_1, \dots, H_n$  náhodné udalosti, také že:

- sú po dvoch disjunktné, tj.  $H_i \cap H_j = \emptyset$  pre každé  $i \neq j$ ;  $i, j = 1, \dots, n$
- $\bigcup_{i=1}^n H_i = \Omega$

Potom ich nazývame úplný systém udalostí.

V tejto práci budeme vychádzať z Bayesovej vety. Bayesova veta je definovaná nasledovne [6]: Nech  $(\Omega, \mathcal{F}, \mathbb{P})$  je pravdepodobnostný priestor a nech náhodné udalosti  $H_1, \dots, H_n$  tvoria úplný systém udalostí. Nech ďalej  $A \in \mathcal{F}$  je náhodná udalosť s kladnými pravdepodobnosťami. Potom

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(A|H_i) \cdot \mathbb{P}(H_i)}{\mathbb{P}(A)} \quad (2.10)$$

## 2.4 Diskriminačná analýza

V tejto sekcii opíšeme čo je diskriminačná analýza, tiež sa budeme venovať pojmom lineárna diskriminačná analýza (LDA) a kvadratická diskriminačná analýza (KDA).

Cieľom diskriminačnej analýzy je zaradiť objekty do tried podľa nejakých vlastností. Napríklad, pri rastlinách môžeme sledovať veľkosť, tvar a farbu kvetu, tvar listov (vlastnosti) pričom ich chceme zaradiť do druhov, rodov či čeladi (triedy).

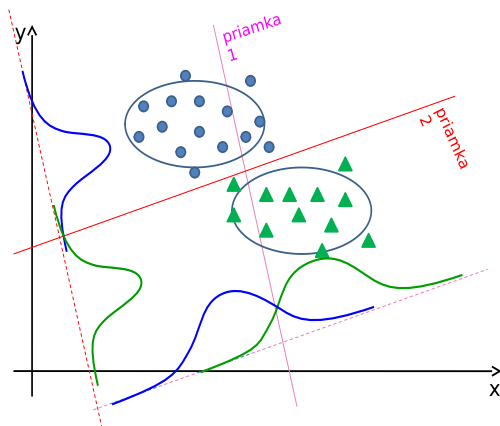
Lineárna diskriminačná analýza a kvadratická diskriminačná analýza vychádzajú zo vzťahov uvedených v predchádzajúcej sekcii, ktoré predpokladajú Gaussovské rozdelenie prvkov v triedach vzhľadom na ich vlastnosti.

Princípom diskriminačnej analýzy je hľadanie takého matematického modelu, ktorý maximalizuje pravdepodobnosť pridelenia (správnej) kategórie pre neznámy objekt. Modelom je rozdelenie pravdepodobností príslušnosti objektu do jednotlivých tried vo viacrozmernom priestore. V tomto priestore sú dimenziami priestoru vlastnosti, podľa ktorých ideme klasifikovať a body v tomto priestore predstavujú jednotlivé objekty, z ktorých každý patrí do nejakej triedy.

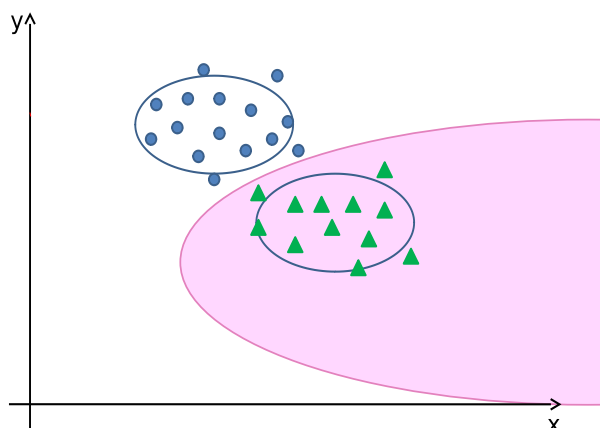
Úlohou klasifikátora je nájsť rozdelenia pravdepodobností príslušnosti známych objektov do tried tak, aby sa čo najmenej “prekrývali” a teda aby bolo možné čo najlepšie oddeliť “príslušníkov” jednej triedy od príslušníkov iných tried. Následne sa tieto rozdelenia použijú na klasifikáciu neznámych objektov.

Tento prístup v sebe latentne zahŕňa efekt “oddeľovacích čiar”, teda akýchsi rozhraní medzi rozdeleniami jednotlivých tried. Predpis týchto čiar sa pri klasifikácii nevyužíva ale dobre ilustruje spôsob, akým klasifikátor pracuje. Pri LDA pritom ide o priamky, pri KDA parabolické krivky (čo je spôsobené tým, že LDA predpokladá rovnaké kovariancie medzi atribútmi v rozdeleniach všetkých tried a KDA zisťuje kovariancie pre každú triedu zvlášť).

Ak by sme napríklad zobrali prípad separácie objektov patriacich do dvoch tried, potom LDA hľadá takú priamku, pri ktorej je rozptyl vzdialeností objektov triedy od priamky minimálny, čo je dôsledkom nájdenia takých rozdelení oboch tried, ktoré sa prekrývajú minimálne. Najlepšie to ilustruje obrázok 2.4. Tento spôsob sa aplikuje aj pri KDA, len s tým rozdielom, že jednotlivé triedy nese-  
parujeme lineárne ale kvadraticky (teda oddeľujúcimi tvarmi sú elipsy). KDA je znázornená na obrázku 2.5.



Obr. 2.4: Na obrázku vidíme dva spôsoby akými pomyselná priamka rozdeľuje dve triedy. Prvá priamka je nevyhovujúca, pretože body z jednej triedy sa nachádzajú v priestore druhej triedy a naopak. Druhá priamka separuje triedy oveľa lepšie.



Obr. 2.5: Na obrázku je znázornená separácia tried pre Kvadratickú diskriminačnú analýzu.

LDA aj KDA pracujú s pravdepodobnosťami príslušnosti objektov do tried (teda napriek ilustračnému obrázku nerozdeľujú priestor “ostro”, ale určujú pravdepodobnosti, s akými objekty patria do jednotlivých tried). Pravdepodobnosť je hodnota vyčísľujúca istotu resp. neistotu výskytu určitej udalosti, čiže je to počet pravdivých udalostí k počtu možných udalostí. V tejto práci sa zaoberáme pravdepodobnosťou podľa Bayesa  $\rightarrow$  Bayesovo pravidlo. Podľa tohto pravidla sa priradí objekt do skupiny s najvyššou podmienenou pravdepodobnosťou. To znamená, že ak máme namerané  $x$  neznámej triedy (napr. vektor nameraných hodnôt (pri RGB súradniciach máme tieto hodnoty 3)), tak si model vypočíta pravdepodobnosť (zo vstupov, ktoré si vysvetlíme nižšie) pre každú známu triedu, a zaradí toto namerané  $x$  do triedy, pre ktorú vyšla najvyššia pravdepodobnosť. Povedzme, že máme  $g$  tried, potom Bayes-ovo pravidlo zaradí neznáme  $x$  do triedy  $i$ , pretože platí  $P(i|x) > P(j|x), \forall j \neq i$ . V tomto vzťahu sú  $i$  a  $j$  triedy a  $x$  je zoznam vlastností.

V skutočnosti je však veľmi náročné získať  $P(i|x)$ . Čo však vieme získať je  $P(x|i)$ , z ktorého sa pomocou Bayesovej vety vieme dostať na  $P(i|x)$ . Vysvetlíme si to konkrétnejšie. Ak poznáme typ celulózového vlákna (napr. drevovina) - triedu, vieme aké súradnice R, G a B - zoznam vlastností, prislúchajú tomuto vláknu. To je výraz  $P(x|i)$ . My však chceme vedieť  $P(i|x)$ . To znamená, že máme zoznam vlastností (súradnice R,G,B) a chceme zistiť akej triede patria (typ celulózového vlákna). Na získanie  $P(i|x)$  môžeme použiť Bayesovu vetu:

$$P(i|x) = \frac{P(x|i)P(i)}{\sum_{\forall j} P(x|j)P(j)} \quad (2.11)$$

kde:

- $P(i|x)$  - je pravdepodobnosť, že objekt je z triedy  $i$ , pokiaľ sme namerali vlastnosti  $x$  (pravdepodobnosť, že ide o triedu  $i$ , pokiaľ skúmaný objekt má vlastnosti  $x$ )
- $P(x|i)$  - pravdepodobnosť že získame vlastnosti  $x$ , keď vieme, že objekt je v triede  $i$  (vieme, že objekt je z triedy  $i$ , tak s akou pravdepodobnosťou bude mať vlastnosti  $x$ )
- $P(x|j)$  - pravdepodobnosť že získame vlastnosti  $x$ , keď vieme, že objekt je v triede  $j$
- $P(i)$  - apriórna pravdepodobnosť o výskyte triedy  $i$ , ktorá je známa bez toho aby sme uskutočnili akékoľvek merania (väčšinou definovaná ako pomer počtu objektov z triedy  $i$  ku všetkým objektom).
- $P(j)$  - apriórna pravdepodobnosť o výskyte triedy  $j$ , ktorá je známa bez toho aby sme uskutočnili akékoľvek merania (väčšinou definovaná ako pomer počtu objektov z triedy  $j$  ku všetkým objektom).

Problémom je, že na to, aby sme mohli prakticky určiť túto pravdepodobnosť  $P(x|i)$  potrebujeme veľmi veľa meraní a to je dosť veľká nevýhoda. Ak však vieme predpokladať, že hodnoty  $x$  majú normálne rozdelenie (najpočetnejší výskyt znakov má hodnotu priemeru, a početnosť výskytu nad a podpriemerných hodnôt rovnakej odchýlky od priemeru je rovnaký) tak nám stačí oveľa menej meraní (a vieme určiť priemer a odchýlku). Ak vieme získať priemer (strednú hodnotu) a odchýlku od strednej hodnoty, vieme určiť distribúciu. Teraz už vieme pre ľubovoľné  $x$  vypočítať pravdepodobnosť výskytu v danej triede. Najskôr si vysvetlíme LDA. Platí vzťah (jeho podrobné odvodenie nájdete v prílohe B):

$$h_i = \mu_i C^{-1} x_k^T - \frac{1}{2} \mu_i C^{-1} \mu_i^T + \ln(P_i) \quad (2.12)$$

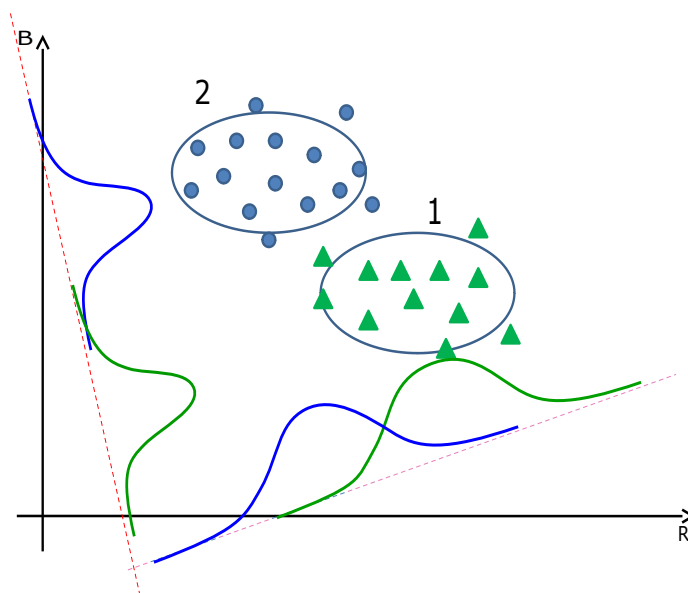
kde:

- $h_i$  - pravdepodobnosť, že  $x_k$  je z triedy  $i$

- $\mu_i$  - stredná hodnota pre triedu  $i$  (alebo vektor stredných hodnôt pre triedu  $i$ )
- $x_k$  - meranie vlastností pre objekt  $k$
- $C$  - kovariančná matica vlastností pre všetky triedy
- $P(i)$  - pravdepodobnosť o výskyte triedy  $i$ , ktorá je dopredu známa bez toho aby sme uskutočnili akékoľvek merania

Pre daný vstup (dané meranie) sa vypočíta s akou pravdepodobnosťou sa meranie nachádza v danej triede - prvej triede, druhej triede až  $g$ -tej triede, t.j. vypočíta sa  $h_1, h_2, \dots, h_g$  a zoberie sa to najvyššie  $h$  a do tej triedy sa dané meranie zaradí.

Pre lepšie pochopenie majme situáciu ako je znázornené na obrázku 2.6.



Obr. 2.6: Na obrázku sú dve triedy (trieda 1 a trieda 2) a obe sú zložené z objektov reprezentovaných dvojzložkovými farebnými vektormi. Na osi  $x$  je červená farba ( $R = \text{red}$ ), na osi  $y$  je modrá farba ( $B = \text{blue}$ ).

Kovariančné matice pre tieto triedy vyzerajú nasledovne:

$$C_1 = \begin{bmatrix} C_1(1,1) & C_1(1,2) \\ C_1(2,1) & C_1(2,2) \end{bmatrix} = \begin{bmatrix} \sigma_{R1}^2 & \sigma_{R1}\sigma_{B1} \\ \sigma_{B1}\sigma_{R1} & \sigma_{B1}^2 \end{bmatrix} \quad (2.13)$$

$$C_2 = \begin{bmatrix} C_2(1,1) & C_2(1,2) \\ C_2(2,1) & C_2(2,2) \end{bmatrix} = \begin{bmatrix} \sigma_{R2}^2 & \sigma_{R2}\sigma_{B2} \\ \sigma_{B2}\sigma_{R2} & \sigma_{B2}^2 \end{bmatrix} \quad (2.14)$$

My však chceme získať maticu  $C$ , ktorú používame vo vzťahu 2.12. Na to aby sme ju získali potrebujeme vedieť najskôr vektor  $p$ . Vypočítame ho ako:

$$p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \frac{n_1}{N} \\ \frac{n_2}{N} \end{bmatrix} \quad (2.15)$$

kde:

- $N$  - celkový počet objektov
- $n_1$  - počet objektov v triede 1
- $n_2$  - počet objektov v triede 2

Pre náš prípad je  $N = 27$ ,  $n_1 = 12$  a  $n_2 = 15$ .

Výslednú maticu  $C$ , ktorá sa nachádza vo vzťahu 2.12 získame z čiastkových matíc (t.j. z matíc  $C_1$  a  $C_2$ ) nasledovne:

$$C = \begin{bmatrix} C(1,1) & C(1,2) \\ C(2,1) & C(2,2) \end{bmatrix} \quad (2.16)$$

kde:

- $C(1,1) = p_1 C_1(1,1) + p_2 C_2(1,1)$
- $C(1,2) = p_1 C_1(1,2) + p_2 C_2(1,2)$
- $C(2,1) = p_1 C_1(2,1) + p_2 C_2(2,1)$
- $C(2,2) = p_1 C_1(2,2) + p_2 C_2(2,2)$

Ak to zovšeobecníme, platí nasledovné:

$$C(r,s) = \frac{1}{N} \sum_{i=1}^g n_i C_i(r,s) \quad (2.17)$$

### 3. Praktická časť

Základným východiskom našej práce je postup laboratórnej identifikácie drevných vlákien (Jablonský a kol. [3]):

1. Izolovaná vzorka sa naфарbí pomocou rozlišovacieho farbiva.
2. Vyhотовí sa mikrofotografia vzorky.
3. Na mikrofotografii sa určí obdĺžnikový región, ktorý reprezentatívne zachytáva časť odfoteného vlákna.
4. Z vybraného regiónu sa ďalej náhodným výberom vyberie určitý počet pixelov.
5. Z vybraných pixelov sa vytvoria vzorky reprezentujúce farebnosť vlákna.
6. Zo vzorky sa zvoleným spôsobom automatickej klasifikácie určí typ vlákna (pričom klasifikátor bol natrénovaný na vzorkách vytvorených rovnakým postupom).

Prínosy našej práce sú v kontexte tohto postupu nasledovné:

- Krok 3 postupu je vo východiskovej práci vykonávaný manuálne. Je teda úlohou laboranta, aby na základe svojho uváženia na fotografiách regióny vybral. Naším prínosom je automatizácia tohto kroku, teda **automatická extrakcia farebnej vzorky z fotografie**. Vytvorili sme dva algoritmy, ktoré kombináciou rôznych techník spracovania obrazu a optimalizácie identifikujú vhodný priečny rez vláknom na fotografii. Prvý algoritmus je založený na hľadaní najkratšej spojnice medzi regiónmi pozadia fotografií, druhý sa snaží nájsť vhodné miesto vzorky vpisovaním elipsy do vlákna. Opísané a experimentálne overené sú v sekcii 3.1.

- **Skúmame efekt použitia rôznych farebných schém pre reprezentovanie vzoriek.** Sekcia 3.2 obsahuje opis experimentov, v ktorých vyššie uvedený postup vykonávame nad sadou fotografií drevných vlákien. Replikujeme experimenty z východiskovej práce Jablonského. V rámci alternatívnych nastavení zapájame aj naše extrakčné metódy. Najmä však porovnávame úspešnosť použitia dvoch rôznych farebných schém – RGB a Helmholtz systémov – pri reprezentácii farieb v obrázkoch. Voľba farebnej schémy totiž určuje, akým spôsobom sú získavané črty pre klasifikáciu, čím ju zásadne ovplyvňuje.
- **Vytvorili sme praktický, znovupoužiteľný nástroj pre prostredie Matlab,** ktorý realizuje kroky 3 až 6 postupu identifikácie a pomocou ktorého sme aj uskutočnili naše experimenty. V sekcii 3.3 sa venujeme opisu tohto nástroja.

### 3.1 Automatický výber regiónov pixelov

Navrhli sme dve metódy, prostredníctvom ktorých sa vyberajú regióny pixelov na fotografiách automaticky. Táto úloha sa dá voľne definovať ako hľadanie pričného rezu vláknom na mieste, kde je fotografia primerane ostrá, kde nie je vlákno ohnuté či prekryté iným vláknom kde nie sú nečistoty, bubliny a podobne. Po nájdení pričny rez slúži ako os obdĺžnika (s danou šírkou), z vnútra ktorého vyberáme pixely do vzorky (pozri obrázok 3.1).

Obidve metódy zdieľajú podobný začiatok svojho postupu: pomocou prahového rozdelenia a morfológického vyhladzovania identifikujú masku obrázka, ktorá určuje kde sa na obrázku nachádza vlákno a kde “pozadie”. Líšia sa až v spôsobe, akým sa na identifikovanom vlákne nájde vhodné miesto pričného rezu.

1. **Metóda krátkej spojnice** je založená na hľadaní najkratšej spojnice medzi regiónmi pozadia (teda plochami pozadia, ktoré sú vláknom od seba oddelené). Ideou metódy je, že minimálna spojnica by mala vlákno rozdeľovať prične (práve preto, že je minimálna) a že minimalizácia tiež bude preferovať hľadanie spojnic medzi susednými regiónmi, čím sa zníži riziko nájdenia rezu na miestach, kde sa prekrývajú viaceré vlákna.
2. **Metóda vpísanej elipsy** hľadá maximálnu vpísanú elipsu vo vlákne, pričom po jej nájdení sa za rez prehlási kratší polomer (pás) tejto elipsy. Ideou metódy je, že vpisovanie čo najväčšej elipsy do podlhovastého vlákna prirodzene orientuje jej kratší polomer prične k tomuto vláknu.



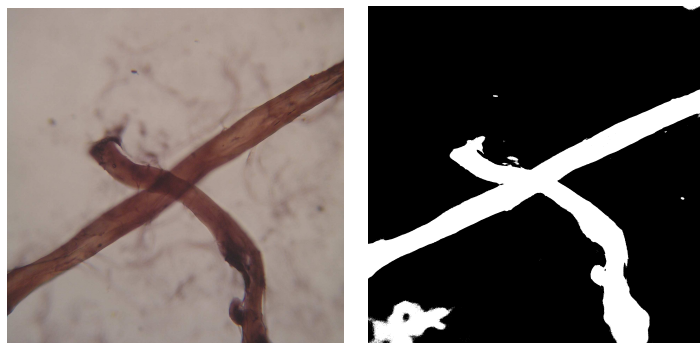


Obr. 3.1: Mikrofotografia drevného vlákna. Modrou farbou je vyznačený možný reprezentatívny región. Cieľom našich metód je takýto región vybrať automaticky.

Nižšie opisujeme presné postupy obidvoch metód, ich vstupné podmienky a výhody a nevýhody, ktoré tieto metódy so sebou prinášajú. Keďže ide o náročný problém, ani jedna z metód nezaručuje nájdenie reprezentatívneho regiónu. Obe metódy sme experimentálne overili, sústrediac sa na ich úspešnosť nachádzania vhodných miest vo vláknach a časovú náročnosť.

### 3.1.1 Metóda krátkej spojnice

1. Načíta sa fotografia. *Výstupom kroku je matica  $M \times N \times 3$  hodnôt, ktoré charakterizujú pixely obrázka.*
2. Pomocou *Otsuovej metódy* sa ohraničí, ktoré pixely obrázka zobrazujú drevné vlákno a ktoré pozadie. Táto metóda pracuje tak, že rozdelí pixely do dvoch tried prostredníctvom prahu svetlosti. V našom prípade jedna trieda má byť triedou pixelov pozadia a druhá trieda má byť triedou pixelov vlákna. Metóda teda predpokladá, že pixely pozadia budú spravidla svetlejšie ako pixely vlákna. Otsuova metóda hľadá prah tak, že minimalizuje varianciu svetlosti pixelov v rámci triedy. *Výstupom kroku je binárna matica  $M \times N$ , predstavujúca masku určujúcu, kde je pozadie a kde vlákno.*

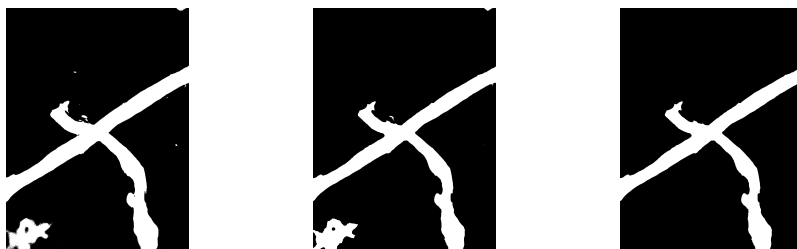


Obr. 3.2: Vľavo: pôvodná fotografia, vpravo: vytvorená maska

3. Pomocou metód *morfologický uzáver* a *morfologická erózia* vyhladíme niektoré lokálne nepresnosti masky ktoré vznikli v kroku 2. Morfologický uzáver nachádza v maske miesta kde pozadie tvorí vo vlákne drobné „trhliny“ alebo „diery“ a vyplňa ich. Morfologická erózia robí to isté ale z pohľadu pozadia – odstraňuje drobné „špinky“ a „výrastky“ vlákna. *Výstupom kroku je upravená maska vlákna, na ktorej sú vyhladené jeho okraje a odstránené drobné diery. To pomôže znížiť počet regiónov, s ktorými sa pracuje v nasledujúcich krokoch.*
4. Vyberie sa najväčšie vlákno. Použije sa algoritmus rozliatia (angl. *flood-fill*<sup>1</sup>), ktorý identifikuje najväčší súvislý región vlákna v maske. Tento krok

<sup>1</sup>[https://en.wikipedia.org/wiki/Flood\\_fill](https://en.wikipedia.org/wiki/Flood_fill)

robíme pretože očakávame, že najlepšia vzorka bude z najväčšieho vlákna. *Výstupom kroku je upravená maska, v ktorej ako jediný zostal najväčší súvislý región vlákna.*



Obr. 3.3: Vľavo: vytvorená maska, v strede: upravená maska, vpravo: vybrané najväčšie vlákno

5. Určia sa regióny pozadia, teda oblasti, ktoré sú oddelené vláknom. Určia sa tiež ich hranice (pixely, ktoré ich ohraničujú) a obsah. Algoritmus ďalej pracuje s prvými najviac piatimi najväčšími regiónmi. Pri hľadaní vzoriek sa vychádza z predpokladu, že vzorka leží medzi dvoma rôznymi regiónmi pozadia. *Výstupom kroku je zoradený zoznam najväčších regiónov pozadia. Každý región je určený zoznamom pixelov, ktoré ho ohraničujú.*

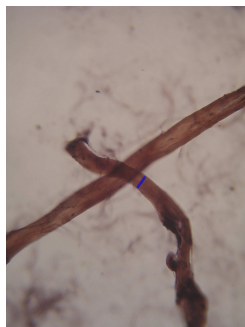


Obr. 3.4: Hranice regiónov pozadia

6. Zo získaných regiónov sa vytvoria dvojice (každý s každým). *Výstupom je zoznam dvojíc regiónov.*

Táto metóda je stochastická a nemusí vždy dospieť k rovnakej spojnici (globálnemu minimu), to však nie je prekážkou pre použitie tohto algoritmu, očakávame, že nájdená spojnica bude prepájať vlákno priečne.

7. Pre každú dvojicu regiónov sa nájde najkratšia úsečka, ktorá tieto regióny spája. Tým sa dosiahne vzorka priečne prechádzajúca vláknom. Aby bol algoritmus časovo efektívny a nebolo potrebné vyhodnocovanie všetkých možných spojnic, použije sa optimalizačná metóda *Luus-Jaakola*. Táto metóda je stochastická a nemusí vždy dospieť k rovnakej spojnici (globálnemu minimu), to však nie je prekážkou pre použitie tohto algoritmu, očakávame, že nájdená spojnica bude prepájať vlákno priečne. *Výstupom kroku je zoznam úsečiek spájajúcich regióny v každej dvojici.*
8. Zo zoznamu sa vyberie najkratšia úsečka, ktorá však musí dosahovať minimálnu, nami zadanú dĺžku. Minimálna dĺžka ošetruje prípad, kedy je vlákien na obrázku viac (no tvoria súvislý región) a niektoré z týchto vlákien je príliš tenké. *Výstupom je konečný výber miesta vzorky (daného úsečkou).*
9. Vyberú sa vzorkové pixely v okolí vybranej úsečky. Okolie je dané obdĺžnikom, ktorého osou je úsečka a ktorý má vopred danú hrúbku (10 pixelov). *Výstupom kroku je konečný výber pixelov do vzorky.*



Obr. 3.5: Fotka s vybraným regiónom

### 3.1.2 Metóda vpísanej elipsy

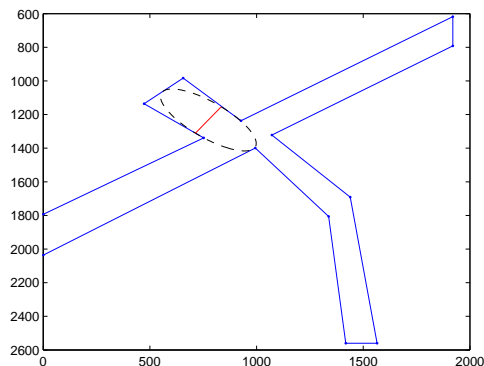
1. Prvé štyri kroky sú analogické ako pri prvej metóde, t.j.
  - načíta sa fotografia
  - pixely obrázka sa pomocou *Otsuovej metódy* rozdelia do dvoch tried – vznikne *maska*
  - vyhladáme niektoré lokálne nepresnosti pomocou *morfologického uzáveru* a *morfologickej erózie*
  - vyberie sa najväčšie vlákno

*Výstupom po všetkých týchto úkonoch je rovnako ako pri prvej metóde upravená maska, v ktorej ako jediný zostal najväčší súvislý región vlákna.*

2. Zjednoduší sa hranica vlákna, kvôli skráteniu času hľadania vpísanej elipsy v nasledujúcom kroku. Na zjednodušenie hranice možno použiť ľubovoľný algoritmus na zjednodušenie polygónov (prehľad rôznych prístupov ponúka Luebke [7]), za predpokladu, že zachová približné obrysy vlákna. My sme použili tzv. *decimáciu*, pričom sme využili existujúcu Matlab implementáciu<sup>2</sup>. Princípom decimácie je postupné odstraňovanie vrcholov a premostovanie vzniknutých dier. Odstraňujú sa pritom tie vrcholy, ktorých odstránením utrpí celkový tvar polygónu čo najmenej (najmenej sa zmení jeho obsah). *Výstupom je zjednodušený polygón ohraničujúci vlákno.*
3. Do polygónu sa vpíše elipsa a nájde sa jej úzky polomer. Pokiaľ je polygónom ohraničené vlákno, teda podlhovastý útvar, očakáva sa, že úzky polomer vpísanej elipsy bude zhruba zodpovedať kolmému rezu týmto vláknom. Hľadanie maximálnej vpísanej elipsy nekonvexného polygónu je výpočtovo ťažký problém. Našťastie však nepotrebujeme vypočítať globálnu maximálnu elipsu, stačí nám aj lokálne optimum. Rozhodli sme sa preto hľadať elipsu pomocou optimalizácie kvadratickým programovaním. Účelovou funkciou je obsah elipsy, ohraničenia vyplývajú z hrán polygónu a pozícií bodov obvodu elipsy. *Výstupom je konečný výber miesta vzorky (daného úsečkou).*
4. Tak ako pri prvej metóde, vyberú sa vzorkové pixely v okolí vybranej úsečky. Okolie je dané obdĺžnikom, ktorého osou je úsečka a ktorý má vopred danú hrúbku (10 pixelov). *Výstupom kroku je konečný výber pixelov do vzorky.*

---

<sup>2</sup><http://www.mathworks.com/matlabcentral/fileexchange/45342-polygon-simplification>



Obr. 3.6: Elipsa vpísaná do polygónu. Červená úsečka predstavyje úzky polomer elipsy a zároveň miesto výberu vzorky pixelov.

### 3.1.3 Porovnanie metód výberu regiónov

Na to aby sme mohli regióny pixelov vyberať z fotografií automaticky prostredníctvom vyššie opísaných metód, musí byť splnených niekoľko podmienok. Každá z metód definuje iné podmienky z čoho vyplýva odlišná použiteľnosť metód pre rôzne situácie.

**Podmienky, ktoré musia byť splnené pri použití prvej metódy:**

Vlákno na fotografii sa musí dotýkať aspoň dvoch okrajov fotografie tak, že sa vytvorí aspoň dva regióny pozadia. Ak napríklad vlákno vyplní celú dolnú tretinu fotografie, čiže sa dotýka až troch strán obrázka, ale je k dispozícii len jeden región pozadia, táto metóda sa nedá použiť. Ďalej, vlákno na fotografii musí byť dobre zaostrené, pretože metóda nevie posúdiť hĺbku ostrosti.

Výhody:

- časovo menej náročné (oproti druhej metóde)
- výrazne znižuje pravdepodobnosť nájdenia úseku na prekryve dvoch vlákien
- je možné použiť ju aj na zložitejšie fotografie kde je viacero rôzne sa prekrývajúcich vlákien

Nevýhody:

- ak je k dispozícii len jeden región pozadia, je hľadanie prostredníctvom tejto metódy nerealizovateľné

**Podmienky, ktoré musia byť splnené pri použití druhej metódy:**

Vlákno na fotografii musí byť čo najjednoduchšie aby bolo možné vytvoriť čo najjednoduchší polygón. S narastajúcou zložitou polygónu sa predlžuje výpočtový čas, a pri veľmi zložitých tvaroch hrozí, že výpočet neprebehne v prakticky použiteľnom čase. Rovnako ako pri prvej metóde vlákno musí byť dobre zaostrené, aby bol výber pixelov reprezentatívny.

Výhody:

- vlákno môže byť uložené na fotografii ľubovoľne (neobmedzuje ju počet regiónov pozadia)
- metóda potenciálne poskytuje možnosť vybrať viacero regiónov pozdĺž elipsy (teda nielen región v úzkom polomere ale aj s ním rovnobežné regióny v iných častiach elipsy)

Nevýhody:

- časovo náročnejšie (oproti prvej metóde)
- vyššia šanca nájdenia regiónu na prekryve viacerých vlákien
- pri zložitej fotografii, kde je viacero vlákien je možné že výpočet neprebehne v prakticky použiteľnom čase

Na podporu týchto tvrdení sme metódy overili experimentálne. Obe metódy sme na vzorke 18 fotografií vlákien spustili 10 krát (teda 180 pokusov). Pritom sme merali, či metóda úspešne skončila hľadanie vzorky (či vrátila akýkoľvek výsledok), čas za aký sa k nemu dostala a či bol výsledný obdĺžnik na nevhodnom mieste (teda na mieste, kde sa dve vlákna prekrývajú). V nasledujúcej tabuľke vidieť výsledky experimentu po jednotlivých fotografiách. V prípade času uvádzame priemerný čas hľadania. V prípade prekrývajúcich sa vlákien počet takýchto prípadov. Prípady, kedy metóda nebola schopná dospieť k riešeniu ani v jednom z desiatich pokusov, uvádzame v tabuľke slovom “neprebehlo”.

Ako vidno z tabuľky, metóda krátkej spojnice je výhodnejšia z hľadiska časovej náročnosti ako aj nulovou chybovosťou čo sa týka výskytu rezov na nevhodných miestach. Taktiež je úspešnejšia pri dokončení hľadania ako takom, neúspech v prípade 15-tej fotografie bol spôsobený práve tým, že sa na nej nachádzal len jediný región pozadia. Práve v takýchto prípadoch by ju mohla nahrádzať metóda vpísanej elipsy.

číslo fotky	čas[s]		počet na prekryvoch	
	k. spoj.	v. elipsa	k. spoj.	v. elipsa
1	6.97	20.28	0	0
2	7.24	31.83	0	0
3	5.85	neprebehlo	0	neprebehlo
4	7.92	28.55	0	0
5	6.35	21.26	0	10
6	7.10	20.67	0	1
7	6.48	neprebehlo	0	neprebehlo
8	6.55	32.18	0	2
9	7.33	29.53	0	6
10	8.95	23.55	0	0
11	6.61	23.35	0	4
12	6.17	18.83	0	4
13	6.84	26.67	0	0
14	5.28	20.17	0	1
15	neprebehlo	21.45	neprebehlo	0
16	6.57	19.70	0	3
17	6.65	neprebehlo	0	neprebehlo
18	7.41	27.18	0	0

Tabuľka 3.1: Výsledky spustenia metód pre 18 rozličných fotografií drevných vlákien.

### 3.2 Vyhodnotenie úspešnosti klasifikácie pri rôznych nastaveniach

V tejto sekcii opisujeme experimenty, pri ktorých sme s rôznymi nastaveniami, vstupnými údajmi a algoritmami vykonávali klasifikáciu drevných vlákien analýzou mikrofotografií prostredníctvom diskriminačnej analýzy. Všeobecným cieľom experimentov bolo preskúmať vlastnosti (najmä však presnosť) rôznych variánt klasifikácie. Vo všetkých experimentoch sme vychádzali zo scenára experimentu Jablonského a kol [3], ktorý sme pre rôzne varianty modifikovali. Jednotlivé varianty vychádzali z alternatív vo viacerých hľadiskách:

- *Vstupné údaje.* Porovnali sme presnosť klasifikácie s použitím pôvodnej vzorky Jablonského, vzorky rozšírenej o ďalšie fotografie a vzorky, v ktorej trénovacia a testovacia množina príkladov pochádzala z odlišných častí fotografií (pôvodný Jablonského prístup vytváral trénovaciu a testovaciu vzorku z rovnakých regiónov).



- *Typ diskriminačnej analýzy.* Skúmali sme úspešnosť klasifikácie pri použití lineárnej a kvadratickej diskriminačnej analýzy.
- *Spôsob výberu vzorkových regiónov z fotografií.* Prvým krokom pri tvorbe vzoriek pre tréning a testovanie klasifikácie je vyznačenie oblúčkového regiónu, na ktorom sa nachádza reprezentatívna časť drevného vlákna. Toto mohlo byť vykonané buď manuálne alebo automaticky s použitím našich metód opísaných v predchádzajúcej sekcii.
- *Systém reprezentácie farieb.* Skúmali sme úspešnosť klasifikácie pri použití RGB a Helmholtzovho systému farieb.
- *Spôsob vytvorenia tréningových príkladov.* Pri príprave tréningových vzoriek z vybraných regiónov na fotografiách bolo možné buď priamo vybrať jednotlivé pixely (s tým, že sme tým pádom akceptovali pomerne veľký rozptyl ich farebnosti aj v rámci jednotlivých tried) alebo vyberať skupiny pixelov, priemerovať ich farebnosť a až tieto priemery používať ako vstupy do tréningu.

### 3.2.1 Scenár a nastavenie experimentov

Vstupom pre naše experimenty bola sada 39 mikrofotografií drevných vlákien (ich vytvorenie nebolo predmetom našej práce). Vlákna patrili do 4 tried: drevovina; sulfátová buničina; Wathman papier; a handrovina. Zafarbenie a fotenie prebiehalo spôsobom ako je opísané v článku [3]. Činidlo, ktoré bolo použité na farbenie je zinok-chlorido-jodidové farbivo (Herzbergovo farbivo). Z pôvodného experimentu v článku boli k dispozícii fotografie nerovnakých počtov pre dané typy vlákien (drevovina – 1, sulfátová buničina – 8, wathman papier – 7, handrovina – 1). Tieto fotografie sme použili pre prípravu dát v experimentoch 1 a 2, aby sme pôvodný experiment čo najvernejšie replikovali. Pre experiment 3 sme použili rozšírenú sadu fotografií, s počtami: drevovina – 8, sulfátová buničina – 13, wathman papier – 11, handrovina – 7.

Výstupom každého experimentu bola presnosť, teda pomer správne určených testovacích prípadov ku všetkým testovacím prípadom.

Experimenty sme realizovali v prostredí Matlab automaticky. Postup experimentu (spoločný pre všetky nastavenia) bol nasledovný :

1. Fotografie sme náhodne rozdelili do tréningovej (22) a testovacej množiny (17).

2. Na každej fotografii trénovacej množiny sme zvolenou metódou výberu na každej fotografii určili obdĺžnikový región, ktorý pokrýval vlákno na reprezentatívnom mieste.
3. Z množiny pixelov, ktoré zvolený región pokrýval, sa vybrali farebné vzorky. Vo všetkých experimentoch (okrem prípadu, kedy sme porovnávali RGB a Helmholtzov systém) sme použili RGB reprezentáciu farieb. Výber vzoriek z jedného regiónu sme uskutočnili buď (1) výberom 10 skupín 100 náhodne vybraných pixelov a priemerovaním ich farieb (čím vzniklo 10 vzoriek na fotografii) alebo (2) výberom 1000 jednotlivých pixelov (chceli sme porovnať úspešnosť oboch prístupov).
4. Zo vzniknutých vzoriek sme natrénovali klasifikátor (vo všetkých experimentoch lineárnu diskrimináciu, okrem prípadu kedy sme chceli porovnať ju s kvadratickou).
5. Pre testovacie fotografie sme rovnakou metódou ako pri trénovacích určili obdĺžnikové regióny.
6. Z každého regiónu sme (za použitia rovnakého farebného systému) pripravili 10 testovacích vzoriek náhodným výberom a priemerovaním 100 pixelov.
7. Vzorky testovacej množiny sme postupne skúšali klasifikovať.
8. Celkovú úspešnosť (presnosť) klasifikácie sme vypočítali ako podiel počtu úspešne určených tried a počtu všetkých testovacích pokusov.

Keďže postup je vo viacerých krokoch nedeterministický (využíva náhodné výbery a pod.), každý experiment sme s rovnakými nastaveniami zopakovali 20 krát a celkový výsledok je tak priemerná presnosť klasifikácie.

### 3.2.2 Presnosť pre rôzne vstupné údaje

Prvé experimenty sme zamerali na overenie výsledkov experimentov, ktoré sú opísané v článku [3, 1]. Autori použili lineárnu diskriminačnú analýzu, RGB reprezentáciu farieb a manuálny výber regiónov, z ktorých vyberali vzorky výberom skupín pixelov a ich priemerovaním. Autori ako prínos prezentovali predovšetkým zníženie počtu farieb na prípravu vzoriek na jednu (Herzbergovo farbivo) a nepotrebnosť použitia informácií o morfológií vlákien. Diskriminačná sila ich prístupu je podľa autorov 95-100% správne identifikovaných neznámych

vzoriek s jedným farebným vektorom R alebo B; keď bola kombinácia dvoch alebo troch vektorov tak diskriminačná sila – 100%.

Reprodukovat tento postup sme sa rozhodli preto, že sme ho chceli preveriť na rôznorodejších vzorkách. Autori totiž trénovacie aj testovacie vzorky vytvárali z rovnakých regiónov na fotografiách, čo mohlo ich výsledky pozitívne skresliť. V našich experimentoch sme pre trénovanie a testovanie vyberali vždy odlišné regióny, resp. fotografie.

Samotný manuálny výber regiónov na fotografiách prebiehal tak, že z fotografií sme pre účely tvorby vzorky vybrali určité regióny, ktoré dobre reprezentovali vizuálne vlastnosti vlákien. Celkový charakter fotografií je totiž taký, že vlákna tieto fotografie nevyplňajú celé, ale nachádzajú sa iba v niektorých častiach. Navyše sú na fotografiách prítomné aj nečistoty, bubliny či miesta, kde sa viaceré vlákna prekrývajú či ohýbajú, čo spôsobuje výrazné farebné skreslenia v týchto miestach. Vyberali sme miesta, ktoré boli vhodne zaostrené, dobre vyfarbené a bez výrazných kazov.

Pre získanie tabuľky údajov z jednotlivých fotografií sme použili program *WoodRecog*. Vstupom pre tento program sú samotné fotografie. Výstupom sú pixely vybraných regiónov.



Obr. 3.7: Fotografia s vybraným regiónom

Uskutočnili sme 3 experimenty (poznámka: tieto experimenty sme uskutočňovali v skorších etapách riešenia projektu, kedy sme ešte nemali ustálený postup experimentovania; preto sa v nasledujúcich experimentoch používajú iné počty vzoriek):

1. Pri prvom experimente sme použili vzorku údajov, ktorá bola použitá pri experimente uskutočnenom v článku vyššie. Z každého typu vlákna sme náhodne vybrali 500 vzoriek na tréovanie, a následne 100 vzoriek na testovanie. Každú vzorku sme vytvorili náhodným výberom 100 pixelov zo všetkých pixelov príslušného typu vlákna, ktoré sme následne spriemerovali (do tzv. priemerného pixelu).
2. Pri druhom experimente sme za použitia tých istých fotografií vytvorili množinu na testovanie, pričom už existujúcu množinu sme použili na tréovanie. Testovaciu množinu sme vytvorili tak, že na každej fotografii z pôvodnej sady sme vybrali nové náhodné miesto o šírke 7px a dĺžke samotného pásika a z neho vygenerovali množinu údajov. Z každého regiónu sme vytvorili 10 vzoriek náhodným výberom a spriemerovaním 100 pixelov. Rozloženie testovacích vzoriek v tomto prípade nebolo rovnomerné medzi typmi vlákien, pretože sme nemali k dispozícii rovnaký počet fotografií pre každý typ vlákna. V experimente sme teda použili celkovo 170 testovacích vzoriek (17 fotografií).
3. Pri treťom experimente sme použili rozšírenú sadu fotografií a rozdelili ju na dve skupiny. Prvú skupinu fotografií sme použili len na tréovanie a druhú len na testovanie. Ďalej sme postupovali rovnako ako v experimente 2.

Sumárne výsledky týchto troch experimentov uvádzame v tabuľke 3.2 (výsledky všetkých dvadsiatich opakovaní experimentov uvádzame v prílohe A v tabuľkách A.1, A.2 a A.3). Priemerná úspešnosť v prvom experimente vychádza 100%, no v ďalších dvoch nastaveniach už taká úspešná nebola: v 2. aj 3. experimente len približne 83% (pre lineárnu diskriminačnú analýzu).

### **3.2.3 Úspešnosť lineárnej a kvadratickej diskriminačnej analýzy**

Experimenty sme v predchádzajúcich nastaveniach vykonali pre dva klasifikátory: (1) lineárnu a (2) kvadratickú diskriminačnú analýzu. V pôvodnej práci bola použitá lineárna diskriminačná analýza, my sme však chceli preskúmať, či

Experiment	Lineárna DA	Kvadratická DA
1. Experiment: Pôvodná sada údajov	100.00%	100.00%
2. Experiment: Pôvodné fotografie, rozličné regióny	83.05%	69.44%
3. Experiment: Rozšírená sada fotografií, oddelené tréningové a testovacie vzorky	82.91%	67.73%

Tabuľka 3.2: Porovnanie presnosti klasifikácie pre tri rôzne vstupné dátové sady a dva rôzne klasifikátory

kvadratická nebude dosahovať lepšie výsledky, keďže ide o klasifikátor vyššieho rádu, ktorý by mohol rozdiely medzi triedami lepšie rozlišovať.

Z výsledkov (uvedených taktiež v tabuľke 3.2) vyplynulo, že lineárny variant na daných vzorkách fungoval lepšie. V prvom experimente boli ešte oba prístupy rovnako dobré vďaka zvolenej vzorke údajov. V ostatných dvoch experimentoch kvadratická diskriminačná analýza prekvapivo zaostávala (v 2. experimente dosiahla len 69% presnosť oproti 83% lineárnej a v 3. experimente iba 68% oproti 83%). Odhadujeme, že horšie výsledky sú dôsledkom pretrénovania, keďže vzorka údajov (pár desiatok fotografií) môže byť pre kvadratickú diskriminačnú analýzu príliš malá. Na základe týchto výsledkov sme v ďalších experimentoch už kvadratickú diskriminačnú analýzu nepoužívali.

### 3.2.4 Zvyšné hľadiská (metódy výberu vzoriek a farebné systémy)

Ďalej sme uskutočnili porovnanie presnosti klasifikácie v zvyšných troch hľadáiskách:

1. Presnosť pri rôznych spôsoboch výberu vzorkových regiónov (metódou krátkej spojnice, metódou vpísanej elipsy).
2. Presnosť pri farebných systémoch RGB a Helmholtz.
3. Presnosť pri rôznych spôsoboch vytvorenia tréningových príkladov (priemery skupín alebo surové pixely).

Vo všetkých troch hľadáiskách boli k dispozícii 2 alternatívne nastavenia. Z toho vyplynulo 8 rôznych nastavení experimentov, ktoré sme uskutočnili a v každom jednom vypočítali presnosť. Prehľad výsledkov ponúka tabuľka 3.3. Tieto experimenty nezahŕňajú možnosť manuálneho výberu regiónov, pretože tá už bola

odmeraná v predchádzajúcich experimentoch, ich výsledok však v tabuľke pre prehľadnosť uvádzame. Podrobnejšie výsledky, rozpísané po dvadsiatich jednotlivých opakovaniach uvádzame v prílohe A v tabuľkách A.4 a A.5.

	Krátka spojnica		Vpísaná elipsa		Manuálny výber
	RGB	HH	RGB	HH	RGB
Priemery skupín pixelov	76.11%	70.50%	74.76%	66.47%	82.91%
Surové pixely	72.82%	64.41%	73.26%	64.35%	-

Tabuľka 3.3: Prehľad výsledných presností klasifikácie pre rôzne alternatívy v troch rôznych hľadiskách.

Vyššie uvedené experimenty ukázali úspešnosť klasifikácie pri použití RGB a Helmholtzovho systému farieb. Okrem toho sme chceli preskúmať, akú odlíšiteľnosť majú dvojice tried drevných vlákien pri použití týchto farebných systémov. Uskutočnili sme preto analýzu pomocou indexu rozlíšiteľnosti. Index rozlíšiteľnosti sa pozerá na rozloženie farebných vzoriek v priestore črt príslušného farebného systému a berie do úvahy vzdialenosť stredov zhlukov vzoriek tried a veľkosť týchto zhlukov. Čím väčšia vzdialenosť a čím menšia veľkosť zhlukov, tým vyšší index rozlíšiteľnosti. Presne je index rozlíšiteľnosti definovaný nasledovným vzorcom [9]:

$$V_{ij} = \frac{d(S_i, S_j)}{\sqrt{(1 + r_i)(1 + r_j)}} \quad (3.1)$$

kde:

- $V_{ij}$  - index rozlíšiteľnosti medzi triedami  $i$  a  $j$
- $S_i, S_j$  - stredy (centroidy) zhlukov vypočítané ako priemerné pozície vzoriek zhliuku
- $d$  - vzdialenosť medzi bodmi v priestore (v tomto prípade centroidmi dvoch zhlukov)
- $r_i, r_j$  - rozsahy zhlukov, vypočítané ako priemer rozdielov maximálnych a minimálnych pozícií vzoriek v jednotlivých dimenziách

Vypočítali sme hodnoty indexu rozlíšiteľnosti pre všetky dvojice tried drevných vlákien. Pritom sme raz použili RGB a raz Helmholtzov farebný systém (HH). Výsledné hodnoty možno vidieť v tabuľke 3.4, pričom môžeme pozorovať, že farebné systémy sa mierne líšia v tom, ktoré dvojice rozlišujú lepšie oproti iným.

Trieda	Trieda	RGB rozlíšiteľnosť	HH rozlíšiteľnosť
F1	F2	1,5654	1,9629
F1	F3	0,9076	1,7404
F1	F4	0,9940	0,5878
F2	F3	0,4974	0,6040
F2	F4	0,6040	1,0045
F3	F4	0,0343	0,7097

Tabuľka 3.4: Indexy rozlíšiteľnosti pre všetky dvojice tried, vypočítané s použitím RGB a Helmholtzovho systému farieb.

### 3.2.5 Zhrnutie experimentov a diskusia

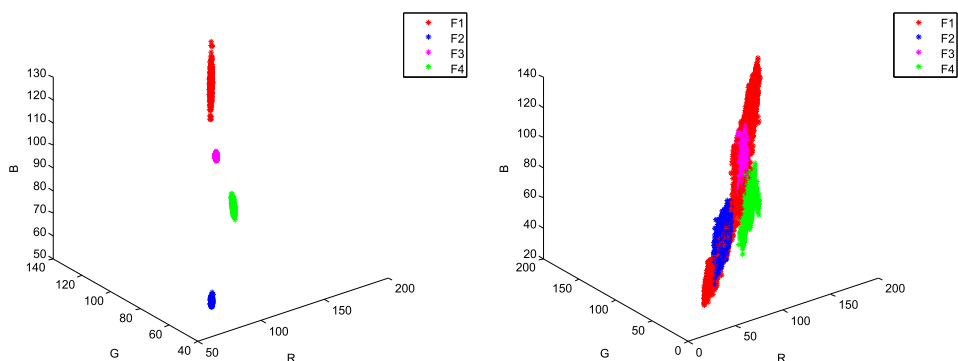
Uskutočnili sme sériu experimentov, pri ktorej sme s rôznymi nastaveniami, vzorkami a použitými algoritmami vykonávali identifikáciu drevných vlákien z mikrofotografií. Výsledky experimentov môžeme zhrnúť nasledovne:

- Zopakovaním experimentov Jablonského a kol. [3] sme ukázali na rezervy použitia lineárnej diskriminačnej analýzy pre rozlišovanie tried drevných vlákien s použitím farebnej informácie z fotografií. Experiment sme replikovali s rovnakou vzorkou údajov a rovnakou metódou aká bola použitá v pôvodnej práci, pričom výsledky úspešnosti boli 100%. V ďalších variantoch, pri ktorých sme použili rozšírenú dátovú vzorku (viac fotografií) a najmä sme trénovacie a testovacie vzorky pripravili z viac vzdialených regiónov fotografií, už boli výsledky horšie (tabuľky A.1, A.2 a A.3). Práve oddelenie trénovacej a testovacej vzorky zhoršilo výsledok metódy najviac, no zároveň bolo z hľadiska metodológie správnym krokom.
- Porovnali sme úspešnosť lineárnej a kvadratickej diskriminačnej analýzy. Úspešnosť bola prekvapivo vyššia pre lineárnu analýzu. Stalo sa tak napriek tomu, že kvadratická je vyššieho rádu a mala by mať tým pádom väčšiu schopnosť oddeliť od seba triedy. Domnievame sa, že kvôli pomerne malému množstvu fotografií (teda príkladov vlákien jednotlivých tried) mohlo dôjsť k pretrénovaniu kvadratického diskriminačného klasifikátora.
- Zmeraním úspešnosti klasifikácie sme nepriamo overovali správnosť dvoch našich algoritmov automatického výberu pixelov z fotografií do vzoriek (metóda krátkej spojnice, metóda vpísanej elipsy). Úspešnosť klasifikácie vyšla nižšia, ako pri manuálnej príprave vzoriek (čo sa dalo čakať), no rozdiel nebol veľký a dá sa povedať, že oba algoritmy sú pre túto úlohu

použiteľné.

- Algoritmy automatického výberu pixelov sme prostredníctvom merania úspešnosti klasifikácie porovnávali aj navzájom. V tomto ohľade medzi nimi neboli zásadné rozdiely. Zároveň však platia zistenia z predchádzajúcej sekcie o odlišnej schopnosti vôbec nejakú vhodnú vzorku na fotografii nájsť a čase potrebnom na výpočet riešení, ktoré ukazujú metódu krátkej spojnice ako použiteľnejšiu.
- Meraním úspešnosti klasifikácie sme porovnali reprezentáciu farieb pomocou RGB a Helmholtzovho systému. RGB systém sa pre klasifikáciu ukázal byť lepší. Ukázalo sa tiež, že ich relatívna rozlišovacia schopnosť pre jednotlivé dvojice tried je odlišná (teda zatiaľ čo jeden systém pre určitú dvojicu zlyháva, druhý ju rozlišuje celkom dobre a zlyháva pri inej).
- Porovnali sme tiež dva rôzne prístupy vytvorenia finálnych príkladov pre trénovanie klasifikátora. Jeden prístup tvorí trénovaciu vzorku opakovaným náhodným výberom skupín pixelov a ich následným priemerovaním, druhý vyberá na trénovanie "surové" pixely. Mierne úspešnejšia bola klasifikácia natrénovaná na základe priemerov. Domnievame sa, že tento výsledok nastal z dôvodu väčšej kompaktnosti zhlukov jednotlivých tried v priestore čírt, ktoré sa pri priemeroch menej prekrývali, zatiaľ čo pri surových pixeloch boli prekryté viac, čo ukazuje aj obrázok 3.8. Opakovaný náhodný výber a priemerovanie spôsobili, že do trénovacej množiny (zhlukov) sa dostali najčastejšie sa vyskytujúce vzorky pixelov. Často vyskytujúce vzorky pixelov sú zároveň aj charakteristické pre danú skupinu. Domnievame sa, že vďaka použitiu týchto charakteristických vzoriek bola úspešnosť klasifikácie vyššia.





Obr. 3.8: Zhluky v priestore (RGB) zodpovedajúce jednotlivým triedam. Možno vidieť, že vzorky pripravené priemerovaním skupín pixelov (vľavo) sú kompaktnejšie ako vzorky pozostávajúce zo surových pixelov (vpravo).

### 3.3 Nástroj pre identifikáciu drevných vlákien pre prostredie Matlab

Prakticky použiteľným výstupom našej práce je nástroj pre prostredie Matlab, ktorý sme vytvorili ako grafické používateľské rozhranie, sprístupňujúce algoritmy a nastavenia pre klasifikáciu drevných vlákien na základe mikrofotografií. Nástroj podporuje načítanie tréningových vzoriek zo súborov obrázkov, nastavenie parametrov tréningovania, výber použitých algoritmov automatickej extrakcie vzorkových pixelov (metóda krátkej spojnice a metóda vpísanej elipsy), výber farebnej schémy (RGB alebo Helmholtz) a samotné tréningovanie. Po natrénovaní nástroj umožňuje načítanie testovacích fotografií (resp. fotografií s neznámymi vláknami) a ich podrobenie sa klasifikácii, ktorá je možná aj s prípadným vyhodnotením úspešnosti (pokiaľ aj k testovacím fotografiám sú známe triedy, do ktorých by ich klasifikátor mal zaradiť).

Nástroj je možné spustiť v Matlabe nasledovným spôsobom:

1. Rozbaliť obsah súboru `identifikacia_vlaken.zip` do ľubovoľného priečinka.
2. Nastaviť pracovný adresár Matlabu na tento priečinok.

3. Pridať podadresár kategorizátor do cesty (pravým kliknutím na priečnik, Add to path, Selected folders and subfolders).
4. Otvoriť a spustiť súbor `classifier.m`.

**Dataset**

	Category	File name	File p
1	1	1_3_Herzberg 40x_P1011026.JPG	C:\Users\Timi\Desktop\nova
2	1	1_4_Herzberg 40x_P1011126.JPG	C:\Users\Timi\Desktop\nova
3	1	1_pokus2_cropped.jpg	C:\Users\Timi\Desktop\nova
4	1	1_pokus3_cropped.jpg	C:\Users\Timi\Desktop\nova
5	1	1_pokus4_cropped.jpg	C:\Users\Timi\Desktop\nova
6	2	2_10_Herzberg 40x_P1010198.JPG	C:\Users\Timi\Desktop\nova
7	2	2_1_Herzberg 40x_P1010160.JPG	C:\Users\Timi\Desktop\nova
8	2	2_2_Herzberg 40x_P1010169.JPG	C:\Users\Timi\Desktop\nova

☒ Attempt to extract categories from file names.  
 (for example "C23\_sample.jpg" will have category 23)

**Training**

Pixel selection method: River crossing  
 Features to consider: RGB  
 Classifier type: Linear discriminant  
 Number of pixels per image: 100  
 Selection rectangle width: 10

Obr. 3.9: Nástroj pre identifikáciu drevných vlákien pre Matlab. Rozhranie pre tréovanie klasifikátora.

Na obrázku 3.9 sa nachádza časť rozhrania nástroja pre tréovanie klasifikátora. Rozhranie obsahuje nasledovné dôležité prvky, ktorých nastavenie by mal používateľ zvážiť:

- Výber tréovacej množiny fotografií (*Dataset*) obsahuje tabuľku so zoznamom súborov fotografií a ich kategórií, ktoré majú byť použité na tréovanie klasifikátora. Do tabuľky možno pridať súbory stlačením voľby *Add files* a následným výberom súborov v dialógu výberu súborov.

- Pokiaľ pri pridávaní používateľ označí možnosť *Attempt to extract categories from file names*, importovací skript sa pokúsi určiť číslo kategórie obrázkov zo začiatkov názvov súborov (za číslo sa bude považovať prvé súvislé číslo nájdené v prefixe názvu súboru, pričom prefix je oddelený od zvyšku názvu podtrhovníkom). Túto možnosť sme pridali z praktických dôvodov, keďže kategória je vo vzťahu k fotografii nemenná a ako informácia je so súborom spätá. Pokiaľ používateľ možnosť neoznačí, musí neskôr doplniť čísla kategórií ručne. Poznámka: pre jednoduchosť pracujeme s kategóriami ako s celými číslami. Preklad týchto čísiel na slovné názvy kategórií aplikácia nerieši.
- Nastavením výberu *Pixel selection method* (časť *Training*) rozhoduje používateľ o tom, aká metóda automatickej extrakcie vzoriek sa použije pri tréovaní a testovaní. K dispozícii sú implementácie našich metód krátkej spojnice (v programe označovanej ako *River crossing*) a vpísanej elipsy (*inscribed ellipse*).
- Nastavením výberu *Features to consider* vyberá používateľ systém farieb, ktorý má byť použitý pri reprezentácii farebnosti vzoriek. K dispozícii sú voľby *RGB* a *Helmholtz*.
- Nastavením výberu *Classifier type* vyberá používateľ typ klasifikátora, ktorý má byť na identifikáciu natrénovaný. Na výber je lineárna a kvadratická diskriminačná analýza.
- Ďalším parametrom v sekcii *Training* je *Number pixels per image*. Určuje, koľko pixelov má byť náhodne vybraných z každej fotografie pri tvorbe tréovacej vzorky.
- Nastavením parametra *Selection rectangle width* špecifikuje používateľ šírku obdĺžnika, z ktorého sa majú pixely do vzorky vyberať (poznámka: os tohto obdĺžnika určí algoritmus automatickej extrakcie vzoriek).
- Stlačením tlačidla *Train* sa uskutoční tréovanie. Jeho priebeh možno sledovať v Matlab konzole.

Classify images

	Categories	File name	File ...	Accuracy	Cat. probabilities
8	[2 2 2 2 2]	2_7_Herzberg 4...	C:\Us...	100.00	[0 95 4 1]
9	[2 2 2 2 2]	2_pokus12_cro...	C:\Us...	100.00	[0 83 15 2]
10	[2 2 2 2 2]	2_pokus14_cro...	C:\Us...	100.00	[0 63 31 6]
11	[4 4 4 4 4]	3_1_Herzberg 4...	C:\Us...	0.00	[1 5 5 90]
12	[3 3 3 3 3]	3_4_Herzberg 4...	C:\Us...	100.00	[0 2 97 1]
13	[3 3 3 3 3]	3_pokus14_cro...	C:\Us...	100.00	[0 5 95 0]
14	[4 4 4 4 4]	3_pokus15_cro...	C:\Us...	0.00	[0 27 16 56]
15	[3 3 3 3 3]	4_1_Herzberg 4...	C:\Us...	0.00	[0 8 91 2]

< >

Add files Clear all

Number of test samples per image: 5

Number of pixels in one sample: 100

☒ Also evaluate classifier accuracy  
(using reference categories extracted from file names)

Classify

Obr. 3.10: Nástroj pre identifikáciu drevných vlákien pre Matlab. Rozhranie pre vykonanie a testovanie klasifikácie.

Na obrázku 3.10 sa nachádza časť rozhrania pre obsluhu klasifikácie, resp. testovania jej úspešnosti. Rozhranie obsahuje nasledovné dôležité prvky:

- Výber fotografií pre klasifikáciu/testovanie v podobe tabuľky. Pri importe súborov pomocou tlačidla *Add files* sa tabuľka vyplní názvami súborov a cestami k nim (druhý a tretí stĺpec). Prvý a posledné dva stĺpce pri importe zostávajú prázdne.
- Nastavením hodnoty *Number of test samples per image* určí používateľ počet opakovaní klasifikácie na každú fotografiu (viacnásobné opakovanie má

zmysel nakoľko algoritmy výberu pixelov nedajú vždy rovnaký výsledok).

- Nastavením hodnoty *Number of pixels in one sample* určí používateľ počet pixelov, ktorý sa z fotografie má zobrať na vytvorenie priemernej farby použitej ako testovacia vzorka.
- Stlačením tlačidla *Classify* sa spustí algoritmus klasifikácie každej z importovaných fotografií (vrátane automatického výberu vzoriek). Priebeh klasifikácie možno sledovať v Matlab konzole. Po skončení klasifikácie sa jej výsledky objavia v tabuľke v prvom stĺpci, kde budú zobrazené výsledky (kategórie) zistené pri všetkých testoch (príklad: ak zadal používateľ 10 opakovaní pre každú fotografiu, výsledkom bude desať čísiel označujúcich výsledné kategórie umiestnených vedľa seba v riadku). V poslednom stĺpci si môže používateľ pozrieť priemerné pravdepodobnosti, s ktorými klasifikátor priradil vzorku jednotlivým triedam.
- Používateľ sa pred klasifikáciou môže rozhodnúť označiť voľbu *Also evaluate classifier accuracy*. V takom prípade klasifikačný skript navyše extrahuje čísla kategórií z názvov testovaných súborov (samozrejme, za predpokladu, že sú tam prítomné) a porovná ich s výsledkami, ktoré vyjdú v klasifikácii. Výslednú úspešnosť pre každú fotografiu zapíše do tabuľky do štvrtého stĺpca.

## 4. Zhodnotenie

Experimenty, ktoré sme uskutočnili potvrdzujú, že objektívna (automatická) analýza mikrofotografií má potenciál konkurovať iným typom metód identifikácie drevných vlákien ako sú subjektívne metódy či metódy založené na chemickom rozbere. V tejto práci sme identifikáciu uskutočnili klasifikáciou pomocou diskriminačnej analýzy na základe farebných vzoriek.

Ukázali sme rôznu úspešnosť klasifikácie pre viaceré nastavenia. Experimentovali sme s dvoma farebnými schémami, RGB a Helmholtzovým farebným systémom, pričom použitie RGB sa ukázalo ako lepšie. Skúmali sme tiež presnosť pri použití lineárneho a kvadratického klasifikátora, pričom ako lepší vyšiel prekvapivo lineárny. Ukázali sme tiež, že úspešnosť klasifikácie je značne ovplyvnená spôsobom, akým pripravujeme tréningové a testovacie vzorky a že pri ich príprave treba poctivo dbať na zachovanie dostatočnej rozmanitosti vzoriek, inak môže dôjsť k pretrénovaniu klasifikátora.

Pôvodne manuálnu prípravu farebných vzoriek z mikrofotografií sme sa pokúsili automatizovať vytvorením dvoch algoritmov (metóda krátkej spojnice a metóda vpísanej elipsy). Tieto algoritmy sa v experimentálnom overení ukázali byť slubnou náhradou manuálnej prípravy.

Praktickým prínosom našej práce je aj vizuálny nástroj pre prostredie Matlab, pomocou ktorého je možné trénovanie klasifikačných modelov a ich následné použitie.

# Literatúra

- [1] Lenka Dubinyová, Michal Jablonský, Štefan Varga, Miroslav Fikar, and Svetozár Katuščák. Cellulose materials identification: The effect of dimensionality of colour photography data. *BioResources*, 11(1), 2015.
- [2] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7):179–188, 1936.
- [3] Michal Jablonský, Lenka Dubinyová, Štefan Varga, Katarína Vizárová, Jozef Šima, and Svetozár Katuščák. Cellulose fibre identification through color vectors of stained fibre. *BioResources*, 10(3), 2015.
- [4] S. Katuščák. *Farba, koloristika, farebné videnie, vnímanie a komunikácia*. Slovenská Technická Univerzita v Bratislave, 1994.
- [5] J. Kopačka. Metódy sekvenčného kvadratického programovania, 2004. diplomová práca.
- [6] Aleš Kozubík. Teória pravdepodobnosti, Február 2015. prednáška.
- [7] David P. Luebke. A developer’s survey of polygonal simplification algorithms. *IEEE Comput. Graph. Appl.*, 21(3):24–35, May 2001.
- [8] R. Luus and T. H. I. Jaakola. Optimization by Direct Search and Systematic Reduction of the Size of the Search Region. *American Institute of Chemical Engineers (AIChE) Journal*, 19:760–766, 1973.
- [9] Š. Varga. *Mathematical Statistics*. Slovak Technical University Press, Bratislava, Slovakia, 2012. [In Slovak].