

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA

## BAKALÁŘSKÁ PRÁCE

Metody segmentace obrazu



**Katedra matematické analýzy a aplikací matematiky**

Vedoucí bakalářské práce: **RNDr. Tomáš Fürst, Ph.D.**

Vypracoval: **Adam Pinkas**

Studijní program: B1101 – Matematika

Studijní obor Matematika a její aplikace

Forma studia: prezenční

Rok odevzdání: 2016

## BIBLIOGRAFICKÁ IDENTIFIKACE

**Autor:** Adam Pinkas

**Název práce:** Metody segmentace obrazu

**Typ práce:** Bakalářská práce

**Pracoviště:** Katedra matematické analýzy a aplikací matematiky

**Vedoucí práce:** RNDr. Tomáš Fürst, Ph.D.

**Rok obhajoby práce:** 2016

**Abstrakt:** Cílem bakalářské práce bude prozkoumat několik metod segmentace obrazu a tyto metody využít na návrh a implementaci programového kódu, který automaticky vyhledá a rozpozná jednotlivá obarvená buněčná jádra. Téma je součástí rozsáhlejšího výzkumu integrity genomu, který probíhá ve spolupráci s Ústavem molekulární a translační medicíny Lékařské fakulty Univerzity Palackého v Olomouci.

**Klíčová slova:** Zpracování obrazu, segmentace obrazu, fitování elips

**Počet stran:** 36

**Počet příloh:** 1

**Jazyk:** český

## BIBLIOGRAPHICAL IDENTIFICATION

**Author:** Adam Pinkas

**Title:** Image segmentation methods

**Type of thesis:** Bachelor's

**Department:** Department of Mathematical Analysis and Application of Mathematics

**Supervisor:** RNDr. Tomáš Fürst, Ph.D.

**The year of presentation:** 2016

**Abstract:** The aim of the Thesis is to master several image segmentation methods. These methods will be implemented in scope of a routine which automatically identifies cellular nuclei on images from high-throughput microscopy. The topic is a part of a broader research of genome integrity in cooperation with the Institute for Molecular and Translational Medicine of the Palacky University in Olomouc.

**Key words:** Image processing, image segmentation, ellipse fitting

**Number of pages:** 36

**Number of appendices:** 1

**Language:** Czech

### **Prohlášení**

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně pod vedením pana Tomáše Fürsta a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne .....

.....

podpis

# Obsah

Úvod	7
<b>1 Prokládání dat elipsou</b>	<b>9</b>
1.1 Základní vlastnosti elips . . . . .	9
1.1.1 Analytické vyjádření elipsy . . . . .	10
1.1.2 Vzdálenost bodu od elipsy . . . . .	12
1.2 Řešení minimalizační úlohy s kvadratickou podmínkou pomocí zobecněných vlastních čísel . . . . .	14
1.3 Fitzgibbonova metoda fitování elips . . . . .	18
<b>2 Implementace programu pro rozdělování dvojic buněčných jader</b>	<b>23</b>
2.1 Popis algoritmu pro dělení dvojic jader . . . . .	24
2.2 Implementace algoritmu . . . . .	27
2.2.1 Funkce SPLIT . . . . .	27
2.2.2 Funkce DRAW_LINE . . . . .	28
2.2.3 Funkce GET_BOUNDARY_POINTS . . . . .	28
2.2.4 Funkce GET_MAX_DIST . . . . .	29
<b>3 Vlastnosti algoritmu pro rozdělování dvojic buněčných jader</b>	<b>30</b>
3.1 Srovnání s Watershed segmentací . . . . .	32
<b>4 Závěr</b>	<b>33</b>
<b>Literatura</b>	<b>34</b>
<b>Přílohy</b>	<b>36</b>

## **Poděkování**

Rád bych poděkoval Tomáši Fürstovi za skvělé nápady a trpělivost, které vedly k dokončení této bakalářské práce.

# Úvod

Metody segmentace digitálního obrazu se používají k rozdělení obrazu na segmenty, jejichž pixely reprezentují objekty zájmu. Díky tomu jsme schopni najít a analyzovat prvky, které nás opravdu zajímají. Příkladem může být snímání očí při zkoumání reakce člověka na reklamu. Uplatnění těchto metod je opravdu rozsáhlé [7].

V následujících kapitolách se budu zabývat aproximací bodů v rovině elipsou a jejím použitím při sestavení metody segmentace obrazu, která bude rozdělovat dotýkající se dvojice buněčných jader na snímcích z laserového mikroskopu. Tyto snímky, jež poskytla laboratoř genomové integrity Ústavu molekulární a translační medicíny Lékařské fakulty Univerzity Palackého v Olomouci [8], jsou součástí výzkumu integrity genomu.

Při výzkumu integrity genomu se buňky obarví pomocí DAPI [9], což je fluorescenční barvivo, které dobře proniká buněčnou membránou a následně obarví buněčné jádro. Díky tomu můžeme snadno rozlišit buněčná jádra od pozadí. Mikroskop snímá jeden vzorek dvakrát a pokaždé se zaměřuje na jinou vlnovou délku. První snímek ukazuje obarvená buněčná jádra. Druhý snímek obsahuje oblasti v jádrech, kde byla narušena DNA. Přilehlá buněčná jádra tedy potřebujeme rozdělit, protože chceme analyzovat počet míst s porušenou DNA a kvalifikovat počet těchto zlomů vzhledem k jednomu jádru.

Buněčná jádra mají přibližně tvar elipsy. Proto se jejich dělení pomocí vhodně vložených elips jeví jako logické.

V první kapitole se věnuji základním vlastnostem elips a teorii řešení minimalizační úlohy pomocí zobecněných vlastních čísel. Dále popíši Fitzgibbonovu

metodu prokládání dat elipsou [1].

Druhá kapitola obsahuje popis implementace segmentační metody používající Fitzgibbonovu metodu pro rozdělování dvojic buněčných jader. Tato implementace je provedená v prostředí MATLAB a je přiložena k této bakalářské práci.

V poslední kapitole analyzují použití nové segmentační metody na snímcích z výzkumu integrity genomu.

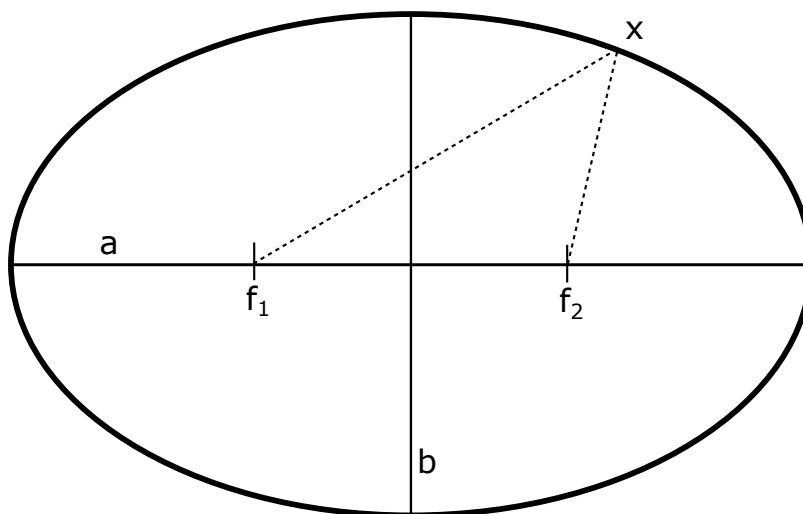


# Kapitola 1

## Prokládání dat elipsou

### 1.1. Základní vlastnosti elips

Elipsy jsou plošné křivky patřící mezi kuželosečky. Elipsu můžeme definovat jako uzavřenou křivku, uvnitř které jsou dva body  $\mathbf{f}_1$  a  $\mathbf{f}_2$  takové, že pro libovolný bod  $\mathbf{x}$  ležící na elipse je součet vzdálenosti  $\|\mathbf{x} - \mathbf{f}_1\| + \|\mathbf{x} - \mathbf{f}_2\|$  konstantní. Jestliže tyto dva body splývají v jeden, hovoříme o kružnici.



Obrázek 1.1: Elipsa

- Body  $\mathbf{f}_1$  a  $\mathbf{f}_2$  se nazývají ohniska.
- Bod, ležící v polovině úsečky spojující ohniska, se nazývá střed elipsy.

- Nejdelší úsečku spojující střed elipsy a bod na elipse nazýváme hlavní poloosa. Její délku značíme  $a$ .
- Nejkratší úsečku spojující střed elipsy a bod na elipse nazýváme vedlejší poloosa. Její délku značíme  $b$ .
- Vzdálenost ohniska a středu  $e = \sqrt{a^2 - b^2}$  se nazývá excentricita.
- Plochu elipsy můžeme vypočítat pomocí vztahu  $S = \pi ab$ .

Tyto informace jsem čerpal z [5].

V následujícím textu budeme používat výše uvedené značení hlavních poloos.

### 1.1.1. Analytické vyjádření elipsy

**Definice 1.1** *Nechť  $A, B, C, D, E$  a  $F$  jsou reálné konstanty a  $\mathbf{x} = (x, y)$  je bod v rovině. Potom rovnice*

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \quad (1.1)$$

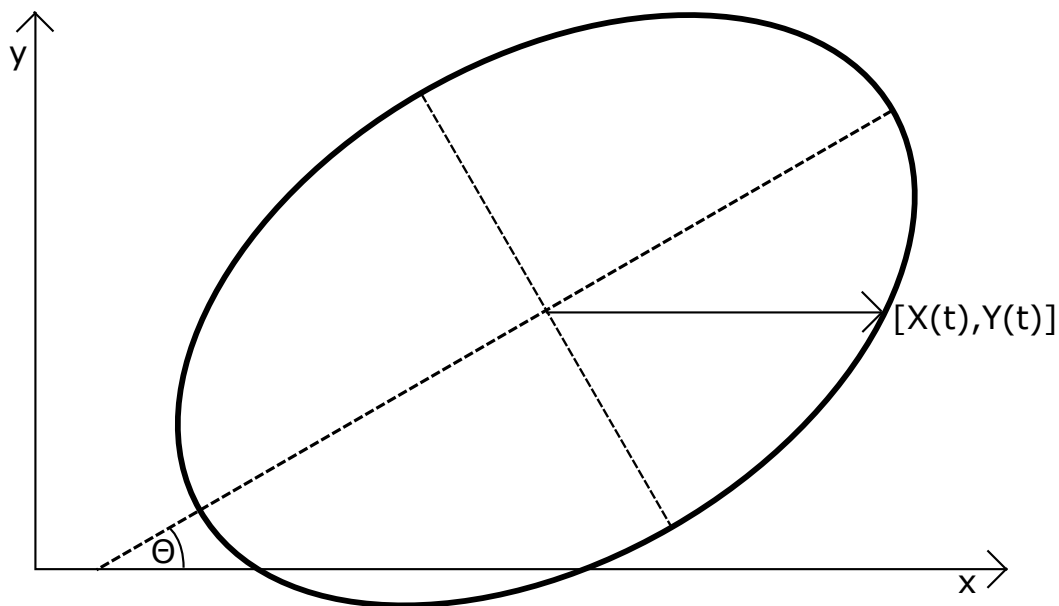
*se nazývá obecná rovnice kuželoseček.*

*Její diskriminant značíme  $\Delta = B^2 - 4AC$ .*

Jestliže diskriminant  $\Delta$  je záporný, rovnice (1.1) reprezentuje elipsu. Důkaz [10] spočívá v tom, že  $\Delta$  je invariantní vůči rotaci. Jestliže má elipsa osy rovnoběžné s osami souřadnicové soustavy, pak  $B$  je rovno nule. Pro tuto elipsu je tedy  $\Delta = -4AC$ . Z obecné rovnice dále plyne, že pro každou elipsu s  $B = 0$  platí  $AC > 0$ , a proto pro každou elipsu je  $\Delta$  záporný.

**Definice 1.2** *Nechť  $t$  je parametr od 0 do  $2\pi$ ,  $x_c$  a  $y_c$  jsou souřadnice středu elipsy a  $\Theta$  je úhel, který svírá hlavní poloosa s osou  $x$ . Pak parametrická rovnice elipsy je ve tvaru*

$$\begin{aligned} x(t) &= x_c + a \cos t \cos \Theta - b \sin t \sin \Theta \\ y(t) &= y_c + a \cos t \sin \Theta + b \sin t \cos \Theta, \end{aligned} \quad (1.2)$$



Obrázek 1.2: Znázornění parametrické rovnice

Parametrická rovnice se používá při vykreslování elipsy. Můžeme ji také použít při hledání přibližné vzdálenosti bodu od elipsy, kdy si do paměti uložíme body ležící na elipse pomocí parametrické rovnice a následně hledáme bod s minimální vzdáleností od pevně zvoleného bodu.

Koeficienty obecné rovnice můžeme snadno získat dosazením délek poloos, středu elipsy  $(x_c, y_c)$  a úhlu rotace  $\Theta$  do následujícího vztahu:

$$\begin{aligned}
 A &= a^2 \sin^2 \Theta + b^2 \cos^2 \Theta \\
 B &= 2(b^2 - a^2) \sin \Theta \cos \Theta \\
 C &= a^2 \cos^2 \Theta + b^2 \sin^2 \Theta \\
 D &= -2Ax_c - By_c \\
 E &= -Bx_c - 2Cy_c \\
 F &= Ax_c^2 - Bx_c y_c + Cy_c^2 - a^2 b^2
 \end{aligned}
 \tag{1.3}$$

Jestliže hlavní poloosy jsou rovnoběžné s osami kartézské souřadnicové soustavy, hovoříme o takzvané elipse v kanonickém tvaru. Pro elipsy v kanonickém tvaru lze odvodit zjednodušené verze vztahů v této sekci (viz např. [5]). My je však

v dalším textu nebudeme potřebovat.

### 1.1.2. Vzdálenost bodu od elipsy

**Definice 1.3** *Nechť  $\mathbf{x} = (x, y)$  je libovolný bod v rovině a vektor  $\mathbf{a} = (A, B, C, D, E, F)$ , kde  $A, B, C, D, E, F$  jsou konstanty použité v obecné rovnici (1.1), reprezentuje elipsu. Potom reálnou funkci*

$$F(\mathbf{x}, \mathbf{a}) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F \quad (1.4)$$

*nazýváme algebraická vzdálenost bodu  $\mathbf{x}$  od elipsy  $\mathbf{a}$ .*

**Definice 1.4** *Nechť  $\mathbf{x} = (x, y)$  je libovolný bod v rovině a  $\mathbf{C}$  je množina bodů ležících na elipse  $\mathbf{a}$ . Potom geometrickou vzdálenost bodu  $\mathbf{x}$  od elipsy  $\mathbf{a}$  definujeme jako*

$$Dist(\mathbf{x}, \mathbf{a}) = \min_{\mathbf{c} \in \mathbf{C}} \|\mathbf{x} - \mathbf{c}\|, \quad (1.5)$$

*kde  $\|\cdot\|$  je euklidovská norma.*

Obecně pro elipsy neexistuje explicitní vzorec pro výpočet geometrické vzdálenosti bodu od elipsy. V praxi se pro odhad této vzdálenosti používají iterační metody. Zajímavou metodu výpočtu aproximace vzdálenosti bodu od elipsy lze najít v [4]. Tato metoda funguje na principu transformace elipsy na kružnici, kdy najdeme nejkratší vektor spojující kružnici a transformovaný bod. Tento vektor pak transformujeme do původní soustavy a vypočteme zde jeho délku.

Geometrická vzdálenost je výpočetně náročná, proto její použití v algoritmech pro fitování elips není moc vhodné. Místo toho je výhodné v těchto algoritmech používat algebraickou vzdálenost, která je výpočetně nenáročná.

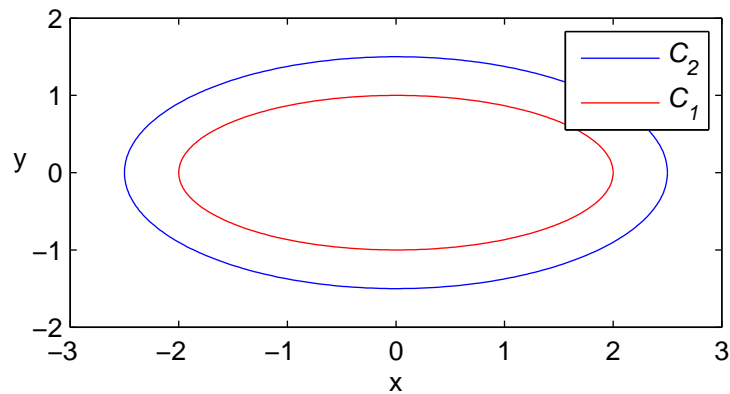
Naproti tomu není výhodné používat algebraickou vzdálenost pro analýzu vhodnosti fitu elipsy, jelikož hodnoty této vzdálenosti nemají vlastnosti geometrické vzdálenosti. Proto je mnohem lepší analyzovat vhodnost fitu pomocí přibližně určených geometrických vzdáleností.

Obě vzdálenosti jsou invariantní vůči rotaci a posunutí.

**Příklad 1.1** Vezměme dvě elipsy  $C_1$  a  $C_2$  takové, že jejich hlavní a vedlejší poloosy leží na osách  $x$  a  $y$ . Délky hlavní a vedlejší poloosy elipsy  $C_1$  jsou 2 a 1. Délky hlavní a vedlejší poloosy elipsy  $C_2$  jsou 2,5 a 1,5. Vypočítáme souřadnice bodů  $\mathbf{x}_i$ ,  $i=1, \dots, 360$ , ležících na elipse  $C_2$  pomocí její parametrické rovnice

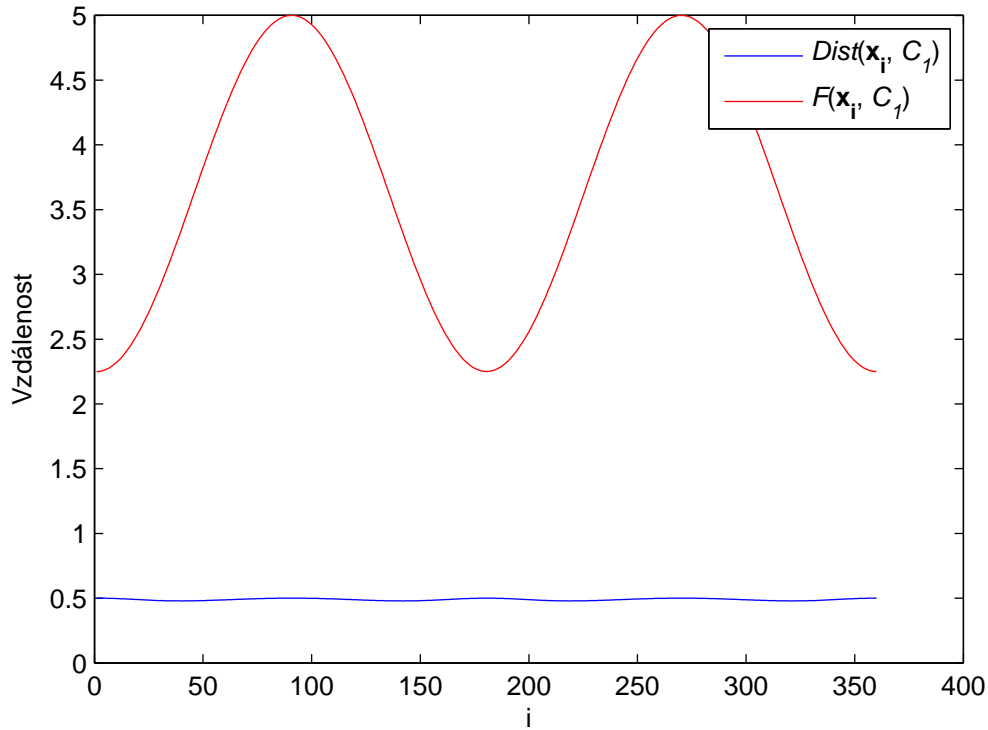
$$X(t) = a \cos t$$

$$Y(t) = b \sin t,$$



Obrázek 1.3: Elipsy

Následně spočítáme algebraické vzdálenosti  $F(\mathbf{x}_i, C_1)$  a geometrické vzdálenosti  $Dist(\mathbf{x}, C_1)$  těchto bodů od elipsy  $C_1$ . Tyto vzdálenosti zobrazíme v grafu, kde na horizontální ose jsou indexy  $i$  a na vertikální ose jsou hodnoty  $F(\mathbf{x}_i, C_1)$  a  $Dist(\mathbf{x}, C_1)$ .



Obrázek 1.4: Výsledný graf

## 1.2. Řešení minimalizační úlohy s kvadratickou podmínkou pomocí zobecněných vlastních čísel

V této kapitole se budeme věnovat základním pojmům a tvrzením, která budeme potřebovat k pochopení principu Fitzgibbonovy metody fitování elips[1].

Nechť  $\mathbf{A}$  je matice typu  $n \times m$ ,  $n \geq m \geq 2$ ,  $\mathbf{C}$  je symetrická matice typu  $m \times m$  a  $d \in \mathbb{R}$  je kladné číslo. Označme  $F(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2$ . Hledáme  $\mathbf{x}$  takové, že

$$\mathbf{x} \text{ minimalizuje } F(\mathbf{x}), \text{ za podmínky } \mathbf{x}^T \mathbf{C} \mathbf{x} = d, \quad (1.6)$$

Dále budeme předpokládat, že matice  $\mathbf{S} = \mathbf{A}^T \mathbf{A}$  je regulární.

Následující definice a věty v této kapitole jsem čerpal z [2], kde jsou uvedené i všechny důkazy.

**Věta 1.1** Jestliže  $\mathbf{S} \in \mathbb{R}^{m \times m}$  je regulární, pak existuje globální řešení úlohy (1.6).

**Definice 1.5** Necht'  $\mathbf{S}, \mathbf{C} \in \mathbb{R}^{m \times m}$ ,  $\mathbf{x} \in \mathbb{C}^m$  a  $\lambda \in \mathbb{C}$ . Potom dvojici  $(\lambda, \mathbf{x})$  nazýváme zobecněné vlastní číslo a jemu příslušející zobecněný vlastní vektor  $\mathbf{x}$ , jestliže platí

$$\mathbf{S}\mathbf{x} = \lambda\mathbf{C}\mathbf{x} \quad (1.7)$$

a tuto rovnici nazýváme problém zobecněných vlastních čísel.

Zobecněná vlastní čísla se počítají analogicky jako vlastní čísla pomocí rovnice

$$\det(\mathbf{S} - \lambda\mathbf{C}) = 0$$

V Matlabu se pro výpočet zobecněných vlastních čísel používá funkce *eig* s parametry  $\mathbf{S}$  a  $\mathbf{C}$ .

**Věta 1.2** Jestliže  $\mathbf{x}_0$  je extrémem  $F(\mathbf{x})$  vzhledem k podmínce  $\mathbf{x}^T\mathbf{C}\mathbf{x} = d$ , pak existuje kladné zobecněné vlastní číslo  $\lambda$ , které spolu se zobecněným vlastním vektorem  $\mathbf{x}_0$  řeší problém zobecněných vlastních čísel (1.7). Dále platí, že

$$F(\mathbf{x}_0) = \lambda d \quad (1.8)$$

**Věta 1.3** Jestliže matice  $\mathbf{S}$  je regulární a matice  $\mathbf{C}$  je symetrická, pak všechny zobecněné vlastní hodnoty řešící (1.7) jsou reálné a různé od nuly.

**Definice 1.6** Množina všech vlastních čísel matice  $\mathbf{C}$  se nazývá spektrum a značíme ji  $\sigma(\mathbf{C})$ . Množinu všech zobecněných vlastních čísel řešící problém zobecněných vlastních čísel (1.7) nazveme spektrum problému zobecněných vlastních čísel a značíme ji  $\sigma(\mathbf{S}, \mathbf{C})$ . Množinu všech kladných zobecněných vlastních čísel z  $\sigma(\mathbf{S}, \mathbf{C})$  značíme  $\sigma^+(\mathbf{S}, \mathbf{C})$ .

**Věta 1.4** Znaménka zobecněných vlastních čísel z  $\sigma(\mathbf{S}, \mathbf{C})$  jsou stejná až na pořadí jako známénka vlastních čísel matice  $\mathbf{C}$ .

**Věta 1.5** Necht'  $\mathbf{S} - \lambda\mathbf{C}$  je pozitivně semidefinitní matice pro  $\lambda \in \sigma^+(\mathbf{S}, \mathbf{C})$ . Potom zobecněný vlastní vektor příslušící  $\lambda$  je lokálním minimem (1.6).

**Věta 1.6** Matice  $\mathbf{S} - \lambda\mathbf{C}$ , kde  $\lambda$  je nejmenší prvek z  $\sigma^+(\mathbf{S}, \mathbf{C})$ , je pozitivně semidefinitní.

**Věta 1.7** Pro nejmenší zobecněné vlastní číslo  $\lambda \in \sigma^+(\mathbf{S}, \mathbf{C})$  existuje zobecněný vlastní vektor  $\mathbf{x}$ , který řeší úlohu (1.6).

**Poznámka 1.1** Jestliže máme nejmenší  $\lambda \in \sigma^+(\mathbf{S}, \mathbf{C})$  a jemu příslušný zobecněný vlastní vektor  $\mathbf{x}_0$ . Pak také  $\lambda$  a  $\alpha\mathbf{x}_0$ , kde  $\alpha \in \mathbb{R}$ , řeší problém (1.7). Dosazením  $\alpha\mathbf{x}_0$  do podmínky  $\mathbf{x}^T\mathbf{C}\mathbf{x} = d$  dostaneme

$$\alpha^2\mathbf{x}_0^T\mathbf{C}\mathbf{x}_0 = d$$

$$\alpha^2 = \frac{d}{\mathbf{x}_0^T\mathbf{C}\mathbf{x}_0}$$

$$\alpha = \pm\sqrt{\frac{d}{\mathbf{x}_0^T\mathbf{C}\mathbf{x}_0}}$$

Jelikož platí, že  $\mathbf{S} = \lambda\mathbf{C}$ , potom  $\alpha$  můžeme získat ve tvaru

$$\alpha = \pm\sqrt{\frac{\lambda d}{\mathbf{x}_0^T\mathbf{S}\mathbf{x}_0}}.$$

Tím pádem získáváme výsledný vektor ve tvaru

$$\mathbf{x} = \pm\sqrt{\frac{\lambda d}{\mathbf{x}_0^T\mathbf{S}\mathbf{x}_0}}\mathbf{x}_0. \quad (1.9)$$

**Příklad 1.2** Minimalizujte  $F(\mathbf{x}) = 4x^2 + 4xy + 2y^2$  za podmínky  $x^2 + 4y^2 = 1$ . Tedy hledáme minimum funkce  $F(\mathbf{x})$  za podmínky, že body  $\mathbf{x}$  leží na elipse se středem v počátku, délkou hlavní poloosy 1 a délkou vedlejší poloosy  $\frac{1}{2}$ .

Úlohu můžeme ekvivalentně zapsat ve tvaru

$$\min F(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2 \text{ za podmínky } \mathbf{x}^T\mathbf{C}\mathbf{x} = 1,$$



kde  $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$  a  $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ . Ihned vidíme, že  $\sigma(\mathbf{C}) = \{1, 4\}$ .

Nejprve vypočteme matici  $\mathbf{S} = \mathbf{A}^T \mathbf{A}$

$$\mathbf{S} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$$

Nyní budeme hledat zobecněná vlastní čísla, která řeší  $(\mathbf{S} - \lambda \mathbf{C}) \mathbf{x} = \mathbf{0}$  tak, že vypočteme rovnici  $\det(\mathbf{S} - \lambda \mathbf{C}) = 0$ .

$$\begin{vmatrix} 4 - \lambda & 2 \\ 2 & 2 - 4\lambda \end{vmatrix} = 0 \text{ pro } \lambda_1 \cong 0,23 \text{ a } \lambda_2 \cong 4,27$$

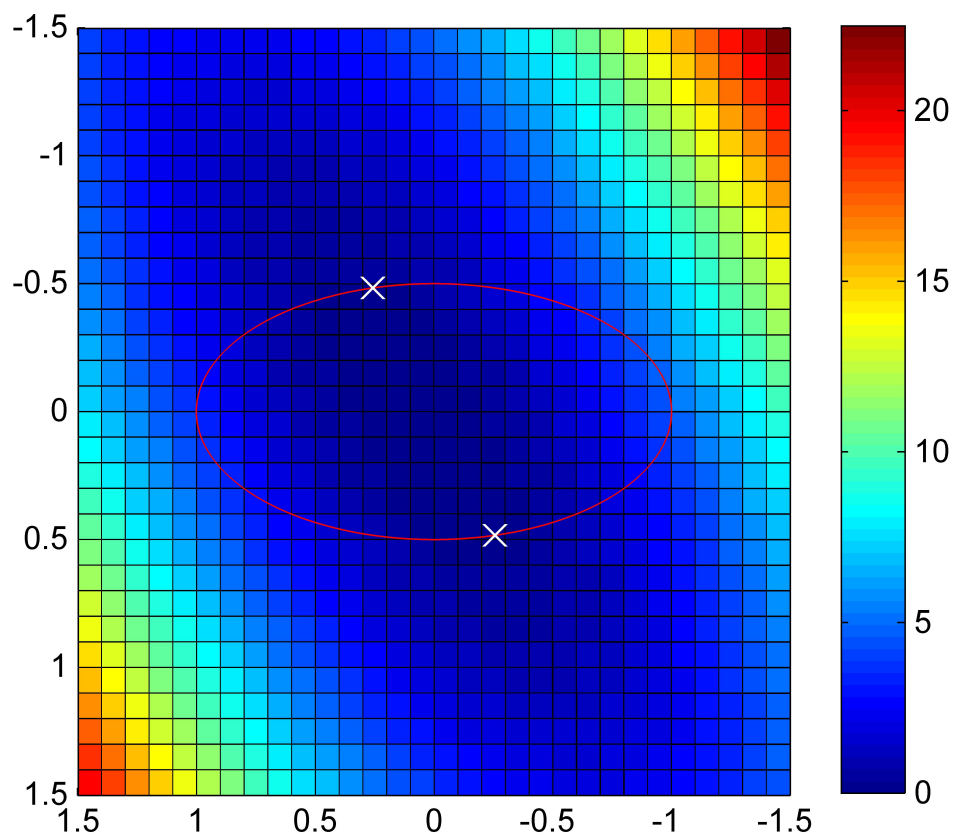
Podle věty 1.4 mají být znaménka zobecněných vlastních čísel  $\lambda_1$  a  $\lambda_2$  stejná (až na pořadí) jako znaménka vlastních čísel matice  $\mathbf{C}$ , což je v našem případě splněno.

Nejmenší kladné zobecněné vlastní číslo je  $\lambda_1$ . Minimem funkce  $F(\mathbf{x})$  za podmínky  $\mathbf{x}^T \mathbf{C} \mathbf{x} = 1$  je tedy podle vět 1.2 a 1.7 hodnota  $\lambda_1$ , protože  $d = 1$ .

Pro určení polohy minima na elipse musíme nejprve určit zobecněný vlastní vektor příslušející zobecněnému vlastnímu číslu  $\lambda_1$  dosazením  $\lambda_1$  do problému zobecněných vlastních čísel (1.7). Řešíme tedy rovnici

$$\left( \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} - \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \right) \mathbf{x} = \mathbf{0}$$

Řešením je například vektor  $\mathbf{x} = (0,26, -0,48)^T$ . Dosazením vektoru  $\mathbf{x}$  do vzorce (1.9). Dostáváme výsledné vektory  $\mathbf{x}_1 = (0,26, -0,48)^T$  a  $\mathbf{x}_2 = (-0,26, 0,48)^T$ .



Obrázek 1.5: Grafické znázornění řešení

### 1.3. Fitzgibbonova metoda fitování elips

Donedávna si matematici mysleli, že prokládat data elipsou lze pouze použitím iteračních metod. Potom však přišel Andrew Fitzgibbon se svou metodou fitování elips [1], která je poměrně jednoduchá a výpočetně nenáročná.

Nechť  $\mathbf{x}_i = (x_i, y_i)$ ,  $i = 1, \dots, n$ , je  $n$  bodů v rovině. Tyto body chceme proložit elipsou  $\mathbf{a} = (A, B, C, D, E, F)$  tak, aby součet čtverců reziduí  $\sum_{i=1}^n F(\mathbf{x}_i, \mathbf{a})^2$ , kde  $F$  je algebraická vzdálenost bodu od elipsy, byl co nejmenší.

**Příklad 1.3** V rovině jsou dány body  $(0,0)$ ,  $(1,1)$ ,  $(2,4)$ ,  $(-1,1)$ ,  $(-2,4)$  a  $(3,9)$ . Pokusíme se do nich vepsat elipsu pomocí obecné rovnice (1.1).

Jednotlivé body dosadíme do této rovnice a dostaneme šest rovnic o šesti neznámých  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$  a  $F$ , které po řádcích zapíšeme do matice

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 8 & 16 & 2 & 4 & 1 \\ 9 & 27 & 81 & 3 & 9 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 4 & -8 & 16 & -2 & 4 & 1 \end{pmatrix}.$$

První sloupec má stejné prvky jako pátý sloupec. Jedním z řešení je tedy vektor  $(1, 0, 0, 0, -1, 0)^T$ , což odpovídá obecné rovnici  $x^2 - y = 0$ , která reprezentuje parabolu. Vidíme, že chceme-li proložit data elipsou, nemůžeme se spoléhat pouze na obecnou rovnici kuželoseček.

Řešení této minimalizační úlohy pouze pomocí metody nejmenších čtverců nám však nezaručí, že výsledný útvar bude elipsou (viz předchozí příklad). Proto je potřeba nějakým způsobem zohlednit podmínku  $B^2 - 4AC < 0$ .

Například F. Bookstein [6] používá podmínku  $A^2 + \frac{B^2}{2} + C^2 = 1$  a minimalizační úlohu řeší iteračně.

A. Fitzgibbon však minimalizační úlohu řeší pomocí podmínky  $4AC - B^2 = 1$ . Snadno se přesvědčíme, že tato podmínka vyhovuje podmínce  $B^2 - 4AC < 0$ . Fitzgibbonovu podmínku můžeme zapsat i maticově

$$\mathbf{a}^T \mathbf{C} \mathbf{a} = 1, \text{ kde } \mathbf{C} = \begin{pmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Minimalizační úlohu převedeme na tvar úlohy (1.6). Hledáme tedy elipsu

$\mathbf{a} = (A, B, C, D, E, F)^T$  takovou, že

$$\mathbf{a} \text{ minimalizuje } \sum_{i=1}^n F(\mathbf{x}_i, \mathbf{a})^2 = \|\mathbf{A}\mathbf{a}\|^2 \text{ za podmínky } \mathbf{a}^T \mathbf{C} \mathbf{a} = 1, \quad (1.10)$$

$$\text{kde } \mathbf{A} = \begin{pmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & & & & & \\ x_n^2 & x_n y_n & y_n^2 & x_n & y_n & 1 \end{pmatrix} \text{ je datová matice.}$$

Matice  $\mathbf{A}$  je sestavena tak, aby výsledkem skalárního součinu vektoru jejího  $i$ -tého řádku a vektoru elipsy  $\mathbf{a}$  byla hodnota algebraické vzdálenosti  $F(\mathbf{x}_i, \mathbf{a})$ .

Snadno můžeme ověřit, že  $\sigma(\mathbf{C}) = \{-2, -1, 2, 0, 0, 0\}$ . Díky tomu víme, že minimalizační úloha (1.10) bude mít právě jedno řešení, protože matice  $\mathbf{C}$  má pouze jedno kladné vlastní číslo. Úloha je tedy dobře definovaná.

Pokračujeme s řešením problému zobecněných vlastních čísel

$$(\mathbf{S} - \lambda \mathbf{C}) \mathbf{a} = \mathbf{0}, \text{ kde } \mathbf{S} = \mathbf{A}^T \mathbf{A} \text{ musí být regulární.} \quad (1.11)$$

Budeme hledat jediné kladné zobecněné vlastní číslo  $\lambda_0$  a jemu příslušný zobecněný vlastní vektor  $\mathbf{a}_0$ .

Problém (1.11) řeší  $\lambda_0$  s  $\alpha \mathbf{a}_0$ , kde  $\alpha$  může být libovolné reálné číslo. Dosazením  $\alpha \mathbf{a}_0$  do podmínky  $\mathbf{a}^T \mathbf{C} \mathbf{a} = 1$  dostaneme

$$\alpha = \pm \sqrt{\frac{\lambda}{\mathbf{a}_0^T \mathbf{S} \mathbf{a}_0}}.$$

Vzhledem k povaze úlohy stačí, když budeme uvažovat pouze kladné  $\alpha$ , protože obecná rovnice elipsy vynásobená číslem  $-1$  je ekvivalentní původní rovnici.

Řešením minimalizační úlohy je vektor

$$\mathbf{a} = \sqrt{\frac{\lambda}{\mathbf{a}_0^T \mathbf{S} \mathbf{a}_0}} \mathbf{a}_0. \quad (1.12)$$

Implementaci funkce pro fitování elips v Matlabu, která má jako vstupní parametr souřadnice bodů v rovině, do kterých chceme fitovat elipsu, a jejím výstupem je vektor  $\mathbf{a} = (A, B, C, D, E, F)$ , můžeme shrnout do následujících kroků:

1. Sestavení matice  $\mathbf{S}$  pomocí souřadnic vstupních bodů
2. Sestavení matice  $\mathbf{C}$

3. Výpočet zobecněných vlastních čísel a jim příslušejících zobecněných vlastních vektorů pomocí volání funkce  $\text{eig}(\mathbf{S}, \mathbf{C})$
4. Vyhledání největšího zobecněného vlastního čísla (jen jedno je kladné) a jemu příslušejícího vlastního vektoru
5. Výpočet výsledného vektoru  $\mathbf{a}$  dosazením do rovnice (1.11)

Fitzgibbonův algoritmus je implementován v příloženém souboru fitellipse.m.

Metodu snadno můžeme transformovat na metodu pro fitování hyperbol změnou podmínky například na  $4AC - B^2 = -1$ .

**Příklad 1.4** *Použitím Fitzgibbonovy metody zkusíme proložit elipsou body z příkladu 1.3.*

*Sestavíme matici  $\mathbf{A}$  pomocí bodů  $(0,0)$ ,  $(1,1)$ ,  $(2,4)$ ,  $(-1,1)$ ,  $(-2,4)$  a  $(3,9)$ .*

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 8 & 16 & 2 & 4 & 1 \\ 9 & 27 & 81 & 3 & 9 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 4 & -8 & 16 & -2 & 4 & 1 \end{pmatrix}$$

*Dále budeme potřebovat matici*

$$\mathbf{S} = \mathbf{A}^T \mathbf{A} = \begin{pmatrix} 115 & 243 & 859 & 27 & 115 & 19 \\ 243 & 859 & 2187 & 115 & 243 & 27 \\ 859 & 2187 & 7075 & 243 & 859 & 115 \\ 27 & 115 & 243 & 19 & 27 & 3 \\ 115 & 243 & 859 & 27 & 115 & 19 \\ 19 & 27 & 115 & 3 & 19 & 6 \end{pmatrix}$$

*Vidíme, že matice  $\mathbf{S}$  má první řádek stejný jako pátý, a tudíž není regulární. V tomto případě není existence a jednoznačnost řešení zaručena, a proto použití Fitzgibbonovy metody není vhodné.*

**Příklad 1.5** *V rovině máme body  $(-2,0)$ ,  $(2,0)$ ,  $(0,1)$ ,  $(0,-1)$ ,  $(1,1)$  a  $(-1,1)$ . Proložíme tyto body elipsou pomocí Fitzgibbonovy metody.*

Nejprve sestavíme matice  $\mathbf{A}$  a  $\mathbf{S}$ .

$$\mathbf{A} = \begin{pmatrix} 4 & 0 & 0 & -2 & 0 & 1 \\ 4 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \end{pmatrix}, \quad \mathbf{S} = \mathbf{A}^T \mathbf{A} = \begin{pmatrix} 34 & 0 & 2 & 0 & 2 & 10 \\ 0 & 20 & 2 & 0 & 0 & 0 \\ 2 & 0 & 4 & 0 & 2 & 4 \\ 0 & 2 & 0 & 10 & 0 & 0 \\ 2 & 0 & 2 & 0 & 4 & 2 \\ 10 & 0 & 4 & 0 & 2 & 6 \end{pmatrix}$$

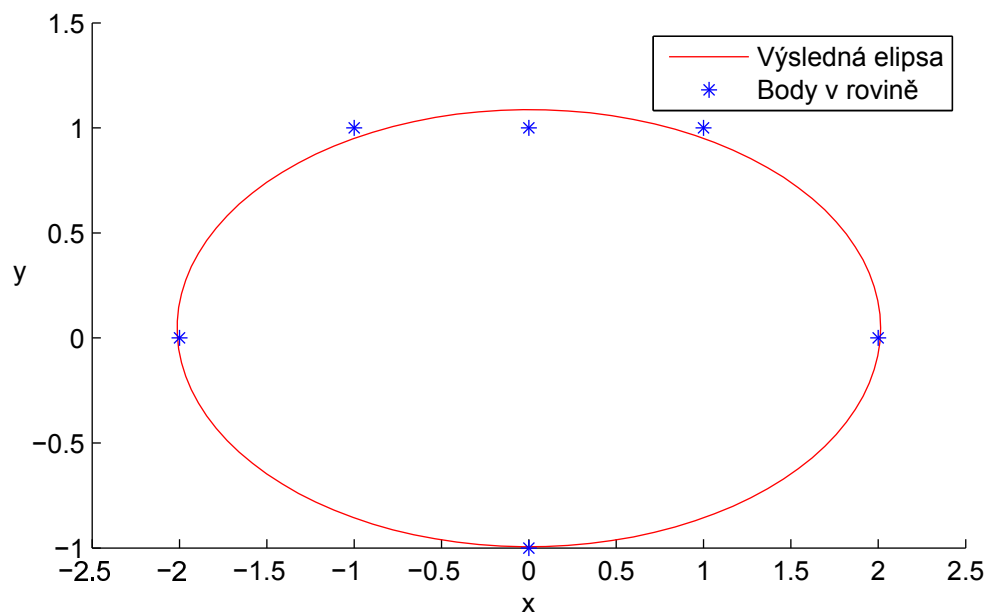
Matice  $\mathbf{S}$  je regulární a můžeme tedy použít Fitzgibbonovu metodu.

Použitím funkce  $\text{eig}(\mathbf{S}, \mathbf{C})$  v Matlabu najdeme nejmenší kladné zobecněné vlastní číslo  $\lambda_0$  takové, že spolu s jemu příslušejícím zobecněným vlastním vektorem  $\mathbf{a}_0$  řeší problém zobecněných vlastních čísel

$$\mathbf{S}\mathbf{a} = \lambda\mathbf{C}\mathbf{a}$$

$$\lambda_0 \cong 0,045; \quad \mathbf{a}_0 \cong (0,247; 0; 0,925; 0; -0,086; -1)^T$$

Dosažením vektoru  $\mathbf{a}_0$  do vztahu (1.12) dostaneme vektor  $\mathbf{a} = \mathbf{a}_0$ , který reprezentuje výslednou elipsu.



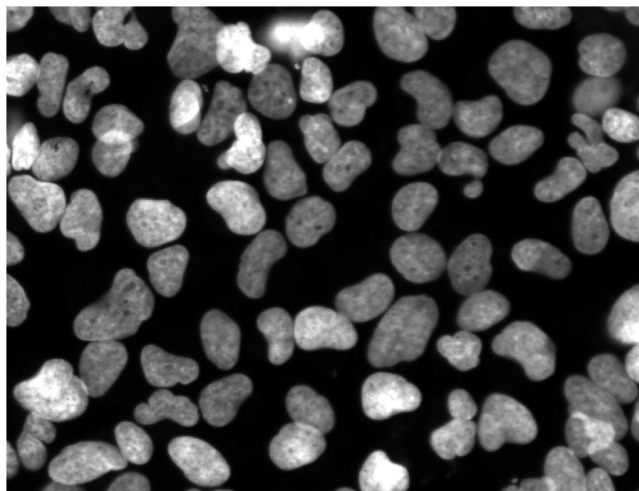
Obrázek 1.6: Výsledná elipsa

## Kapitola 2

# Implementace programu pro rozdělování dvojic buněčných jader

Nyní se zaměříme na popis programu pro rozdělování dvojic buněčných jader pomocí fitování elips metodou uvedenou v předchozí kapitole. Program je implementován v prostředí Matlab a používá některé standardní funkce z Matlab Image Processing Toolbox [11].

Na počátku máme černobílý snímek ve formátu TIFF s hloubkou 16 bitů a rozlišením 1360x1024, který typicky vypadá jako obrázek 2.1.



Obrázek 2.1: Snímek z mikroskopu

Tento snímek převedeme Otsuovou metodou prahování [12] na binární snímek, jež je v Matlabu reprezentován maticí s prvky nabývající hodnot 1 a 0. Jakmile máme snímek v binární podobě, odstraníme z něj objekty, které jsou příliš malé nebo se dotýkají jeho hran.



Obrázek 2.2: Binární snímek

Aplikací tohoto postupu nám z obrázku 2.1 vznikl obrázek 2.2. Teď je potřeba roztrždit objekty na snímku do třech skupin:

- objekty reprezentující jedno jádro
- dotýkající se dvojice jader
- shluky obsahující nejméně tři jádra

My se dále zaměříme pouze na skupinu dotýkajících se dvojic buněčných jader.

## 2.1. Popis algoritmu pro dělení dvojic jader

Vstupním parametrem pro algoritmus je matice reprezentující binární snímek s dvojicemi dotýkajících se jader (viz například obrázek 2.3).





Obrázek 2.3: Příklad vstupních dat

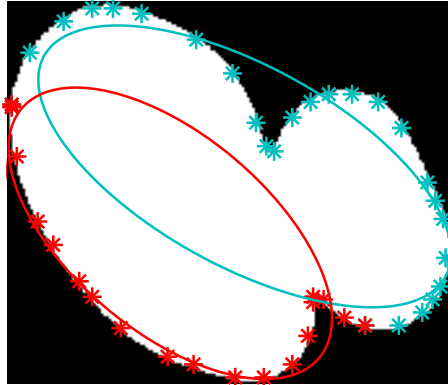
V této matici najdeme pomocí standardní funkce *bwboundaries* uspořádaný seznam souřadnic pixelů ležících na hranici jader. Z hranice vybereme tzv. reprezentanty. To jsou pixely tvořící vrcholy polygonální aproximace hranice.

Předpokládejme, že máme  $n$  reprezentantů. Vytvoříme z nich  $\frac{n(n-1)}{2}$  dvojic tak, že každá dvojice dělí hranici na dvě části, které obsahují nejméně šest reprezentantů (dělicí dvojice náleží oběma částem). Jestliže se tohle dělení nezdaří pro nedostatečný počet reprezentantů, nemá smysl dál pokračovat.



Obrázek 2.4: Rozdělení hranice na dvě části

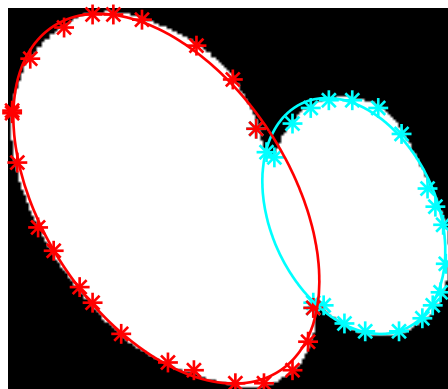
Hranice je tedy každou touto dvojicí rozdělena na dvě části. Do každé části nafitujeme pomocí jejích reprezentantů elipsu.



Obrázek 2.5: Proložení částí hranice elipsami

Vypočítáme geometrickou vzdálenost každého reprezentanta od elipsy aproximující tu část hranice, do které reprezentant patří. Pro každou elipsu najdeme k ní příslušícího reprezentanta, jehož vzdálenost je od dané elipsy největší. Součet těchto dvou maximálních vzdáleností uložíme do paměti spolu s referencí na dělicí dvojici.

Jakmile dokončíme fitování elips pro každou dělicí dvojici, najdeme tu z nich, pro niž bylo fitování elips nejvhodnější. Jinými slovy najdeme dělicí dvojici s nejnižší maximální odchylkou fitovaných elips od dat.



Obrázek 2.6: Proložení částí hranice elipsami

Porovnáme-li snímky 2.5 a 2.6, tak jsme okamžitě schopni říct, že elipsy na snímku 2.6 lépe aproximují tvar jader.

Posledním krokem algoritmu je rozdělení dvojice buněk úsečkou, která spojuje dělicí dvojici reprezentantů s nejnižší maximální odchylkou fitovaných elips od dat.



Obrázek 2.7: Výsledný snímek

## 2.2. Implementace algoritmu

V následujících sekcích uvedeme popis funkcí, které jsou přiloženy k této práci.

### 2.2.1. Funkce SPLIT

#### Vstup:

- `i_bwimage`: matice binárního snímku, který obsahuje dotýkající se dvojice jader

#### Výstup:

- `r_bwimage`: matice binárního snímku, kde je dvojice jader rozdělena

Funkce SPLIT je hlavní funkcí a volá ostatní funkce uvedené v této sekci. Algoritmus je detailně popsán v sekci 2.1.

## 2.2.2. Funkce DRAW\_LINE

### Vstup:

- `i_bwimage`: matice binárního snímku
- `i_point1`: souřadnice prvního dělicího bodu
- `i_point2`: souřadnice druhého dělicího bodu

### Výstup:

- `r_bwimage`: matice binárního snímku, kde mezi body `i_point1` a `i_point2` je vykreslena černá úsečka

Funkce `DRAW_LINE` slouží k vykreslení černé úsečky mezi body `i_point1` a `i_point2`.

Nejprve je matice `i_bwimage` uložena do matice `r_bwimage`. Pomocí parametrické rovnice úsečky změňme hodnoty pixelů, které leží na úsečce spojující body `i_point1` a `i_point2`, na hodnotu 0 (na binárním snímku vznikne černá čára).

## 2.2.3. Funkce GET\_BOUNDARY\_POINTS

### Vstup:

- `i_boundary`: souřadnice pixelů ležících na hranici jader

### Výstup:

- `r_points`: souřadnice reprezentantů hranice

Funkce `GET_BOUNDARY_POINTS` vybere z hranice tzv. reprezentanty. To jsou pixely tvořící vrcholy polygonální aproximace hranice.

Označme počet pixelů v `i_boundary` písmenem  $n$  a body z `i_boundary` označme  $\mathbf{x}_i, i = 1, \dots, n$ . Dále budeme předpokládat, že body v `i_boundary` jsou uspořádány tak, jak leží na hranici.

Nejprve vezmeme  $\mathbf{x}_1$  a uložíme ho do `r_points`. Potom postupně procházíme body  $\mathbf{x}_i, i \geq 2$ , až narazíme na bod  $\mathbf{x}_j$  takový, že pro některý bod  $\mathbf{x}_i, 2 \leq i < j$ ,

je jeho vzdálenost od úsečky určené body  $\mathbf{x}_1$  a  $\mathbf{x}_j$  větší než 1. Bod  $\mathbf{x}_j$  uložíme do `r_points`.

Analogicky určíme další body patřící do `r_points`. Algoritmus se zastaví, jakmile při procházení bodů  $\mathbf{x}_i$  narazíme na bod  $\mathbf{x}_1$ .

## 2.2.4. Funkce GET\_MAX\_DIST

### Vstup:

- `i_points`: souřadnice reprezentantů hranice
- `i_ellipse`: vektor obsahující souřadnice středu elipsy, délky její hlavní a vedlejší poloosy a úhel rotace elipsy

### Výstup:

- `r_max_dist`: maximum z geometrických vzdáleností bodů z `i_points` od elipsy `i_ellipse`

Pomocí parametrů ve vektoru `i_ellipse` sestavíme parametrickou rovnici elipsy a uložíme si souřadnice 360 bodů, které jsou na této elipse rovnoměrně rozprostřeny. Pro každý bod z `i_points` určíme jeho nejmenší vzdálenost od bodů ležících na elipse, a poté z těchto vzdáleností uložíme maximální hodnotu do proměnné `r_max_dist`.

## Kapitola 3

# Vlastnosti algoritmu pro rozdělování dvojic buněčných jader

Implementace algoritmu byla otestována na snímcích, jež poskytla laboratoř genomové integrity Ústavu molekulární a translační medicíny Lékařské fakulty Univerzity Palackého v Olomouci [8],

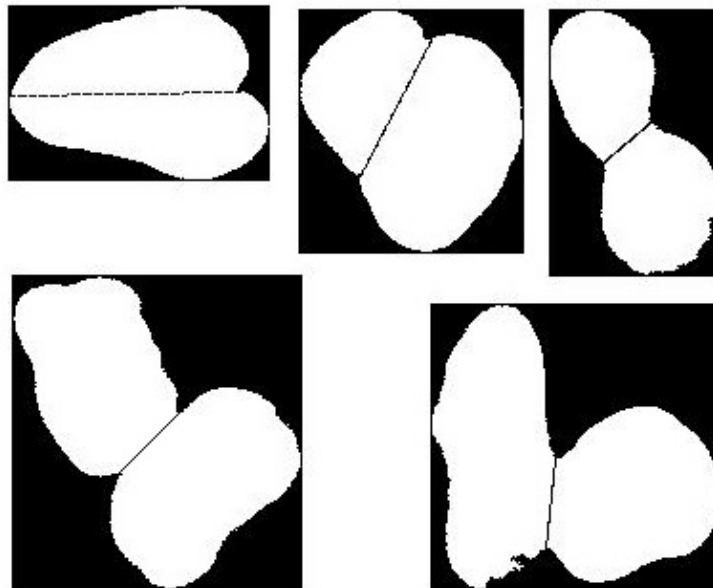
Na náhodném výběru 100 dvojic algoritmus selhal 7x. Tyto snímky jsou dostupné jako příloha.

Algoritmus selhával pouze v případě, kdy byl tvar buněčných jader zdeformován a nepodobal se elipse. Avšak vzhledem ke konstrukci algoritmu jsou tyto chyby očekávatelné. Navíc tato jádra by měla být úplně vyřazena z dalších analýz výzkumu integrity genomu.



Obrázek 3.1: Špatně rozdělená buněčná jádra

Naproti tomu výsledky dělení dvojic jader, které se svým tvarem dobře aproximují elipsu, jsou výborné.



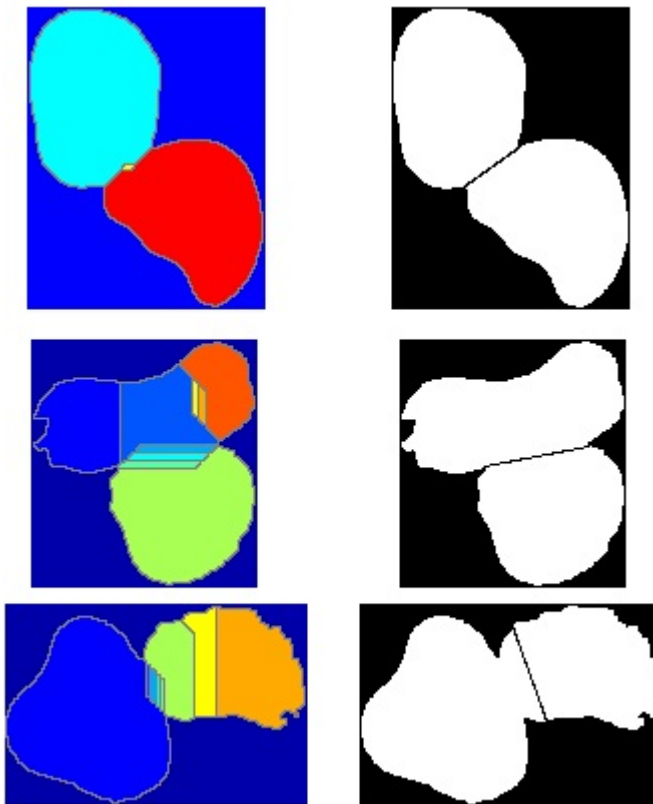
Obrázek 3.2: Správně rozdělená buněčná jádra

Nevýhodou algoritmu je časová náročnost zejména v případě, kdy hranice jader není hladká. To vede k určení několikanásobně více reprezentantů hranice, než v případě kdy je hranice hladká. Hranici lze však snadno vyhladit pomocí standardních funkcí z Matlab Image Processing Toolbox.

Za nevýhodu také můžeme považovat to, že dělení proběhne vždy. Je tedy nutné správně určit, zda vstupní data, která předáváme algoritmu, opravdu reprezentují dvojice dotýkajících se jader.

### 3.1. Srovnání s Watershed segmentací

Použijeme-li na stejný výběr 100 snímků segmentační algoritmus Watershed [13], který implementujeme podle [14], dostaneme nekonzistentní výsledky. Často dojde ke správnému oddělení pouze jednoho jádra. Druhé jádro je rozděleno na několik segmentů. Správně byly dvojice rozděleny v 9 případech.



Obrázek 3.3: Srovnání výsledků

Výhodou Watershed algoritmu je jeho nízká výpočetní náročnost. Rozdělení 100 dvojic metodou fitování elips trvalo přibližně 186 sekund. Watershed algoritmus zvládl zpracovat 100 stejných dvojic přibližně za 2 sekundy.



# Kapitola 4

## Závěr

Cílem této práce bylo především navrhnout a implementovat algoritmus, který pomocí fitování elips bude schopen rozdělit dvojice dotýkajících se buněčných jader na snímcích, které jsou součástí výzkumu integrity genomu laboratoře genomové integrity Ústavu molekulární a translační medicíny Lékařské fakulty Univerzity Palackého v Olomouci.

Nejprve bylo nutné najít vhodnou metodu prokládání bodů v rovině elipsou, která by byla dostatečně stabilní a rychlá. Obě tyto vlastnosti má Fitzgibbonova metoda fitování elips. Proto jsem se v první kapitole zaměřil na popis základních vlastností elips, a také jsem zde uvedl poměrně elegantní způsob řešení minimalizační úlohy s kvadratickou podmínkou pomocí zobecněných vlastních čísel, který byl nezbytný k pochopení Fitzgibbonovy metody. Na závěr první kapitoly jsem popsal Fitzgibbonovu metodu.

Druhá kapitola je věnována samotnému návrhu algoritmu pro dělení dvojic přilehlých buněčných jader pomocí fitování elips do jejich hranice.

Ve třetí kapitole jsem zkoumal vlastnosti tohoto nového algoritmu a provedl jeho srovnání se segmentačním algoritmem Watershed. Nový algoritmus prokázal vysokou stabilitu a dokázal rozdělit mnohem více dvojic než algoritmus Watershed. Proto považuji cíl této práce za splněný.

Vyvýjení nového segmentačního algoritmu mi napomohlo k zlepšení se v programování v prostředí Matlab, a také prohloubilo mé znalosti v oblasti zpracování digitálního obrazu.

# Literatura

- [1] Andrew W. Fitzgibbon, Maurizio Pilu, and Robert B. Fisher: *Direct Least Squares Fitting of Ellipses* IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(5), 476–480, May 1999.
- [2] R. Schöne, T. Hanning: *Least Squares Problems with Absolute Quadratic Constraints* Journal of Applied Mathematics, vol. 2012, Article ID 312985, 12 pages, 2012.
- [3] MathWorks: *Matlab Documentation* [online]. [cit. 18. května 2016]. Dostupné z: <http://www.mathworks.com/help/matlab/>.
- [4] K. Zimmermann, T. Svoboda: *Approximation of Euclidean Distance of Point from Ellipse* Research Report CTU–CMP–2005–23, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, August 2005.
- [5] Wikipedia: *Ellipse* [online]. [cit. 18. května 2016]. Dostupné z: <https://en.wikipedia.org/wiki/Ellipse>
- [6] F. Bookstein: *Fitting Conic Sections to Scattered Data* Computer Graphics and Image Processing, no. 9, pp. 56-71, 1979.
- [7] Wikipedia: *Image segmentation* [online]. [cit. 18. května 2016]. Dostupné z: [https://en.wikipedia.org/wiki/Image\\_segmentation](https://en.wikipedia.org/wiki/Image_segmentation)
- [8] Ústav molekulární a translační medicíny - LF - UP: <http://www.imtm.cz/>
- [9] Wikipedia: *DAPI* [online]. [cit. 18. května 2016]. Dostupné z: <https://en.wikipedia.org/wiki/DAPI>
- [10] Proof for Classifying Conics by Using the Discriminate [online]. [cit. 18. května 2016]. Dostupné z: [http://www.mesacc.edu/~mic2191380/calculus2/tilli\\_treats/Proof%20Classifing%20Conics.pdf](http://www.mesacc.edu/~mic2191380/calculus2/tilli_treats/Proof%20Classifing%20Conics.pdf)
- [11] MathWorks: *Image Processing Toolbox* [online]. [cit. 18. května 2016]. Dostupné z: <http://www.mathworks.com/products/image/>

- [12] N. Otsu: *A Threshold Selection Method from Gray-Level Histograms* IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, 1979, pp. 62-66.
- [13] Wikipedia: *Watershed (image processing)* [online]. [cit. 18. května 2016]. Dostupné z: [https://en.wikipedia.org/wiki/Watershed\\_\(image\\_processing\)](https://en.wikipedia.org/wiki/Watershed_(image_processing))
- [14] MathWorks: *Watershed* [online]. [cit. 18. května 2016]. Dostupné z: <http://www.mathworks.com/help/images/ref/watershed.html>

# Přílohy

Součástí této práce je CD se složkou DATA a soubory:

- draw\_line.m
- fitellipse.m
- get\_boundary\_points.m
- get\_max\_dist.m
- process\_images.m
- split.m
- watershed\_split.m

Složka DATA obsahuje připravené binární snímky dvojic přilehlých buněčných jader ve formátu TIFF, na kterých jsme ve třetí kapitole testovali segmentační metodu Watershed (funkce watershed\_split.m) a novou metodu pro dělení dvojic jader pomocí fitování elips (funkce split.m).

Funkce watershed\_split.m a split.m se spouští pomocí funkce process\_images.m. Vstupní parametr této funkce je adresa složky s binárními snímky dvojic přilehlých buněčných jader ve formátu TIFF (např.: D:\DATA\). Po spuštění funkce process\_images.m se ve vstupním adresáři vytvoří složky ELLIPSE a WATERSHED, které budou obsahovat výsledné snímky s rozdělenými dvojicemi buněčných jader.