

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Jiří Palas

Studijní program: Otevřená informatika (magisterský)

Obor: Počítačové vidění a digitální obraz

Název tématu: Automatizovaná analýza částic v mikroskopických snímcích

Pokyny pro vypracování:

1. Seznamte se se stávajícími metodami pro automatickou detekci a klasifikaci částic v dvourozměrných mikroskopických snímcích.
2. Na základě provedené analýzy navrhněte a implementujte vhodný způsob (metodu, algoritmus) pro automatizovaný předvýběr hledaných částic v mikroskopickém snímku. Současně navrhněte a implementujte způsob, jakým může laborant ručně opravit automatizovaně provedený výběr a označit jednotlivé třídy částic.
3. Navrhněte atributy (příznaky) popisující částice, které mohou sloužit pro automatizovanou identifikaci různým tříd částic, a implementujte metody pro jejich extrakci.
4. Implementujte vhodný klasifikátor (např. K-NN, Bayesův klasifikátor, SVM, ...), který umožní automatizovaně klasifikovat částice do příslušných tříd na základě extrahovaných atributů v dalších mikroskopických snímcích.
5. Implementujte statistické metody, které budou následně aplikovány na vybrané částice, pro vypočet požadovaných charakteristik, jako je např. počet částic ve třídách, zda částice tvoří/netvörí shluky, kde se částice nachází v prostoru ad.
6. Pro implementaci metod použijte programovací jazyk Java v prostředí platformy NetBeans a využijte knihovnu OpenCV.
7. Navržené a implementované metody ověřte na reálných datech a získané výsledky vyhodnoťte. Navrhněte možná další zlepšení.

Seznam odborné literatury:

- [1] Szeliski R. 2010. Computer Vision: Algorithms and Applications (1. vydání). Springer-Verlag New York, Inc., New York, NY, USA.
[2] Ashbrook A., Thacker N.A. Tutorial: Algorithms for 2-Dimensional Object Recognition. Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester, 1998.

Vedoucí diplomové práce: doc. RNDr. Ing. Marcel Jiřina, Ph.D.

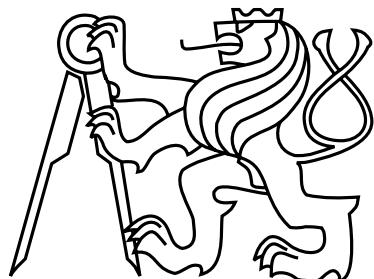
Platnost zadání: do konce letního semestru 2015/2016

L.S.

doc. Dr. Ing. Jan Kybic
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra kybernetiky



Diplomová práce

**Automatizovaná analýza částic v mikroskopických
snímcích**

Bc. Jiří Palas

Vedoucí práce: doc. RNDr. Ing. Marcel Jiřina, PhD.

Studijní program: Otevřená informatika, Magisterský

Obor: Počítačové vidění a digitální obraz

9. května 2015

Poděkování

Na tomto místě bych rád poděkoval maminec za vaření obědů a večeří během krušných hodin psaní.

Prohlášení

Prohlašuji, že jsem práci vypracoval samostatně a použil jsem pouze podklady uvedené v přiloženém seznamu.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 11. 5. 2015

Abstract

Abstrakt v angličtině.

Abstrakt

TODO Ještě nějak lépe přeformulovat

Tato práce si klade za cíl návrh a implementaci software zaměřeného na automatizovanou analýzu částic v mikroskopických snímcích. Práce se zabývá analýzou současných metod zpracování obrazu a strojového učení využitelnými pro automatizovanou detekci, klasifikaci a analýzu mikroskopických částic. Na jejím základě je navržen a implementován algoritmus pro automatizovanou detekci a klasifikaci částic. V rámci práce jsou navrženy a implementovány metody vizuální a interaktivní analýzy detekovaných a klasifikovaných částic.

Obsah

0.1 Motivace	1
0.2 Přehled	1
1 Analýza	3
1.1 Specifikace	3
1.2 Data	4
2 Detekce nanočástic	7
2.0.1 Úvod do problematiky	7
2.0.2 Popis použitého řešení	8
2.0.3 Výsledky segmentace	9
3 Klasifikace	11
3.0.4 Úvod do problematiky	11
3.1 Trénovací množina	11
3.2 Volba klasifikačního algoritmu	12
3.3 Výsledky klasifikace	14
3.3.1 Vstupní data	14
3.3.2 Vzorek 1	15
3.3.3 Vzorek 2	16
3.3.4 Vzorek 3	17
3.3.5 Vzorek 4	18
3.3.6 Vzorek 5	19
3.3.7 Vzorek 6	20
3.3.8 Celkový přehled	21
4 Design a implementace	23
4.1 Způsob práce s daty	23
4.2 Uspořádání aplikace	24
4.2.1 Editační panel	25
4.2.1.1 Editor snímků	25
4.2.1.2 Panel analýzy	25
4.2.2 Přehled projektů	26
4.2.3 Přehledu klasifikátorů	27
4.2.4 Inspekční panel	27
4.3 Uchování dat	28
4.3.0.1 Ukládání dat snímků	28

4.3.0.2 Ukládání dat klasifikátoru	29
4.4 Implementace	30
5 Závěr	31
5.1 Diskuze	31
5.2 Shrnutí práce	31
5.3 Další směrování	31
A Seznam použitých zkratek	35
B Obsah přiloženého CD	37

Seznam obrázků

1.1	Příklad vyhodnocovaného snímku.	4
2.1	Příklad výstupu segmentačního algoritmu na vzorku 5.	10
2.2	Příklady chybně provedené segmentace.	10
3.1	Množina trénovacích dat.	15
3.2	Výsledek klasifikace částic pro vzorek 1.	15
3.3	Výsledek klasifikace částic pro vzorek 2.	16
3.4	Výsledek klasifikace částic pro vzorek 3.	17
3.5	Vyhodnocení klasifikace částic pro vzorek 4.	18
3.6	Výsledek klasifikace částic pro vzorek 5.	19
3.7	Výsledek klasifikace částic pro vzorek 6.	20
4.1	Přehled rozložení aplikace.	24
4.2	Ukázka panelu přehledem projektů a snímků.	26
4.3	Ukázka panelu s přehledem klasifikátorů.	27
4.4	Příklad souborové a složkové struktury projektu.	28

Seznam tabulek

2.1 Výsledky segmentace.	9
3.1 Vyhodnocení klasifikace částic pro vzorek 1.	16
3.2 Vyhodnocení klasifikace částic pro vzorek 2.	17
3.3 Vyhodnocení klasifikace částic pro vzorek 3.	18
3.4 Vyhodnocení klasifikace částic pro vzorek 4.	19
3.5 Vyhodnocení klasifikace částic pro vzorek 5.	20
3.6 Vyhodnocení klasifikace částice pro vzorek 6.	21
3.7 Vyhodnocení klasifikace pro všechny vzorky.	21

Úvod

0.1 Motivace

V současné době se čistý a aplikovaný výzkum v biologii, medicíně a nanotechnologiích opírá o pozorování procesů uvnitř buněk, tkání a orgánů. Mezi takové procesy patří pozorování distribuce léčiv a genů, lokalizaci nádorů, toxikologickému testování, prověřování zdravotních rizik, reakce na chirurgické implantáty atd. [13] [12] [5].

Pro pozorování těchto procesů se využívají nano-částice (NP). NP jsou částice o rozměrech mezi 1-100 nm. Mezi nejčastěji používané NP patří částice vytvořené chemických sloučenin s kovy. Různé typy NP se odlišují buď tvarem nebo velikostí. Na základě svých vlastností pak NP vykazují specifické chování ve výše zmíněných procesech [8].

NP mají velice malý rozměr na to aby se daly snímat pomocí světelné mikroskopie. Proto se k jejich pozorování používají biologické zobrazovací metody pracující ve velikém rozlišení. Mezi takové metody patří například transmisní elektronová mikroskopie (TEM), rastrovací elektronová mikroskopie (SEM) a atd. [13] [12] [5]. Výsledkem těchto zobrazovacích metod je snímek nebo série snímků.

Snímky jsou následně vědci či laboranty vyhodnocovány. Pracovníci UAKČR UMG uvedli, že postrádají kvalitní, specializovaný a přitom jednoduchý nástroj pro vyhodnocování snímků s nanočásticemi.

Tato práce si klade za cíl vytvořit aplikaci pro vyhodnocování snímků. Práce klade důraz na vytvoření co největší úrovně automatizace v aplikaci. Tak aby došlo k celkovému zkvalitnění a zrychlení vyhodnocování jednotlivých snímků. Základem pro vyhodnocování snímků je lokalizace jednotlivých částic. Dále jejich úspěšná klasifikace do jednotlivých tříd (kategorií) podle typů NP. Následná expertní analýza laborantem.

0.2 Přehled

Ke každé kapitole v krátkosti napsat co tam je.

Kapitola 1

Analýza

1.1 Specifikace

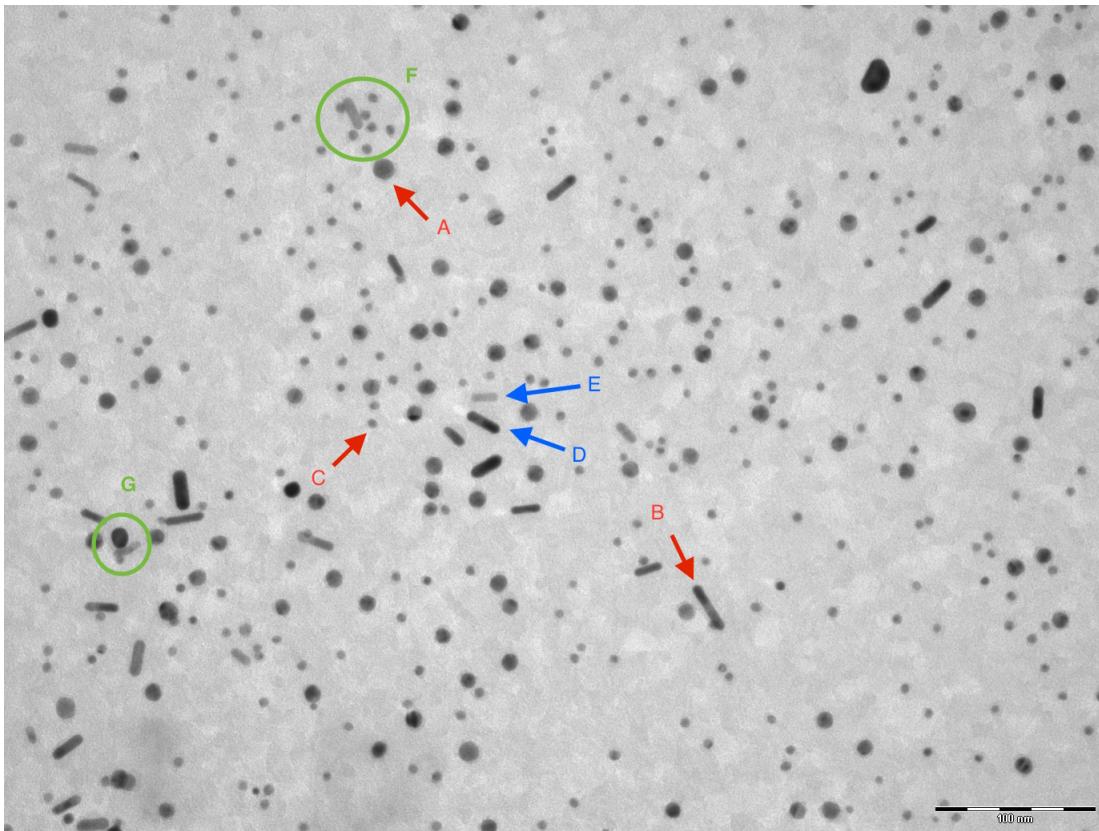
Na základě návštěvy pracovišť UAKČR UMG jsou popsány požadavky na aplikaci pro automatizovanou analýzu mikroskopických snímků obsahující nanočástice:

Typická práce laboranta vyhodnocujícího NP se sestává z následujících částí:

- příprava biologického vzorku
- nasnímání vzorku pomocí elektronového mikroskopu
- identifikace částic
- klasifikace (kategorizace) vzorků
- analýza vzorku

Práce se zaměřuje na automatizaci nebo částečnou automatizaci bodů 3-5. Ty se pak podrobněji zahrnují

- načítání mikroskopické obrázky ve formátech tiff, jpeg
- zobrazení načtených obrázků
- schopnost detekovat NP v obrázku
- schopnost nalézt těžiště NP
- zobrazení detekovaných NP v obrázku
- možnost označit jednotlivé třídy NP
- automatizace označování NP
- vizuální shluková analýza NP



Obrázek 1.1: Příklad vyhodnocovaného snímku. Červené šipky ukazují různé tvary vyhodnocovaných NP. Tvar A: Koláč. Tvar B: Tyčinka. Tvar C: Kolečko. Částice E a D demonstруjí možné rozdíly v jasech částic. Region G zobrazuje příklad překryvu částic. Region F zobrazuje příklad shlukování částic.

1.2 Data

Aplikace se zabývá vyhodnocováním NP obsažených ve snímcích jako na obr. 1.1. Snímky jsou 8-bitové jednokanálové (šedotónové) obrázky s rozlišením mezi 2 a 3 megapixely.

Částice ve snímku můžeme pozorovat jako kontrastní útvary s nízkým jasem (tmavě šedé až černé) na světlém pozadí.

Snímané vzorky jsou v mikroskopu nerovnoměrně osvětleny. To se projevuje na nerovnoměrném rozložení jasu v obrázku. Viz. levý horní roh a pravý horní roh na obr. 1.1. Dalším negativním jevem, který znesnadňuje detekci je tendence NP tvořit shluky viz. region F na obr. 1.1 a překrývat se viz. region G na obr. 1.1. Navíc ani částice nemají jednotnou úroveň jasu viz. částice E a D na obr. 1.1.

Práce rozlišuje tři typy částic A, B, C viz. obr. 1.1. Pracovní názvy těchto částic jsou koláč (A), tyčinka (B) a kolečko (C). K rozlišení těchto částic uvažu-

jeme informaci o tvaru a velikosti. Jednotlivé částice lze považovat za konvexní útvary [18]. NP mohou být v obrázku libovolně orientované. V určitých částech snímku splývají s pozadím více viz. a v některých méně viz. Částice nemají výraznou texturu. Snímek typicky obsahuje 300 - 700 částic.

Na základě provedené charakteristiky obrazu a častic je vybrán vhodný detekční a klasifikační algoritmus viz kapitola 2 a kapitola 3.

Kapitola 2

Detekce nanočástic

2.0.1 Úvod do problematiky

Cílem detekčního algoritmu je úspěšně nalézt jednotlivé částice v obraze. Dále je třeba nalezené částice popsat modelem na jehož základě bude částice možné klasifikovat pomocí klasifikačního algoritmu a dále vyhodnocovat.

Pro úspěšnou detekci částic je nejprve potřeba jejich segmentace. Segmentace je proces při kterém jsou ze snímku odděleny jednotlivé objekty a části od pozadí. Jedná se často o nejsložitější a nejdůležitější krok při analýze obrazu.

Navzdory důležitosti, existuje pouze limitované množství zdrojů o automatické segmentaci nanočástic využitelných pro náš případ. Všechny nalezené zdroje se zabývají převážně detekcí NP kruhovitého tvaru [?][?][4]. Na druhou stranu ale existuje rozsáhlý výzkum zabývající se segmentací biologických buněk. V následujícím textu je shrnuta literatura zabývající se segmentací biologických buněk. Tento směr výzkumu je přímo spjat s jedním z cílů této práce [?].

Velice často používanou metodou pro segmentaci buněk je prahování založené na charakteristikách histogramu obrázku. Přehled prahovacích technik uvádí [15]. Pro obdržení uspokojivého výsledku segmentace za pomoci prahování je ale zapotřebí pozadí obrázku s neměnnou intenzitou. Je také třeba pamatovat na to, že malá změna v úrovni prahu může mít negativní dopad na další analýzu obrázku [?].

Adaptivní prahování nebo také lokální automatické prahování může být řešení pro prahování obrázku ve kterém intenzita pozadí kolísá. Tato metoda ale stále neřeší problémy spojené s segmentací objektů které tvoří shluky. Prahování nemusí být finálním krokem v segmentačním procesu. Naopak často bývá využíván právě jako první krok před dalším zpracováním např. pomocí morfologických operací [?] [?].

Další hojně užívanou metodou pro segmentaci buněk je tzv. *watershed* a jeho odvozeniny. Metoda nejprve nalezne body (markery) ukazující přibližnou polohu buněk a následně segmentuje obrázek na několik ovlivněných markery [11] [?] [?] [?].

Graph-cut metody jsou také aplikovány na segmentaci buněk. Metoda zkonstruuje graf z obrázku kde každý pixel odpovídá jednotlivým vrcholům grafu. Každý pár vrcholů je spojen se sousedními vrcholy hranou jejíž váha odpovídá podobnosti intenzit sousedních pixelů. Segmentace pak hledá normalizovaný minimální řez grafem [?]. Avšak tento přístup špatně separuje překrývající se buňky, obzvláště pokud je jejich intenzita podobná. Existuje úprava která tento problém řeší, funguje však pouze na částice buňky kruhového tvaru [?].

Mezi další segmentační metody používané při segmentaci buněk se řadí **aktivní kontury** [?] [?]. Aktivní kontury jsou originálně navržené k segmentaci jednoho objektu s komplikovaným tvarem. Avšak po úpravě vícefázový algoritmus může být použit k segmentaci více nepřekrývajících se objektů. Článek [?] na základě dalších použití na segmentaci překrývajících se buněk nedoporučuje.

2.0.2 Popis použitého řešení

Na základě předchozí analýzy byla ozkoušena segmentace metodami prahování *watershed* a adaptivního prahování.

Adativní prahování překvapivě produkuje od začátku dobrou míru segmentace. Avšak nedokáže oddělit částice, které se dotýkají nebo překrývají. Proběhl pokus zpřesnit výsledky pomocí watershed algoritmu. Avšak výsledek ke zpřesnění výstupu segmentace nedošlo, protože nebyl nalezen korektní způsob umístění markerů. Bylo vyzkoušeno umístění na základě distanční transformace [10] a na základě *ultimátní eroze pro konvexní sety* [?]. Současné řešení se tedy ustálilo na níže popsaném řešení.

Snímky s NP obsahují malé množství šumu, proto je obrázek předzpracován mediánového filtru s oknem o velikosti 3×3 pixely.

Z důvodů problémů s jasem popsaných dříve je pro úvodní segmentaci vybrán algoritmus lokálního adaptivního prahování. Algoritmus volí práh pro každý pixel individuálně na základě rozsahu hodnot intenzity v lokálním okolí [14]. Výstupem algoritmu je pak binární obrázek.

V této aplikaci se hodnota prahu v bodě (x, y) získává jako:

$$t(x, y) = \sum_{i=x-(B-1)/2}^{x+(B-1)/2} \sum_{j=y-(B-1)/2}^{y+(B-1)/2} f(i, j) - C \quad (2.1)$$

Kde $t(x, y)$ je hodnota prahu v bodě (x, y) , $f(i, j)$ hodnota intenzity v bodě (i, j) , B velikost okna a C konstanta. Velikost okna musí být liché číslo.

Binární obrázek získaný prahováním je následně upraven pomocí morfológické operace otevření s jádrem:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (2.2)$$

V binární obrázek, upraveným morfologickou operací jsou následně prostřednictvím algoritmu [16] nalezeny obrysy jednotlivých NP. Obrys je reprezentován vektorem bodů, kterého ho tvoří.

K detekci byla využita implementace adaptivní filtrace a algoritmu [16] z OpenCV[?]. Adaptivní filtrace využívá funkce *adaptiveThresh* a algoritmus na vyhledávání obrysů pak funkci *findContours*.

2.0.3 Výsledky segmentace

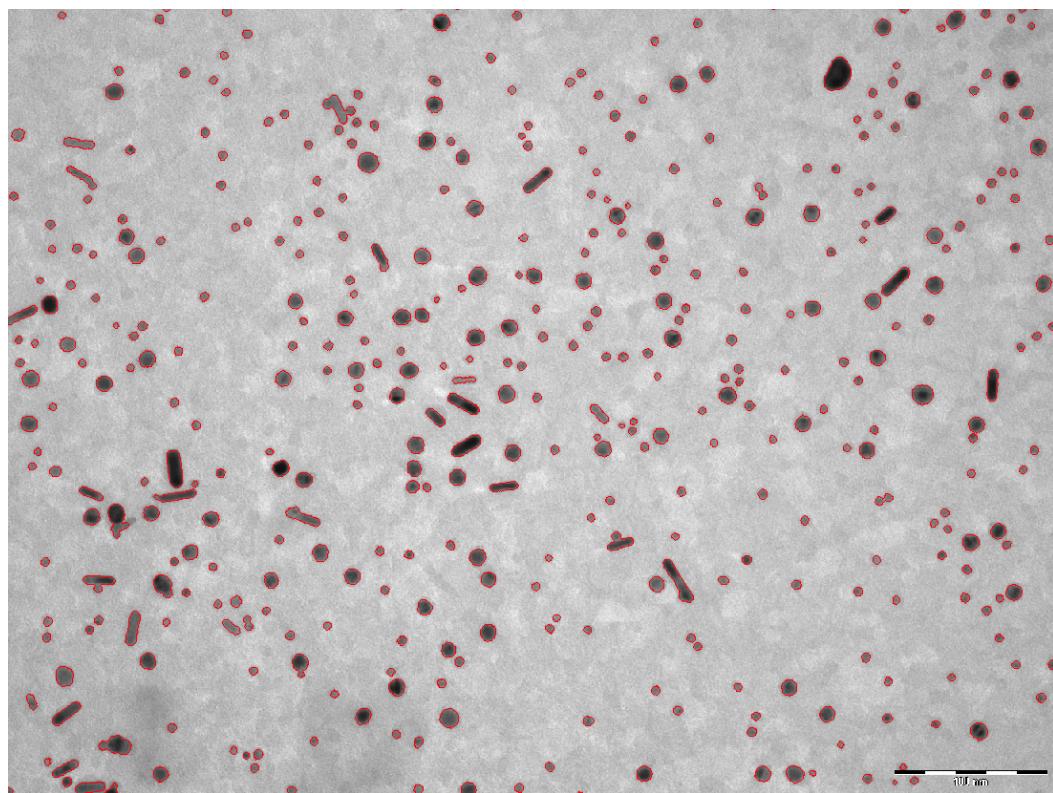
Segmentace a její vyhodnocení výsledků probíhalo na datech snímcích pocházejících z UMG. Ze snímků které jsou k dispozici je zvoleno 7 mikroskopických snímků obsahujících variující tvary částic. Výstupy segmentace algoritmu jsou prezentovány pomocí barevného značení, kde kontura nalezené částice je zobrazena červeně.

Příklad výstupu segmentačního algoritmu je možné pozorovat na obr. 2.1. Detaily chybně provedené segmentace jsou znázorněny na obr. 2.2.

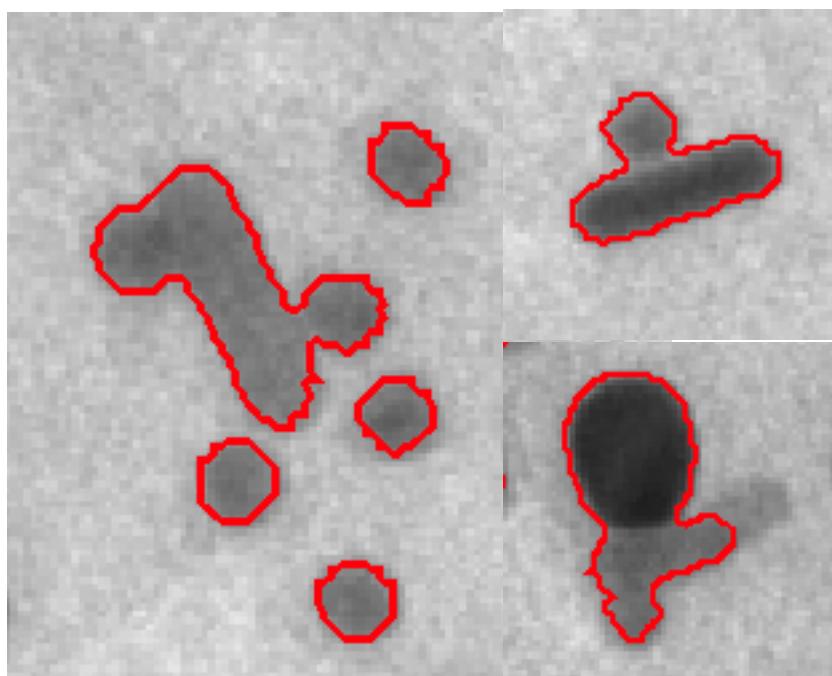
Vyhodnocení výstupu segmentace je možné pozorovat v tabulce 2.1. Nejčastějším problémem při segmentaci je segmentace spojených částic typu tyčinka a kolečko. Ve vzorcích 1, 2, 7 se vyznačují malou mírou shlukování částic. Napak ve vzorcích 3, 4, 5, 6 jsou částice poskládány blíže sobě a některé se překrývají.

Vzorek	Chybná segmentace	Očekáváno	Přesnost
1	1	353	0.997
2	1	383	0.997
3	29	435	0.932
4	31	458	0.932
5	27	426	0.936
6	8	411	0.980
7	1	338	0.997
Celkem	98	2809	0.965

Tabulka 2.1: Výsledky segmentace.



Obrázek 2.1: Příklad výstupu segmentačního algoritmu na vzorku 5.



Obrázek 2.2: Příklady chybně provedené segmentace.

Kapitola 3

Klasifikace

Tato kapitola se pojednává o klasifikaci NP do jednotlivých tříd A, B nebo C. Vstupem pro klasifikátor je množina laborantem ručně klasifikovaných NP na jejímž základě probíhá zařazení nezařazených NP do příslušných tříd. Trénovací množinu určuje uživatel prostřednictvím rozhraní aplikace.

3.0.4 Úvod do problematiky

Protože je trénovací množina určena uživatelem je využito algoritmu učení s učitelem. Mezi takové algoritmy patří např. rozhodovací stromy, Support Vector Machine (SVM), naivní Bayes, neuronové sítě nebo algoritmus k-nejbližších sousedů (K-NN). Jednotlivé klasifikační metody shrnuje [?]

Typické kroky při tvorbě klasifikátoru učení s učitelem jsou:

- Určení trénovací množiny dat
- Získ trénovací množiny dat
- Tvorba příznaků
- Volba trénovacího algoritmu
- Vyhodnocení výsledků

3.1 Trénovací množina

Trénovací množinu dat tvoří data jednotlivé částice s přiřazenou příslušností do A, B nebo C. Částice jsou získávány jako výstup detekčního algoritmu. Jednotlivé částice od sebe odlišujeme tvarem, definovaným vektorem bodů, a velikostí plochy v pixelech.

K vytvoření trénovací množiny se v aplikaci nachází interaktivní uživatelské rozhraní, které umožňuje přiřadit jednotlivým částicím příslušnost do tříd a vložit množinu trénovacích dat do klasifikátoru.

3.2 Volba klasifikačního algoritmu

Na základě malého počtu příznaků a průběžných experimentálních výsledků byl ke klasifikaci je vybrán algoritmus **k-nejbližších sousedů (K-NN)**.

Jak bylo zmíněno algoritmů strojového učení je celá řada. Jak tedy vybrat ten správný. Správným klasifikátorem se miní takový který bude úspěšně klasifikovat NP s 80% přesnosti. Pro výběr klasifikátoru byla zvolena experimentální strategie. Tedy implementovat klasifikátor otestovat s jakou přesností klasifikuje. Pokud se klasifikační přesnost pohybuje pod danou hranicí pak . Přibližná hranice byla stanovena na 80% přesnosti klasifikace.

Jako první byl zvolen implementačně nenáročný K-NN klasifikátor. K-NN je neparametrická metoda využívaná pro klasifikaci a regresi. Spočívá ve vyhledávání k nejbližších sousedů v prostoru příznaků. K-NN potřebuje pro svou funkci definovat funkci vzdálenosti. Existuje řada metrik jak vzdálenost měřit. Mezi takové metriky patří např. Minkowského metrika také uváděná jako L_k norma. Jejími podmnožiny pak tvoří např. Manhattanská či Eukleidovská vzdálenost [?].

Práce využívá pro měření vzdálenosti eukleidovský metriku na vektor o dvou příznacích. Prvním příznakem je plocha částice a druhým tvarová podobnost.

Tvarovou podobnost částice porovnáváme na základě Hu-momentů[?]. Jedná se o momenty pomocí nichž dokážeme tvar identifikovat. Výsledkem je deskripitor invariantní vůči translaci, rotaci z změně měřítka. Moment tvaru v binárním obrázku o rozměrech $M \times N$ je popsán jako:

$$u_{pq} = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} i^p j^q f(i, j) \quad (3.1)$$

Kde $f(x, y)$ je intenzita pixelu (v binárním obrázku 1 nebo 0) na souřadnicích (x, y) a $p + q$ udává stupeň momentu.

Protože výpočet momentu je funkce vzdálenosti mezi pixely tvaru a středem všech měření vztaženými relativně k těžišti tvaru (x', y') za účelem odstranění stupně volnosti v posuvu. Souřadnice těžiště pak lze vypočítat za použití rovnice 3.1 jako:

$$i' = \frac{u_{10}}{u_{00}} \quad j' = \frac{u_{01}}{u_{00}} \quad (3.2)$$

Odtud pak relativní momenty:

$$u_{pq} = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (i' - i)^p (j' - j)^q f(i, j) \quad (3.3)$$

Jednotlivé momenty nejsou samy o sobě dostatečně popisné aby dokázaly reprezentovat libovolné tvary, nebo udržovaly invariantní charakteristiky tvaru. Na jejich základě je odvozeno sedm invariantních momentových funkcí jejichž forma je již v hodná pro reprezentaci tvaru. Tyto momenty jsou:

$$m_1 = (u_{20} + u_{02}) \quad (3.4)$$

$$m_2 = (u_{20} - u_{02})^2 + 4u_{11}^2 \quad (3.5)$$

$$m_3 = (u_{30} - 3u_{12})^2 + (3u_{21} - u_{30})^2 \quad (3.6)$$

$$m_4 = (u_{30} + u_{12})^2 + (u_{21} + u_{03})^2 \quad (3.7)$$

$$\begin{aligned} m_5 = & (u_{30} - 3u_{12})(u_{30} + u_{12})((u_{30} + u_{12})^2 - 3(u_{21} + u_{03})^2) \\ & + (3u_{21} - u_{03})(u_{21} + u_{03})(3(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2) \end{aligned} \quad (3.8)$$

$$\begin{aligned} m_6 = & (u_{20} - u_{02}) - (u_{30} + u_{12})^2 - (u_{21} + u_{03})^2 \\ & + 4u_{11}(u_{30} + 3u_{12})(u_{21} + u_{03}) \end{aligned} \quad (3.9)$$

$$\begin{aligned} m_7 = & (3u_{21} - u_{03})(u_{30} + u_{12})((u_{30} + u_{12})^2 - 3(u_{21} + u_{03})^2) \\ & - (u_{30} - 3u_{12})(u_{21} + u_{03})(3(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2) \end{aligned} \quad (3.10)$$

Podobnost jednoho tvaru s druhým pak vypočítáme jako:

$$s(A, B) = \sum_{i=1 \dots 7} |m_i^A - m_i^B| \quad (3.11)$$

Kde $s(A, B)$ je podobnost tvarů A a B, m_i je jeden ze sedmi Hu-momentů z rovnic 3.4-3.10. Výsledkem je hodnota rostoucí od 0 a platí, že čím větší menší podobnost tím větší hodnota.

Jeden z hlavních nedostatků K-NN při výpočtu vzdálenosti od trénovacích dat je příklad kdy jednotlivé příznaky mají rozdílné škály nebo pokud se jedná o směs numerických a kategorických hodnot. Toto je příklad mnou zvolených příznaků. Velikost plochy částic se pohybuje mezi desítkami až stovkami pixelů. Zatímco podobnost se pohybuje v přibližném rozpětí mezi 0 a 2. Hodnoty je proto třeba normalizovat. Jinak by klasifikátor klasifikoval částice pouze podle velikosti [6] [2].

Každý příznak x_{ij} ($i = 1 \dots k, j = 1 \dots n$), kde k je velikost vektoru příznaků a n velikost trénovací množiny, je pak normalizován jako:

$$x'_{ij} = \frac{x_{ij} - \min_i}{\max_i - \min_i} \quad (3.12)$$

Kde $\min_i = \min_{i=1 \dots k} (x_{ij})$ a $\max_i = \max_{j=1 \dots k} (x_{ij})$.

Normalizace se provádí vždy při přidání nové částice mezi trénovací data. V případě, že příznaky nově přidané částice posouvají \min_i nebo \max_i jsou přeskálovány i ostatní hodnoty podle nového rozpětí.

Pro výpočet podobnosti tvarů byla využita funkce `matchShapes` z OpenCV[?].

Prezentovaná metoda volby příznaků je náchylná na uživatelskou chybu špatného zařazení při učení.

3.3 Výsledky klasifikace

3.3.1 Vstupní data

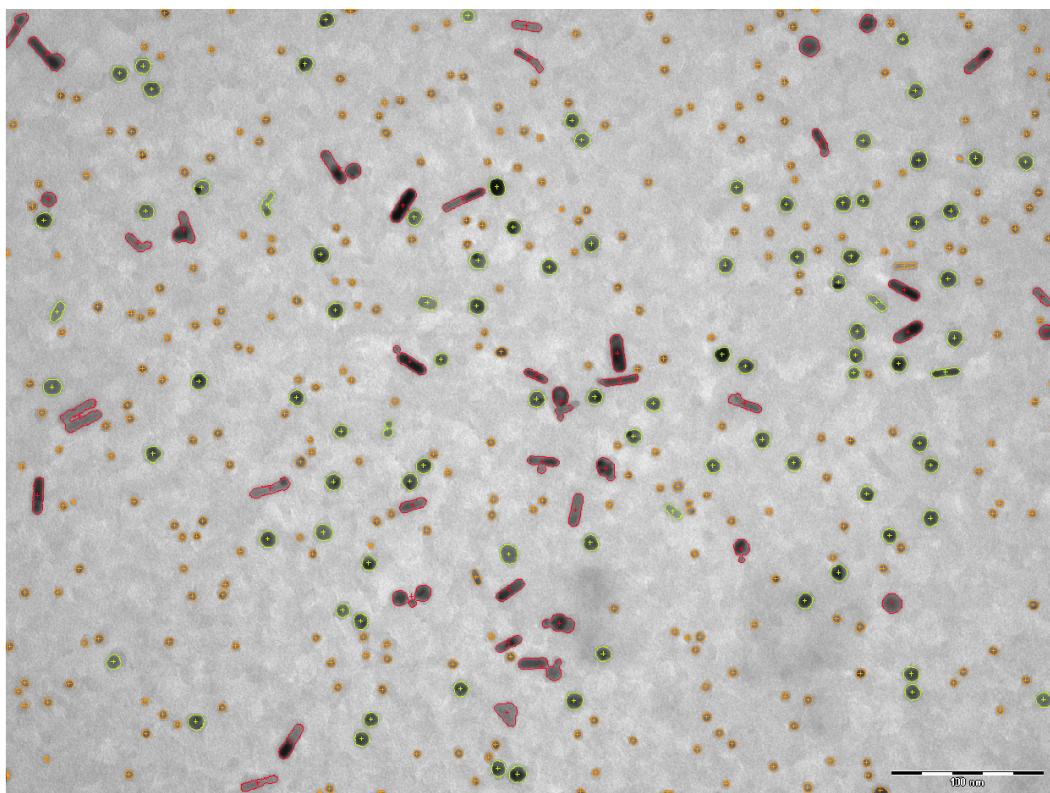
Trénování a vyhodnocení výsledků probíhalo na datech snímcích pocházejících z UMG. Ze snímků které jsou k dispozici je zvoleno 7 mikroskopických snímků obsahujících variující tvary nanočástic. Výstupy klasifikačního algoritmu jsou prezentovány pomocí barevného značení, kde oražová barva značí částici typu kolečko, červená částici typu tyčinka a žlutozelená typ koláč.

Výstup automatické klasifikace je porovnáván vůči výstupu jak by částice klasifikoval laborant.

Vstupní trénovací množina je zobrazena na obr. ???. Zvolený algoritmus

Nastavení parametrů pro adaptivní prahování:

- Velikost okna: 51
- C: 21



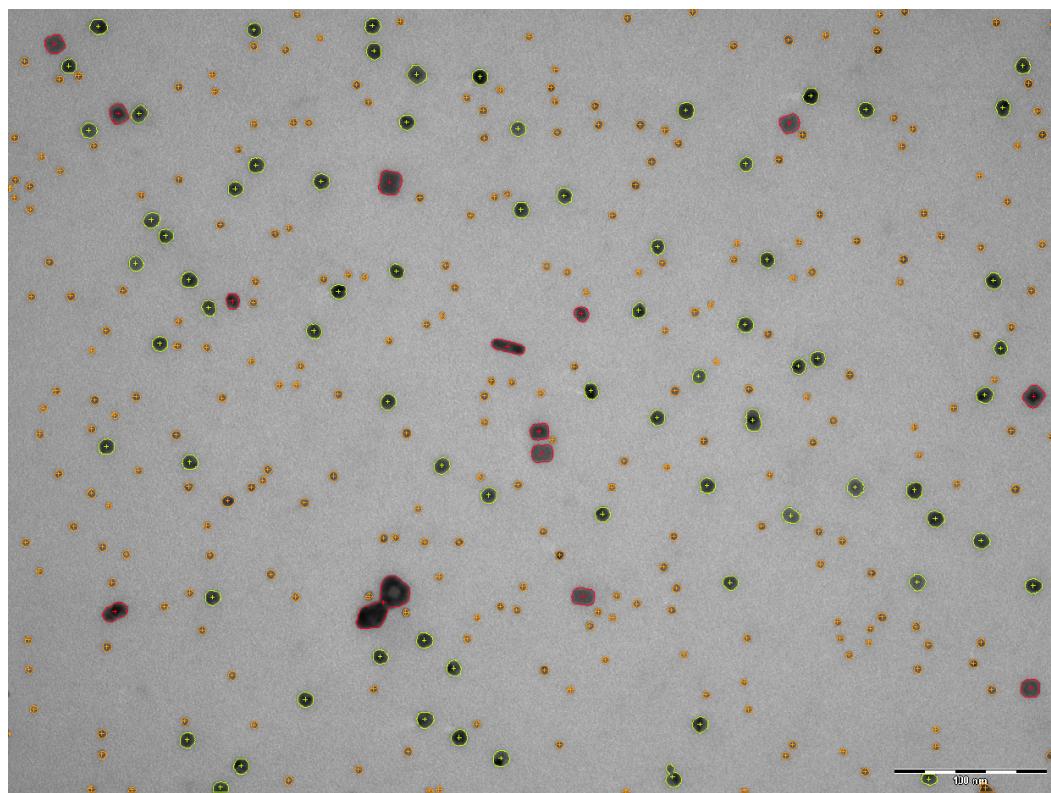
Obrázek 3.1: Množina trénovacích dat.

3.3.2 Vzorek 1

V této části se nachází zjištěné hodnoty výsledků klasifikace pro vzorek 1.

Třída	Dobře	Špatně	Celkem	Přesnost
-1	0	2	2	0,000
0	2	0	2	100,000
1	267	0	267	100,000
2	72	11	83	86,747
Celkem	341	11	352	96,875

Tabulka 3.1: Vyhodnocení klasifikace částic pro vzorek 1.



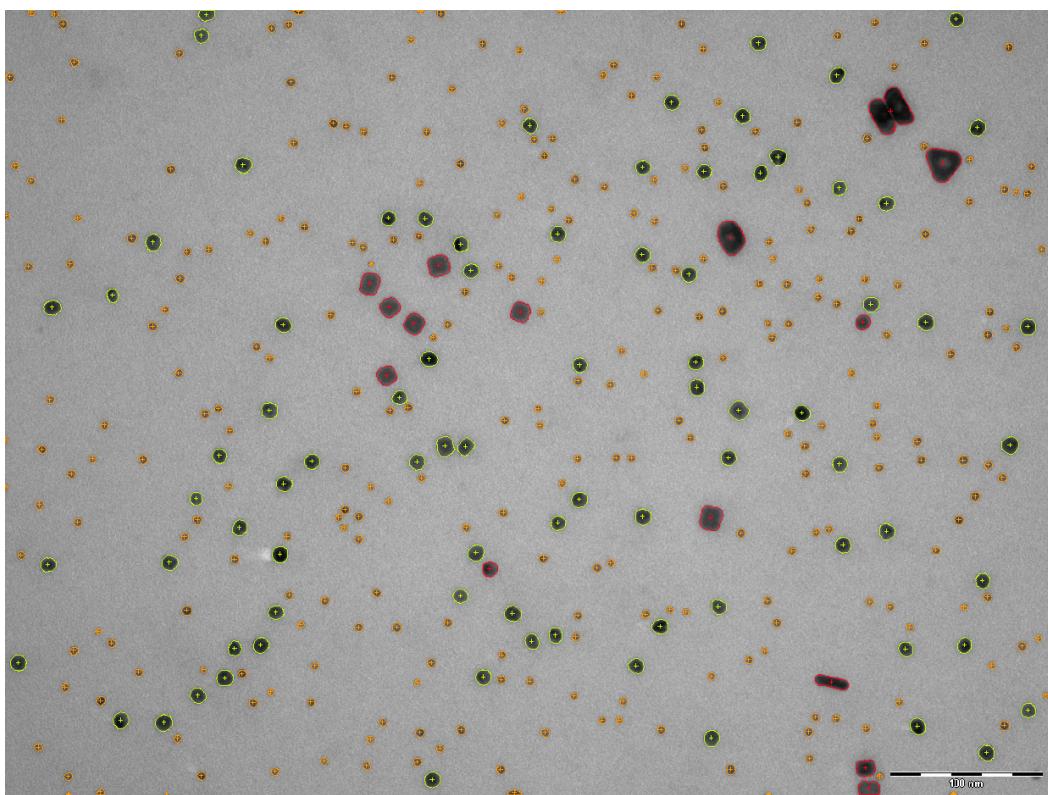
Obrázek 3.2: Výsledek klasifikace částic pro vzorek 1.

3.3.3 Vzorek 2

V této části se nachází zjištěné hodnoty výsledků klasifikace pro vzorek 2.

Třída	Dobře	Špatně	Celkem	Přesnost
	0	2	2	0,000
	2	0	2	100,000
	267	0	267	100,000
	72	11	83	86,747
Celkem	341	11	352	96,875

Tabulka 3.2: Vyhodnocení klasifikace částic pro vzorek 2.



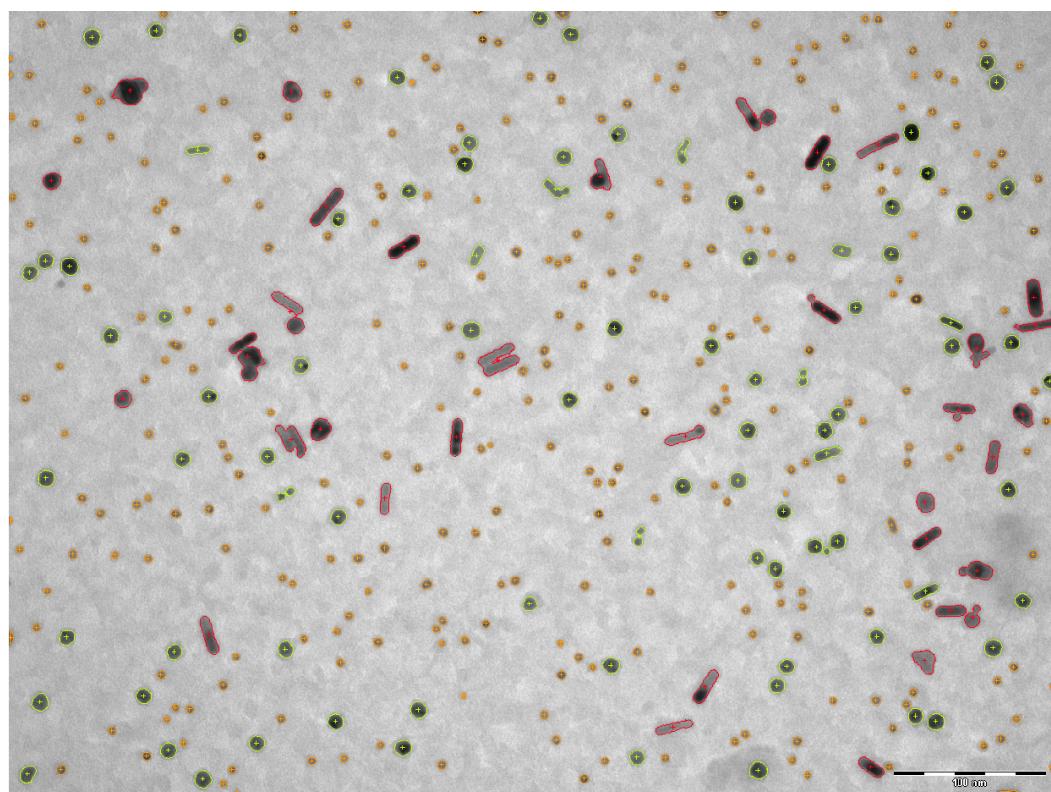
Obrázek 3.3: Výsledek klasifikace částic pro vzorek 2.

3.3.4 Vzorek 3

V této části se nachází zjištěné hodnoty výsledků klasifikace pro vzorek 3.

Třída	Dobře	Špatně	Celkem	Přesnost
-1	0	5	2	0,000
0	20	2	22	90,909
1	296	0	296	100,000
2	86	10	96	89,583
Celkem	402	12	414	97,101

Tabulka 3.3: Vyhodnocení klasifikace částic pro vzorek 3.



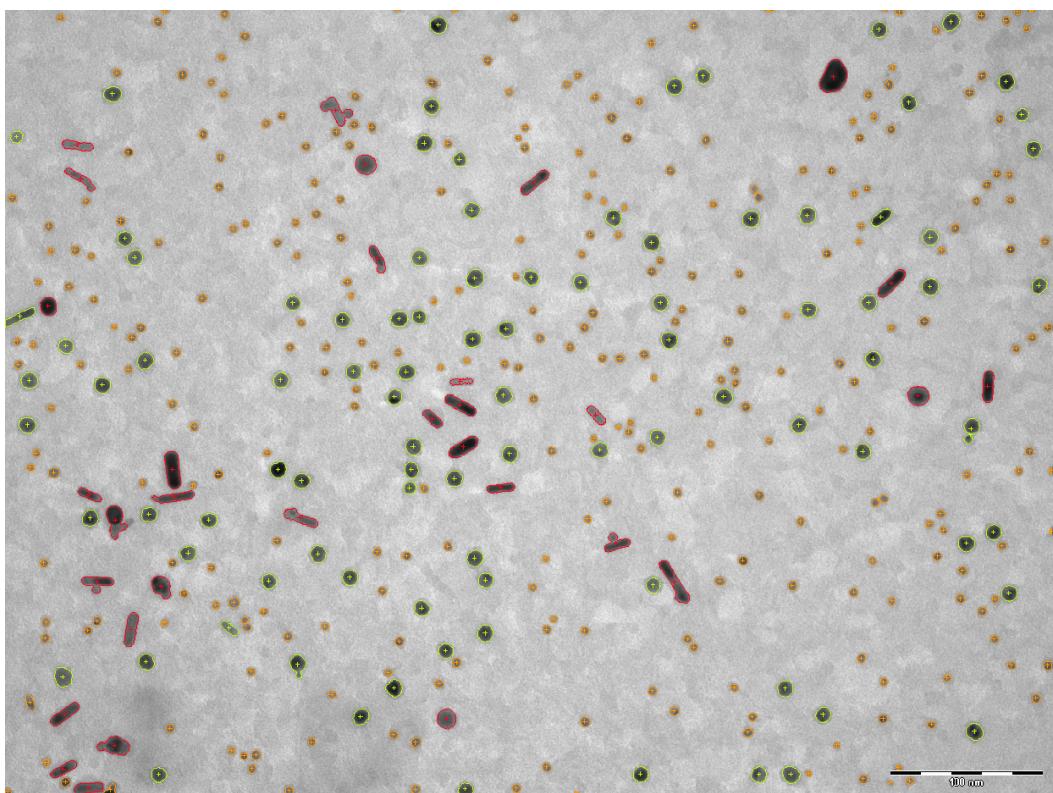
Obrázek 3.4: Výsledek klasifikace částic pro vzorek 3.

3.3.5 Vzorek 4

V této části se nachází zjištěné hodnoty výsledků klasifikace pro vzorek 4.

Třída	Dobře	Špatně	Celkem	Přesnost
-1	0	3	2	0,000
0	17	3	20	85,000
1	289	2	291	99,313
2	93	9	102	91,176
Celkem	399	14	413	96,610

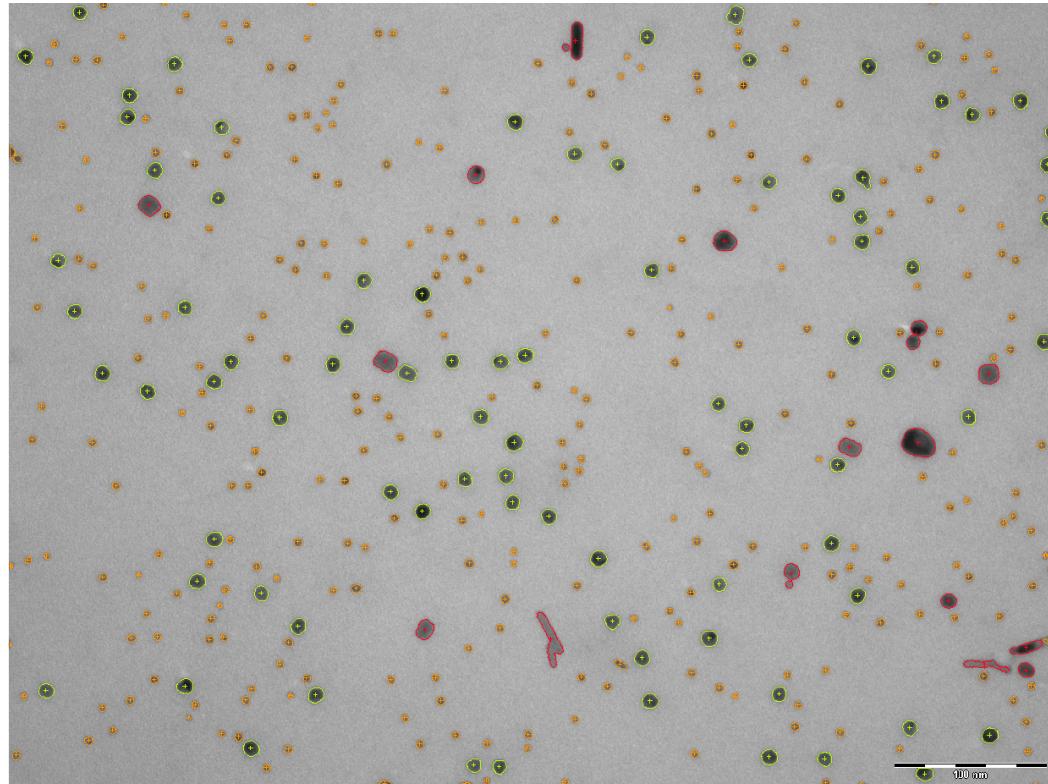
Tabulka 3.4: Vyhodnocení klasifikace částic pro vzorek 4.



Obrázek 3.5: Vyhodnocení klasifikace částic pro vzorek 4.

3.3.6 Vzorek 5

V této části se nachází zjištěné hodnoty výsledků klasifikace pro vzorek 5.



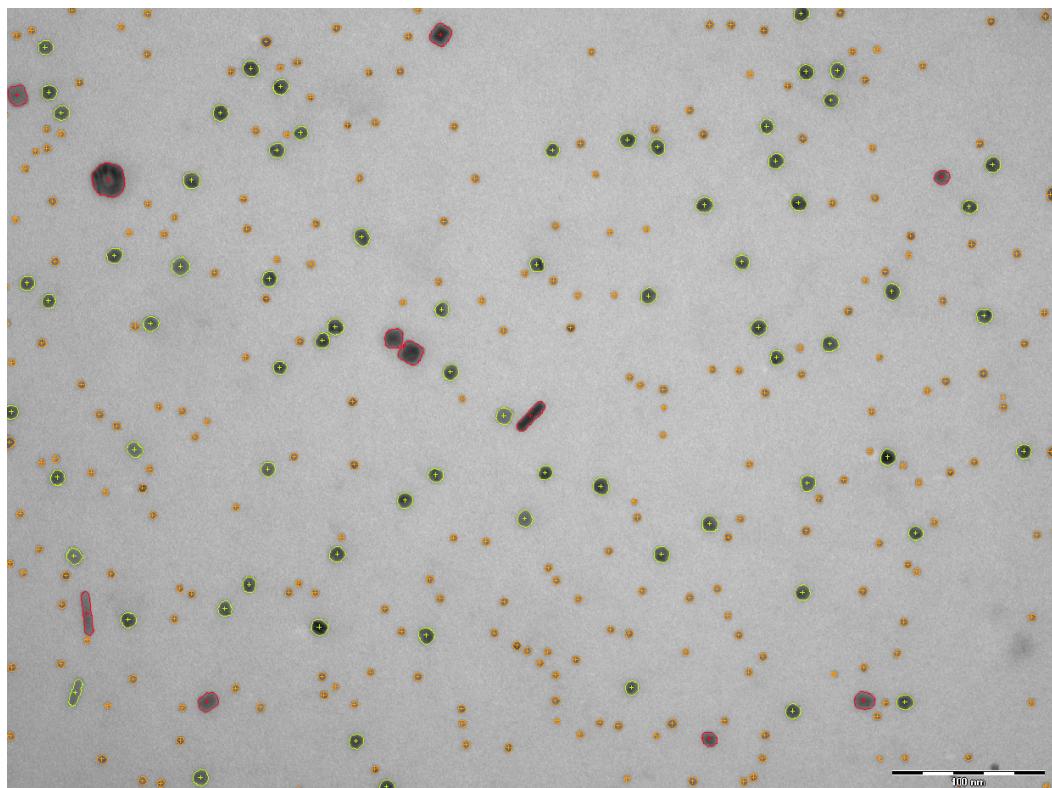
Obrázek 3.6: Výsledek klasifikace částic pro vzorek 5.

Třída	Dobře	Špatně	Celkem	Přesnost
-1	0	3	2	0,000
0	4	0	4	100,000
1	307	0	307	100,000
2	85	9	94	90,426
Celkem	396	9	405	97,778

Tabulka 3.5: Vyhodnocení klasifikace částic pro vzorek 5.

3.3.7 Vzorek 6

V této části se nachází zjištěné hodnoty výsledků klasifikace pro vzorek 6.



Obrázek 3.7: Výsledek klasifikace částic pro vzorek 6.

Třída	Dobře	Špatně	Celkem	Přesnost
-1	0	2	2	0,000
0	2	1	3	66,667
1	254	0	254	100,000
2	74	4	78	94,872
Celkem	330	5	335	98,507

Tabulka 3.6: Vyhodnocení klasifikace částice pro vzorek 6.

3.3.8 Celkový přehled

V této části se nachází shrnutí výsledků pro všechny klasifikované vzorky.

Třída	Dobře	Špatně	Celkem	Přesnost
-1	0	22	14	0,000
0	73	12	85	85,882
1	1983	2	1985	99,899
2	570	64	634	89,905
Celkem	2626	100	2718	96,615

Tabulka 3.7: Vyhodnocení klasifikace pro všechny vzorky.

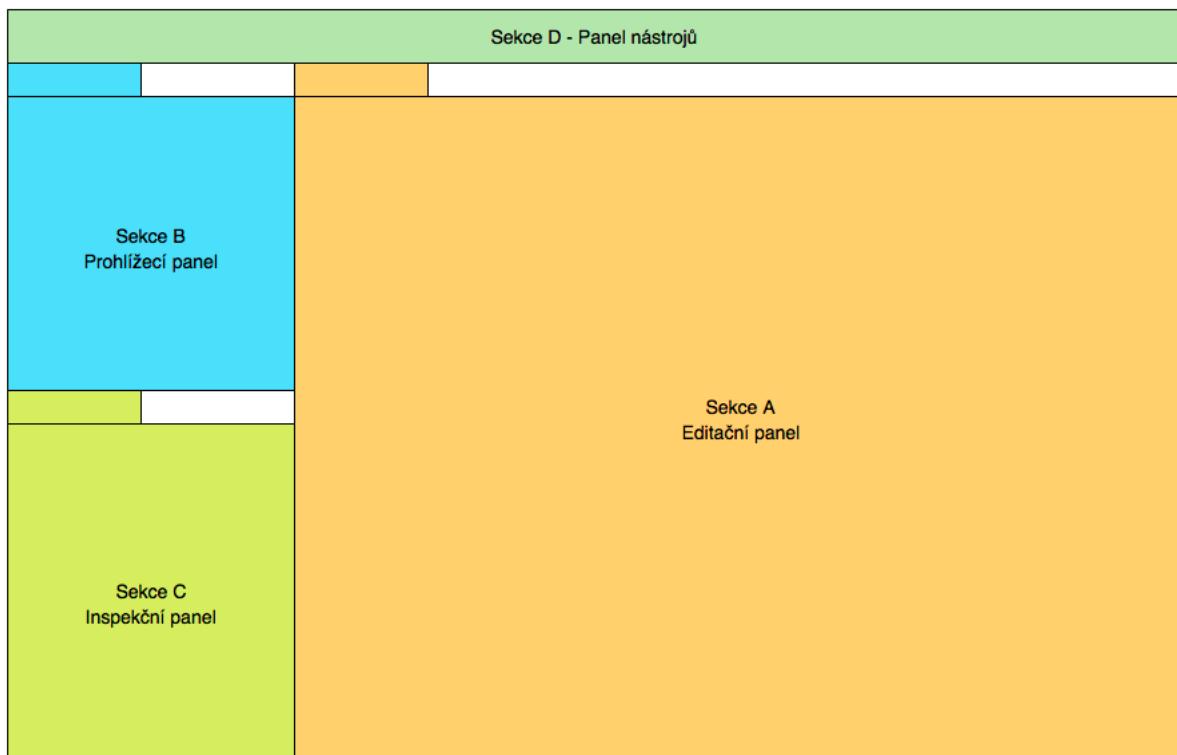
Kapitola 4

Design a implementace

4.1 Způsob práce s daty

Zpracování mikroskopických snímků v laboratoři probíhá typicky po dávkách. Za tímto účelem bylo navrženo rozhraní které umožňuje seskupovat snímky do skupin. Tyto skupiny nazýváme v aplikaci *Projekty*. Projekt je stromová struktura složek obsahující soubory se snímky a uloženými daty. Více v sekci o uchování dat 4.3 Typický způsob práce s aplikací je předpokládán následujícím způsobem:

1. Pořízení skupiny snímků.
2. Vytvoření projektu a načtení skupiny obrázků.
3. Detekce častic ve vybraném obrázku.
4. Možná úprava parametrů detektoru při špatně provedené detekci a opětovná detekce
5. Ruční označení vybraného počtu častic.
6. Přiřazení označených častic klasifikátoru jako trénovací množinu.
7. Klasifikace dosud neoznačených častic natrénovaným klasifikátorem.
8. Oprava nebo trénovací množiny dat.
9. Aplikace detekce a klasifikace na další snímky
10. Analýza snímků s česticemi.



Obrázek 4.1: Přehled rozložení aplikace.

4.2 Uspořádání aplikace

Na základě těchto požadavků bylo navrženo rozložení aplikace viz [obr!] Hlavní části okna aplikace jsou:

Sekce A - Hlavní editační panel

Hlavní editační panel je standardně umístěno uprostřed aplikačního okna a zastává jeho největší část. Panel je určen k zobrazování hlavního obsahu aplikace jako jsou mikroskopické obrázky a analýzy. Jednotlivé položky je možné zobrazit v záložkách.

Sekce B - Prohlížecí panel

Prohlížecí panel je standardně umístěn v horní levé části aplikačního okna. Komponenty zobrazují jednotlivé projekty a klasifikátory.

Sekce C - Inspekční panel

Inspekční panel je standardně umístěn v dolní levé části hlavního okna. Slouží k zobrazování doplňkových informací o mikroskopickém obrázku nebo o datech v něm umístěných.

Sekce D - Panel nástrojů

Panel nástrojů je umístěn v horní části okna aplikace. Obsahuje kontrolní prvky pro manipulaci s daty v aplikaci.

4.2.1 Editační panel

V editačním panelu je zobrazen buď *editor snímků* nebo *panel analýzy*.

4.2.1.1 Editor snímků

Editor snímků je hlavní a nejdůležitější součást aplikace. Slouží primárně k práci obrazovými daty jako je detekce a klasifikace. Hlavní funkcí panelu je zobrazování obrazových dat a nanočástic.

Ten umožňuje práci ve třech režimech.

- výběr
- značení
- tvoření

První režim *výběr*, je určený k interaktivnímu výběru detekovaných částic. To umožňuje uživateli jak interaktivně zkoumat data tak opravit výsledky detekce.

Druhý režim *značení*, je určený k přiřazování částic do jednotlivých tříd. V tomto režimu uživatel buď poskytuje trénovací data pro klasifikátor, aplikuje klasifikátor na aktuálně zobrazené nanočástice nebo může upravovat výsledek klasifikace.

Třetí režim *tvoření* umožňuje manuální přidávání částic. Tento režim umožňuje umístit novou částici na libovolné místo v obraze.

Nedílnou součástí editace částic je jejich vizualizace. Vizualizace částic realizována přímo pomocí Java 2D Graphics. Jednotlivé částice jsou zobrazeny na obrazovém podkladě. Zobrazení lze přiblížovat, oddalovat a volit různé režimy zobrazení částic.

Částice je možné zobrazení s/bez obrysu a s/bez zobrazeného středu.

4.2.1.2 Panel analýzy

Panel analýzy umožňuje analyzovat data pomocí hierarchického shlukování a výsledek vizualizovat pomocí dendrogramu.

Hierarchické shlukování je sekvence vnořených rozkladů, která na jedné straně začíná triviálním rozkladem, kdy každý objekt dané množiny objektů

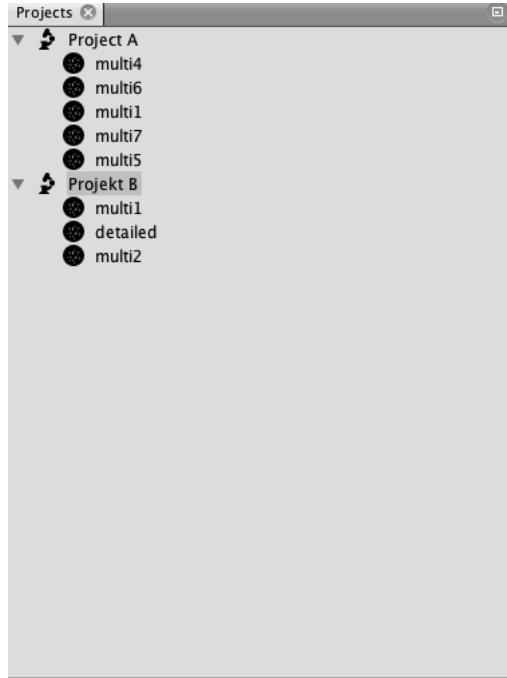
tvoří jednoprvkový shluk, a na druhé straně končí triviálním rozkladem s jedním shlukem obsahujícím všechny objekty. Podle směru postupu při shlukování dělíme metody hierarchického shlukování na aglomerační a divizivní [9].

Dendrogram je binární strom znázorňující hierarchické shlukování. V aplikaci je dendrogram vizualizován tak, že každý uzel tohoto stromu představuje shluk. Vertikální řezy dendrogramem jsou rozklady ze shlukovací sekvence. Horizontální směr v dendrogramu představuje vzdálenost mezi shluky (rozklady).

Dendrogramem ve možné vést řezy danou vzdáleností, které rozdělí dendrogram na dílčí shluky. V tomto rozdelení je každý shluk barevně oddělen. Výsledek vzniklého rozdelení má pak uživatel možnost sledovat v původním snímku v NP, kde každá NP má barvu podle příslušnosti do shluku.

Implementace konstrukce dendrogramu vychází z [3]. Původní řešení je modifikováno pro účely aplikace. Byla přidána funkcionality tvorby řezů v histogramu.

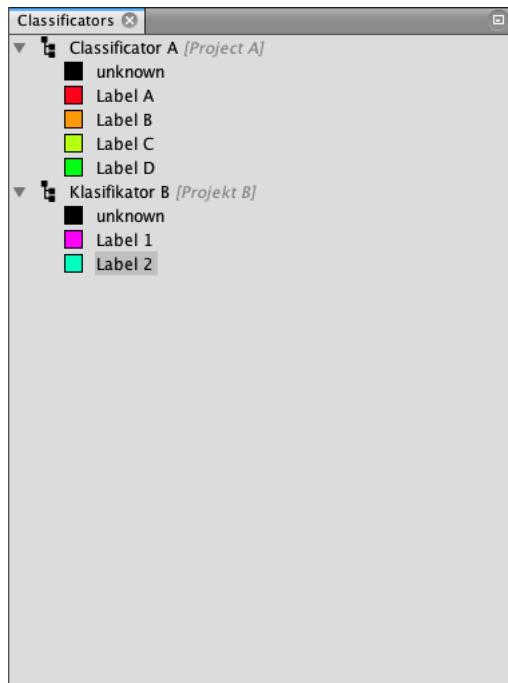
4.2.2 Přehled projektů



Obrázek 4.2: Ukázka panelu přehledem projektů a snímků.

Přehled otevřených projektů je zobrazen na panelu přehledu projektů. Panel poskytuje uživateli přehled o otevřených projektech a obrázcích v přehledné stromové struktuře. Kořen stromu tvoří projekt. Listy projektu pak znázorňují jednotlivé snímky přiřazené k projektu. Panel je také zodpovědný o přiřazování jednotlivých snímků k projektu a jejich ukládání jednotlivých obrázků.

4.2.3 Přehledu klasifikátorů



Obrázek 4.3: Ukázka panelu s přehledem klasifikátorů.

Panel klasifikátoru je standardně součástí prohlížecího panelu. Jeho funkcí je zobrazovat vytvořené klasifikátory, vytvářet klasifikátory nové a poskytovat uživateli možnost s klasifikátorem manipulovat.

Klasifikátoru lze vytvořit jednotlivé třídy pro zařazení částic. Každou třídu je možné pojmenovat a přiřadit k ní barvu. Přiřazená barva klasifikátoru ovlivňuje jakou barvou jsou částice, přiřazené do dané třídy, zobrazovány v editoru obrázků.

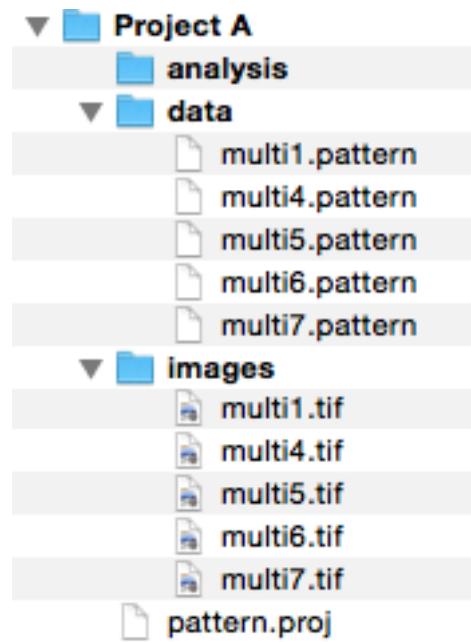
Klasifikátor je možné

4.2.4 Inspekční panel

Inspekční panel obsahuje v základní verzi dva panely. Panel s informacemi o snímku a inspekční panel.

Panel s *informacemi o snímku* zobrazuje uživateli obecné informace o snímku jako je jeho rozlišení, absolutní počet detekovaných částic, počet částic v jednotlivých třídách atd.

Inspekční panel zobrazuje informace o částicích které má uživatel vybrané. Takovou informací je např. počet vybraných částic, příslušnost částic do tříd, těžiště atd. Vybrané částice je pak možné exportovat do formátu CSV k další analýze.



Obrázek 4.4: Příklad souborové a složkové struktury projektu.

4.3 Uchování dat

Pro uchovávání dat je v aplikaci navržena struktura složek a souborů nazvaná *projekt*. Struktura projektu se sestavá ze tří složek, souboru *pattern.proj* a souborů s daty viz obr. 4.4.

Aplikace umožňuje uchovávat:

- pro každý snímek informace o detekovaných a klasifikovaných hodnotách NP.
- informace o klasifikátoru projektu

4.3.0.1 Ukládání dat snímků

Data pro jednotlivé snímky jsou uchovávány ve složce *data* v souborech s příponou *.pattern*. Název souboru s daty pak odpovídá názvu souboru snímku ke kterému data patří. Pro ukládání dat je využito formátu JSON [1]. Příklad struktury ukládaného dokumentu je následující:

```

1 {
2     "name": "multi2",
3     "image": "multi2.tif",
4     "particles": [
5         {

```

```

6         "id": 0,
7         "cog": {
8             "x": 1003.0,
9             "y": 1029.5
10        },
11        "contour": [
12            {
13                "x": 1001.0,
14                "y": 1029.0
15            },
16            {
17                "x": 1001.0,
18                "y": 1030.0
19            },
20            {
21                "x": 1005.0,
22                "y": 1030.0
23            },
24            {
25                "x": 1005.0,
26                "y": 1029.0
27            }
28        ]
29    }

```

4.3.0.2 Ukládání dat klasifikátoru

Data pro klasifikátor se ukládají přímo do souboru *pattern.proj*, který se nachází v kořenovém adresáři projektu. Příklad struktury ukládaných dat ve formátu JSON

```

1 {
2     "info": {
3         "id": "Classifier A",
4         "path": "path/to/pattern.proj"
5     },
6     "labels": [
7         {
8             "id": -1,
9             "name": "unknown",
10            "color": {
11                "value": -16777216,

```

```
12     "falpha": 0.0
13   }
14 }
15 ...
16 ]
17 "selectedLabel": 2,
18 "examples": [
19   {
20     "id": 47,
21     "cog": {
22       "x": 841.6866566716642,
23       "y": 918.4822588705647
24     },
25     "contour": [ ... ]
26   }
27 ]
28 }
```

4.4 Implementace

K implementaci je využito netbeans

Kapitola 5

Závěr

5.1 Diskuze

Zhodnocení výsledků detekce. Pořád to není úplně ono navržené metody stále selhávají např. na obr. Důvodem je malý nejčastěji kontrast částic s okolím. Nebo jejich těsná blízkost

Zhodnocení výsledků klasifikace. Byl ozkoušeno testování pomocí konvexity ale nepodařilo se dotáhnout do konce.

5.2 Shrnutí práce

Čeho jsem dosáhl. Detekce je taková a taková. Klasifikátor dobře klasifikuje malé částice. Kterých je v obrázcích z TEM většinové množství.

Poskytl jsem metody pro analýzu dat pomocí hierarchického shlukování.

5.3 Další směřování

Aplikace nyní míří do testovacího provozu na UMG, kde bude vyhodnocen její přínos. Další možný vývoj na aplikaci se bude skládat z vylepšování uživatelského rozhraní a doplnění o další funkctionalitu, která vyplyne z testovacího provozu.

Detekční a klasifikační algoritmus bude testován v reálném provozu, kde se ověří jeho funkčnost na větším počtu snímku. V případě nefunkčnosti některé částic algoritmu na budoucích datech budou selhání vyhodnocena a algoritmus upraven.

Literatura

- [1] Json. Dostupné z: <<http://www.json.org/>>. [software].
- [2] ARSHBROOK, A. – THACKER, N. Tutorial: Algorithms For 2-Dimensional Object Recognition. 1998. Dostupné z: <<http://www.tina-vision.net/docs/memos/1996-003.pdf>>.
- [3] BEHENKE, L. *Hierarchical clustering Java* [online]. Dostupné z: <<https://github.com/lbehnke/hierarchical-clustering-java>>.
- [4] CHEN, L. – HO, C. Development of Nanoparticle Shape Measurement and Analysis for Process Characterization of TiO₂ Nanoparticle Synthesis. *Advanced Material Science*. 2008, 18, 8, s. 677–684.
- [5] D'AMICO, F. – SKARMOUTSOU, E. Clustering and colocalization in transmission immunoelectron microscopy: A brief review. *Micron*. 2008, 39, 1, s. 1 – 6. ISSN 0968-4328. doi: <http://dx.doi.org/10.1016/j.micron.2007.09.005>. Dostupné z: <<http://www.sciencedirect.com/science/article/pii/S0968432807001400>>.
- [6] FLUSSER, J. On the independence of rotation moment invariants. *Pattern Recognition (PR)*. 2000, 33, 9, s. 1405–1410. Dostupné z: <<http://library.utia.cas.cz/prace/20000033.pdf>>.
- [7] GONZALEZ, R. C. – WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA : Prentice-Hall, Inc., 2006. ISBN 013168728X.
- [8] HORIKOSHI, S. – SERPONE, N. *Microwaves in nanoparticle synthesis: fundamentals and applications*. John Wiley & Sons, 2013.
- [9] JAN, K. – ŠILHÁN, D. *Shluková analýza* [online]. Dostupné z: <<http://www.fd.cvut.cz/personal/nagyivan/Projekty/Classification/ShlukovaAnalyza.pdf>>.
- [10] K., A. M. . A. *Image segmentation with wattershed algorithm* [online]. 2014. [cit. 8. 5. 2015]. Dostupné z: <<http://opencv-python-tutroals>>.

readthedocs.org/en/latest/py_tutorials/py_imgproc/py_watershed/py_watershed.html.

- [11] MALPICA, N. et al. Applying Watershed Algorithms to the Segmentation of Clustered Nuclei. *Cytometry Part A*. 1997, 28, 4, s. 289–297.
- [12] MAYHEW, T. M. et al. A review of recent methods for efficiently quantifying immunogold and other nanoparticles using {TEM} sections through cells, tissues and organs. *Annals of Anatomy - Anatomischer Anzeiger*. 2009, 191, 2, s. 153 – 170. ISSN 0940-9602. doi: <http://dx.doi.org/10.1016/j.aanat.2008.11.001>. Dostupné z: <<http://www.sciencedirect.com/science/article/pii/S094096020800157X>>.
- [13] OSTROWSKI, A. et al. Overview about the localization of nanoparticles in tissue and cellular context by different imaging techniques. *Beilstein Journal of Nanotechnology*. 2015, 6, s. 263–280. ISSN 2190-4286. doi: 10.3762/bjnano.6.25.
- [14] R. FISHER, A. W. S. P. – WOLFART, E. *Adaptive thresholding* [online]. 2003. [cit. 8. 5. 2015]. Dostupné z: <<http://homepages.inf.ed.ac.uk/rbf/HIPR2/adptrsh.htm>>.
- [15] SAHOO, P. K. et al. A Survey of Thresholding Techniques. *Comput. Vision Graph. Image Process.* February 1988, 41, 2, s. 233–260. ISSN 0734-189X. doi: 10.1016/0734-189X(88)90022-9. Dostupné z: <[http://dx.doi.org/10.1016/0734-189X\(88\)90022-9](http://dx.doi.org/10.1016/0734-189X(88)90022-9)>.
- [16] SUZUKI, S. – ABE, K. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*. 1985, 30, 1, s. 32–46.
- [17] WÄHLBY, C. et al. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *Journal of Microscopy*. 2004, 215, 1, s. 67–76.
- [18] WORTIS, M. Chemistry and Physics of Solid Surfaces. 1988.

Příloha A

Seznam použitých zkratek

NP - Nanočástice, z aglického *nanoparticles*

TEM - Transmisní elektronová mikroskopie

SEM - Rastrovací elektronový mikroskop

UMG - Ústav molekulární genetiky, Ústav Akademie Věd České Republiky

Příloha B

Obsah přiloženého CD

V této kapitole naleznete obsah CD přiloženého k této diplomové práci.

- readme.txt - popis obsahu CD s dalšími informacemi
- pattern/ - zdrojové kódy aplikace v jazyce Java
- text/
 - src/ - zdrojový text v LaTeXu, včetně obrázků a šablon
 - palasjur_2015dip.pdf - text diplomové práce ve formátu PDF