



Faculty of electrical engineering
University of Ljubljana

Med  Tech

Zentralinstitut für Medizintechnik



Technical University München

MASTER THESIS

ANALYSIS OF CELL CULTURE IMAGES

Andreja Jarc Dipl.Ing.

Supervisors: Prof. Dr. Stanislav Kovačič Dipl. Ing.
Dr. med. Markus Eblenkamp

MÜNCHEN, MAY 2005

UNIVERZA V LJUBLJANI
FAKULTETA ZA ELEKTROTEHNIKO

UNIVERZITETNI ŠTUDIJ
MAGISTRSKO DELO

ANALIZA SLIK CELIČNIH KULTUR

ANDREJA JARC UNIV. DIPL. INŽ.

MENTOR: PROF. DR. STANISLAV KOVAČIČ
SOMENTOR: DR. MARKUS EBLENKAMP

LJUBLJANA, maj 2005

KAZALO

I. UVOD	III
II. MATERIALI IN METODE	VI
III. REZULTATI.....	VII
III.I. RAZVRŠČANJE SLIK NA OSNOVI KOFLUENTNOSTI	VII
III.II. RAZVRŠČANJE SLIK NA OSNOVI MORFOLOŠKIH ZNAČILNOSTI	VIII
IV. RAZPRAVA	VIII
V. ZAKLJUČEK.....	IX

Opomba avtorice: delo je bilo celoti izdelano v sklopu mednarodne študijske izmenjave na Tehnični univerzi v Münchnu. Magistrsko delo se sestoji iz obširnejšega povzetka dela v slovenskem jeziku, temu pa sledi originalna verzija magistrske naloge, napisana v angleškem jeziku.

Ključne besede: celične kulture, računalniški vid, avtomatsko razvrščanje.

I. UVOD

Gojenje človeških in živalskih celic ter tkiv je prisotno v številnih znanstvenih panogah. Pojavlja se na področjih celične in molekularne biologije, kakor tudi na hitro rastočem področju farmacevtskih in biotehnoloških znanosti.

Seznam celic, ki jih gojijo v celičnih kulturah, obsega celice vezivnega tkiva - fibroblastne celice, osteoblastne celice v kosteh, srčne in gladke mišične celice, endotelijske celice na notranji strani žil in druge. Celice lahko odstranimo iz organizma ali prvotnega tkiva in jih vzgojimo *in vitro*. Celice so tako gojene v celičnih kulturah (ali celičnih kolonijah). Za uspešno rast in razmnoževanje celic *in vitro* so potrebni ustrezni okoliški pogoji gojenja, kakor tudi primerni gojitveni medij, kamor celice nasadimo. **Primarne celične kulture** imenujemo tiste, ki so prvič uspešno presajene *in vitro* neposredno iz tkiva. **Sekundarne celične kulture** sledijo iz primarnih. Celice večinoma ohranjajo tkivno-specifične lastnosti v prvih presaditvah, medtem ko jih kasneje spremenijo ali celo izgubijo. Gojenje celic v sekundarni kulturah (ali celičnih linijah) omogoča veliko količino enotnega materiala za nadaljnjo uporabo. V farmacevtski industriji v večini primerov nadomesti poskuse na živalih, hkrati pa je tudi čedalje bolj prisotno na področju tkivnega inženirstva.

Pri gojenju celic v celičnih kulturah je mikroskop nepogrešljivo orodje. Skozi mikroskop redno opazujemo in nadzorujemo:

- morebitno okuženost celične kulture z bakterijami,
- stopnjo razraščенosti (konfluentnost) celične kulture na gojišču¹ in

¹ Konfluentnost je izraz, ki pove stopnjo razraščенosti celic na gojišču. Če je kultura konfluentna, pomeni, da celice nimajo več prostora, da bi se naprej razmnoževale. Če je kultura sub-konfluentna, pomeni, da je v posodi še dovolj prostora za celično rast in razmnoževanje.

- tipičnost celic po morfoloških² značilnostih ali morebitno morfološko spremenjenost.

Glavna težava pri opazovanju celic pod mikroskopom je, da ne obstajajo objektivni oz. standardizirani kriteriji, po katerih bi se opazovalec odločil, kako naj naprej ravna s celično kulturo. Vse nadaljnje obdelave celic so odvisne izključno od subjektivne presoje in izkušenosti opazovalca.

Tako smo na Inštitutu za medicinsko tehniko na Tehnični univerzi v Münchnu prišli na idejo, da bi razvili postopek za objektivno avtomatsko prepoznavanje celičnih kultur na osnovi njihovih mikroskopskih posnetkov. Ta postopek naj bi se v prihodnosti vgradil v avtomatsko enoto, ki bi bila sposobna izvrševati nekaj rutinskih opazovanj (štetje celic, ugotavljanje stopnje konfluentnosti, ugotavljanje morfoloških značilnosti,...) in nekaj osnovnih obdelav celičnih kultur.

Namen naloge je bil razviti postopek za avtomatsko prepoznavanje celičnih kultur, ki bi bil sposoben pravilno razvrstiti mikroskopske slike kultur po:

- stopnji konfluentnosti (to pomeni, da bi postopek odkril, kako goste so razraščene celice v kulturi) in
- po morfoloških značilnostih celičnih kultur.

Postopek razvrščanja smo razvili na primerih slik endotelijskih in fibroblastnih celičnih kultur.

Endotelijske celice (glej **Slika 1**, str. 8)³ obraščajo notranjost žil in tako preprečujejo, da bi kri odtekala iz žil. Na **Sliki 1** vidimo značilno konfluentno (povsem razraščeno) kulturo endotelijskih celic. Celice se razraščajo v značilnem »tlakovanem«⁴ vzorcu. So okrogle oblike, naključno orientirane, v povprečju velike okrog 10-20 µm.

Fibroblastne celice (glej **Slika 2**, str. 9)⁴ so sestavni del vezivnega tkiva. Fibroblastne celične kulture tvorijo značilen vzorec v obliki vlaken. Na sliki konfluentne fibroblastne kulture (**Slika 2**) vidimo, da so celice podolgovate, rastejo v določeni smeri in tvorijo periodičen vzorec. Po

² Morfološke značilnosti celic pomenijo njihovo zunanjo podobo.

³ Referenca se nanaša na Figure 1, str. 8 v originalnem, angleškem delu magistrske naloge.

⁴ Referenca se nanaša na Figure 2, str. 9 v originalnem, angleškem delu magistrske naloge.

⁵ Referenci se nanašata na Figure 3, str. 10 in Figure 4, str. 11 v originalnem, angleškem delu magistrske naloge.

velikosti so nekoliko večje od endotelijskih celic. Na **Sliki 3** in **Sliki 4**⁵ vidimo sub-konfluentni celični kulturi zgoraj omenjenih celic. Za sub-konfluentne kulture velja, da še niso povsem razraščene, zato imajo na gojišču še dovolj prostora za nadaljnjo rast in razmnoževanje. Tako je na slikah s sub-konfluentnimi celičnimi kulturami močno prisotno enakomerno ozadje posode, v kateri so celice gojene.

Glavni cilji in naloge magistrskega dela so bili:

- Ob uporabi mikroskopa za opazovanje celičnih kultur sistematično spreminjati parametre, ki vplivajo na kakovost mikroskopskih slik (odprtina zaslonke, ostrina, čas osvetlitve, resolucija slik). Poskušali smo podati standardizirani postopek mikroskopiranja in zajemanja slik, na podlagi katerega bi razviti postopek razvrščanja optimalno deloval.
- Poiskati v literaturi primerne matematične metode, ki bi s slik celičnih kultur izločile značilne informacije (značilke), po katerih bi slike pravilno razvrstili. V prvem delu smo najprej želeli ločiti slike sub-konfluentnih in konfluentnih celičnih kultur. Te naloge smo se lotili na naslednje načine:
 - detekcija robov s pomočjo Sobelovega filtra (poglavje 2.2.1.1.)⁶. Gostoto robov, njihovo dolžino in obliko smo primerjali na slikah posameznih razvrstitvenih razredov;
 - statistični pristop opisa teksture slik (poglavje 2.2.1.2.)⁷, ki temelji na porazdelitvi sivinskih vrednosti na sliki; in
 - uporaba Haralickovih teksturnih koeficientov (poglavje 2.2.1.2.)⁸, ki temeljijo na matrikah vezanih verjetnosti.

V drugem delu je sledilo razvrščanje slik konfluentnih endotelijskih in fibroblastnih celičnih kultur na osnovi njihovih morfoloških značilnosti. Razvrščanja slik s konfluentnimi kulturami smo se lotili na naslednje načine:

^{6, 7, 8}Reference se nanašajo na istoštevilna poglavja v originalnem delu magistrske naloge.

- spektralna analiza, ki temelji na Fourierovi transformaciji (poglavje 2.2.2.1.)⁹. S to metodo smo numerično ovrednotili in primerjali orientiranost ter periodičnost vzorcev, ki jih kažejo vzorci obeh konfluentnih celičnih kultur;
 - lokalna prostorska in spektralna analiza s pomočjo nabora Gaborjevih filtrov (poglavje 2.2.2.2.)¹⁰; in
 - analiza segmentiranih slikovnih objektov s pomočjo programske opreme Cellenger® (poglavje 2.2.2.3)¹¹.
- Zadnji zastavljeni cilj magistrske naloge je bil ovrednotiti dobljene rezultate razvrščanja slik s pomočjo Fisherjevega optimizacijskega kriterija (poglavje 2.3.1.)¹². Kriterij temelji na domnevi, da so razredi, v katere smo želeli pravilno razvrstiti slike na podlagi značilk, med seboj ločeni s hiper-ravninami. Značilke v vektorskem prostoru, za katere je vrednost Fisherjeve optimizacijske funkcije največja, predstavljajo lastnosti, po katerih lahko razrede slik med seboj najbolj ločimo.

II. MATERIALI IN METODE

Slike, ki smo jih uporabili za zgoraj opisane naloge, smo pridobili s fazno kontrastno tehniko mikroskopiranja, z inverznim mikroskopom. Skupna povečava celičnih kultur na mikroskopu je bila 100×. Slike smo posneli s CCD kamero, v resoluciji 1030× 1300 pikslov. Vsega skupaj smo posneli 40 slik (10 iz vsakega razreda). Vsaka slika znotraj enega razreda je bila posneta z različno osvetljenostjo. Z različnimi osvetljenostmi smo želeli zagotoviti robustnost razvrščevalnih metod, da bi bile neodvisne od osvetljenosti slike.

Posnete slike smo razdelili v štiri razrede, v katere smo želeli avtomatsko razvrščati:

^{9, 10, 11, 12} Reference se nanašajo na istoštevilka poglavja v originalnem delu magistrske naloge.

- Slike s sub-konfluentnimi endotelijskimi celičnimi kulturami;
- Slike s sub-konfluentnimi fibroblastnimi celičnimi kulturami;
- Slike s konfluentnimi fibroblastnimi celičnimi kulturami; in
- Slike s konfluentnimi endotelijskimi celičnimi kulturami.

III. REZULTATI

III.I. RAZVRŠČANJE SLIK NA OSNOVI KONFLUENTNOSTI

V **Tabeli 1** so predstavljeni povzetki rezultatov metod za razvrščanje slik na osnovi konfluentnosti celičnih kolonij.

*	Sub-konfluentna kultura	Konfluentna kultura	Vrednost Fisherjeve optimizacijske funkcije
Gostota robov na sliki	Nizka	Visoka	0.59
Homogenost slik (Statistični momenti)	Visoka	Nizka	0.24
Haralickovi teksturni koeficienti	Visoki	Nizki	4.81

Tabela 1: Povzetek značilnih rezultatov razvrščanja po kriteriju konfluentnosti. *Vrednosti so signifikantno značilne. Najvišjo vrednost Fisherjeve optimizacijske funkcije smo dosegli z metodo Haralickovih teksturnih koeficientov.

III.II. RAZVRŠČANJE SLIK NA OSNOVI MORFOLOŠKIH ZNAČILNOSTI

V **Tabeli 2** so zbrani povzetki rezultatov metod razvrščanja slik na osnovi morfoloških značilnosti celičnih kultur.

*	Fibroblastne celice	Endotelijske celice	Vrednost Fisherjeve optimizacijske funkcije
Orientiranost vzorca	Visoka	Nizka	8.06
Periodičnost vzorca	Opazna	Neopazna	7.31
Oblika segmentiranih objektov	Podolgovati	Okrogli	9.99

Tabela 2: Povzetek značilnih rezultatov razvrščanja po morfoloških značilnostih konfluentnih celičnih kultur. *Vrednosti so signifikantno značilne. Najvišjo vrednost Fisherjeve optimizacijske funkcije smo dosegli z metodo primerjave oblik segmentiranih objektov s programom Cellenger®.

IV. RAZPRAVA

Na podlagi povzetih rezultatov lahko zaključimo, da smo zastavljeni nalogi razvrščanja slik celičnih kultur po kriterijih konfluentnosti in po morfoloških značilnostih uspešno rešili. Slike celičnih kultur po kriteriju konfluentnosti glede na vrednost Fisherjevega kriterija najbolje razvrstimo s **Haralickovimi teksturnimi koeficienti**. Slike celičnih kultur po morfoloških značilnostih pa najbolje razvrstimo glede na **obliko segmentiranih objektov** na danih slikah.

Na tem mestu je potrebno dodati, da je predlagane metode potrebno preizkusiti na večjem številu slik in s spreminjajočimi parametri, ki vplivajo na kvaliteto posnetih slik. Predlagane metode bi bile v kombinaciji z ostalimi metodami gotovo robustnejše in uporabne na širših problemih razvrščanja slik celičnih kultur. Metode bi lahko med seboj kombinirali s pomočjo fuzzy logike.

V. ZAKLJUČEK

V magistrskem delu je bila zadovoljivo izpolnjena zastavljena naloga objektivnega razvrščanja slik celičnih kultur. Slike smo uspešno razvrstili glede na kriterij konfluentnosti in morfoloških značilnosti celičnih kultur.

Za bodoče delo bi bilo smiselno preveriti in razširiti robustnost predlaganih metod glede na kvaliteto posnetih slik. Značilke metod bi bilo zanimivo kombinirati med seboj s pomočjo fuzzy logike.

Sistem za razvrščanje, ki smo ga vzpostavili na osnovi dveh celičnih kultur, bi bilo potrebno razširiti še na vrsto ostalih celičnih kultur, ki jih gojijo v bioloških laboratorijih.

“Izjavljam, da sem diplomsko delo izdelala samostojno pod vodstvom mentorja prof. dr. Stanislava Kovačiča univ. dipl. inž. el. in somentorja dr. Markusa Eblenkampa. Izkazano pomoč drugih sodelavcev sem v celoti navedla v zahvali.”

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS	4
ABSTRACT	5
1. INTRODUCTION	6
1.1. BACKGROUND	6
1.1.1. INTRODUCTION TO THE OBSERVED CELL CULTURES	8
1.2. FUNDAMENTALS OF DIGITAL IMAGE PROCESSING	11
1.3. OBJECTIVES AND MOTIVATION.....	13
2. MATERIALS AND METHODS.....	15
2.1. IMAGE ACQUISITION	15
2.1.1. BASICS OF THE SEEDING TECHNIQUE.....	15
2.1.2. INTRODUCTION TO MICROSCOPY.....	16
2.1.2.1. PHASE CONTRAST IN TRANSMITTED LIGHT	17
2.1.3. IMAGE SHOTS.....	19
2.2. IMAGE ANALYSIS.....	20
2.2.1. SEPARATION OF SUB-CONFLUENT AND CONFLUENT IMAGES	20
2.2.1.1. EDGE DETECTION	20
2.2.1.2. TEXTURE RECOGNITION	22
2.2.2. SEPARATION OF CONFLUENT ENDOTHELIAL AND FIBROBLAST IMAGES	29
2.2.2.1. SPECTRAL APPROACH: FOURIER SPECTRUM	29
2.2.2.2. GABOR FILTERING.....	29
2.2.2.3. CELLENGER® APPROACH.....	32
2.3. STATISTICAL PATTERN RECOGNITION	43
2.3.1. FISHER LINEAR DISCRIMINANT.....	43
3. RESULTS.....	45
3.1. SEPARATION OF SUB-CONFLUENT AND CONFLUENT IMAGES.....	46
3.1.1. EDGE DETECTION.....	46
3.1.2. STATISTICAL APPROACH.....	52
3.1.3. HARALICK TEXTURE COEFFICIENTS	56
3.2. SEPARATION OF CONFLUENT ENDOTHELIAL AND FIBROBLAST IMAGES	58
3.2.1. SPECTRAL ANALYSIS	58
3.2.2. GABOR FILTERING	62
3.2.3. CELLENGER® APPROACH.....	71

TABLE OF CONTENTS

4. DISCUSSION.....	77
4.1. EDGE DETECTION.....	77
4.2. TEXTURE	79
4.2.1. STATISTICAL APPROACH	79
4.2.2. HARALICK TEXTURE COEFFICIENTS	80
4.2.3. SPECTRAL ANALYSIS	82
4.2.4. GABOR FILTERING	83
4.3. CELLENGER® APPROACH.....	85
4.4. MICROSCOPY	88
5. CONCLUSION AND OUTLOOK.....	89
6. BIBLIOGRAPHY.....	90
INTERNET SITES.....	92
7. ACKNOWLEDGMENTS.....	94

LIST OF ABBREVIATIONS AND DEFINITIONS

2-D	two-dimensional
N-D	n-dimensional
CELLENGER®	Cellenger platform
CONFLUENCE	final state of growth of a cell culture before it is subcultured
ENDOCONF	image of a confluent endothelial cell culture
ENDOSUB	image of a sub-confluent endothelial cell culture
EQ.	equation
FIBROCONF	image of a confluent fibroblast cell culture
FIBROSUB	image of a sub-confluent fibroblast cell culture
GLCM	Gray Level Co-occurrence Matrix
HUVEC	Human Umbilical Vein Endothelial Cell
MADofMD	Mean Absolute Difference of Mean Direction
PBS	Phosphate-Buffered Saline
RGB	Red Green Blue color system representing an $M \times N \times 3$ array of color pixels, where each pixel is a triplet corresponding to the red, green and blue components of an RGB image at the specific spatial location.
SD	standard deviation
SDofMD	Standard Deviation of Mean Direction
TUM	Technische Universität München
UC1214	Umbilical Cord Cells 1214
ZIMT	Zentral Institut für Medizintechnik

ABSTRACT

Automated feature extraction and object recognition are large research areas in the field of image processing and computer vision. There already exist many automated image analysis methods for acquiring numerical information from medical and biological images, especially information concerning cell counting and quantitative measures of cell structures (e.g., size and area). However, no report on automated classification of cell cultures has been found in the literature. This was aimed and achieved in the present thesis. The image classification of endothelial and fibroblast cell cultures is composed of two main steps. First, the cell culture is classified by its confluence character. Classification is performed through edge detection, statistical moments and Haralick texture coefficients. The method which used Haralick texture coefficients delivered results with the highest separability value. Secondly confluent endothelial cell cultures and fibroblast cell cultures are distinguished from each other. Distinction is performed through Fourier descriptors, Gabor filtering and image processing software Cellenger® (Definiens). The Cellenger® delivered the highest separability value with Fisher linear discriminant criterion. One way for using these descriptors would be to “teach” representative descriptor values for a set of different cell culture morphology features to a system. The features of an unknown culture would subsequently be determined by how closely its descriptors matched those stored in the system memory.

Keywords: cell cultures, pattern recognition, automated classification.

1. INTRODUCTION

1.1. Background

Cultivation of human or animal cells and tissues is a widely used technique in many different disciplines ranging from the basic science of cellular and molecular biology to the rapidly evolving field of biotechnology.

The list of different cell types, which can now be grown in culture, is quite extensive, and includes connective tissue cells such as fibroblasts or osteoblasts, cardiac and smooth muscle cells, endothelial cells lining blood vessels, and others. If a cell is removed from the original tissue or an organism and placed *in vitro*, then the cell is cultivated in a **cell culture**. Environmental conditions and extracellular medium that the cell is exposed to *in vitro* should meet the essential requirements for survival and growth of a cell. Only by this, the cell is able to survive, to proliferate and to differentiate.

Freshly isolated cell cultures from tissue are known as primary cultures until they are passaged or subcultured. Primary cultures are usually heterogeneous and have a relatively low growth fraction, but they are more representative of the cell types in the tissue from which they were derived and express tissue specific properties. Subcultures (also known as a cell line) allow expansion of the culture and the possibility of cloning, characterization and preservation, and greater uniformity, but expansion may cause a loss of specialized cells and differentiated properties [4]. The greatest advantage of subculturing a primary culture into a cell line is the provision of a large amount of consistent material suitable for prolonged use.

Several studies have suggested that more immature cells are able to multiply to a higher degree *in vitro* than fully differentiated cells of specialized tissues [15]. In contrast to the *in vitro* multiplication of fully differentiated cells, these immature cells can be induced to differentiate into functional active cells after several generations *in vitro*.

Cell cultures are mainly used as a test object in both basic cell research and pharmaceutical development. The cell cultivation often serves as an alternative for animal tests. In addition, the expansion of cells is essential for the field of tissue engineering, which is regarded as a promising therapeutic approach in the future.

A microscope is an indispensable tool in cultivating cells. It is necessary to regularly check the following parameters through a microscope:

- if the cell culture is bacterially contaminated;
- if the cell culture is confluent; and
- whether cells look typical or degenerated.

There does not exist a set of objective rules for visual inspection of cell cultures. Thus, any further manipulation of the cells results from an observer's subjective interpretation.

With experimental focus on stem cells (further on described as **fibroblast cells**) and **endothelial cells**, the biological laboratory at the Zentalinstitut für Medizintechnik (ZIMT) of the Technical University of Munich was the initiator of the following work. Their need of objectiveness in cell culture observation induced an idea of image based automated cell culture recognition in order to distinguish between cell types. As a result, many routine laboratory observations (e.g., counting of cells and detecting a degree of cell growth) would be automated and sped up, thus becoming more effective. Moreover, an automated process would deliver reproducible and objective results independent of researchers' skills and experiences. An idea for a fingerprint-based identification of different cell cultures according to their morphology has also been entertained.

1.1.1. INTRODUCTION TO THE OBSERVED CELL CULTURES

Cells can be physically described by their specific morphological features. Knowledge of cell morphology is of great importance in selecting the right set of features to correctly classify cells. Experiments in cell morphology were performed on endothelial and fibroblast cell cultures.

All blood vessels and lymphatics¹ are lined by **endothelial cells**. The layer the endothelial cells constitute is the endothelium.

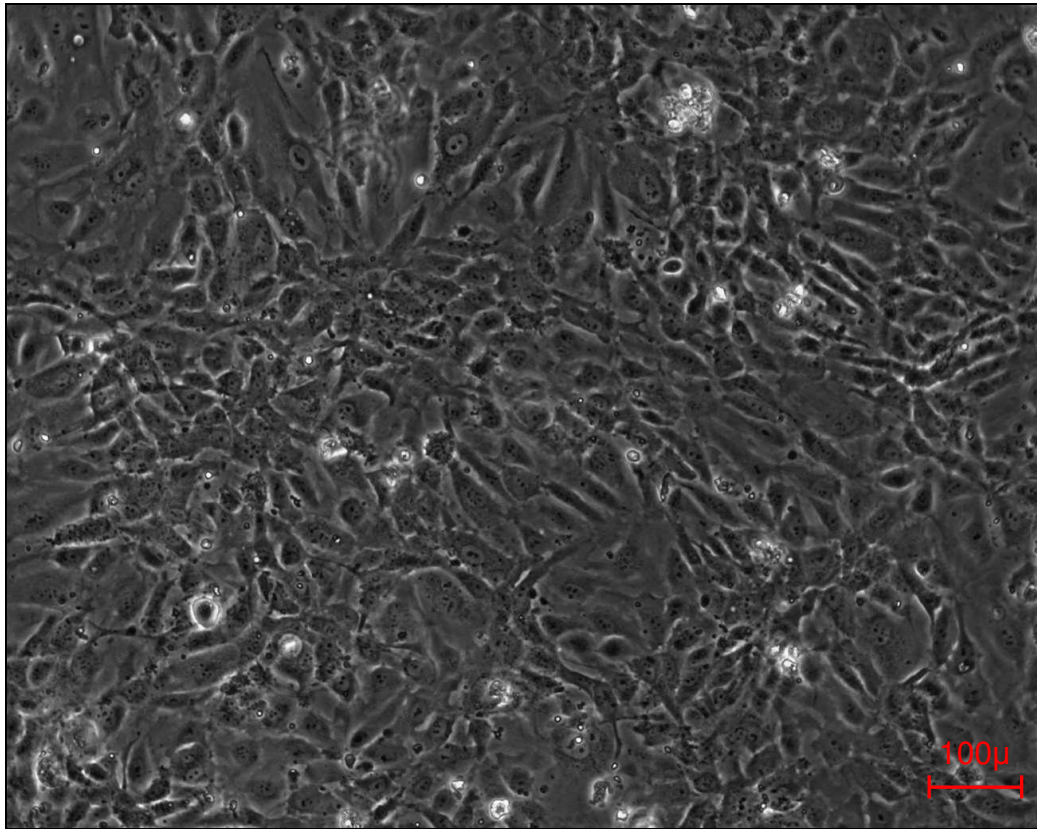


Figure 1: A fully grown or confluent cell culture of HUVEC endothelial cells. Endothelial cells are very flat, round, have a central nucleus, are approximately 1 – 2 μm thick and are approximately 10 – 20 μm in diameter. They form flat, pavement like patterns on the inside of vessels to prevent the blood leakage.

¹ Lymphatics are vessels that convey lymph.

A **fibroblast** is a cell that makes structural fibers² and ground substance of connective tissue.

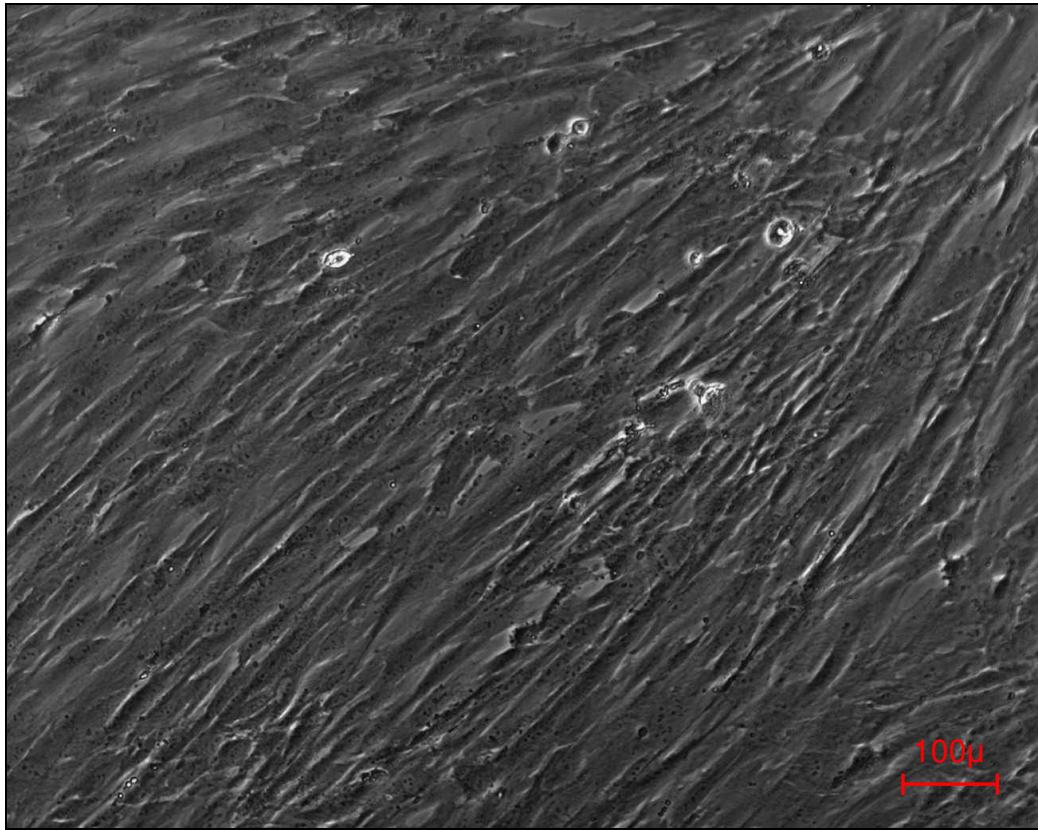


Figure 2: A fully grown or confluent cell culture of UC1214, showing the same morphological characteristics as fibroblast cells. UC1214 are gained from the umbilical cord and have a unique capacity to renew themselves and to give rise to specialized cell types. UC1214 are approximately 1 μm thick and approximately 15 – 30 μm in diameter.

When cells are seeded into a culture vessel they enter a lag period of about 2 – 24 h, followed by a period of exponential growth ('log phase') and finally enter a period of reduced or no growth after they become **confluent** or fully grown ('plateau phase') [4]. These phases have unique characteristics for each cell line and give rise to some important measurements: the length of the lag period, the population doubling time, the saturation density at plateau and others. Before cell growth reaches the plateau phase it is said to be **sub-confluent**, meaning that there is still enough space for cells to proliferate. When they become confluent, they have no more vacant space and the proliferation accordingly stops.

² A fiber is any of the fine, thin structures constituting the outside cell matrix of connective tissue.

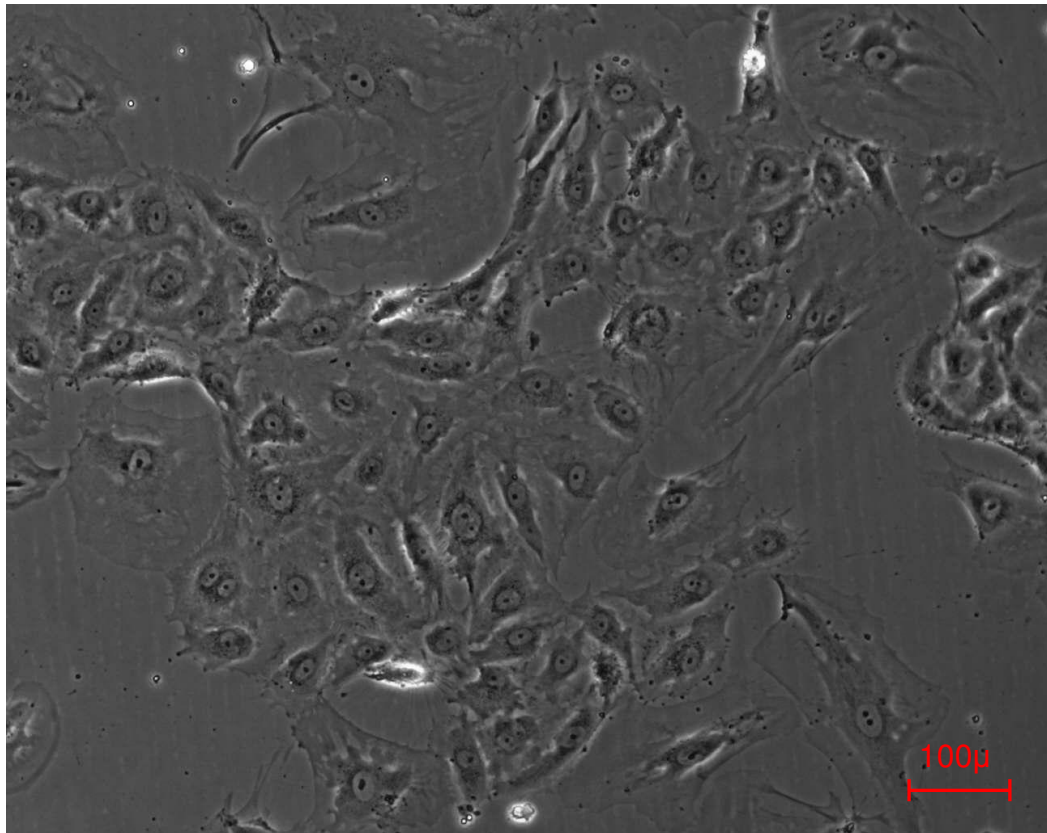


Figure 3: A typical sub- confluent cell culture of HUVEC endothelial cells. In comparison with the confluent endothelial cell culture on Figure 1 some morphological differences can be seen in the images. Cells in this image have a pointed out nucleus, sometimes even two together if cell is about to be divided. Cells cover a large area around their nucleus and have many outgrowths in their surface.

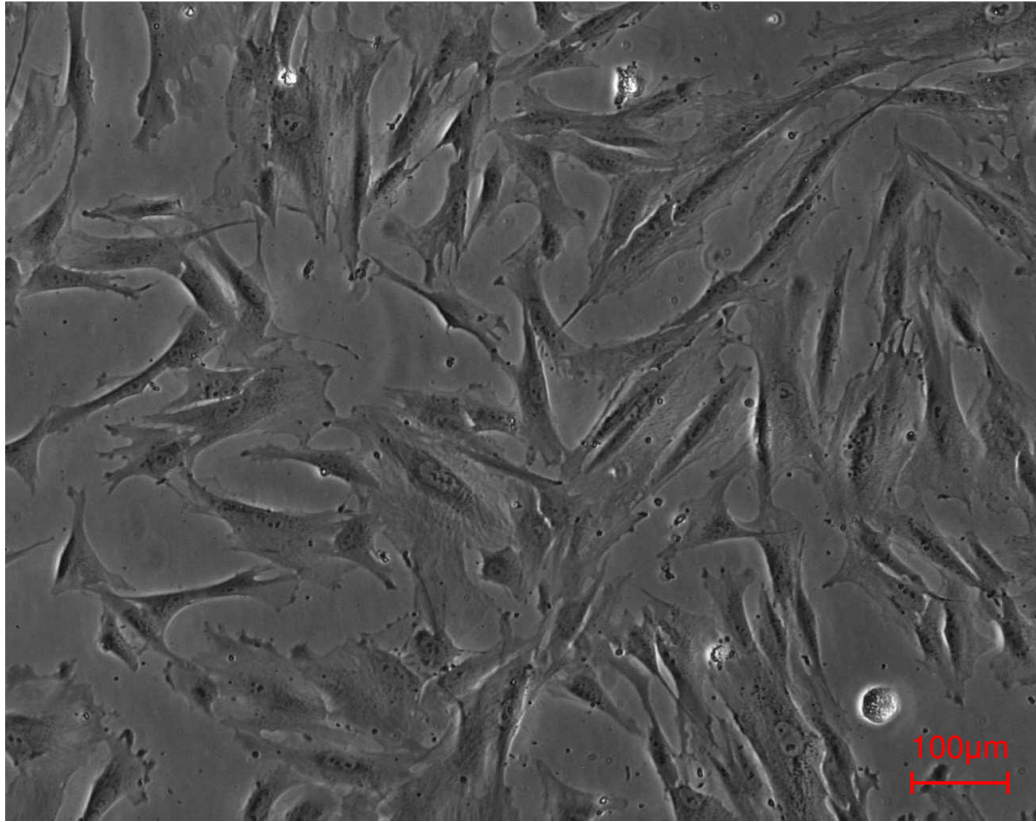


Figure 4: A typical sub-confluent cell culture of UC1214 differentiated into fibroblast cells. In comparison to the confluent UC1214 cells in Figure 2 they do not have a pointed out nucleus, their surface is rather large with many lengthy outgrowths.

1.2. Fundamentals of digital image processing

Due to the digital image processing and analysis focus of this work some basic digital image definitions will be given in this chapter.

An image may be defined as a two – dimensional function $f(x,y)$, where x and y are spatial coordinates, and the amplitude value of f at any pair of coordinates (x,y) is called the intensity or gray level of the image at that point. When x , y and the amplitude values of f are all finite, discrete quantities, we say that an image is a digital image. Therefore a digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as *picture elements*, *image elements*, or (most commonly) *pixels*. An image may be continuous with respect to the x - and y - coordinates and with respect to amplitude values.

INTRODUCTION

Conversion of a continuous image to a digital one requires that the spatial coordinates, as well as amplitude values, are digitized. Digitizing the coordinate values is called *sampling*; digitizing the amplitude values is called *quantization* [5].

The result of digitization is a matrix of real numbers. In digitization, a continuous image is divided into M rows and N columns, creating a digital image of size $M \times N$. The coordinates (x,y) are now discrete quantities. A quantified amplitude value at a coordinate (x,y) represents the *intensity* of the image in that point. Note that the term *gray value* is used often to refer to the intensity of monochrome images.

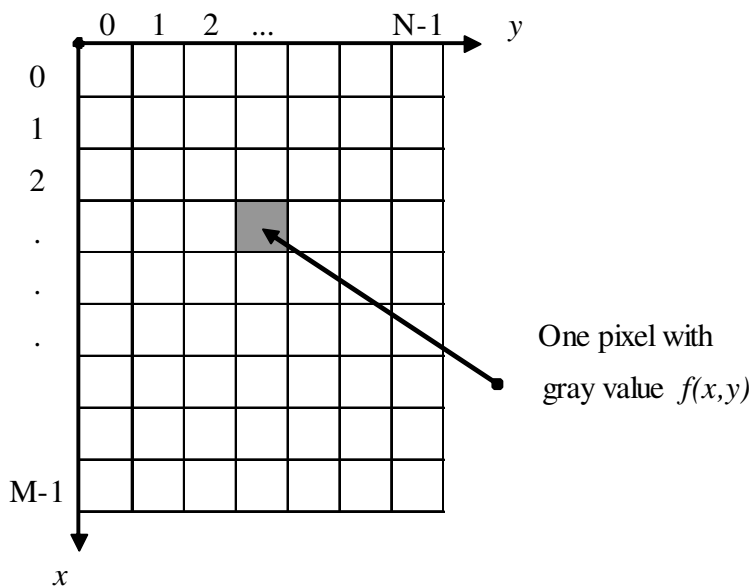


Figure 5: Coordinate conventions used in many image processing books as well as in Matlab 6.5.

Typical numbers of rows, symbol M : 256, 512, 1024, 1030; of columns, symbol N : 256, 512, 1024, 1300 respectively and for gray values, symbol f : 2^k , $k = 1, 8, 16, 32$ -bit representation [24].

On the other hand, color images are formed by a combination of individual monochromatic 2-D images. In the RGB color system, a color image consists of three individual component images: red, green and blue. Therefore, many of the techniques developed for monochrome images can be extended to color images by processing the three component images individually [6].

1.3. Objectives and motivation

As it has been stated in the introductory section, this work was motivated by the need of objectiveness in cell culture observation. The main idea was to take images sub-confluent and confluent endothelial and fibroblast cell cultures, analyze them by using various methods and to propose the method which would deliver the best classification results. All together there were four cell classes to be identified:

- endothelial sub-confluent cells;
- fibroblast sub-confluent cells;
- fibroblast confluent cells; and
- endothelial confluent cells.

The conceptual formulation of this work can be subdivided into three tasks carried out in the framework of the thesis.

First, endothelial and fibroblast cell cultures were seeded and cultivated.

The first task was to try out different microscope parameters in a phase contrast microscopy technique to attempt to obtain clear and sharp images ready for further processing. Different parameters like size of aperture opening, focus, brightness and spatial resolution were systematically changed and their influence on images observed. Rough image pre-processing such as contrast enhancement was achieved with a proper set of microscope parameters. Further image pre-processing was accomplished with Zeiss' image software. Furthermore, a set of images for further analysis had been taken. Accordingly, a standardized procedure of image acquisition for further image processing was attempted to be proposed.

The second task was to review in the literature for the most appropriate methods to analyze images of interest. Different methods were identified as appropriate to analyze textured images of endothelial and fibroblast cell cultures. The first intention was to separate sub-confluent images

from confluent images and trying to further classify the images as fibroblast or endothelial images. Image analysis was carried out with different standard mathematical approaches, such as:

- Spatial statistical analysis based on gray values of each image of the four classes;
- Edge detection and the comparison of edge density, edge shape and length of the four image classes;
- Texture analysis using Haralick texture coefficients, where the image was searched within a specified distance and angle for each spatial gray value. These values were collected in so-called co-occurrence matrices from which Haralick texture coefficients were calculated and compared among images of the four image classes;
- Spectral analysis, where Fourier coefficients were compared for both confluent cell cultures;
- Local spatial spectral analysis using Gabor analytical functions, where Gabor coefficients were compared. (Gabor filtering was applied on both confluent cell cultures) and;
- Cellenger® (section 2.2.2.3.) was the last tool utilized to classify both confluent cell cultures. Images of both classes were first segmented into small object primitives of which shape features were compared to both cell types. Program Cellenger® was also proficient in separating uniform background from cells.

The third task was to summarize the results obtained from the methods mentioned above. With the help of the Fisher's criterion (optimization function), a combination of the smallest possible subset of features that maximizes the classification success rate was searched for. Accordingly, a method was proposed which delivered the most discriminatory subset of features for both types of confluent cell cultures. The proposed analytical method is meant to be integrated into an automatic recognition unit used to classify different cell types cultivated in the biological laboratory at ZIMT. Thus, the three tasks complete the first step of realizing an idea of fingerprint based recognition of different cell types.

2. MATERIALS AND METHODS

2.1. Image acquisition

2.1.1. BASICS OF THE SEEDING TECHNIQUE

Most cultures are propagated as a monolayer, anchored to a glass or plastic substrate. The choice of medium where cells are seeded is still often empirical [4].

In the case of fibroblast cell culture the chosen medium is as detailed in [34] and the medium for endothelial cell culture is the endothelial cell growth medium.

For subculturing a monolayer culture is dissociated by trypsinization, diluted by culture medium and transferred to a new vessel (suspension cultures only need to be diluted). This is best done by rinsing the monolayer with PBS (phosphate buffered saline) to wash out the medium which prohibits the action of trypsin. Afterwards, about 2 ml of trypsin is added for about 60 seconds. Cells are then resuspended in medium, counted and reseeded.

The nature of the substrate to which cells are anchored is determined largely by the type of cell and the project application. Polystyrene is an almost universal substrate [4]. Culture vessels in which the cells were cultivated are optically transparent, with 25 square centimeters of surface area and containing approximately 5 ml of medium. When cells were being subcultured, approximately 2000 cells per square centimeter were being reseeded and placed into a special cell culture incubator with constant temperature of 37 °C, CO₂ atmosphere of 5% and humidity of 100%. After about 2-24 hours in the incubator the cells adhered to the substrate surface and a period of exponential growth began. After about two days, cells were grown enough to provide sufficient quality images of a sub-confluent cell culture. Allowing the cell culture to become

MATERIALS AND METHODS

confluent, the cells enter a period of reduced or no growth. About six days was needed to reach a confluent cell culture.

Cells used for this work were obtained from individual cell lines after being passage from their primary cultures. Fibroblast cells were from the third passage from primary stem cells and endothelial cells were from the second passage derived from the defrosted sixth passage.

2.1.2. INTRODUCTION TO MICROSCOPY

In general light microscopy is divided into two groups: transmitted and reflected – light microscopy.

Transmitted light microscope (inverse microscope) was developed for the examination of living cells in culture vessels.

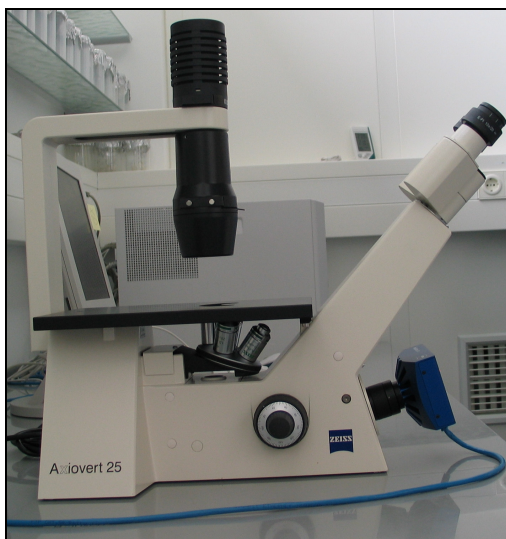


Figure 6: Example of a transmitted light microscope. Note that the objectives are under specimen table and the light comes from below the specimen table in order to observe a cell culture in the vessel. The CCD camera (the blue object located on the right side) is connected to a PC.



Figure 7: Example of a reflected light microscope. Note that the objectives are above the specimen table and light comes from above the specimen table.

Reflected light microscope is on the other hand used for examination of structures of metal samples, the surface of ceramics or any specimen brought to an object glass.

Due to different problem definitions and characteristics of objects of interest, many microscope techniques are used in biology. For example, in fluorescence microscopy the specimens are treated with special reagents (fluorescence dyes). The specimen's individual molecules are able to absorb light for an extremely short time and then to reemit the absorbed light. However, the reemitted light features a wavelength which is viewed only through a barrier filter.

In the present work, a transmitted light microscope Carl Zeiss Axiovert25 using the phase contrast technique was used. The contrast technique will be explained in detail.

2.1.2.1. PHASE CONTRAST IN TRANSMITTED LIGHT

In practical microscopy, not always nicely stained samples which are easy to view in a simple bright field are to be examined. Unstained samples, such as living cell cultures as it was in the case of the present work, absorb practically no light and are barely visible in a bright field. The phase contrast technique allows optical effects not visible to the naked eye- to be translated into intensity changes that can be sensed by the eye.

The contrast technique described by Dutchman Frits Zernike in 1934, not only earned its discoverer the Nobel prize for physics, but also revolutionized biomedical basic research of living i.e. unstained cells.

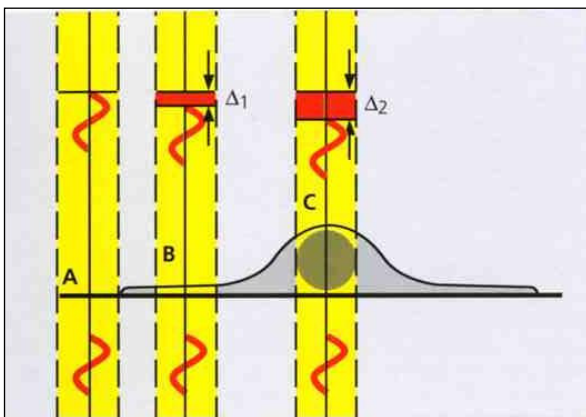
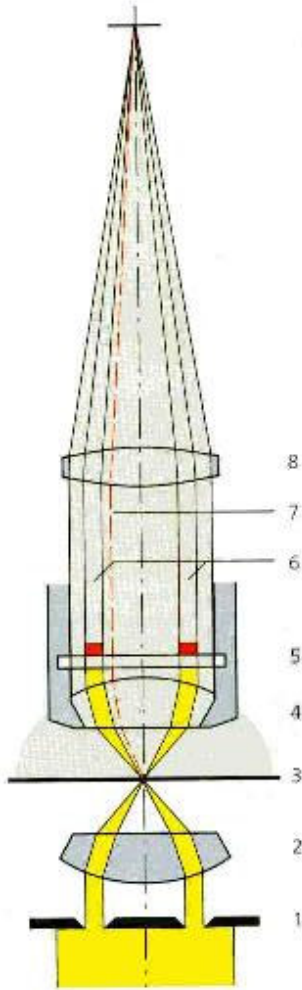


Figure 8: Phase contrast is ideal for thin unstained objects, for example culture cells, which are approx. 5 to 10 μm thick and which barely absorb any light in the visible part of the spectrum. The eye can scarcely see the objects in bright fields and dark fields. However, very small differences exist between the refractive indices of the cells and the surrounding liquid solutions (A) and within the cells between the cytoplasm (B) and the cell nucleus (C) within the cell. Figure source:[12].



Phase contrast makes these tiny differences visible by the use of optical devices – i.e. it translates them into differences in intensity. The optical effect used consists of a shift of phase in plasma or water the light waves are shifted by small degrees, since these media have slightly different refractive indices. The higher the refractive index of a medium, the smaller the velocity of light in the medium. As a result, a light wave which has passed through a cell nucleus lags behind the light waves which only had to pass through water. The amount of lag is called phase shift. Before their entry into the sample, the waves are still in phase, but this is no longer the case when they have passed through the various materials.

The human eye cannot see these phase shifts in the microscope image. It can only distinguish between different intensities and colors. Therefore, the phase contrast technique uses optical tricks to translate phase shifts into gray values.

Figure 9: Phase stop (1) illuminates the sample (3) via condenser optics (2). The entire light enters the objective (4) and an image of the phase stop (1) is created in the objective pupil (5). A “phase ring” is attached to the objective pupil (5) and it does two things: it attenuates the pronounced bright light coming from the phase stop of the condenser, and secondly, it adds a constant phase shift to this light. If a specimen contains objects such as cells and their nuclei, they guide the light from the direct ray to new paths (7). All the partial rays are fused to form the intermediate image (9) by the tube lens (8). Figure source: [12].

The phase contrast technique requires special objectives which are equipped with a phase ring near the pupil. A phase ring is attached to the objective pupil and does two things: it attenuates the pronounced bright light coming from the phase stop of the condenser and secondly it adds a constant phase shift to incoming light.

Figure 10: Image of a phase ring. Figure source: [12].



2.1.3. IMAGE SHOTS

All images used for the purpose of the thesis were taken with a CCD camera (Zeiss' AxioCam digital camera) under magnification of 100 on the microscope. The magnification factor is computed through the following equation:

$$M_{\text{Microscope}} = M_{\text{Objective}} \times M_{\text{Ocular}}$$

where M is the magnification factor. Thus, the overall magnification is 100× (10×10).

For camera viewing, the entire light is redirected from the tube lens to the camera adapter by a beam splitter.

Images' quality was enhanced with Zeiss' AxioVision software module. Images were captured in resolution of 1030 × 1300 pixels and brightness ranging from 50 to 150 ms (exposure time). Afterwards they were saved in '.jpeg' format. Images were classified into four classes:

- endothelial sub-confluent;
- endothelial confluent;
- fibroblast sub-confluent; and
- fibroblast confluent.

In both cell cultures, confluence was measured with Cellenger® (section 2.2.2.3.). For cultures of the sub-confluent classes, the confluence rate is approximately 57%. See definition of confluence in section 1.1.1. For cultures of the confluent classes, the confluence rate is approximately 100 %. Ten images of each class (40 images altogether) were shot. They were shot in the same brightness range of 50-150 ms at an interval of 10 ms of exposure time. Some of the images were shot with the same sample but at different brightness levels. Different increments in brightness are used in order to exclude brightness dependence in the classifying methods.

2.2. Image analysis

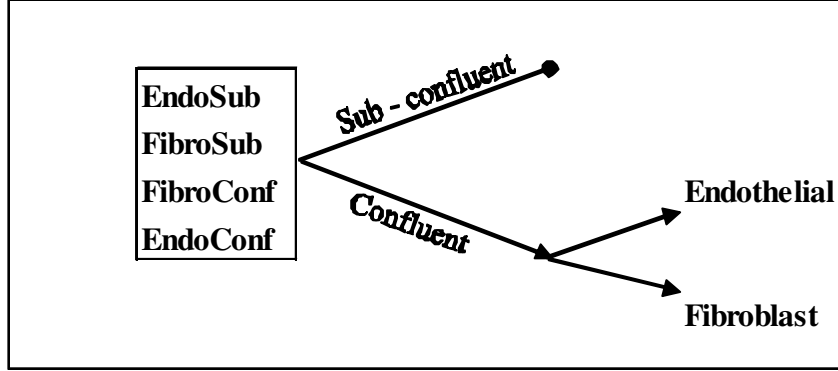


Figure 11: The first objective is to analyze images due to their confluence character. Images are separated into images of sub-confluent cell cultures and images of confluent cell cultures. Furthermore, different algorithms are applied to separate confluent endothelial cell culture images and confluent fibroblast cell culture images. The measure of separability is estimated with Fisher criterion.

2.2.1. SEPARATION OF SUB-CONFLUENT AND CONFLUENT IMAGES

2.2.1.1. EDGE DETECTION

An edge within an image is defined as a set of connected pixels that lie on the boundary between two regions. Edge detection is by far the most common approach for detecting meaningful discontinuities in intensity values. Discontinuities in intensity values are detected by first- and second- order derivatives. In the present image analysis only first-order based derivatives were used to detect edges.

The first-order derivative is the gradient of a 2-D function $f(x,y)$, and is defined as the vector:

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2.1)$$

The magnitude of the vector is:

$$\begin{aligned} \nabla f &= \text{mag}(\nabla f) = [G_x^2 + G_y^2]^{1/2} \\ &= [(\partial f / \partial x)^2 + (\partial f / \partial y)^2]^{1/2} \end{aligned} \quad (2.2)$$

The gradient vector points in the direction of the maximum rate of change of f at coordinates (x,y) . The angle at which this maximum rate of change occurs is:

$$\alpha(x, y) = \tan^{-1} \left(\frac{G_x}{G_y} \right) \quad (2.3)$$

The **Sobel** edge detector was used in the present project. The Sobel uses the following masks to approximate digitally the first derivatives G_x and G_y .

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

Image neighborhood

-1	-2	-1
0	0	0
1	2	1

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)$$

-1	0	1
-2	0	2
-1	0	1

$$G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

The gradient of the center point z_5 in a 3×3 neighborhood is computed by the Sobel edge detector as follows:

$$\begin{aligned} g &= [G_x^2 + G_y^2]^{1/2} \\ &= \left\{ [(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)]^2 + [(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)]^2 \right\}^{1/2} \end{aligned} \quad (2.4)$$

A pixel at location (x,y) is an edge pixel if $g \geq T$, where T is a specified threshold.

2.2.1.2. TEXTURE RECOGNITION

Image texture has two basic properties: the first is concerned with the gray level primitives and the second with the spatial organization of the gray level primitives (pixels). The combination of the two properties yields the texture characteristics of fineness, coarseness, smoothness, granulation, randomness, orientation, periodicity and others. The purpose of texture characterization is to define a set of measurements or features that will identify the relevant properties of a texture.

In order to describe images of the two cell types, the texture content was quantified. Texture information might be important for the problem of distinguishing sub-confluent images from confluent ones. Since each cell type has a unique texture, the purpose of quantification is to detect and mathematically illustrate differences in texture content. In the following two chapters methods for mathematically describing texture through statistical and spectral measures are illustrated.

2.2.1.2.1. Statistical approach

A frequently used approach to texture analysis is based on statistical properties of the intensity histogram [5]. One class of such measures is based on statistical moments. The expression for the n -th moment about the mean is given by:

$$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i),$$

$$z = \{z_i\}$$
(2.5)

where z is a random variable indicating intensity, $p(z)$ is the histogram of the intensity levels in a region, L is the number of possible intensity levels, and m is the mean (average) intensity. The equation to compute m is:

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \quad (2.6)$$

STANDARD DEVIATION is a measure of image contrast. Standard deviation, derivable from the 2nd moment is, defined as:

$$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2} \quad (2.7)$$

Apart from n -th moment functions, the functions of uniformity, entropy, and smoothness are also used for statistical description [5].

UNIFORMITY is a measure of the pixel intensity uniformity on an image. Uniformity is maximized when all pixel gray levels of an image are equal. Uniformity is defined as:

$$U = \sum_{i=0}^{L-1} p^2(z_i) \quad (2.8)$$

ENTROPY is a measure of randomness. Entropy is higher when pixel intensities of an image are more random. Entropy is defined as:

$$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \quad (2.9)$$

SMOOTHNESS is a measure of the relative smoothness of the pixel intensities in a region. Smoothness equals 0 for a region of constant intensities and approaches 1 for a region with large intensity excursions. The variance used in this measure is normalized from 0 to 1 in order to be comparable with other measures.

$$R = 1 - 1/(1 + \sigma^2) \quad (2.10)$$

Evaluated images are first normalized to zero mean and unit variance and then analyzed. See results in section 3.1.2 for comparison of these measures of four different image types.

2.2.1.2.2. Haralick texture coefficients

Texture measurements derived from gray-value histograms using first-order statistics do not carry information regarding the relative position of pixels with respect to one another. To overcome this disadvantage, Haralick proposed an analysis process which considers the positions of pixels with equal or nearly equal intensity values in addition to the distribution of intensities [7]. The motivation for choosing the Haralick texture coefficients was examples from literature, where features based on co-occurrence matrices delivered the best results among different analytic techniques in remote sensing, terrain classifications and various biomedical applications [19].

Spatial gray level co-occurrence estimates are related to the second-order statistics.

Haralick suggested the use of gray level co-occurrence matrices (GLCM), which have become one of the best known and widely used texture features. Entries in the GLCM P_d of size $G \times G$ for a displacement vector $d(\Delta x, \Delta y)$ are defined as follows. $P_d(i, j)$ equals the number of times a pixel with a gray level i is for the displacement vector d apart from a pixel with a gray level j .

To illustrate, the following image containing 3 different gray values is considered:

$$\begin{matrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{matrix}$$

The 3×3 P_d co-occurrence matrix for the image for a displacement vector of $d = (0, 1)$ is:

$$P_d = \begin{bmatrix} 4 & 0 & 2 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The size of P_d is determined by the number of distinct gray levels in the input image. The example has 3 distinct gray levels (0, 1, and 2). The value of $P_d(i, j)$ is the number of occurrences an element with gray level represented by i is offset by (0,1) from an element with gray level

represented by j . For example, $P_d(0,0)$ is 4 because there are four pixel pairs with both pixels of intensity 0 (i and $j=0$ represents intensity equaling 0) that are offset from each other by the displacement vector $d(0,1)$. Examples of P_d - matrices for the other displacement vectors are shown below:

$d(-1,1)$	$\begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$
$d(-1,0)$	$\begin{bmatrix} 4 & 2 & 0 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$
$d(-1,-1)$	$\begin{bmatrix} 2 & 2 & 0 \\ 0 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}$

Displacement operator d , expressed in Cartesian coordinates, can also be expressed as polar coordinates with a displacement parameter D and an angle Θ .

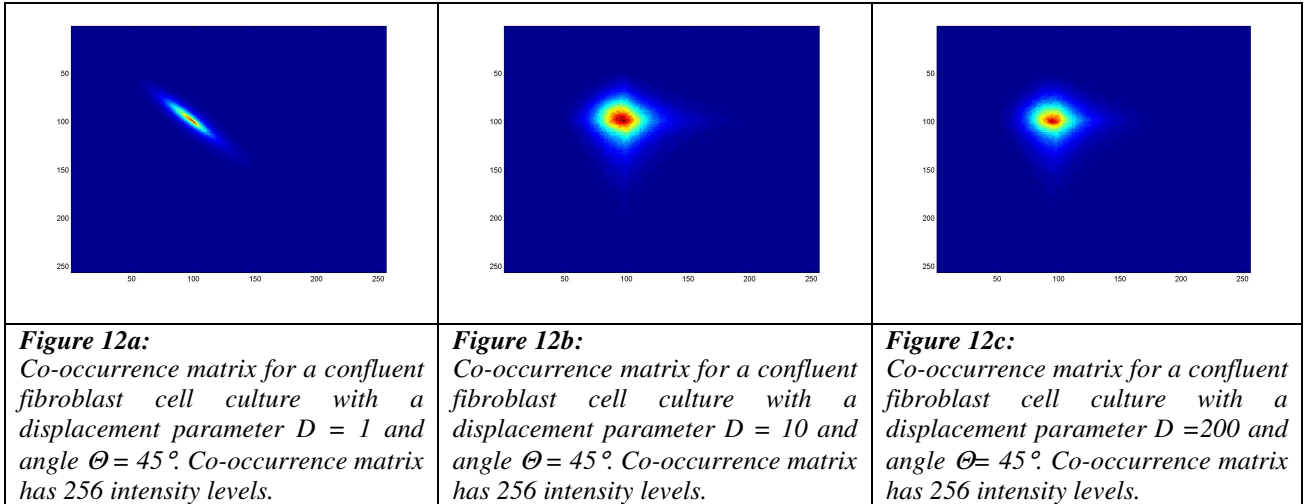
- $(0,D)$ means displacement parameter D in direction of 0° ;
- $(-D,D)$ means displacement parameter D in direction of 45° ;
- $(-D,0)$ means displacement parameter D in direction of 90° ;
- and $(-D,-D)$ means displacement parameter D in direction of 135° .

The right choice of these two parameters depends on the resolution of scale at which the texture is viewed and which provides the best discrimination. This can be achieved by varying displacement parameter D and an angle Θ . As a co-occurrence matrix depends on D and Θ , the presence of a given texture pattern may be detected by choosing an appropriate pair of the parameters.

From the visual inspection of co-occurrence matrices generated with a displacement parameters $D = 1, 50, 70, 100, 200, 300, 500$ and angles $\Theta = 0^\circ, 45^\circ, 90^\circ$ and 135° , respectively; it was

MATERIALS AND METHODS

observed, that matrices with a displacement parameter over 100 do not show much difference. From spectral analysis in section 3.2.1 a peak in frequency spectrum $S(r)$ was detected at approximately 10 units. As a result, an analysis of texture coefficients with a displacement parameter of 10 in all four directions was implemented. Another reason for using a displacement parameter of 10 is that textures in images of confluent cell cultures showed an approximate 10 pixel wide pattern.



Thus, co-occurrence matrices were generated with a displacement parameter value of 10 and angles of 0, 45, 90 and 135°.

Features from co-occurrence matrices are extracted in several ways. Approximately twenty features that can be extracted from co-occurrence matrices appear in the literature [23], but only four of them were used in the present work. The most common method (used in present calculations) to extract features is to apply a weighted function to each element of the co-occurrence matrix and sum the weighted element values. The weighting applied to each element is based on a feature weighting function. By varying the function, different texture information can be extracted from the matrix. The weighting functions fall into two general classes:

- Type 1: Co-occurrence matrix descriptors based on the element's value, and
- Type 2: Co-occurrence matrix descriptors based on the spatial position of the element.

Additional notation for further understanding:

$P(i,j)$ is the (i,j) th element of a normalized co-occurrence matrix P_d . Element values are bounded by $[0,1]$ and the sum of all element values equates to 1.

$$\begin{aligned}
 P_x(i) &= \sum_j P(i, j); & P_y(j) &= \sum_i P(i, j); \\
 \mu_x &= \sum_i i \sum_j P(i, j) = \sum_i iP_x(i); & \mu_y &= \sum_j j \sum_i P(i, j) = \sum_j jP_y(j); \\
 \sigma_x &= \sum_i (i - \mu_x)^2 \sum_j P(i, j). & \sigma_y &= \sum_j (j - \mu_y)^2 \sum_i P(i, j).
 \end{aligned}$$

Type 1 - Weighting dependent on element's value of $P(i,j)$, i.e. $W(i,j) = F\{P(i,j)\}$:

$$\mathbf{Energy} = \sum_i \sum_j P^2(i, j) \tag{2.11}$$

Energy is 1 for a constant image.

Type 2 – Weighting dependent on spatial position, i.e. $W(i,j) = F\{i,j\}$:

$$\mathbf{Correlation} = \sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y)}{\sqrt{(\sigma_x \sigma_y)}} P(i, j) \tag{2.12}$$

Correlation is a statistical measure of the correlation of a pixel to its neighbor over the whole image. Range is $[-1, 1]$. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN for a constant image.

$$\mathbf{Contrast} = \sum_i \sum_j |i - j|^2 P(i, j) \tag{2.13}$$

Contrast is a measure of the intensity contrast between a pixel and its neighbor over the whole image. Contrast is 0 for a constant image.

$$\mathbf{Homogeneity} = \sum_i \sum_j \frac{P(i, j)}{(1 + |i - j|)} \quad (2.14)$$

Homogeneity is a measure of the distribution of elements in the co-occurrence matrix to its diagonal. Homogeneity is 1 for a diagonal co-occurrence matrix.

2.2.2. SEPARATION OF CONFLUENT ENDOTHELIAL AND FIBROBLAST IMAGES

2.2.2.1. SPECTRAL APPROACH: FOURIER SPECTRUM

Spectral measures of texture are based on the Fourier transform, which is ideally suited for describing the directionality of periodic or almost periodic 2 - D patterns in an image. These global texture patterns are easily recognizable as concentrations of high energy bursts in the spectrum. Thus, spectral texture analysis is useful for discriminating between periodic and non-periodic texture patterns. An extended introduction to the Fourier transform and the frequency domain can be found in [6, 8 and 20].

For the present image analysis, interpretation of spectrum features as a simplified expression of the spectrum in polar coordinates was used. The expression is described as the function $S(r, \theta)$, where S is the spectrum function and r and θ are the polar coordinates. Fourier spectrum is useful for texture description because:

- Prominent peaks in the spectrum give the principal direction of the texture patterns and
- The location of the peaks in the frequency plane gives the fundamental spatial period of the patterns.[6]

For each direction of θ , $S(r, \theta)$ may be considered as a 1-D function $S_{\theta}(r)$ and for each frequency r , $S_r(\theta)$ may be considered as a 1-D function. The descriptor typically used for the purpose to characterize the behavior of a texture is the location of the highest values in both of the 1-D functions $S(r)$ and $S(\theta)$.

2.2.2.2. GABOR FILTERING

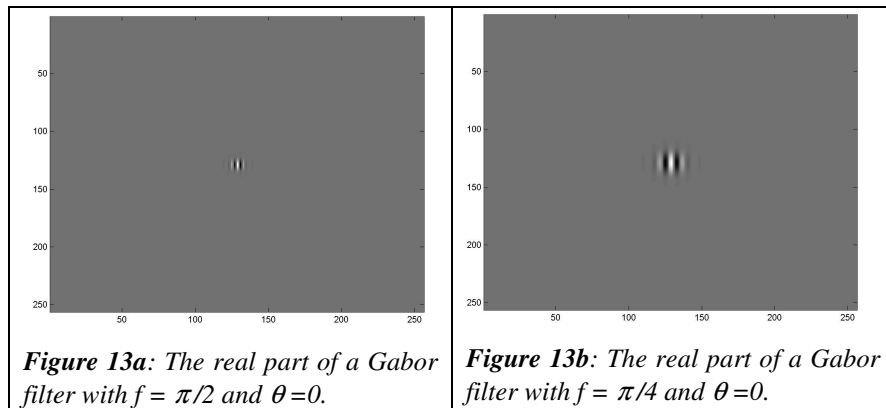
Gabor filters are a widely used feature extraction methods in image analysis. In 1946, Dennis Gabor introduced the elementary functions, which have the smallest joint uncertainty in time and frequency [26]. In image processing, the most attractive and active application for Gabor filters is

texture segmentation [9 and 10]. Besides textures, Gabor filters are used in edge detection, line segmentation, and shape recognition [30].

A two-dimensional complex Gabor function consists of two directional sinusoids of a certain frequency modulated by a Gaussian curve. The two sinusoids correspond to the real and the imaginary parts of the Gabor function. A Gabor function $g(f, \theta, x, y)$ centered at the origin is defined as:

$$g(f, \theta, x, y) = e^{j(x \cdot f \cdot \cos \theta + y \cdot f \cdot \sin \theta) - \frac{f^2(x^2 + y^2)}{2\sigma^2}} \quad (2.15)$$

where x and y are the spatial coordinates. The parameter f denotes the central frequency of the pass band. The central frequency ranges from 0 (DC component) to π . θ is the spatial orientation, and σ determines the bandwidth of the filter. The first component of Eq. 2.15 is the imaginary exponent, which represents the pair of sinusoids having spatial orientation θ and frequency f . The second component is the modulating Gaussian curve. The deviation of the Gaussian curve is normalized with respect to f . The effect of the normalization is that all filters with varying frequencies include the same number of sinusoidal waves in the spatial domain when σ is constant. Therefore, decreasing the frequency f increases the size of the spatial filter.

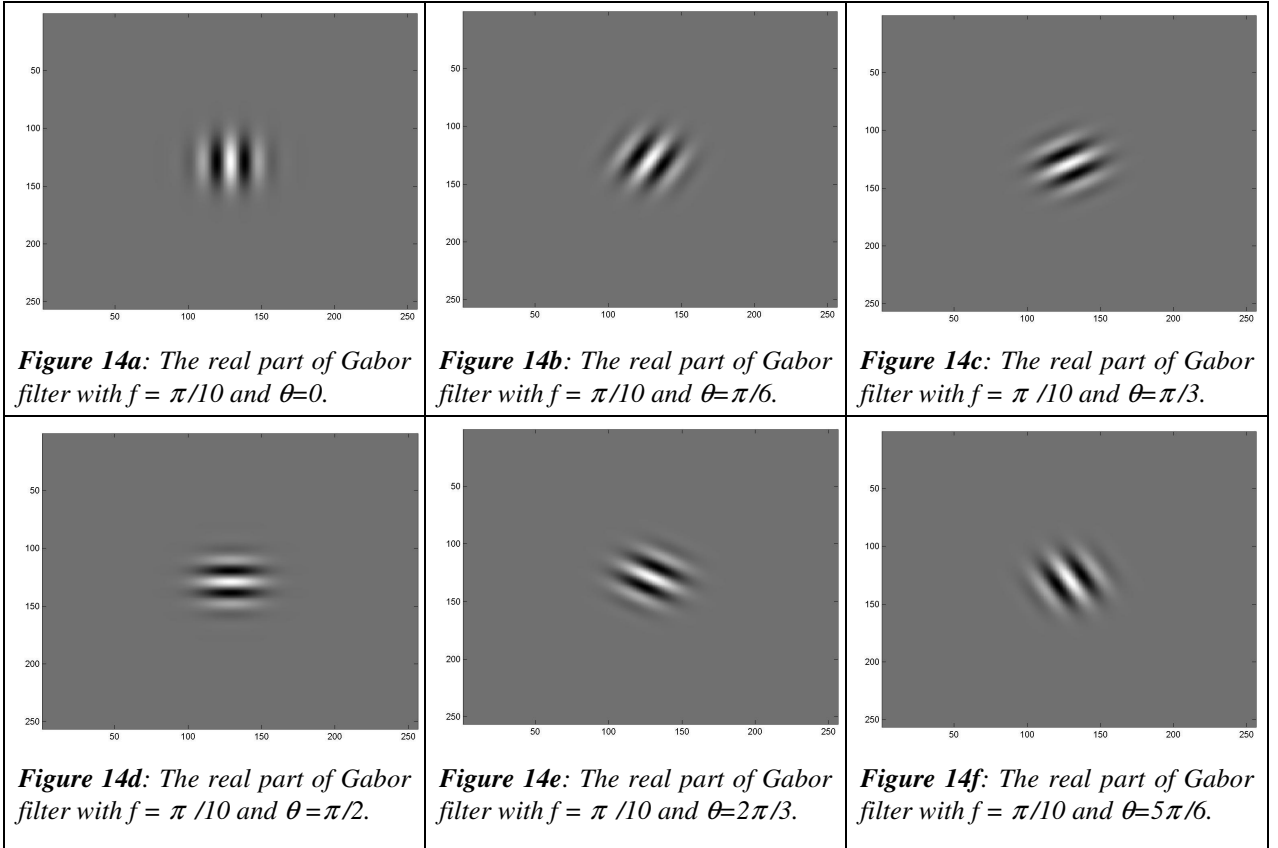


Gabor filters are typically used as filter banks containing filters in several frequencies and orientations. The response of a Gabor filter bank to an image is a set of images that are each

filtered with one of the filters. For this project, a bank of Gabor filters with only one frequency

$f = \frac{\pi}{10}$ and six different orientations $\theta = 0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}$ was used.

Selection of frequency f was based on the spectral analysis of the images (section 3.2.1.). A spike at about 10 units was seen on the frequency spectrum $S(r)$. Thus, it was assumed that this frequency is the optimal frequency for the project's analysis with Gabor filters. The following figures show the real part in spatial domain of the Gabor filters used.



In order to obtain a response from a Gabor filter, an input image $f(x,y)$ (x and y are from the set of image points) is convolved with a two-dimensional Gabor function $g(x,y)$ as follows:

$$r_{f,\theta}(x,y) = f(x,y) * g_{f,\theta}(x,y) = \iint_{x,y} f(\xi,\eta) g_{f,\theta}(x-\xi, y-\eta) d\xi d\eta \quad (2.16)$$

Each image was normalized to zero mean and unit variance before it was filtered.

Typically, for Gabor filtering, a multi-channel filtering scheme is used. For the project, an image is filtered with a set of Gabor filters with a preferred spatial frequency f and six filter orientations θ . The obtained features form a feature vector field, which is later used. The application of the filter bank resulted in a 6-dimensional feature vector at each point of the image (i.e., a 6-dimensional vector field for the entire image).

The filter results can be combined into a single quantity (called Gabor energy) and is related to a model of complex cells in the visual cortex [13]. Gabor energy is defined as:

$$R(f, \theta) = \sum_x \sum_y |r_{f, \theta}(x, y)| \quad (2.17)$$

where x and y are the spatial coordinates of the response. Consequently, the Gabor features form a Gabor feature matrix:

$$R = (R(f, \theta_0) \quad R(f, \theta_1) \quad \dots \quad R(f, \theta_n))$$

For each of the twenty images (ten images of confluent endothelial cell culture and ten of confluent fibroblast cell culture) matrices with Gabor features were obtained. The feature vectors computed for different points of a texture image form clusters in the multi-dimensional feature space. The larger the distance between two clusters, the better the discrimination properties are of the Gabor features. Distances between clusters of the two image types were evaluated with the Fishers criterion.

2.2.2.3. *CELLENGER® APPROACH*

2.2.2.3.1. *Introduction to Cellenger®*

Human visual perception is qualitative and objective. Essential information that the human eye uses in order to understand images is contained both in meaningful image regions and in their neighborhood relations. Essential information is almost never contained in single image points.

The image recognition software Cellenger®, developed by the company Definiens AG (settled in Munich- Germany under the leadership of the Nobel Prize winner Prof. Gerd Binnig), is based on handling and modelling complex image processing data.

2.2.2.3.2. Basic concepts

Image object hierarchy

Similar to human vision, Cellenger's concept of image understanding is based on a proper segmentation of a certain visual image content of interest against other visual image contents. Segmentation is realized by cutting the image into zoned partial areas of differing characteristics. The segments are called image objects. The pixels of the associated region are linked to the image object with an is-part-of link object. Two image objects are neighbored with each other if their associated regions lie within a selected pixel neighbourhood.

An image is divided by image objects. All image objects of such division belong to an image object level. Any output of an arbitrary segmentation algorithm provided in Cellenger® can be interpreted as a valid image object level. The simplest object level is generated through a partition of an image into image objects consisting of a single pixel.

Furthermore, object levels are structured in an image object hierarchy. Different techniques are used by Cellenger® to construct a valid object level and thus a hierarchical network of image objects in different levels. The different levels represent image information in different spatial resolutions.

Image objects are networked. Therefore, each object knows its relation to its neighbour object (horizontal relation) as well as its super-object and sub-object (vertical relation).

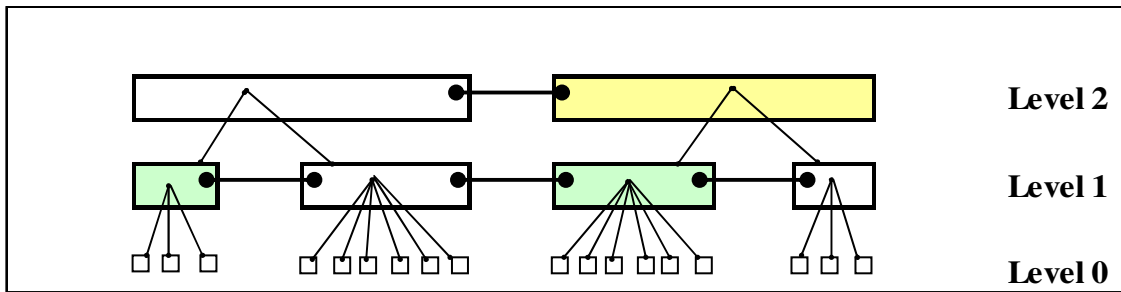


Figure 15: The sequence in which the levels are segmented plays an important role. It makes a difference in determining which level is constructed first. Adjacent sub-objects cannot be merged if they are not sub-objects of the same super-object.

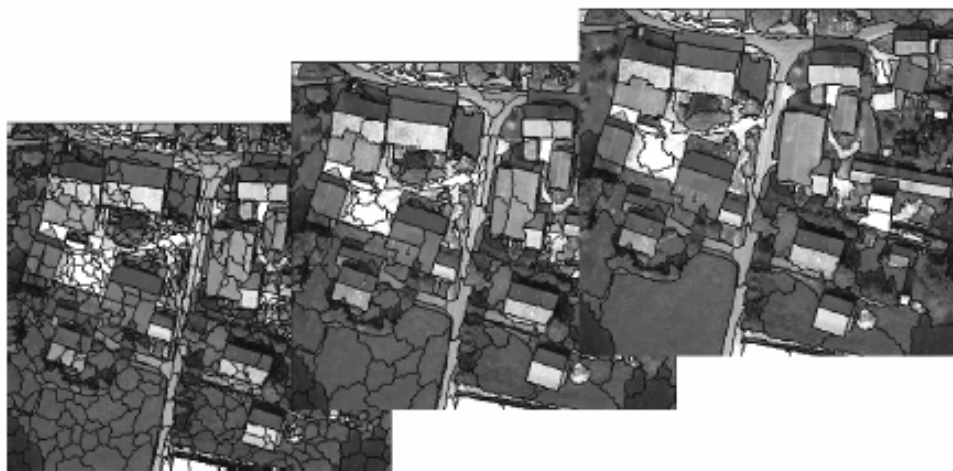


Figure 16: Image objects with a hierarchy of 3 levels. Image objects between levels are connected by hierarchical is-part-of link objects. Figure source: [3].

Features

Each image object can be described with many different features available in Cellenger®. Features are computed by defined algorithms according to the current objects' network situation. Features are divided in two major categories: "Object features" (related to an image object) and "Global features" (related to general information and not connected to an image object). Since regions of the image provide more representative information than single pixels, there is a greater palette of image object features describing color, shape and texture. An important example of such object features is e.g. *Rel. border to neighbouring objects of a given class*.

On the other hand, global features describe objects' network situation in general. For example, information about *Mean value of a given image layer* or *Number of levels in the image object hierarchy* or *Number of objects classified as a given class* can be obtained.

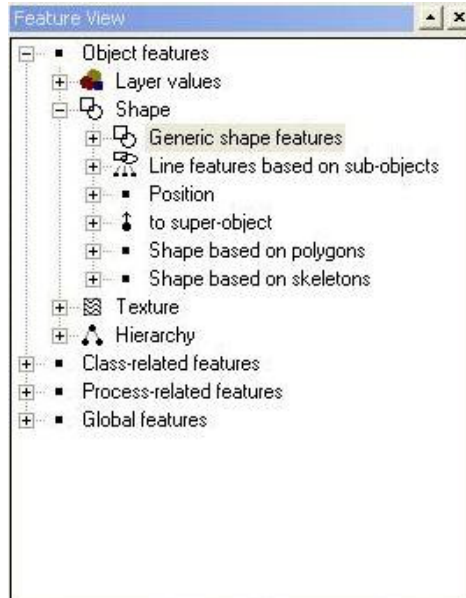
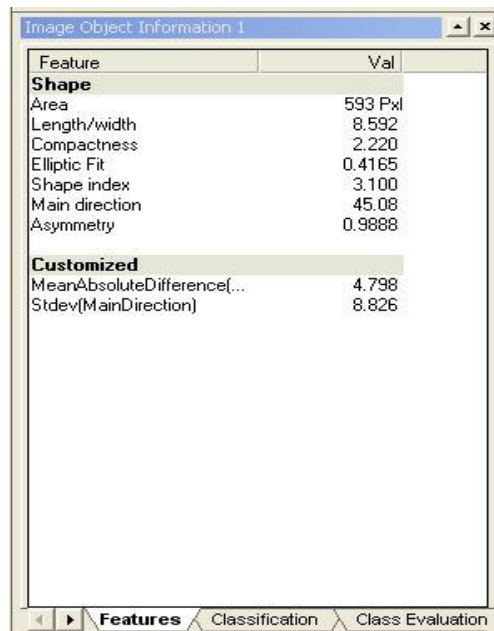


Figure 17: Feature options in Cellenger.

Both types of features “Object features” and “Global features” are accessible anytime in a special Image object information window.



Feature	Val
Shape	
Area	593 Pxl
Length/width	8.592
Compactness	2.220
Elliptic Fit	0.4165
Shape index	3.100
Main direction	45.08
Asymmetry	0.9888
Customized	
MeanAbsoluteDifference(...)	4.798
Stdev(MainDirection)	8.826

Figure 18: A list of features for a selected object.

As seen in *Figure 18* the features: *Length/width*, *Compactness*, *Elliptic fit*, *Shape index*, *Density*, *Main direction*, *Asymmetry*, *Mean Absolute Difference of Main Direction to the neighbour objects* and *Standard Deviation of Main Direction to the neighbour objects* were used in the analysis of image objects after multi-resolution segmentation (detailed in section 2.2.2.3.3.) in order to compare and to recognize significant differences between image objects of confluent endothelial and fibroblast cell cultures.

Features listed above are cited from [3] and defined as follows:

LENGTH / WIDTH

There are two ways in which Cellenger® computes the *Length / width* ratio. Cellenger® takes the smaller of the two results as the feature value.

The *Length / width* ratio is equal to the ratio of eigenvalues of the covariance matrix, with the larger eigenvalue being the numerator of the fraction.

$$\gamma = \frac{l}{w} = \frac{eig_1(S)}{eig_2(S)} \quad (2.18)$$

The *Length / width* ratio is approximated using the bounding box and therefore equals:

$$\gamma = \frac{l}{w} \approx \frac{a^2 + ((1-f)b)^2}{A}, \quad (2.19)$$
$$f = \frac{A}{ab}$$

where a is the length of a bounding box, b is its width, ab is its area and f is degree of filling. The degree of filling f is the area of A covered by the image object divided by the total area ab of the bounding box.

The feature value range = $\{ \geq 0 \}$.

COMPACTNESS

Compactness is calculated by multiplying together the length l and width w of the corresponding image object and dividing by the number of its inner pixels a .

$$c = \frac{l \cdot w}{a} \quad (2.20)$$

The feature value range = $\{ \geq 0 \}$.

ELLIPTIC FIT

The first step in the calculation of elliptic fit is the creation of an ellipse with the same area as the considered object. In the calculation of the ellipse, the length and width of the object is considered. The area of the image object outside the ellipse is then compared with the area inside the ellipse not filled by the image object. A value of 0 represents “no fit”, while a value of 1 represents a perfectly fitted object.

The feature value range = $\{ 0 : 1 \}$.

SHAPE INDEX

Shape index is mathematically expressed as the border length of an image object e divided by four times the square root of its area. The square root is taken to exclude the dependence on the image object’s size.

$$s = \frac{e}{4 \cdot \sqrt{A}} \quad (2.21)$$

The feature value range = $\{ \geq 1 \}$.

DENSITY

Density d is expressed by the area covered by the image object divided by the approximation of the object’s area radius. Cellenger® uses the following implementation:

$$d = \frac{\sqrt{n}}{1 + \sqrt{\text{Var}(X) + \text{Var}(Y)}} \quad (2.22)$$

where n is the number of pixels forming the image object and the radius is approximated using the covariance matrix.

The feature value range = $\{ \geq 0 \}$.

MAIN DIRECTION

The main direction of an image object is the direction of the eigenvector belonging to the larger of the two eigenvectors derived from the covariance matrix of the spatial distribution of an image object.

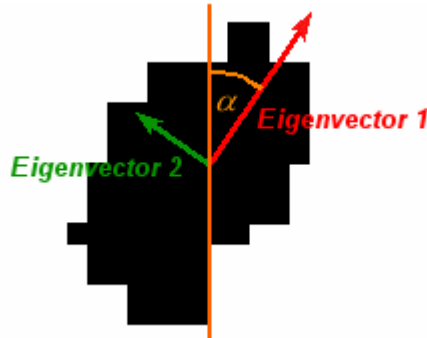


Figure 19: Figure shows the definition of main direction feature. Figure source [3].

The feature value range = $\{ 0 : 180 \text{ degrees} \}$.

ASYMMETRY

Asymmetry of an object increases when the length of the object increases. For an image object, an ellipse is approximated. Asymmetry can be expressed by 1 minus the ratio of the minor axis length n to the major axis length m (see **Figure 20**).

$$\kappa = 1 - \frac{n}{m} \quad (2.23)$$

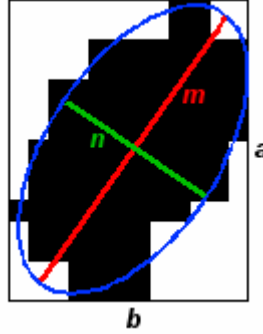


Figure 20: Figure shows the definition of asymmetry. Figure source: [3].

The feature value range = {0 : 1}.

MEAN ABSOLUTE DIFFERENCE OF MAIN DIRECTION

Mean Absolute Difference of Main Direction (MADofMD) is one of the two custom defined features to test the hypothesis that fibroblast cells grow in parallel whereas the growth direction of endothelial cells is indistinctive.

$$MADofMD = \frac{\sum_{j=1}^k |m - m_j|}{k} \tag{2.24}$$

where m stands for the main direction value of the observed image object, m_j represents the main direction value of the j -th neighbour object and k is a number of all neighbour objects.

STANDARD DEVIATION OF MAIN DIRECTION

Standard Deviation of Main Direction (SDofMD) is the next custom defined feature to test significance that fibroblast cells grow in parallel, whereas endothelial cells grow in random directions. The definition of SDofMD is:

$$SDofMD = \frac{\sum_{j=1}^k (m - m_j)^2}{k} \tag{2.25}$$

A low SDofMD is an indication that image objects “grow” in parallel (there does not exist a high deviation in their main direction values).

Classes and classification

Classes describe semantic meaning of image objects. Classes can be combined by inheritance link to inherit class descriptions from a parent class or to combine a group of different classes into a group class. Class definition rules are created via fuzzy logic.

Image objects are classified as a certain class by passing the fuzzy logic threshold which defines class membership. An image object may have any number of classification links. Classification can be performed with any classification algorithm as long as results may be translated into membership values.

2.2.2.3.3. *Multi-resolution segmentation*

A necessary prerequisite for object oriented image processing is successful image segmentation. Segmentation is also one of the main focuses where the images of the two cell types are compared. The segmentation used as a basis for further analysis of image objects or object primitives is multi- resolution segmentation developed and patented by Definiens AG in Munich, Germany. For detailed explanation of the algorithm behind this segmentation, refer to [1].

A short description of important characteristics of other segmentation techniques is given below.

A short review on image segmentation

Segmentation subdivides an image into its constituent regions or objects. The level to which the subdivision is carried out depends on the application. Segmentation should stop when the objects of interest have been isolated. Segmentation of nontrivial images is one of the most difficult tasks in image processing. Segmentation accuracy determines the eventual success or failure of computerized analysis procedures [6].

One of the simplest approaches to segmentation is global thresholding. Another approach involves region growing algorithms, which are clustering pixels, starting the algorithms at a limited number of single seed points. One major problem with this approach is the difficulty in

formulating a stopping rule for the growth of a region [6]. In many operational applications, different types of texture segmentation algorithms where characteristics are extracted from a textured image are used. The characteristics comprise spatial frequencies [2] and co-occurrence matrices [7] to wavelet coefficients [18]. Although texture segmentation leads to excellent and reproducible results for specific applications, texture segmentation is useful for a limited number of textures in image data [1]. The most promising segmentation is watershed segmentation, which involves selecting markers (watersheds) that include a priori knowledge of shape, location, relative distances and texture content. Hence, segmentation by watersheds offers a framework that can make effective use of context knowledge of an image. Thus, watershed segmentation possesses a significant advantage over other segmentation methods [6].

Multi-resolution segmentation developed by Definiens AG, besides the textural information considers also an end size of a segmented image object. The Definiens group wanted to develop a high-quality solution applicable and adaptable to many problems, including textured image data of arbitrary type. The advantage of this method is that it takes into account the scale (size) of objects of interest. Therefore, the scale of resulting image objects is adaptable to fit the scale of the task. Furthermore, structures of similar scale are of comparable quality.

General Concept of multi-resolution segmentation

As with a majority of segmentation techniques, multi-resolution segmentation can be described as a region merging (fusion) technique. The technique starts with each pixel forming one image object or region. A pair of image objects is then merged into one larger object if they satisfy the local homogeneity criterion (if the total “merging cost” is low enough). This step is repeated until the local homogeneity criterion cannot be met. In the simplest case, two single pixels fuse into a new image object of which its gray value is an average of the gray values of the two original pixels.

The merging cost is determined by the weighted sum of the gray value and shape value. The weights accorded the gray value and the shape value are freely adjustable. To illustrate, assume the weight of a gray value is 60 % and the weight of a shape value is 40 %. The sum of the weights always equals 100 %. Through changing the weights, the segmentation algorithm generates whether compact image objects or more fractal shaped ones.

MATERIALS AND METHODS

With local homogeneity criterion, it is assured that only those image objects for which merging costs are locally minimal merge.

The procedure stops when there are no more possible merges. The size of the resulting image object depends on the least degree of fitting value (also termed as the scale parameter and is arbitrary adjustable). A merge with better degree of fitting than the scale parameter is said to fulfil the homogeneity criterion [1].

For the present problem of extremely heterogeneous image data, the gray value weight was set to 100 %, since the shape of image objects was the object of interest after segmentation. In the present case, gray value was the most important criterion to deliver meaningful objects.

Through experience, a scale parameter, which determines end size of a segmented object, was set to value of 20. Hence, the optimal size of image objects was obtained through a compromise between using meaningful image context information and object size. A scale parameter with default value of 10 would deliver objects which might have been too small and thus not useful for comparing their features between cell types.

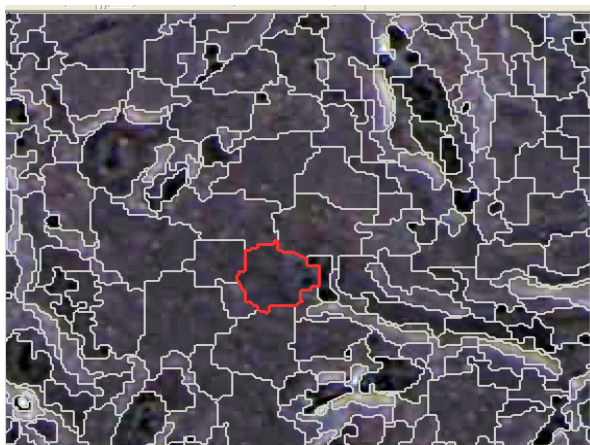


Figure 21: Segmented image objects of a confluent endothelial cell culture after multi-resolution segmentation, scale parameter = 20, gray value weight = 100%.

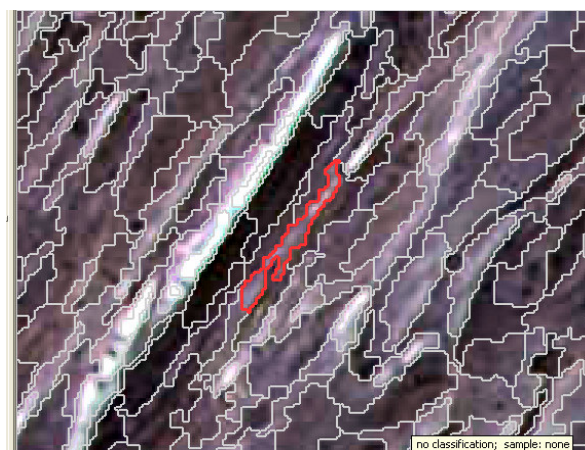


Figure 22: Segmented image objects of a confluent fibroblast cell culture image after multi-resolution segmentation, scale parameter = 20, gray value weight = 100%.

2.3. Statistical pattern recognition

Statistical object description uses elementary numerical descriptions called features, x_1, x_2, \dots, x_n ; in image analysis. The features are derived from object description methods discussed in the section 2.2. The pattern $\mathbf{x} = [x_1, x_2, \dots, x_n]$ (also referred to as pattern vector, or feature vector) that describes an object is a vector of elementary descriptions, and the set of all possible patterns forms the pattern space (also referred to as the feature space). If the elementary descriptions are appropriately chosen, similarity of objects in each class results in the close proximity of their patterns in pattern space. The classes form clusters in the feature space, which can be separated by a discrimination curve (or hyper-surface in a multi-dimensional feature space). If a discrimination hyper-surface exists which separates the feature space such that only objects from one class are in each separated region, the problem is called a recognition task with separable classes. If the discrimination hyper-surfaces are hyper-planes, it is called a linearly separable task.

For statistical evaluation of the project's feature spaces, a Fisher linear discriminant was used. Application of Fisher linear discriminant is one of the simplest and widely used methods to test classification results.

2.3.1. FISHER LINEAR DISCRIMINANT

A cluster comprises a number of similar objects collected or grouped together. According to Jain A. et al.: "a cluster is an aggregation of points in the test space such the distance between any two points in the cluster is less than the distance between any point in the cluster and any point not in it." [11]. Attention was focused on Fisher linear discriminant with which maximization of Fisher criterion decides which features in the feature space best describe the pattern of a specific class. A larger Fisher criterion implies a larger weighted distance between two clusters. Thus, the Fisher criterion expresses the distance between two clusters relative to their compactness in one single quantity. The Fisher criterion is defined as:

$$\tau(w) = \frac{w^T S_b w}{w^T S_w w} \quad (2.26)$$

where S_b is the between-class scatter matrix, and S_w is the within-class scatter matrix defined by:

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \quad (2.27)$$

and

$$S_w = S_1 + S_2$$

where

$$S_i = \frac{1}{N_i - 1} \sum_{x \in C_i} (x - m_i)(x - m_i)^T \quad (2.28)$$

where x_i is a pattern vector belonging to class C_i and m_i is a mean pattern vector of class C_i .

In the case of the Fisher linear discriminant, the parameter vector w of the linear discriminant function is determined to maximize the class separability criterion.

$$w = \arg \max_{w'} \tau(w') = \arg \max_{w'} \frac{w'^T S_b w'}{w'^T S_w w'} \quad (2.29)$$

The classical solution of the Eq. 2.29 is by using the matrix inversion:

$$w = S_w^{-1}(m_1 - m_2). \quad (2.30)$$

The performance of the various image analysis methods was evaluated according to the Fishers criterion by looking at the separability of images of confluent cell cultures from images of sub-confluent cell cultures. Furthermore, the separability of images of confluent endothelial cell cultures from images of confluent fibroblast cell cultures was evaluated according to the Fishers criterion. As previously mentioned, a larger Fishers criterion leads to better separability between the observed classes.

3. RESULTS

The results are divided in two major parts. The first part (section 3.1.) includes graphics and charts of data attempting to depict the division of images of sub-confluent cell cultures from images of confluent cell cultures. Edge detection (section 2.2.1.1.), Haralick texture coefficients (section 2.2.1.2.2.) and previously described statistical approach (section 2.2.1.2.1.) were used to evaluate the confluence criteria.

The results in the second part (section 3.2.) refer to the analysis of images of confluent endothelial cell cultures and confluent fibroblast cell cultures. The objective was to build two clusters of features that characterize the two cell types with a negligibly small probability of misclassification. For this purpose, spectral analysis (section 2.2.2.1.), Gabor filtering (section 2.2.2.2.) and the Cellenger® approach (section 2.2.2.3.) were applied.

All statistics were plotted using the Statistics toolbox of Matlab® version 6.5. Data was usually not in a normal distribution. Therefore, a nonparametric Kruskal – Wallis test was used to verify the significant difference between mean values of the features. The test is based on an analysis of variance using the ranks of the data values.

3.1. Separation of sub-confluent and confluent images

3.1.1. EDGE DETECTION

According to the section 2.2.1.1., a Sobel filter was used to detect edge points which were subsequently connected into edges. After filtering an image with a Sobel mask, a threshold of 0.22 was set in order to obtain a binary image.

Afterwards, a stability test based on different thresholds was carried out. It was discovered that a selected threshold does not influence the results which follow from filtered images (see *Figure 25*).

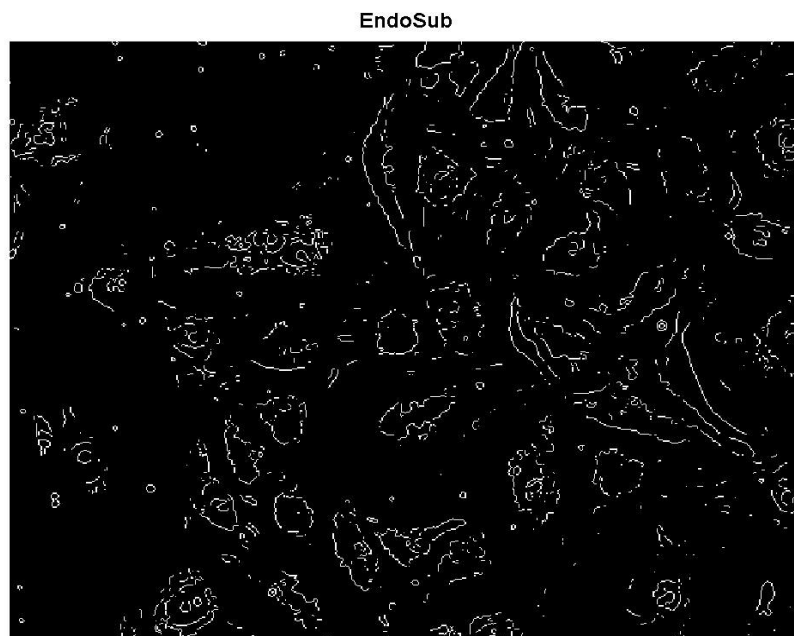


Figure 23a: A filtered image of a sub-confluent endothelial cell culture. Original image was filtered with a Sobel mask in both directions (horizontal and vertical) and afterwards superposed. Threshold was set to 0.22 when values were from 0 to 1. Note that density of edges is low and edges are curved.

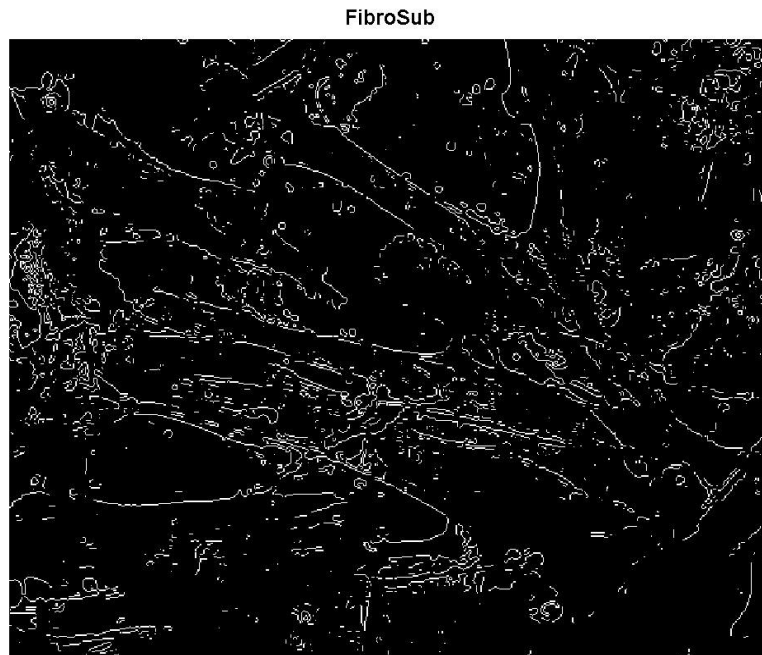


Figure 23b: A filtered image of a sub-confluent fibroblast cell culture. Original image was filtered with a Sobel mask in both directions (horizontal and vertical) and afterwards superposed. Threshold was set to 0.22 when values were from 0 to 1. Note that density of edges is low and edges are elongated.

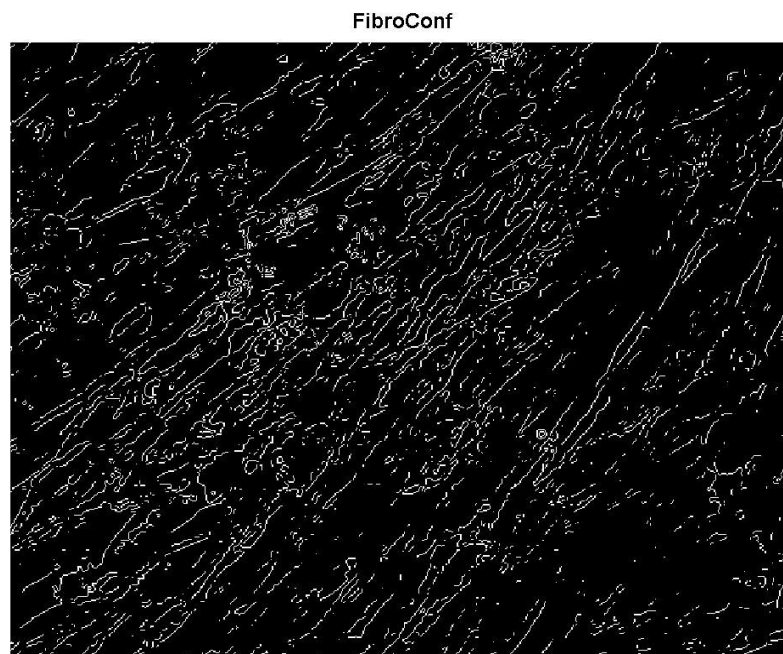


Figure 23c: A filtered image of a confluent fibroblast cell culture. Original image was filtered with a Sobel mask in both directions (horizontal and vertical) and afterwards superposed. Threshold was set to 0.22 when values were from 0 to 1. Note that density of edges is high and edges are elongated.

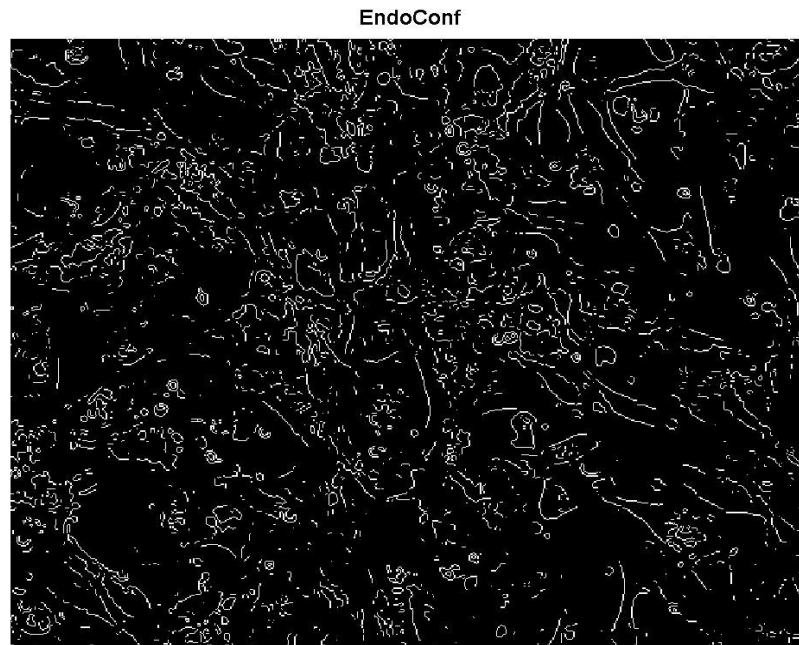


Figure 23d: A filtered image of a confluent endothelial cell culture. Original image was filtered with a Sobel mask in both directions (horizontal and vertical) and superposed afterwards. Threshold was set to 0.22 when values were from 0 to 1. Note that density of edges is high and edges are curved.

The following set of charts depicts features based on the above filtered binary images. The charts illustrate that the number of edges within images of confluent cell cultures is significantly higher than the number of edges within images of sub-confluent cell cultures. The next feature measured is the length of the edges within the binary images. The images illustrate that fibroblast cells exhibit straight and long edges whereas endothelial cells exhibit shorter and curved edges. The shape of edges was measured through asymmetry, which delivers higher values for longer shapes. A lower asymmetry value implies that object is more symmetric. An object with perfect symmetry is a circle (asymmetry value = 0). The number of edges and the length of the edges were computed with Matlab®. Asymmetry was computed with Cellenger®.

Chart 1: Results from statistical analysis of images are displayed as box plots. In all charts, the first box represents the results of images of sub-confluent endothelial cell cultures, the second box represents results of images of sub-confluent fibroblast cell cultures, the third box represents the results of images of confluent fibroblast cell cultures and the fourth box represents the results of images of confluent endothelial cell cultures. The boundary of the box closest to zero indicates the 25th percentile, a line within the box indicates the median, and the boundary of the box farthest from zero indicates the 75th percentile. The top and the bottom of the vertical dashed lines mark the 5th and 95th percentiles respectively.

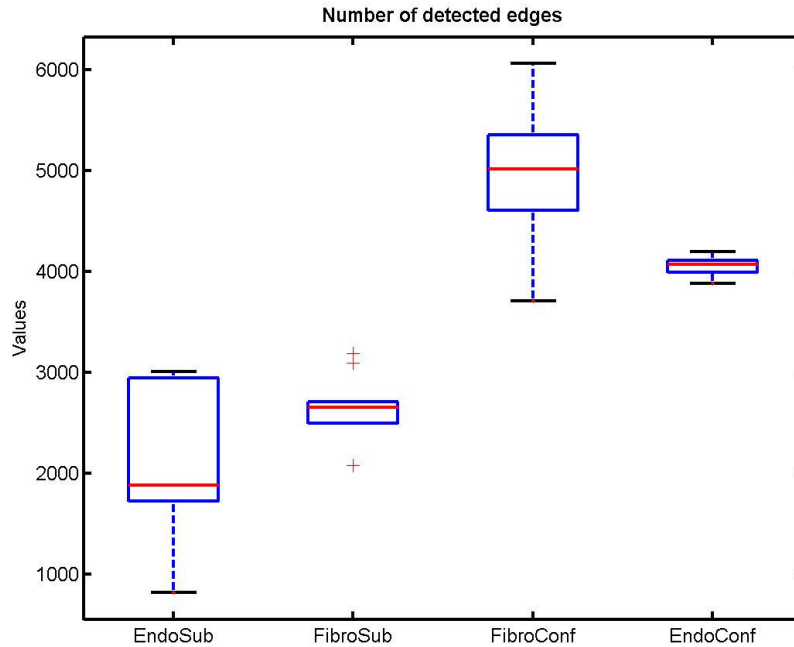


Chart 1a: Chart of density of edge points (=their number) within the images of different cell cultures. Note that the number of edges is significantly higher within images with confluent cell cultures. According to the Kruskal-Wallis test, $p = 5.6 \cdot 10^{-7}$, which implies a significant difference between image groups.

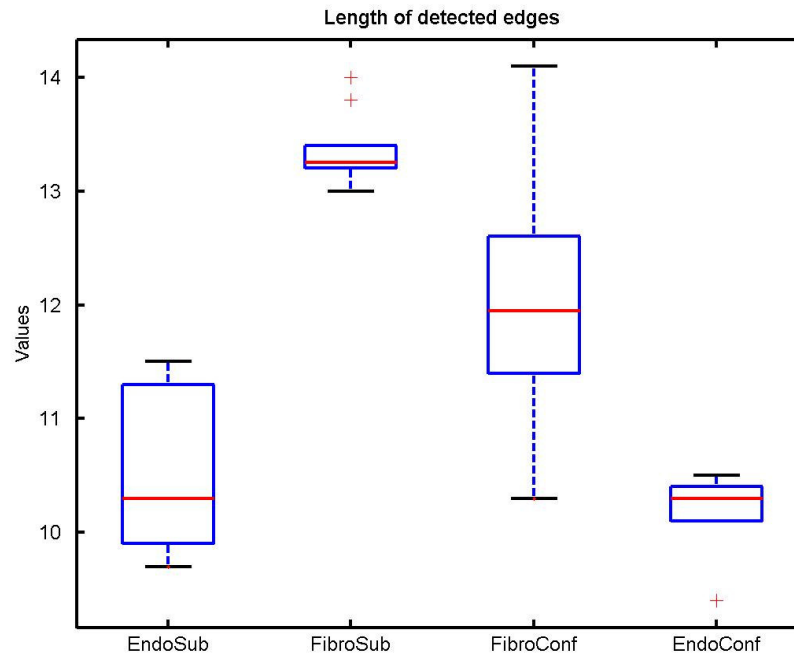


Chart 1b: Chart of the length of detected edges within the images of different cell cultures. Note that the length of edges within the images of fibroblast cell cultures is longer than the length of edges within the images of endothelial cell cultures. According to the Kruskal-Wallis test, $p = 1.6 \cdot 10^{-6}$, which implies significant difference between the image groups, but no significant difference between sub-confluent and confluent images.

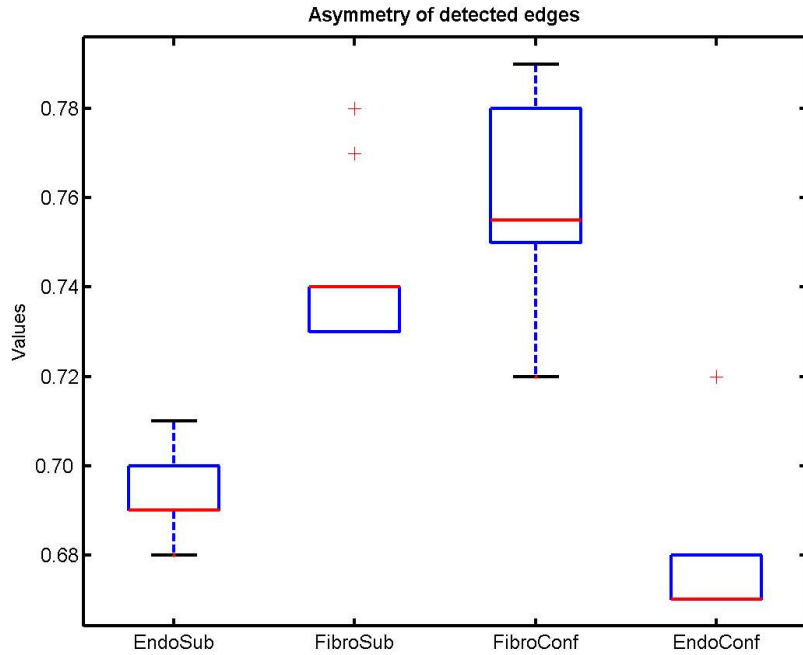


Chart 1c: Chart of the asymmetry of detected edges within the images of different cell cultures. Note that the asymmetry of edges within the images of fibroblast cell cultures is higher than the asymmetry of edges within the images of endothelial cell cultures. According to the Kruskal-Wallis test, $p = 3.7 \cdot 10^{-7}$, which implies significant difference between image groups, but no significant difference between sub-confluent and confluent images.

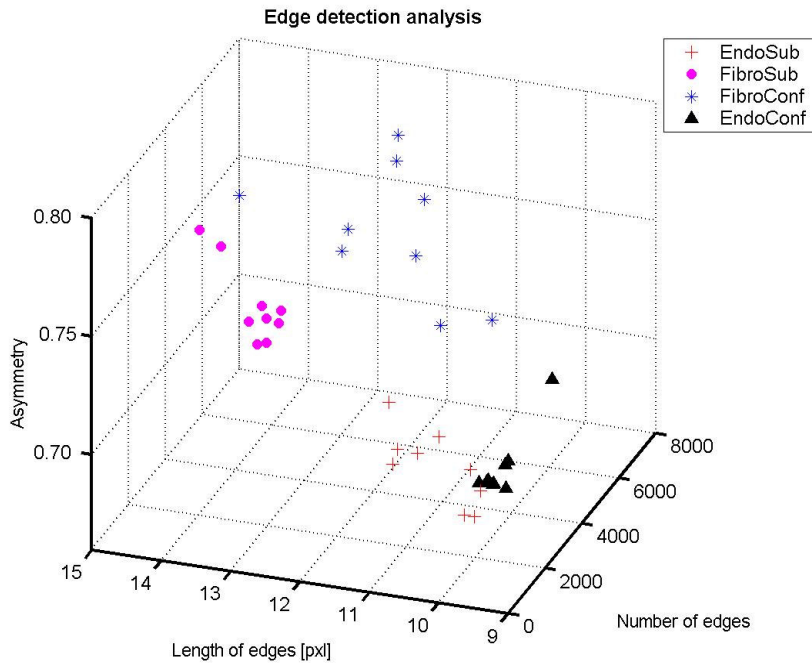


Figure 24: Figure of the three features from filtered binary images. Clusters of the four different image types are obvious and discriminative.

RESULTS

The stability of the Sobel edge detection filter was tested with different values of threshold. The following results were obtained:

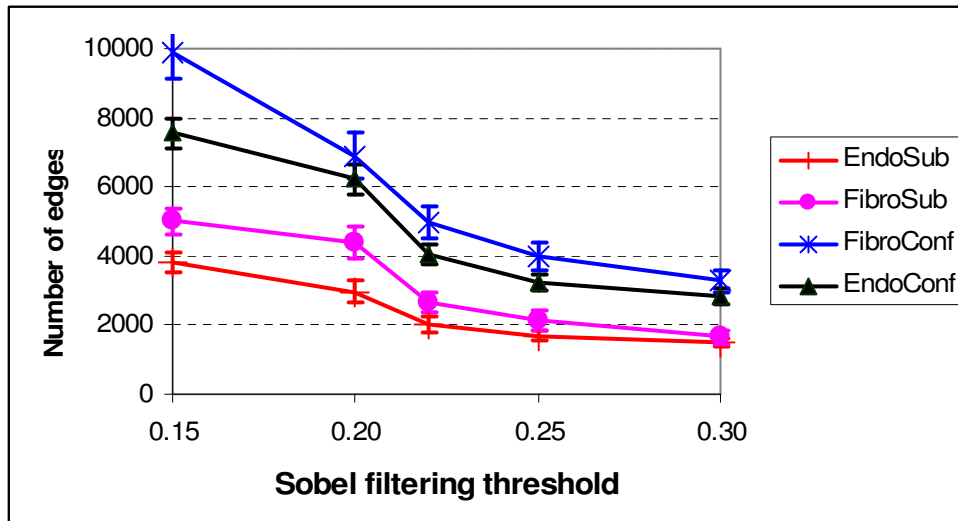


Figure 25: Figure of the stability of the filtering of binary images relative to a chosen threshold value. Different threshold values are shown on the x axis and the number of edges for each image group is represented on the y axis. Note that the number of edges diminishes with about the same tendency for higher threshold values for each image group. This implies an independence of the threshold value from the filtering result.

3.1.2. STATISTICAL APPROACH

The following charts consist of results computed with equations introduced in section 2.2.1.2.1.

Chart 2: Results from statistical analysis of images are displayed as box plots. In all charts the first box stands for the results of sub-confluent endothelial cell images, the second box stands for the results of sub-confluent fibroblast cell images, the third box stands for the results of confluent fibroblast cell images and the fourth box stands for the results of confluent endothelial cell images.

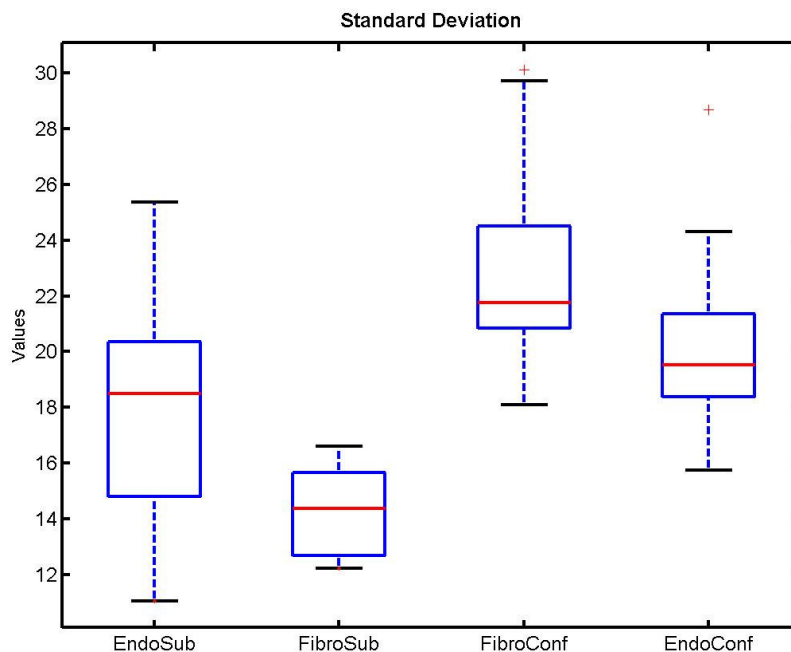


Chart 2a: Chart of the standard deviation of observed image groups. Note that the standard deviation is higher for confluent images because the intensity values change more often than the intensity values of sub-confluent images, where a uniform background dominates. According to the Kruskal-Wallis test, $p = 5.2 \cdot 10^{-6}$, which implies a significant difference between image groups.

RESULTS

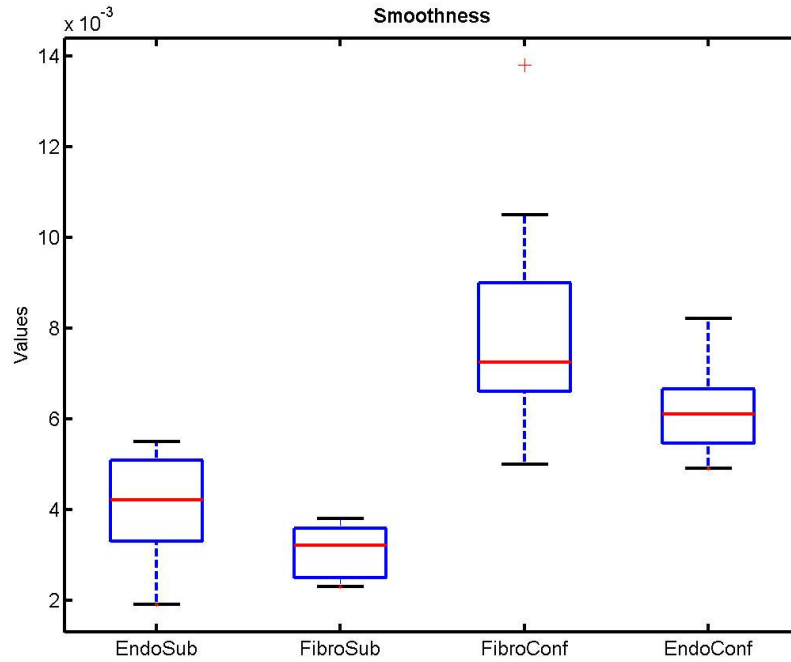


Chart 2b: This chart represents the results of smoothness for observed image groups. Note that the smoothness is higher for confluent images because they are composed of regions with larger excursions of intensity values. The Kruskal-Wallis test delivered $p = 9.4 \cdot 10^{-6}$, which implies a significant difference between groups.

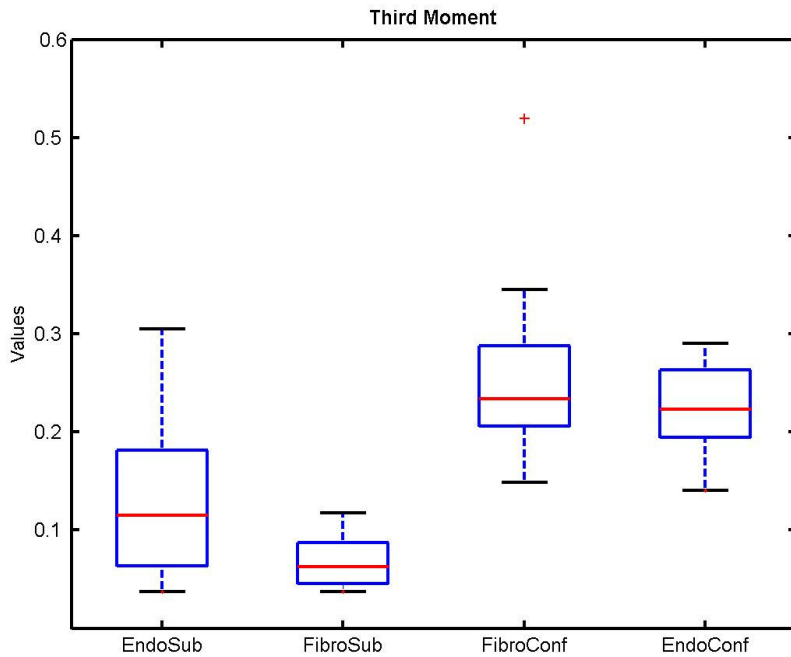


Chart 2c: This chart represents the results of the third moment for the observed image groups. The third moment is a measure of skewness of a histogram. Note that the third moment is higher for confluent images. The Kruskal-Wallis test delivered $p = 2.7 \cdot 10^{-6}$, which implies a significant difference between groups.

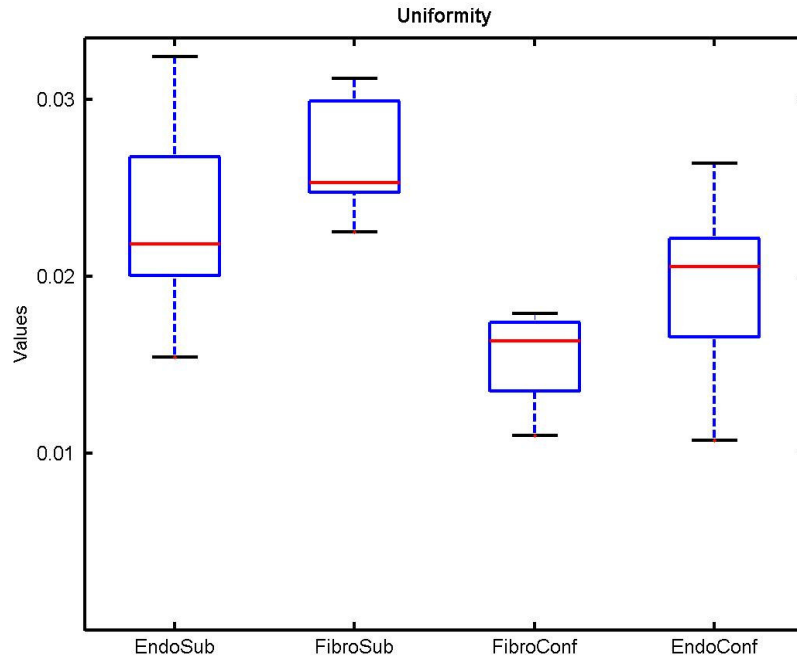


Chart 2d: This chart represents the results of uniformity for the observed image groups. Note that uniformity is higher for sub-confluent images because it is maximal if all gray values are equal. The Kruskal-Wallis test delivered $p = 5.5 \cdot 10^{-6}$, which implies a significant difference between groups.

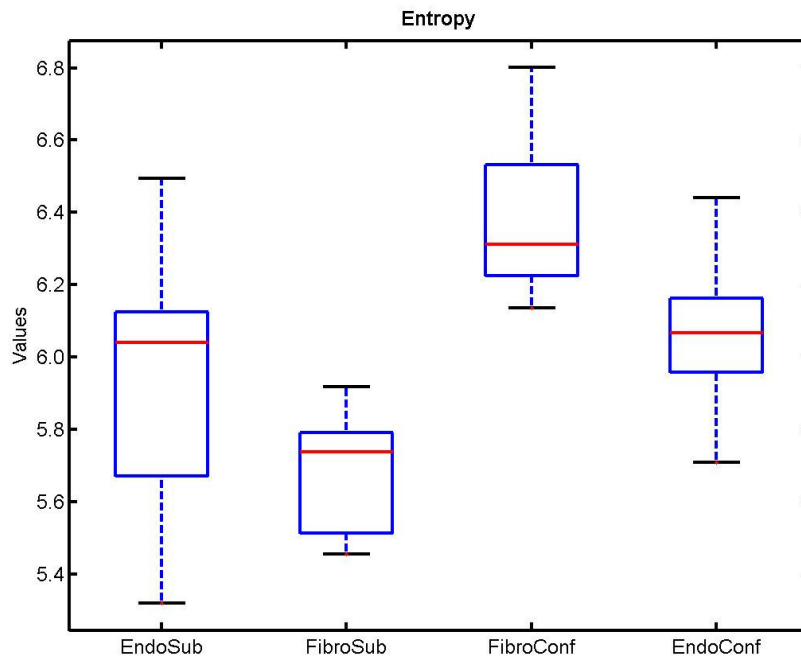


Chart 2e: This chart represents the results of entropy for the observed image groups. Note that entropy is higher for confluent images because it is a measure of randomness. The more random values are within an image the greater is the entropy. The Kruskal-Wallis test delivered $p = 1.8 \cdot 10^{-6}$, which implies a significant difference between groups.

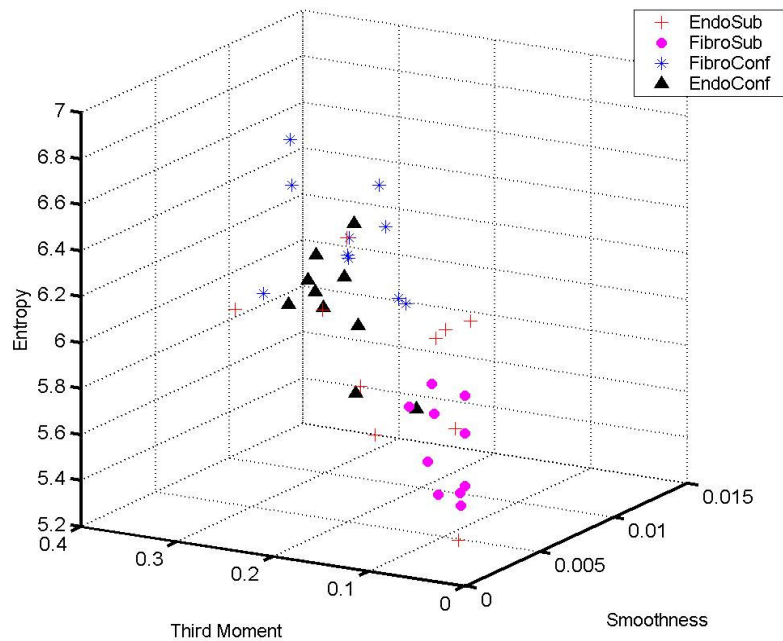


Figure 26: Entropy, Third moment and Smoothness shown in 3-D. Note that the clusters of sub-confluent images and confluent images of both cell types are obvious.

3.1.3. HARALICK TEXTURE COEFFICIENTS

The direction and shape of the edges was mathematically computed with co-occurrence matrices of different orientations. In addition, Haralick texture coefficients were computed.

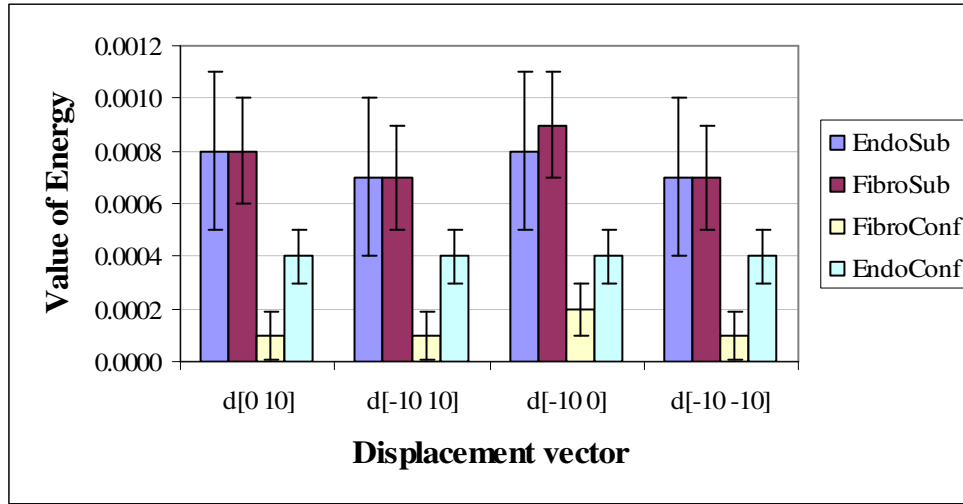


Figure 27a: Figure shows results for Energy: the first of Haralick texture coefficients based on co-occurrence matrix. Note that the difference between images of sub-confluent and confluent cell cultures is distinctive. The height of barks marks the mean value of energy of all ten images belonging to one of the four groups. Vertical lines mark \pm SD.

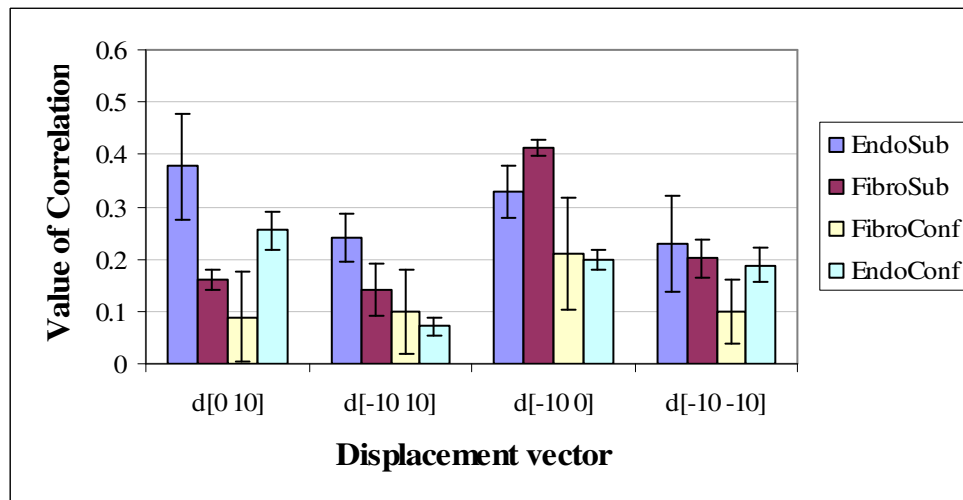


Figure 27b: Figure shows results for Correlation. Note that the difference between images of sub-confluent and confluent cell cultures is not distinctive. The height of barks marks the mean value of correlation and vertical lines mark \pm SD.

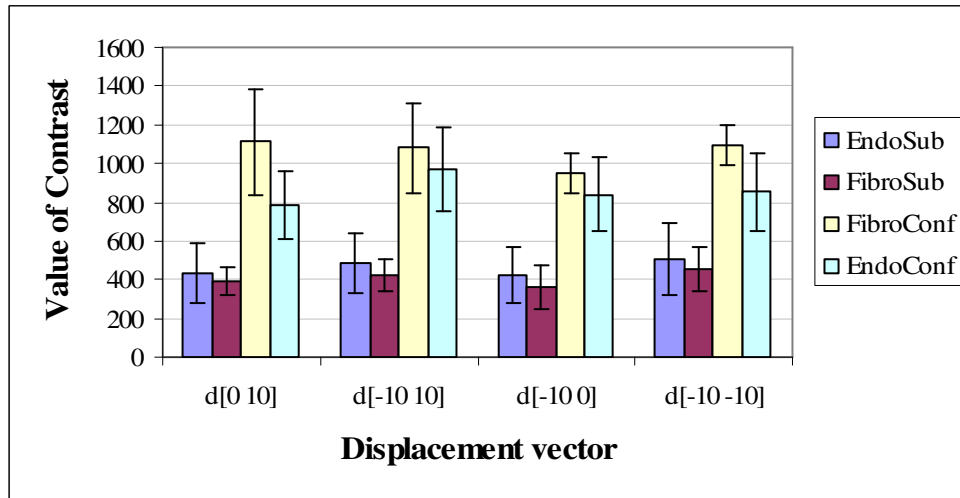


Figure 27c: Figure shows results for Contrast. Note that the difference between images of sub-confluent and confluent cell cultures is distinctive. The height of balks marks the mean value of correlation and vertical lines mark \pm SD.

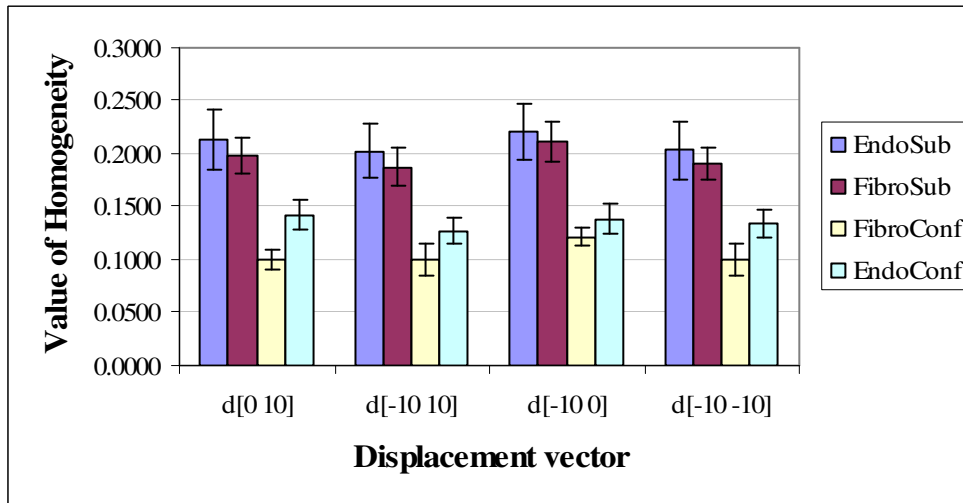


Figure 27d: Figure shows results for Homogeneity. Note that the difference between images of sub-confluent and confluent cell cultures is distinctive. The height of balks marks the mean value of correlation and vertical lines mark \pm SD. Homogeneity is a parameter which is higher for regions with uniform levels of gray values.

3.2. Separation of confluent endothelial and fibroblast images

In this section, results are computed only for images of confluent cell cultures. Different methods were applied to attempt to discriminate images of confluent endothelial cell cultures from images of confluent fibroblast cell culture.

3.2.1. SPECTRAL ANALYSIS

Spectral analysis was applied with the intention of receiving a spectral character of confluent images of interest. Representative parameters were derived from different spectra. The parameters were later used for the needs of Haralick texture coefficients (section 2.2.1.2.2.) and Gabor filtering (section 2.2.2.2.).

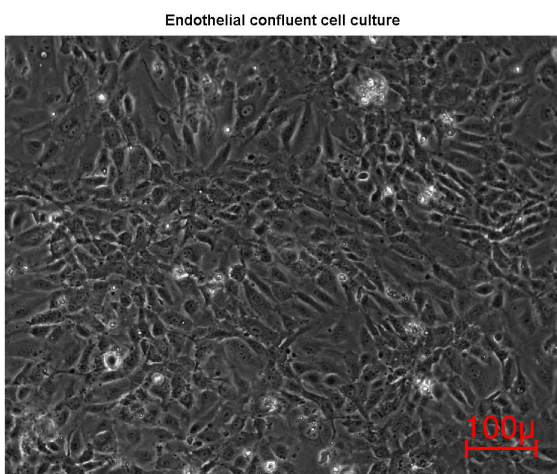


Figure 28: The original image of a confluent endothelial cell culture.

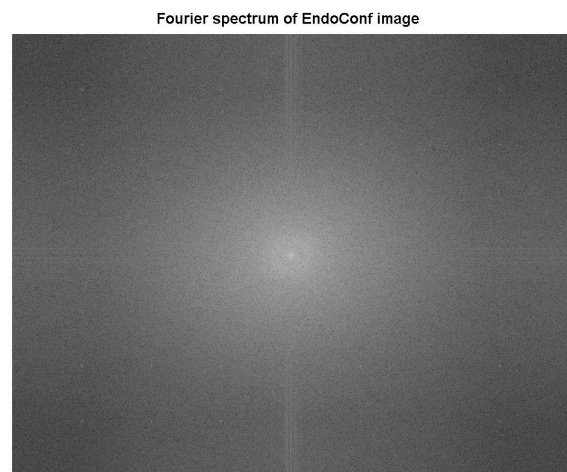


Figure 28a: Corresponding Fourier spectrum. Bursts of energy in perpendicular direction are due to the uniform regions in the image. The other component of the spectra is caused by the random orientation of cells.

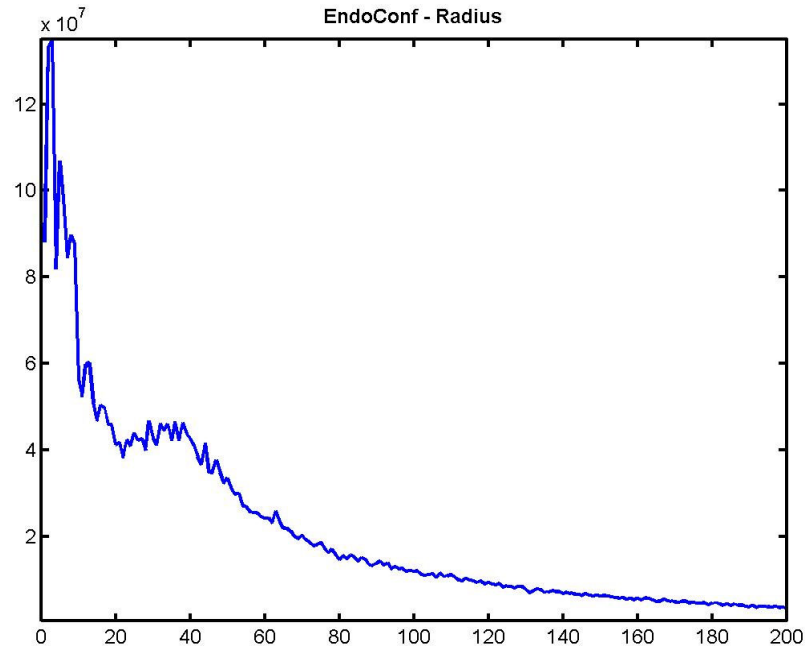


Figure 28b: The plot of the $S(r)$ (the frequency component) of the Fourier spectrum. The peak at the origin corresponds to the DC component. There are no other strong peaks in the spectrum, which implies no periodicity in a pattern built of endothelial cells within the image (**Figure 28**).

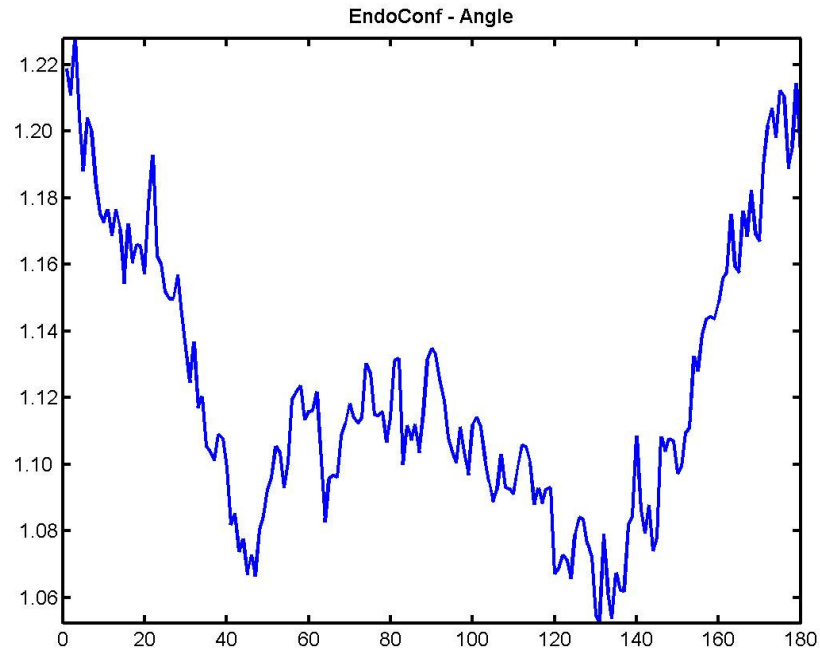


Figure 28c: The plot of the $S(\theta)$ (angle component) of the Fourier spectrum. It shows a random nature of the energy bursts, which is consistent with the energy distribution on **Figure 28a**.

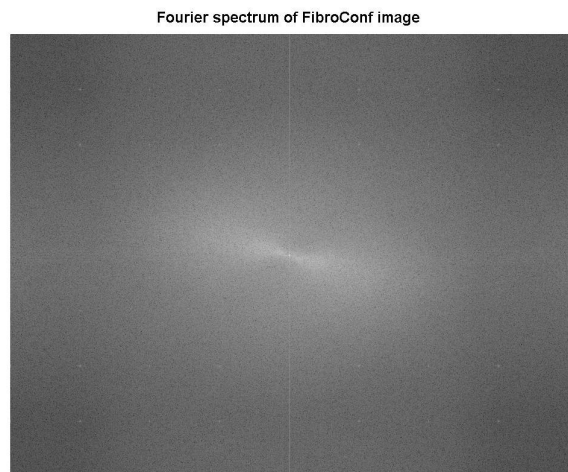
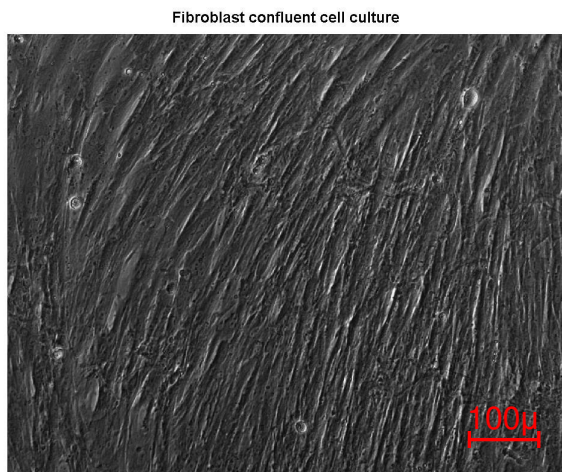


Figure 29: An original image of a confluent fibroblast cell culture.

Figure 29a: Corresponding Fourier spectrum. Bursts of energy in perpendicular direction are due to the uniform regions in the image. The other component of the spectra along the horizontal axis corresponds to the main direction of the texture which is approximately 90° .

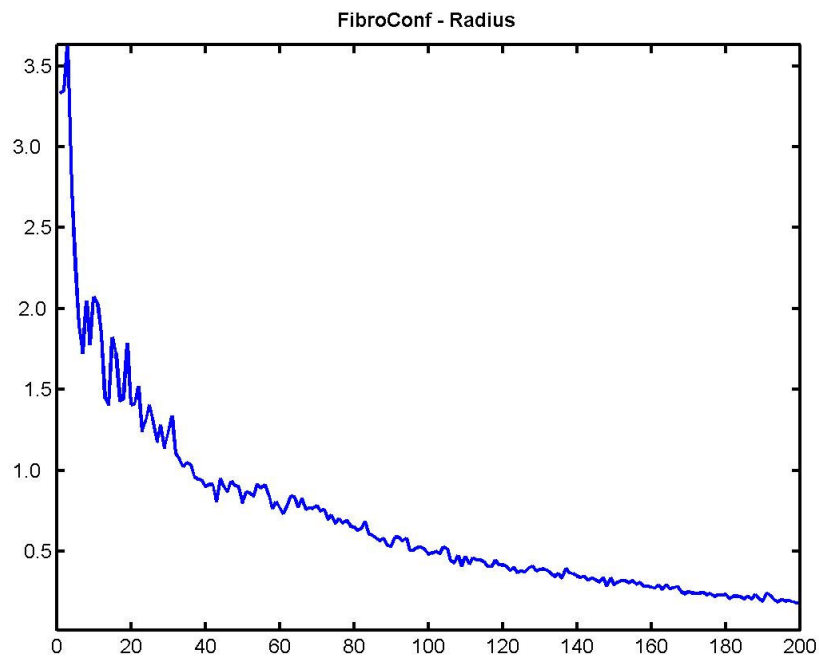


Figure 29b: The plot of the $S(r)$ (frequency component) of the Fourier spectrum. The peak at the origin corresponds to the DC component. The slight peak at approximately $r = 10$ implies periodical pattern built of fibroblast cells within the image (**Figure 29**). The pattern's period is 10 pixels.

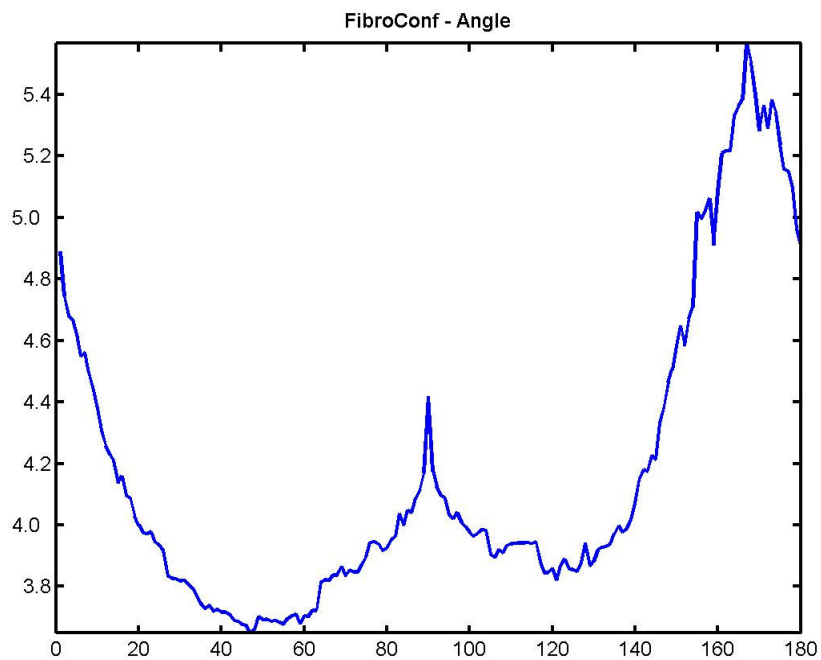


Figure 29c: The plot of the $S(\theta)$ (angle component) of the Fourier spectrum. It shows strong energy bursts in the region near the origin, at 90° and 170° . This is consistent with the energy distribution on the **Figure 29a** and it implies that the pattern of the fibroblast cell culture “grows” in the direction of 90° .

3.2.2. GABOR FILTERING

The following figures show original and Gabor filtered images of confluent endothelial cell cultures.

Endothelial confluent cell culture

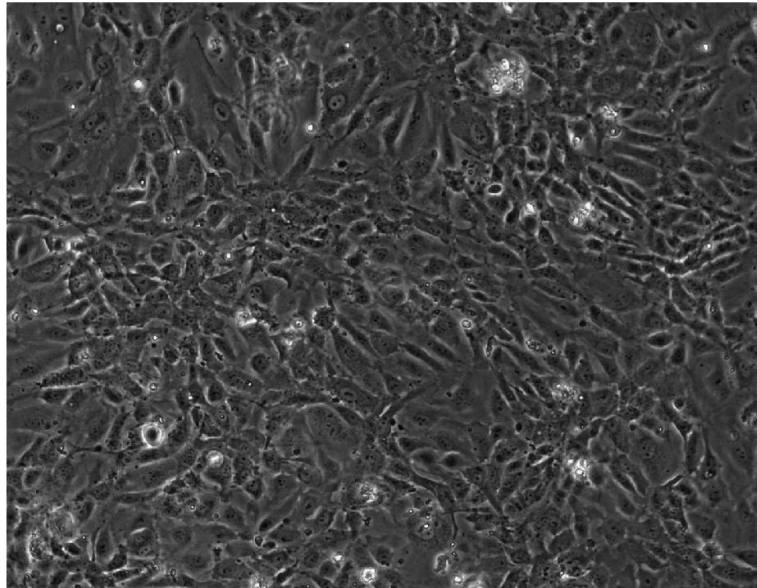


Figure 30: Original image of a confluent endothelial cell culture.

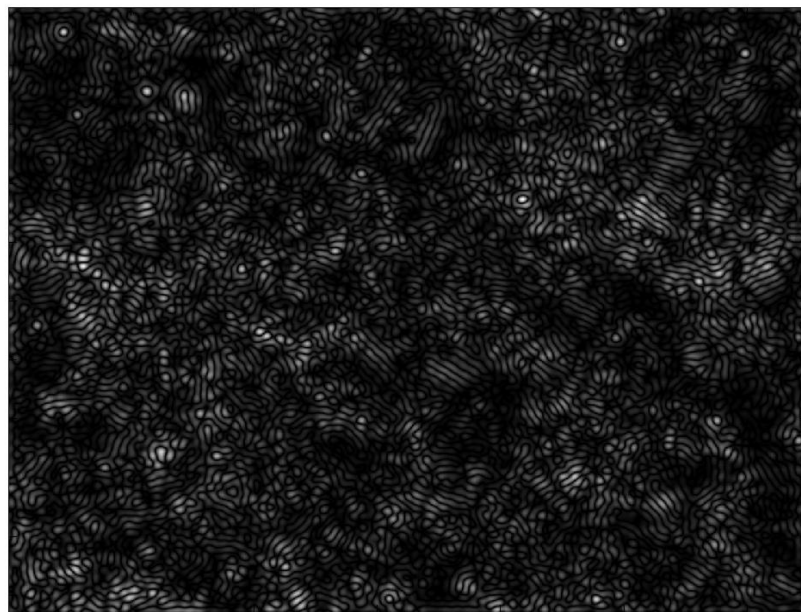


Figure 30a: Image of superimposed Gabor filtering results for all orientations.

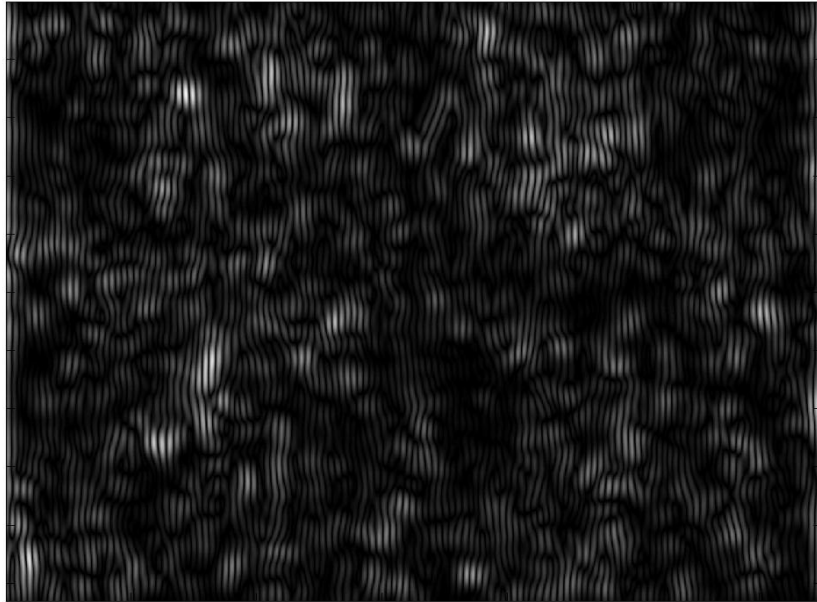


Figure 30b: Image after Gabor filtering with orientation of 0° . Energy of resulting image is $2.0 \cdot 10^6$.

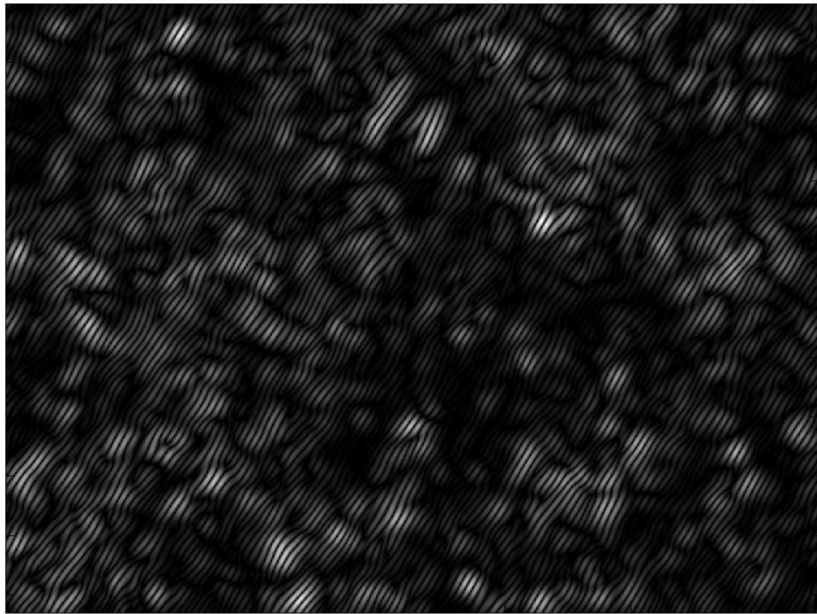


Figure 30c: Image after Gabor filtering with orientation of $\pi / 6 = 30^\circ$. Energy of resulting image is $1.9 \cdot 10^6$.

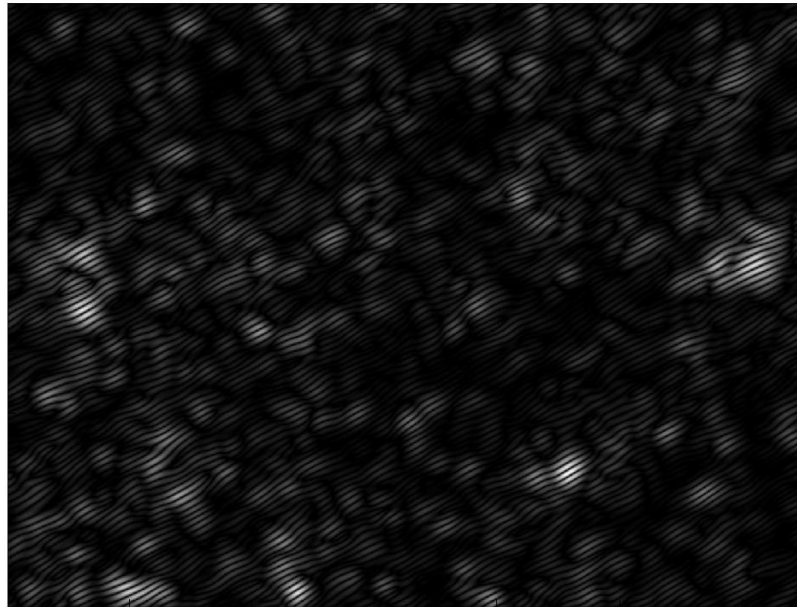


Figure 30d Image after Gabor filtering with orientation of $\pi/3 = 60^\circ$. Energy of resulting image is $1.9 \cdot 10^6$.

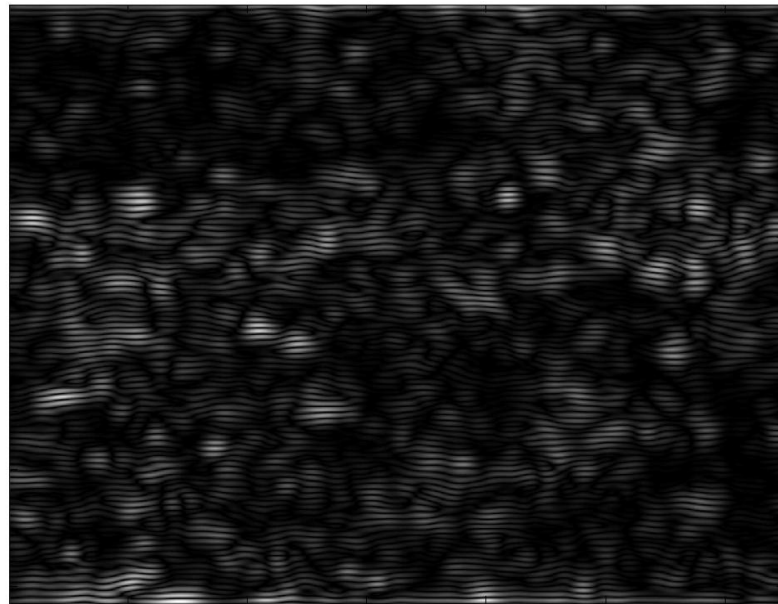


Figure 30e: Image after Gabor filtering with orientation of $\pi/2 = 90^\circ$. Energy of resulting image is $2.1 \cdot 10^6$.

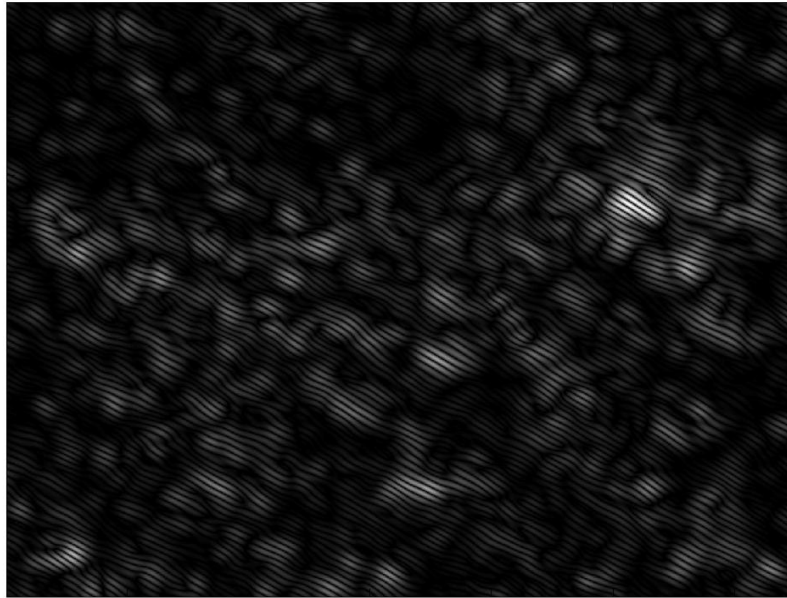


Figure 30f: Image after Gabor filtering with orientation of $2\pi / 3 = 120^\circ$.
Energy of resulting image is $2.1 \cdot 10^6$.

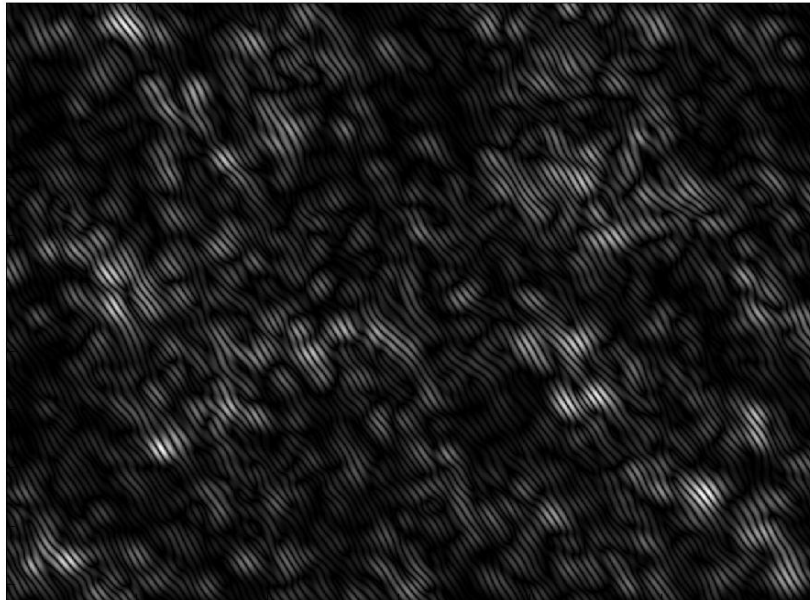


Figure 30g: Image after Gabor filtering with orientation of $5\pi / 6 = 150^\circ$.
Energy of resulting image is $2.0 \cdot 10^6$.

RESULTS

The following figures show original and Gabor filtered images of confluent fibroblast cell cultures.

Fibroblast confluent cell culture

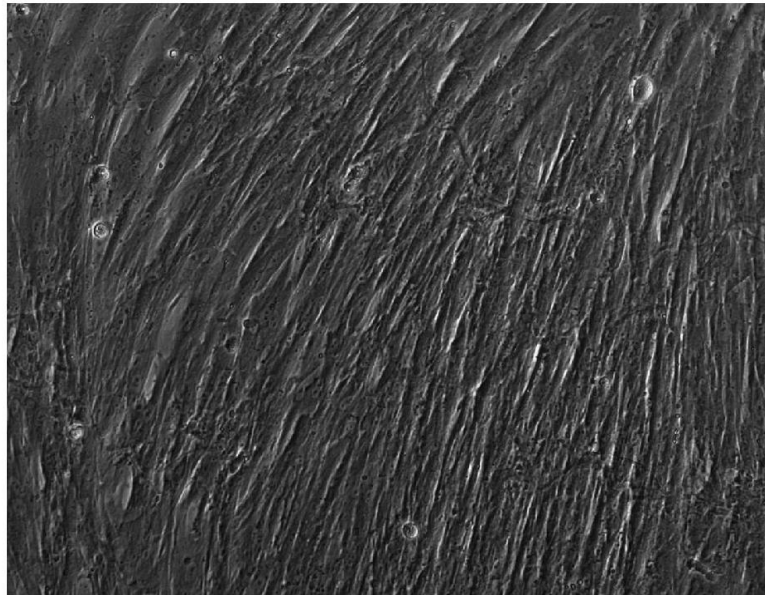


Figure 31: Original image of a confluent fibroblast cell culture.

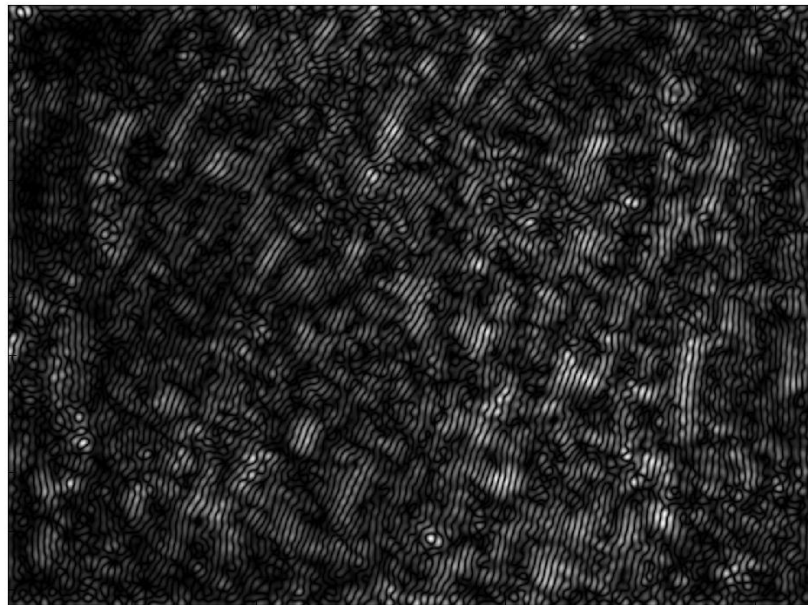


Figure 31a: Image of superimposed Gabor filtering results for all orientations.

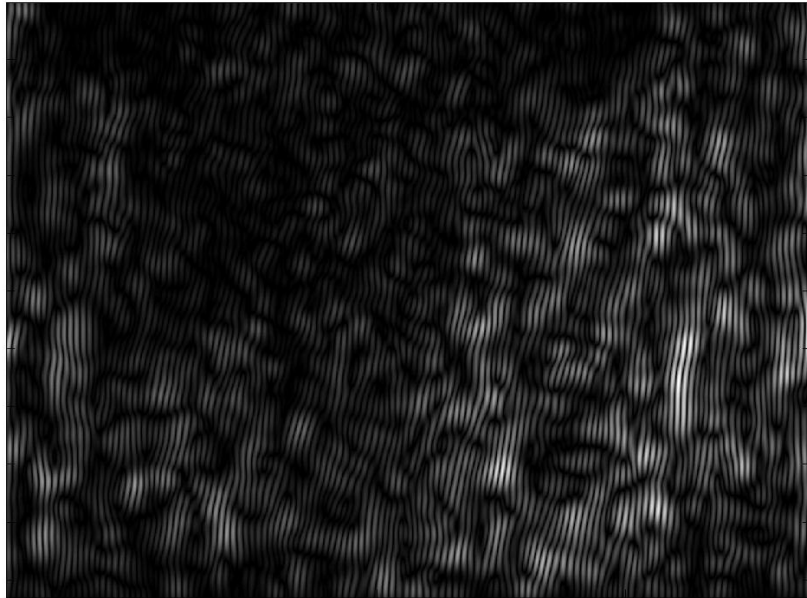


Figure 31b: Image after Gabor filtering with orientation of 0° . Energy of resulting image is $2.9 \cdot 10^6$.

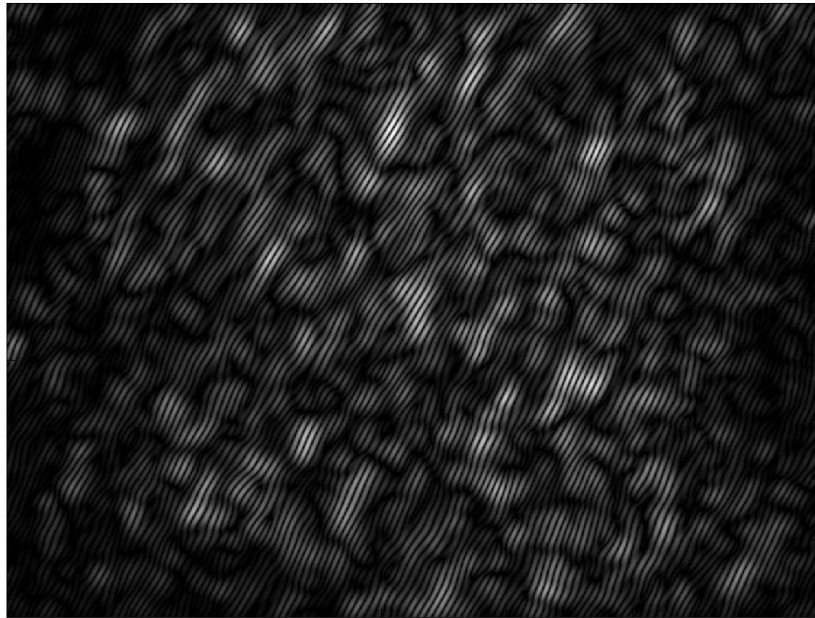


Figure 31c: Image after Gabor filtering with orientation of $\pi / 6 = 30^\circ$. Energy of resulting image is $3.4 \cdot 10^6$.

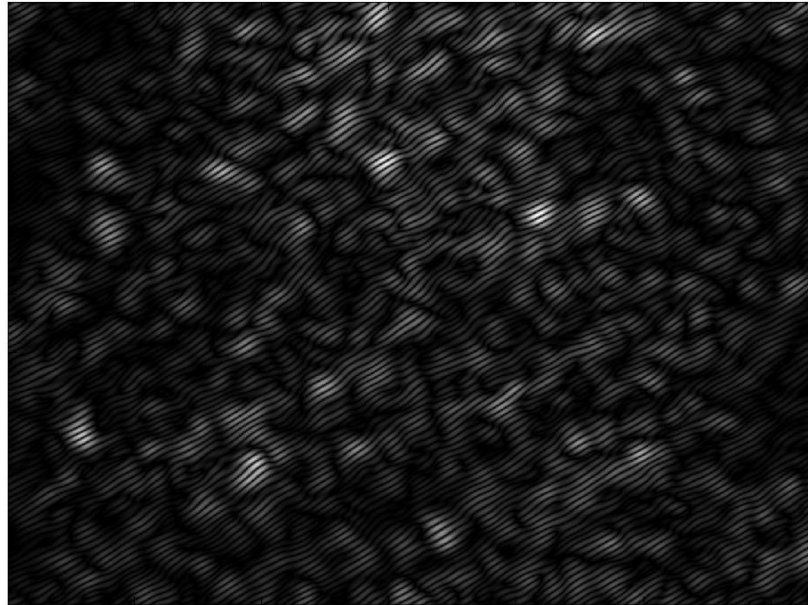


Figure 31d: Image after Gabor filtering with orientation of $\pi / 3 = 60^\circ$.
Energy of resulting image is $1.5 \cdot 10^6$.

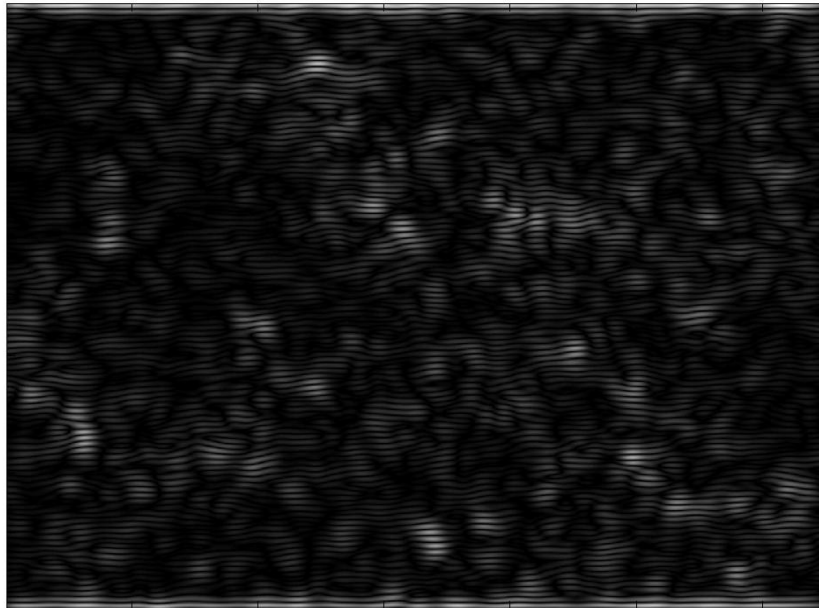


Figure 31e: Image after Gabor filtering with orientation of $\pi / 2 = 90^\circ$.
Energy of resulting image is $9.7 \cdot 10^5$.

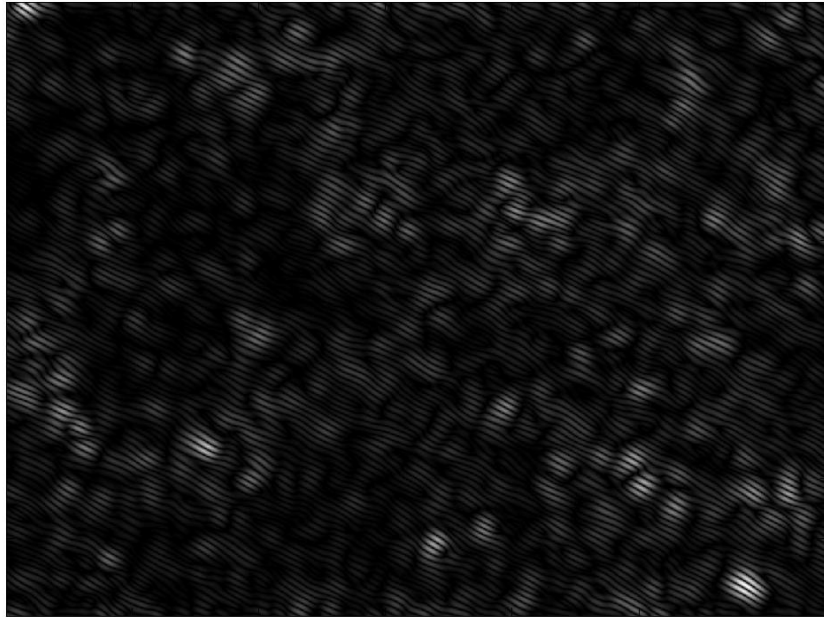


Figure 31f: Image after Gabor filtering with orientation of $2\pi / 3 = 120^\circ$.
Energy of resulting image is $7.0 \cdot 10^5$.

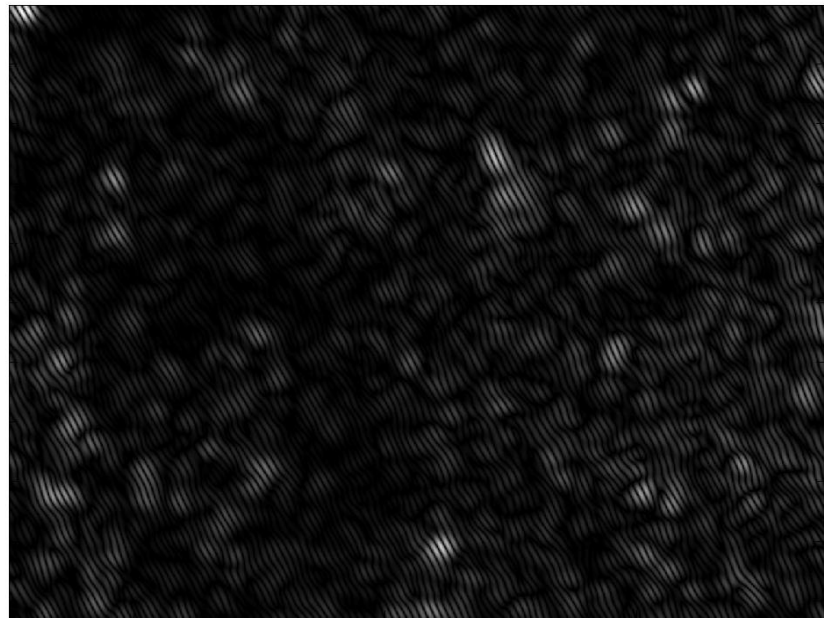


Figure 31g: Image after Gabor filtering with orientation of $5\pi / 6 = 150^\circ$.
Energy of resulting image is $9.5 \cdot 10^5$.

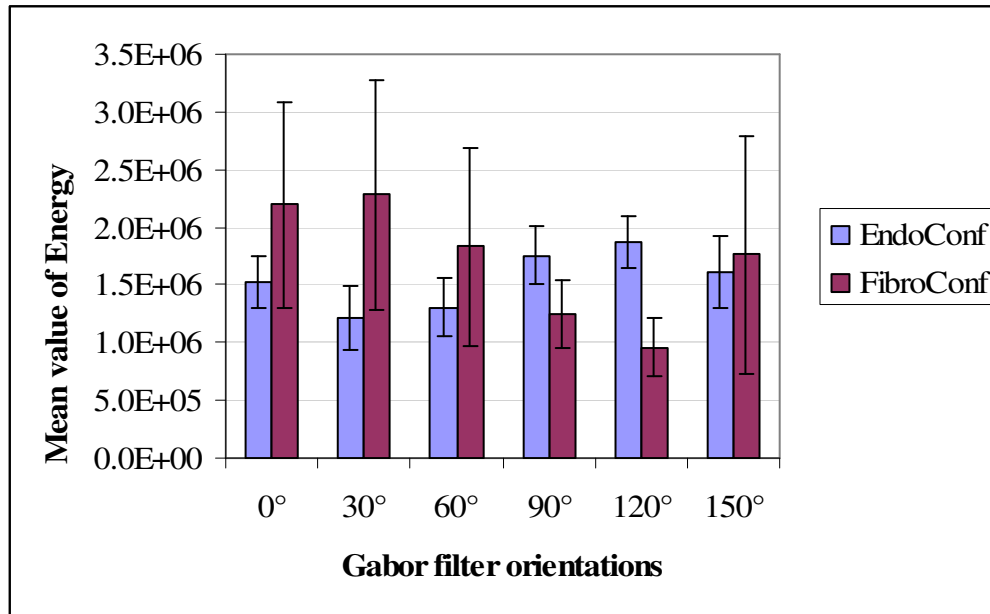


Chart 3: Mean values of Energy of images after Gabor filtering in different orientations. Vertical lines symbolize the standard deviation. Note that values of energy are in general more scattered for images of fibroblast cell cultures. This implies that the direction of texture from fibroblast cells is more distinct than the direction of texture from endothelial cells, which seems random.

3.2.3. CELLENGER® APPROACH

On average, numbers of segmented objects were higher on images of fibroblast cell cultures, but not significantly different than on images of endothelial cell cultures ($p > 0.001$).

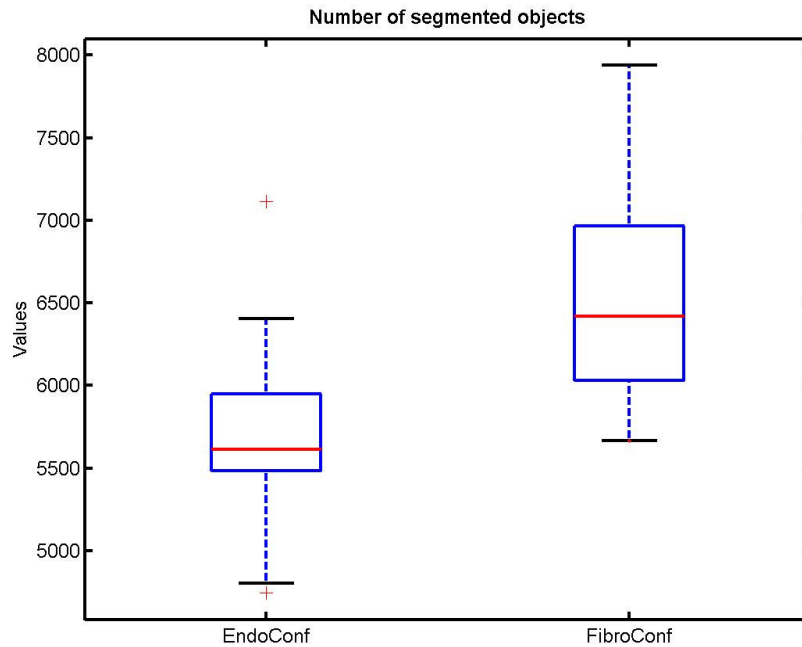


Chart 4a: The chart represents the results of the number of segmented objects for the two image groups. Note that the number of segmented objects is larger for images of fibroblast cell culture, because there are regions with larger excursions of the intensity values. The Kruskal-Wallis test delivered $p = 8.2 \cdot 10^{-3}$, which does not confirm a significant difference between groups.

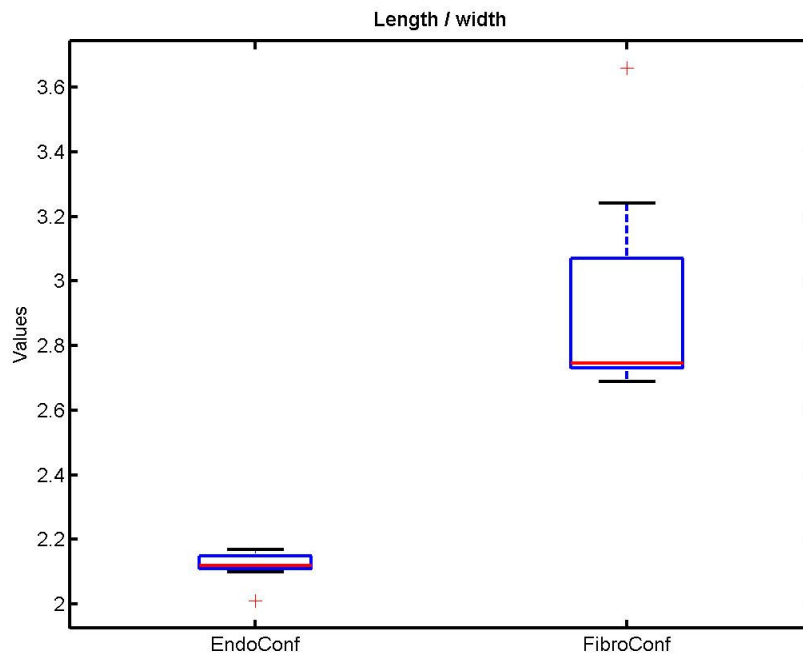


Chart 4b: The chart represents the results of the length/width ratio of the two image groups. Note that the length/width ratio is higher for fibroblast cell culture because the segmented objects are lengthier and thinner than the segmented objects within images of the endothelial cell cultures. The Kruskal-Wallis test delivered $p = 1.5 \cdot 10^{-4}$ which implies a significant difference between groups.

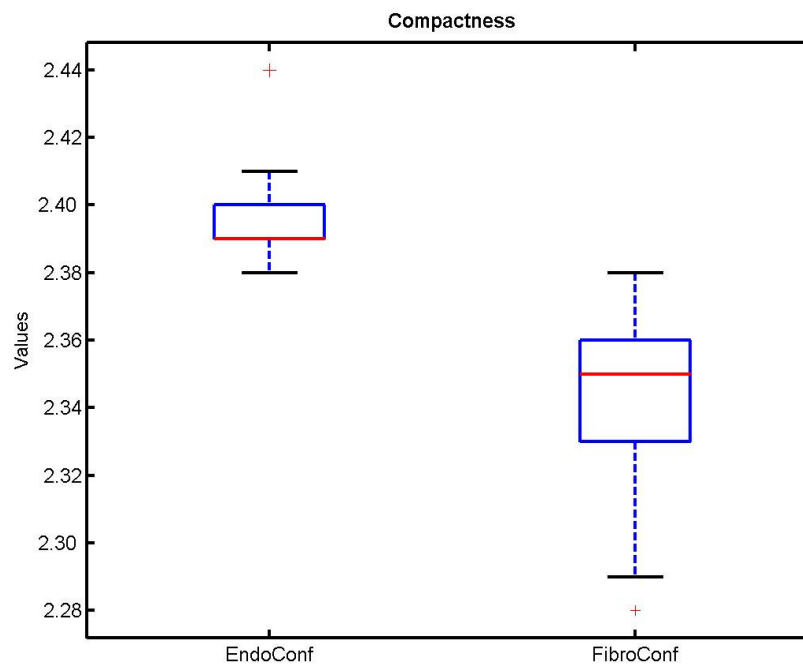


Chart 4c: The chart represents the compactness of the two image groups. Note that the compactness is higher for images of endothelial cell cultures. The Kruskal-Wallis test delivered $p = 2.1 \cdot 10^{-4}$, which implies a significant difference between groups.

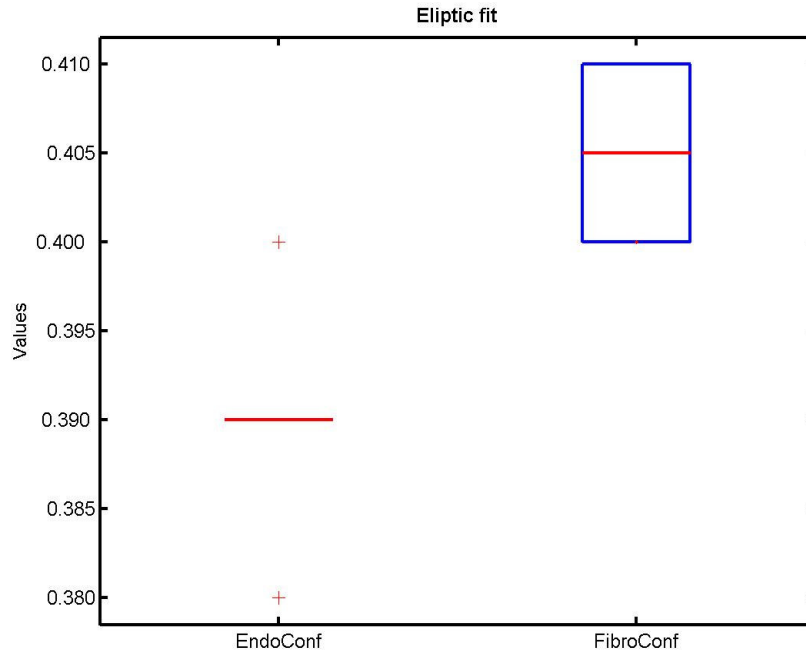


Chart 4d: The chart represents the results of the elliptic fit for the two image groups. Note that the elliptic fit is higher for images of fibroblast cell cultures. The Kruskal-Wallis test delivered $p = 1.2 \cdot 10^{-4}$, which implies a significant difference between groups.

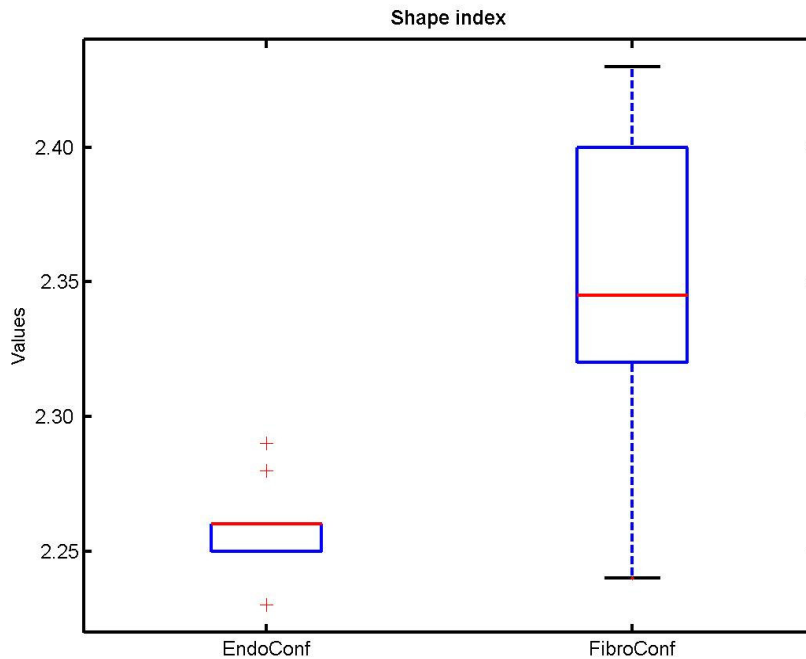


Chart 4e: The chart represents the results of the shape index for the two image groups. Note that the shape index value is higher for images of fibroblast cell cultures. The Kruskal-Wallis test delivered $p = 1.4 \cdot 10^{-4}$, which implies a significant difference between groups.

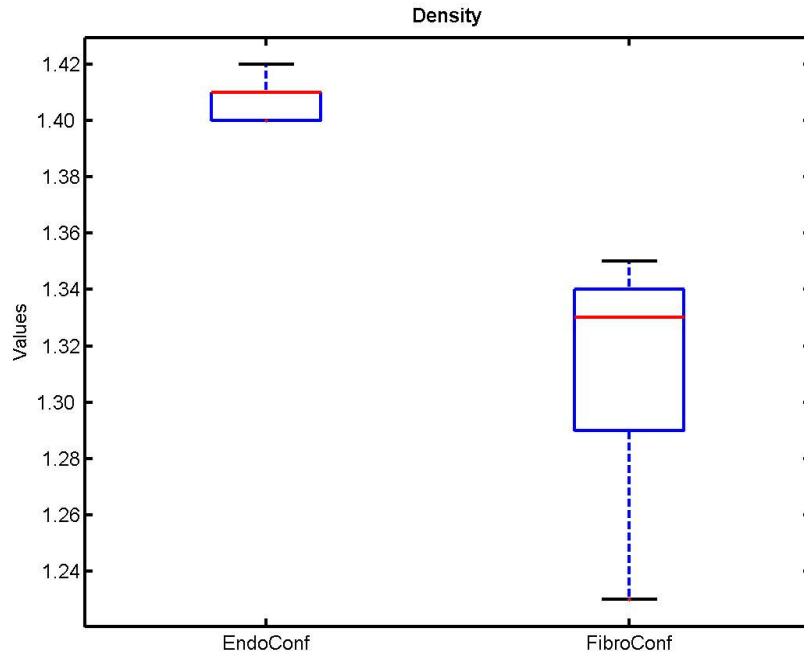


Chart 4f: The chart represents the results of the density for the two image groups. Note that the density is higher for images of endothelial cell cultures. The Kruskal-Wallis test delivered $p = 1.0 \cdot 10^{-4}$ which implies a significant difference between groups.

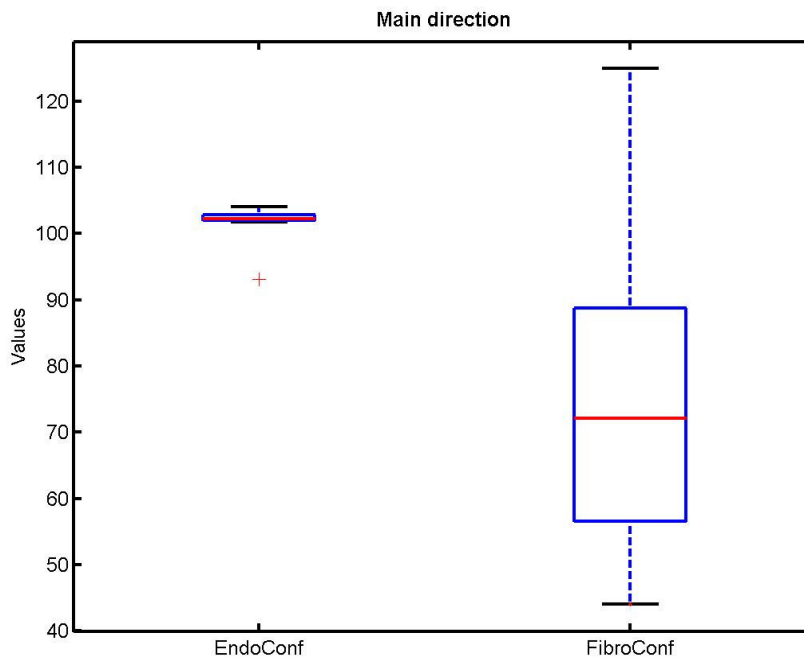


Chart 4g: The chart represents the results of the main direction for the two image groups. The Kruskal-Wallis test delivered $p = 2.3 \cdot 10^{-2}$, which does not confirm significant difference between groups.

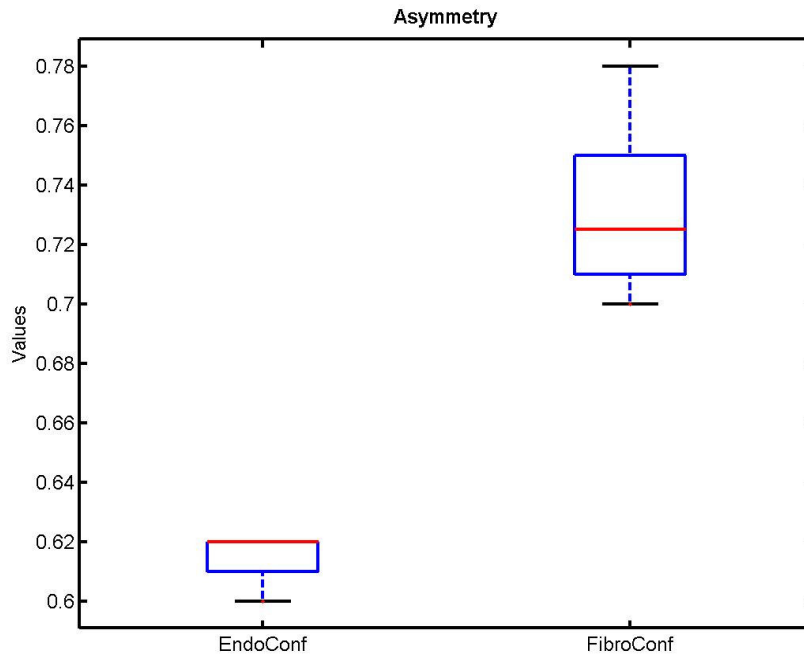


Chart 4h: The chart represents the asymmetry values of the two image groups. Note that the asymmetry value is higher for images of fibroblast cell cultures. The Kruskal-Wallis test delivered $p = 1.1 \cdot 10^{-4}$, which implies a significant difference between groups.

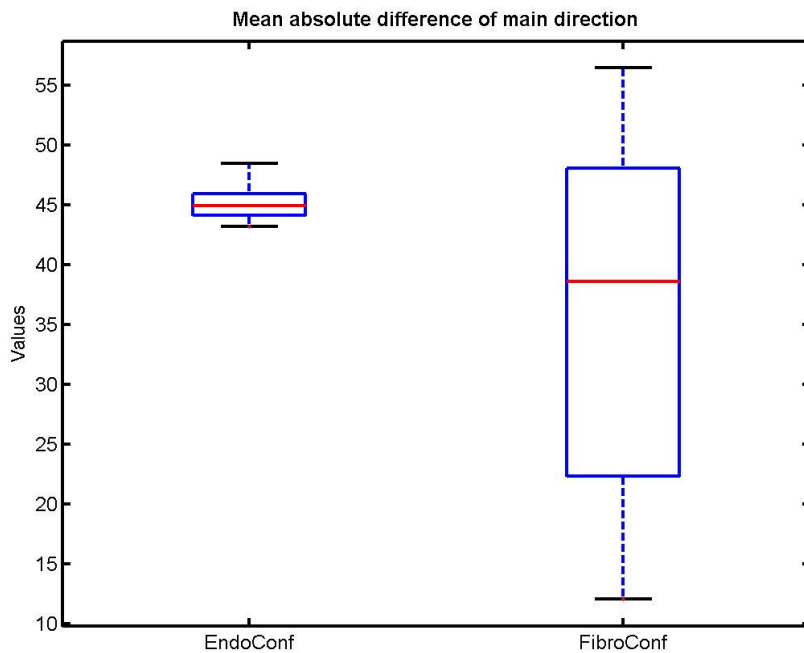


Chart 4i: The chart represents the results of the mean absolute difference of main direction for the two image groups. The Kruskal-Wallis test delivered $p = 3.3 \cdot 10^{-1}$, which does not confirm significant difference between groups.

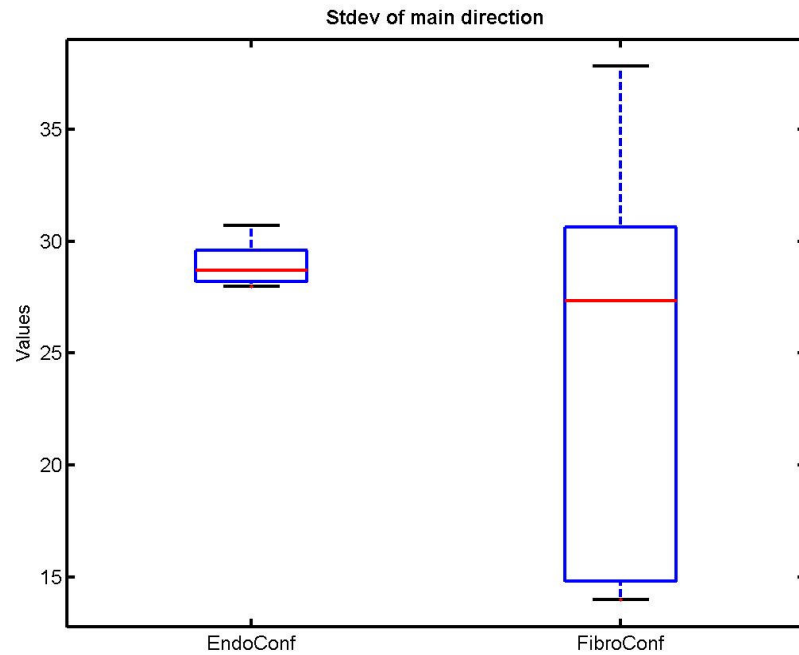


Chart 4j: The chart represents the standard deviation of main direction values for the two image groups. The Kruskal-Wallis test delivered $p = 3.6 \cdot 10^{-1}$, which does not confirm significant difference between groups.

4. DISCUSSION

The primary objective of the thesis was to select, examine and propose methods of image analysis for separating images of sub-confluent cell cultures from images of confluent cell cultures. Separation and classification should have a high probability of success. However, the majority of object classification problems do not have separable classes [20]. Therefore, some objects will always be misclassified. The “quality” of classification between sub-confluent and confluent images was evaluated with the Fisher criterion.

4.1. Edge detection

From the first inspection the images it was observed, that sub-confluent images show high amounts of background area with respect to confluent images. A uniform background was almost nonevident in the confluent images. The Cellenger® program estimated the background of sub-confluent cell cultures to be, on average, 43% of the image area. The program estimated the background of confluent images to be about 1% of the image area.

As a result, edge detection seemed to be straightforward. One objective was to prove that the more confluent the cell culture, the higher number of detected edge points. To prove this, the Sobel filter mask (section 2.2.1.1.) was used and the threshold was set to 0.22 (the mean threshold for all observed filtered images). The stability of the threshold was also tested. From **Figure 25**, it can be seen that the filter result did not depend on the threshold chosen. For future consideration, another filter mask (e.g. Laplacian or Canny) should be used on the same images in order to compare the filter results with those generated with the Sobel filter. Moreover, for further future consideration, a breakpoint of the threshold should be determined. The breakpoint

is where the number of edge points severely drops. The breakpoint of the threshold for the present project is assumed to be less than 0.15.

The filtered images from the four classes show interesting results. **Figure 23a - Figure 23d** show the filtered images. It was believed that the number of edge points that merged into edges is significantly lower on sub-confluent cell culture images than on confluent cell culture images. The belief was numerically verified by counting the edge points. The results are depicted in **Chart 1a**. The number of edges within the sub-confluent cell images range from 2000 to 3000, and the number of edges within the confluent cell images range from 4000 to 5500. The difference in the ranges can be used as a threshold for the separation of sub-confluent cell cultures from confluent cell cultures. If a threshold is set to 3500 (sub-confluent < 3500; confluent > 3500) each of the 40 sample images are correctly classified. Using only the number of edges to separate, though, might be misleading. It is obvious that the number of edges is dependent on the stage of the confluence. Thus, for example, if a sub-confluent culture is only 20% away from being confluent, the method of edge counting would not be sufficient to correctly classify the image. Future consideration should be given to verifying how well the number of edges classifies images correctly due to different stages of confluence.

The next step in the analysis of filtered images is to observe the length of the edges in the filtered images. **Figure 23a – Figure 23d** show that the images of fibroblast cell cultures have longer, straighter and orientated edges and that the images of endothelial cell cultures have curved and shorter edges. The length of the edges was numerically verified. The result can be seen in **Chart 1b**. This feature correctly describes whether an observed image is of an endothelial or fibroblast cell culture, but the feature does describe the confluence character of the cell culture in an image. The feature showed no significant difference between sub-confluent and confluent cell cultures.

The shape of the edges was numerically measured by asymmetry (a feature in the Cellenger® program). Asymmetry ranges from 0 to 1. 0 stands for a round object and 1 stands for a line. **Chart 1c** shows that the edges of endothelial cell cultures show more circular shape than those of fibroblast cell culture. A threshold set to 0.72 would separate endothelial cells from fibroblast cells with a very small misclassification error. However, the objective of the present task is to separate images by their confluence.

With the Fisher criterion in the 3-D feature space, an equation for a hyper-plane $w' * x - b = 0$ was obtained. The hyper-plane divides sub-confluent and confluent cell cultures according to the three features (the number of edges, the length of edges and the asymmetry) from filtered images.

The equation of hyper-plane is: $w' = [-0.0002 \ 0.0990 \ -0.4790]$ and $b = 0.0089$. The optimal separability value calculated from the Fisher criterion is 0.59.

4.2. Texture

4.2.1. STATISTICAL APPROACH

After the images were described according to the detected edges, the texture information was estimated. Texture content is also an important approach for describing a region. The most common approach to texture analysis is based on statistical properties of the intensity histogram. **Chart 2a – Chart 2e** depict the statistical measures of texture content of images.

A statistical approach delivers general information about the images and thus is relatively convenient for describing differences between sub-confluent and confluent cell cultures. Sub-confluent images generally have a higher degree of constant character due to a higher presence of a uniform background. On the other hand, confluent cell cultures have a higher degree of randomness due to a generally higher rate of excursions in gray level values. These descriptors are numerically evaluated through statistical moments defined in section 2.2.1.2.1. Before images are evaluated statistically, normalization is performed on the images. Hence, all images have zero mean and unit variance.

For example, the standard deviation (**Chart 2a**) and the entropy (**Chart 2e**) are higher for confluent cell cultures because the pixel values in those images are more random than the pixel values in images of the sub-confluent cell cultures. The reason is that the constant background lowers the value of entropy and standard deviation in an image. On the other hand, confluent images are the least smooth and the least uniform (shown in **Chart 2b** and **Chart 2d**). The smoothness is a feature that is 0 for a constant area and approaches 1 for a region with large excursions of its intensity levels. From **Chart 2c**, the larger value of the third moment applies to images with confluent character.

The statistical approach delivered good results in distinguishing between sub-confluent and confluent cell cultures. However, it is almost impossible to set thresholds that do not possess any error in misclassification.

With the Fisher criterion in the 5-D feature space, the separability value was worse than the separability value obtained through edge detection. For the statistical approach, the optimal separability value calculated from the Fisher criterion is 0.24. The equation for a hyper-plane which divides sub-confluent and confluent cell cultures is $w' = [-0.0108 \quad -62.1300 \quad -0.5757 \quad 7.7281 \quad 0.3790]$ and $b = -1.8066$.

4.2.2. HARALICK TEXTURE COEFFICIENTS

The motivation to use Haralick texture coefficients for the aim of this work is the bibliography from the early 1990's. During the early nineties, this method was one of the most popular and promising methods for the analysis of the texture content. Haralick texture coefficients are derived from co-occurrence matrices. The distribution of intensity values and the position of pixels with equal or nearly equal intensity values are important factors in deriving the Haralick texture coefficients. The inclusion of spatial information with intensity information make Haralick texture coefficients preferable over the statistical approach mentioned in the section 4.2.1.

The co-occurrence method describes second-order image statistics and works well for a large variety of textures [20]. Helpful properties of the co-occurrence method are the description of spatial relations between tonal pixels and the invariance of some gray-level transformations. On the other hand, the method does not take into account primitive shapes and angles of textures that are not a multiplication of 45° . Therefore this method is not recommended if the texture consists of large primitives and if precise detection of the texture orientation is needed [20].

Values of the elements along the diagonal of the co-occurrence matrix $P(i,i)$ are equal to the area of the regions in the image with brightness i . Thus, the diagonal elements correspond to the values of the histogram. The values of elements off the diagonal of the matrix $P(i,j)$ are equal to the length of the border dividing regions with brightness i and j where $i \neq j$. For example, in an image with low contrast, the elements of the co-occurrence matrix farthest from the diagonal are equal to zero or are very small. For high-contrast images, the opposite is true.

Co-occurrence matrices generated for this work (see **Figure 12a** – **Figure 12c**) show poor contrast of the original images. The values in co-occurrence matrices are grouped at the specific place along the diagonal, whereas off diagonal values are generally very low or zero. Thus, a certain pre-processing enhancement is needed to achieve better results.

A displacement parameter $D = 10$ in directions of 0° , 45° , 90° and 135° was used to generate the co-occurrence matrices. The choice of displacement parameter was based on the spectral results (see **Figure 28b** and **Figure 29b**) where in the frequency spectrum $S(r)$ a peak at about 10 units was detected. For future consideration, the co-occurrence matrices and subsequent Haralick texture coefficients (with displacement parameters from 0 to 100 in steps of 10 in all directions) should be observed. Such thorough analysis of co-occurrence matrices and texture coefficients was not foreseen within the volume of this work.

Energy, the first of the Haralick coefficients, is highest when all $P(i,j)$ are equal (a constant image). From **Figure 27a** it is evident that the Energy value for images of sub-confluent cell cultures is significantly higher than the Energy value for images of confluent cell cultures. The result was expected. The threshold may be set to the mean value of 0.007 ± 0.003 , but it should be noted that some images will be misclassified. No evidence of different orientation is detected from the values of Energy. The height of the barks is almost independent of different orientations of the displacement vector for all four groups of images.

Correlation, the next of the Haralick coefficients, is a measure of linearity, linear directional structures in direction Θ result in large correlation values in this direction [20]. Correlation values do not show any significant difference between sub-confluent and confluent image groups **Figure 27b**, thus no threshold can be set for the purpose to separate them. Interesting to see is that the highest values of correlation are reached at the displacement vector $[0 \ 10]$ for the endothelial cells and at $[-10 \ 0]$ for the fibroblast cells.

Contrast derived from the co-occurrence matrix is a measure of the intensity contrast between a pixel and its neighbours. Contrast is 0 for a constant image and very low if the highest values are near the main diagonal, because the differences $(i-j)$ are smaller along the main diagonal. As expected, the contrast value is higher for images of confluent cell cultures because of their higher local variations in pixel values, as shown in **Figure 27c**. A threshold of 800 ± 200 may be set to distinguish sub-confluent and confluent cell cultures. However, some misclassification errors will probably exist.

Homogeneity is a measure of the distribution of elements to the diagonal in the co-occurrence matrix. This descriptor has the opposite effect of contrast. Homogeneity is 1 for a diagonal co-occurrence matrix. From **Figure 27d** homogeneity values are significantly higher for images with sub-confluent cell cultures. A threshold of 0.17 ± 0.03 may be set to separate both cell cultures according to their confluence character.

With the Fisher criterion in the 4-D feature space, an equation for a hyper-plane ($w' * x - b = 0$) was obtained, which divides sub-confluent and confluent cell cultures according to the four Haralick texture coefficients. $w' = 1.0e+003 * [-7.1489 \ -0.0066 \ -0.0000 \ 0.0731]$ and $b = -2.7295$. The optimal separability value calculated from the Fisher criterion is 4.81.

As can be seen from the discussion of dividing sub-confluent and confluent cell cultures, the highest separability according to the Fisher criterion was achieved by Haralick texture coefficients.

The second goal of the thesis is to propose a method for correctly classifying confluent endothelial and confluent fibroblast cell cultures. There are two main principles used to characterize the texture content of both cell groups. The first principle, Gabor filtering, is based on spectral analysis and enriched with spatial information of the texture. The second principle, the Cellinger® approach, attempts to detect differences of texture primitives generated with multi-resolution segmentation of images of both classes. Multi-resolution segmentation is a special segmentation algorithm explained in detail in [1]. It is patented and owned by the Definiens Group. The “quality” of classification between endothelial and fibroblast images was again evaluated with the Fisher criterion.

4.2.3. SPECTRAL ANALYSIS

Spectral analysis of texture is based on the Fourier transform, which is ideally suited for describing the orientation of periodic or nearly periodic 2-D patterns in an image. From visual inspection of the confluent fibroblast cell culture in **Figure 29**, it can be seen that the pattern is distinctly orientated. In contrast, the pattern of the endothelial cell culture in **Figure 28** is randomly orientated. Spectral analysis numerically proved the results from visual inspection. The angular plot $S(\theta)$ in **Figure 29c** shows strong energy components in the region near the origin, 90° and

170°. This is consistent with the Fourier spectrum of the image of a confluent fibroblast cell culture in **Figure 29a**, where emphasized energy along the horizontal and vertical axes is associated with a portion of the constant regions within the image. But, the main burst of energy along the horizontal axis corresponds to the strong vertical (nearly 90°) fibroblast pattern. On the other hand, the angular plot $S(\theta)$ of confluent endothelial cells in **Figure 28c** shows the random nature (without distinctive peaks) of its pattern. The same can be interpreted from the Fourier spectrum in **Figure 28a**, where, apart from stronger vertical and horizontal energy bursts, the rest of the energy is distributed in no particular direction. From the frequency plots $S(r)$ in **Figure 28b** and in **Figure 29b** a slight peak at about 10 units can be viewed. This information served as an input parameter in co-occurrence matrices and for Gabor filtering. 10 pixels is in the average diameter of a cell in an image taken at resolution 1030×1300 pixels.

For the present project, spectral analysis is only to obtain basic information about the images (e.g., frequency and orientation). However, spectral analysis can also help in classification. For classification, a texture vector may be set up with several values that consider the number of peaks as a function of radius r or angle θ . The aim is to associate a simple number or a vector that is indicative of the type of texture with a pixel or a region in the image. Thus, classification system would exist where a set of known textures could be used to create a feature space in two or more dimensions. If an unknown texture is observed, the texture may be described by the terms of known textures by using the feature space to look for the closest matching texture according to given texture numbers of the unknown texture.

4.2.4. GABOR FILTERING

In practice, two-dimensional Gabor filters are useful for a variety of computer vision problems. In particular, Gabor filtering may be applied in texture analysis [29, 30]. While they are widely used, the methods for selecting a bank of Gabor filters are often vague. For the purpose of the present problem, Gabor spatial frequency of $\pi/10$ and orientations $n \cdot \pi/6$ for $n = [0, 1 \dots 6]$ were used. Thus, a bank of six Gabor filters was selected. Selecting the right parameters for Gabor filters in order to get optimal results is not trivial. For example, if the frequency is too high, the filter is spatially concentrated on a small area. This might be problematic especially for small

images where it is not possible to notice a difference when the orientation change is small [29]. Therefore, to increase angular resolution, lower frequency filters must be used. On the other hand, if the filter frequency is too low, the responses of the filters are more affected by the fundamental shape frequencies of the object. Thus, the dimensions of the object appear as peaks at filter responses. The decision for the parameters used for Gabor filtering was based on the paper from Kämäräinen J. et al. [27].

For future consideration, the Gabor filters could be systematically verified for the set of images with the range of spatial frequencies from 0 to π and with angular resolutions higher than $\pi / 6$.

Before being filtered, all images were normalized to the zero mean and unit variance in order for the filtered results to be comparable. From **Figure 30a** – **Figure 30g** results of images of confluent endothelial cell cultures after Gabor filtering can be seen. White regions show higher signal output and thus higher bursts of energy. Black stripes seen in different orientations result from oriented Gabor filter applied to an image. As previously mentioned, endothelial cell cultures do not tend to orientate in a specific direction. As a result, the signal output on all result images (filtered with differently oriented filters) appears to be equal. On the other hand, result images of fibroblast cell cultures in **Figure 31a** – **Figure 31g** show distinctively stronger output signals on images filtered with Gabor filters in direction 0 and $\pi / 6$. From the original image in **Figure 31**, it can be noticed that the pattern of the fibroblast cell culture orientates vertically and slightly towards an angle of $\pi / 6$. This is how filtered results can be interpreted. **Figure 30a** and **Figure 31a** show the algebraic sum of all endothelial and fibroblast cell culture filtered images, respectively.

In **Chart 3**, mean energy values of the filtered images in each orientation are depicted. In four out of six cases, fibroblast images indicate higher signal energy in comparison to endothelial images. Moreover, fibroblast results show higher scatter over the mean. This results from the pattern orientation of each original image. The fibroblast cell culture has a distinct oriented pattern, which can be in any direction. Thus, high standard deviations among filtered images of fibroblast cell culture exist. Endothelial cells are random (in no distinctive direction). This is why the energy of the output signal is equally distributed around the mean and no great scatter can be noticed.

According to the Fisher criterion in the 6-D feature space, energy of the response on each filter orientation is a feature. An equation of a hyper-plane ($w' * x - b = 0$) that would divide confluent endothelial and fibroblast patterns was attempted to be created. The hyper-plane equation

obtained is: $w' = 1.0e-005 * [-0.2968 \ 0.1838 \ -0.3228 \ 0.1256 \ 0.6013 \ -0.1381]$ and $b = -0.7024$.

The optimal separability value according to the Fisher criterion is 8.0682.

Although Gabor filtering gives very good results in discrimination between textures, the method is computationally expensive. Faster computation was achieved by filtering in the frequency domain because convolution is then simple algebraic multiplication.

4.3. Cellenger® approach

Many of the texture description features presented so far (the Fourier power spectrum, the statistical approach, Haralick texture coefficients as well as Gabor filtering) are interrelated. As mentioned, the Cellenger approach uses another method to distinguish both cell types. First, a multi-resolution segmentation with a scale parameter of 20 and gray value weight of 100% was applied (see **Figure 21** and **Figure 22**). Through multi-resolution segmentation each image was subdivided into 5000-8000 object primitives (**Chart 4a** shows the number of segmented objects), which were thereafter compared according to different features provided by the Cellenger® software. Texture information of images was included in the multi-resolution segmentation where regions with similar pixels were merged into object primitives. Afterwards, only physical features (shapes) were measured by the provided features. Therefore, this method includes texture information and the shape information of smaller units indicative for a certain image (pattern). As later seen, the Cellenger® program achieved the highest separability measure between both confluent cell classes.

But first, some notes on the selection of features to compare object primitives of both confluent cell culture image sets follow. The first feature compared was the number of segmented objects. The segmentation algorithm merged pixels and small objects into larger ones according to their intensity resemblance. From **Chart 4a** it is evident that the number of segmented object is different for both cell cultures. However, the difference is not significant enough to allow placing a simple threshold for separating the cells patterns. The higher number of segmented objects for fibroblast cell culture proves the assumption that its pattern has high intensity value variance. Thus, only smaller segmented objects can be built for the specified merging tolerance. The *Length / width*

ratio is used because the visual inspection of both cell patterns shows that fibroblast cell cultures show lengthier and thinner structures than endothelial cell cultures, which show curved and symmetrical shapes. Afterwards, the length / width ratio should be able to distinguish between the object primitives shapes of both cell cultures. The results in **Chart 4b** numerically prove the assumption beforehand. The ratio is significantly higher for fibroblast cells because the length is longer and the width is shorter than for endothelial cells. A threshold can be set at a value of 2.5. Images over the threshold are classified as fibroblasts. Images below the threshold are classified as endothelial cells.

Compactness is the similarity of an object to a square or rectangular. Each regular square (or rectangular) has a compactness value of 1. As compactness deviates from 1, shape becomes less square-like. Compactness of object primitives is shown in **Chart 4c**. Compactness also delivers significantly different results between both cell groups. A threshold of 2.37 possesses few misclassification errors. It should be noted, though, that the top and the bottom of the vertical dashed lines of box plots mark the 5th and 95th percentiles, respectively.

Elliptic fit is a measure of how well an object fits to an ellipse with the same area as the object. **Chart 4d** shows that fibroblast objects are significantly more elliptical than endothelial objects. It is evident that this feature is discriminative between the two classes. A threshold of 0.395 possesses few misclassification errors.

Shape index describes the fractal characteristic of an object. The more fractal an object appears, the higher its shape index. From **Chart 4e**, it is evident that fibroblast objects have a significantly higher shape index than endothelial objects. The explanation is that fibroblast patterns possess a random character of intensity values. Thus, more fractal objects are generated at segmentation as the algorithm searches for similar pixels at incremental distances until similar pixels are found. Therefore, pixels from the ridge are merged with some of the central pixels.

Density is similar to *Compactness*. The results in **Chart 4f** as compared to the results in **Chart 4c** show the similarity. A threshold of 1.37 may possess a low probability for misclassification.

Main direction is a feature which was many times mentioned in the thesis. The results in **Chart 4g** coincide with previously discussed orientation characteristics of both cell classes (Fourier spectrum and Gabor filtering). Fibroblast cell cultures show a high scatter in the main direction values whereas endothelial objects show a low scatter over the mean. Hence, a threshold may be set to 102 ± 5 , where all objects with a higher scatter are classified of fibroblast cell cultures.

Asymmetry is a feature which is very low for symmetric objects (e.g., a circle, a square). The

DISCUSSION

highest value of 1 is for asymmetric objects (e.g. a line). Therefore, asymmetry is a good measure for the assumption that fibroblast object primitives are lengthier and thinner than endothelial objects. The box plot in **Chart 4h** supports the assumption. Significantly higher asymmetry values were obtained for fibroblast object primitives. As a result, a threshold of 0.66 would possess low misclassification error. Objects below the threshold are probably of an endothelial image and those above the threshold belong to a fibroblast image.

Mean absolute difference of main direction and *SD of main direction* are custom defined features. The reason for the customization is to be able to show, that pattern stripes in fibroblast images are parallel whereas endothelial patterns do not show any distinctive parallelism. With these two features, the information about the main direction of the pattern could also be included. However, the results obtained in **Chart 4i** and **Chart 4j** are not impressive. Features are not distinctive for the cell groups. One explanation for it could be that the two custom features do not exactly measure for what they were customized to measure. Another explanation could be that segmented objects do not contain any information about parallelism.

At this point, information about the Fisher criterion is also given. Fisher criterion is set up in the 10-D feature space. Each of the dimensions represents one of the previously discussed features, respectively. The equation of a hyper-plane which divides both cell groups is: $w' = [0.0019 \quad -2.4772 \quad -7.3031 \quad 97.4784 \quad 11.6643 \quad 108.76 \quad 0.0553 \quad 167.11 \quad 0.0254 \quad -0.0076]$ and $b = -3.1993$. The optimal separability value according to the Fisher criterion is 9.9924.

As it is evident from the given discussion, many features are interrelated and give similar results. Therefore, for needs of classification, a reduced feature space should be proposed. Only those features which delivered significantly different results between cell groups should be used. From those proposed features, only features which describe different physical object characteristics should be used to set up a feature space. However, it has been shown that the Cellenger approach delivered the highest separability value; Gabor filtering gave promising results, but the method needs further experimentation and refinement. Hence, the Cellenger® should be used in an automatic objective image recognition unit. But, the Cellenger® program has some drawbacks. Except for the company's manual and some conference publications, there does not exist relevant documentation about or proper help for using the program. Feature definitions are often vague. Hence, much work was performed on a trial-and-error basis. Some of the features did not show reproducible results when computed with other methods (e.g. classic programming). On the other

hand, Cellenger® is friendly to use and good results can be quickly achieved also without special programming experience. The program is very suitable for biological laboratories where people do not want to waste their time on developing complex solutions for their image recognition problems. They need a program which is user-friendly and gives satisfying results with little effort.

4.4. Microscopy

Images analysed during this work were taken at different levels of brightness. The results were still representative enough to separate images into four different groups. It seems that analysis methods are robust enough to tolerate images with different levels of brightness. However, images should be bright enough (not too dark or overexposed to the light) in order for a healthy human eye to be able to distinguish important structures and objects within an image.

After an image has been shot, it is advisable that a contrast enhancement with e.g. Zeiss' Vision software is applied before image is saved. This pre-processing step is necessary for some image analysis. The image should also be normalized to zero mean and unit variance, especially for later statistical computations.

Images can be taken as RGB or gray-value because the colour information in phase contrast microscopy is not important. It is important, though, that the all images are taken at the same resolution. All images analysed in this work, are in resolution 1030×1300 pixels. No experiments with images at other resolutions were performed.

For future consideration, proposed analysing algorithms for different resolutions should be examined for robustness.

In conclusion, the quality of the image is important. The resolution of an image should be set to default value of 1030×1300 pixels. In addition, the focus and brightness on the microscope should be adjusted in order for a human eye to easily recognize essential image data. After the image is shot, contrast enhancement is performed. Then, the image is saved in any suitable format (e.g., '.jpeg').

5. CONCLUSION AND OUTLOOK

Automated feature extraction and object recognition are large research areas in the field of image processing and computer vision. Many uses for automated image analysis methods for acquiring numerical information from medical and biological images have also been reported [19 and 22]. However, no report on automated classification of cell cultures has been found in the literature. Automated objective classification is attempted and achieved in the present thesis.

No human or automated recognition is possible without preliminary knowledge. Decisions about classes or groups to classify recognized objects into are based on such knowledge. Knowledge about cell morphology and experiences based on visual inspection of endothelial and fibroblast cell cultures were required to develop different description techniques for extraction of a sufficiently large collection of feature descriptors from their images. One approach for using these descriptors is to “teach” representative descriptor values for a set of different cell culture morphology features to a system. The features of an unknown culture are subsequently determined by how closely its descriptors match those stored in the system memory.

Many description techniques were proposed and tested in the frame of this work. In the first stage of classification, the aim was to separate cell culture images according to their confluence character. From all methods proposed, Haralick texture coefficients delivered the highest separability value. In the next step, classifying images of confluent endothelial and fibroblast cell cultures was attempted. For this purpose, the methods of Fourier descriptors, Gabor filtering and shape description technique using Cellenger® were applied. The latter method delivered results with the highest separability value. However, its effectiveness would increase when fused with the results of other techniques by means of fuzzy logic.

The outlook for further work is to experiment and verify the robustness of the proposed classification system. Moreover, the feature space should be completed for other cell types and integrated into an automated recognition unit.

6. BIBLIOGRAPHY

1. Baatz M., Schäpe A. [2002]. "Multi-resolution segmentation: an optimization approach for high quality multi-scale image segmentation", *XII. AGIT- Symposium*, Salzburg, Karlsruhe, Herbert Wichmann Verlag, pp. 12-23.
2. Bovik A., Clark M. [1987] "Texture segmentation using Gabor modulation/demodulation", *Pattern Recognition Letters*, Vol. 6, Sept., pp.261-267.
3. Definiens. [2001]. *Cellenger Developer Studio 4.02*, User guide.
4. Freshney R. I. [1986]. *Animal cell culture: a practical approach*, IRL Press Limited.
5. Gonzalez R. C., Woods, R. E. and Eddins L. S. [2004]. *Digital image processing using Matlab*, Prentice Hall, Upper Saddle River, NJ.
6. Gonzalez R. C., Woods R. E. [2002]. *Digital image processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.
7. Haralick R., Shanmugan K. and Dinstein I. [1973]. "Textural features for image classification", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, Nr. 1, pp. 610- 621.
8. Jain A. K. [1989]. *Fundamentals of digital image processing*, Prentice Hall, Englewood Cliffs, NJ.
9. Jain A. K., Farrokhnia, F. [1991]. "Unsupervised texture segmentation using Gabor filters", *Pattern Recognition*, Vol. 24, Nr. 12, pp. 1167-1186.
10. Jain A. K., Ratha N. K. and Lakshmanan S. [1997]. "Object detection using Gabor filters", *Pattern Recognition*, Vol. 30, Nr. 2, pp.295-309.
11. Jain A. K. and Dubes R. C. [1988]. *Algorithms for clustering data*, Prentice Hall, Englewood Cliffs, NJ.
12. Kapitza H. G. [1997]. *Mikroskopieren vom Anfang an*, Carl Zeiss Verlag, Jena GmbH.

13. Kruizinga P., Petkov N. and Grigorescu S.E. [1999]. "Comparison of texture features based on Gabor filters", *Proceedings on Image Analysis and Processing*, 10th International Conference Venice, Italy, pp. 142-147.
14. Kyrki V., Kamarainen J. and Kälviäinen H. [2001]. "Content-based image matching using Gabor filtering", *Advanced Concepts for Intelligent Vision Systems Theory and Applications*, 3rd International Conference Baden-Baden, Germany, pp. 45-49.
15. Lanza R. P., Langer R. and Vacanti J. [2000]. *Principles of tissue engineering*, 2nd ed., Academic Press.
16. Mohr R. [2003]. *Statistik für Ingenieure und Naturwissenschaftler: Grundlagen und Anwendung statistischer Verfahren*, Expert Verlag, Renningen.
17. Murray R. S. [1990]. *Statistik*, 2nd ed., Schaum's Outline, McGraw Hill Book Company GmbH, Hamburg.
18. Salari E. and Ling Z. [1995]. "Texture segmentation using hierarchical Wavelet Decomposition", *Pattern Recognition*, Vol. 28, Nr. 12, pp. 1819-1824.
19. Smutek D. et al. [2003]. "Image texture analysis of sonograms in chronic inflammations of thyroid gland", *Ultrasound in Med. and Biol.*, Vol. 29, No. 11, pp 1531-1543.
20. Sonka M., Hlavac, V. and Boyle R. [1999]. *Image processing, analysis, and machine vision*, 2nd ed., International Thompson Publishing Company, USA.
21. Szeto L. K., Liew A. W., Yan H., Tang S. [2003]. "Gene expression data clustering and visualization based on a binary hierarchical clustering framework", *1st Asia-Pacific Bioinformatics Conference*, Adelaide, Australia.
22. Urbani M. [2004]. *Computerunterstützte und automatische Aufnahmen von Transmissions-Elektronenmikroskopischen Aufnahmen der Leber*, Dissertation, Ludwig-Maximilians-University, München, Germany.
23. Walker R. F., Jackway P. and Longstaff I. D. [1995]. "Improving co-occurrence matrix feature discrimination", *Proceedings of Digital Image Computing: Techniques and Applications*, 3rd International Conference Brisbane, Australia.
24. Young I. T., Gerbrands J. J. and Van Vliet L. J. [1998]. *Fundamentals of image processing*, Delft University of Technology.

INTERNET SITES

25. <http://teaching.anhb.uwa.edu.au/mb140/MoreAbout/Endothel.htm> accessed on 17/03/2005.
26. Kämäräinen J. K. [2003]. *Feature extraction using Gabor filters*, Dissertation, Lappeenranta University of Technology. Viewed at <http://www2.lut.fi/~jkamarai/publications/downloads/thesis.pdf>, accessed on 11/04/2005.
27. Kämäräinen J. K., Kyrki V. and Kälviäinen H. *Fundamental frequency Gabor Filters for Object Recognition*, Department of Information Technology, Lappeenranta University of Technology. Viewed at <http://www2.lut.fi/~jkamarai/publications/publications.html>, accessed on 29/03/2005.
28. Kämäräinen J. K., Kyrki V. and Lindth T. *Signal discrimination based on power spectrum of filter response*, Department of Information Technology, Lappeenranta University of Technology. Viewed at <http://www2.lut.fi/~jkamarai/publications/publications.html>, accessed on 19/03/2005.
29. Kyrki V., Kälviäinen H. and Kämäräinen J. K. *Invariant Shape Recognition Using Global Gabor Features*, Department of Information Technology, Lappeenranta University of Technology. Viewed at <http://www2.lut.fi/~jkamarai/publications/publications.html> accessed on 19/03/2005.
30. Kyrki V., Kämäräinen J. K. *Simple Gabor feature space for invariant object-recognition*, Department of Information Technology, Lappeenranta University of Technology. Viewed at <http://www2.lut.fi/~jkamarai/publications/downloads/laitosrap83.pdf>, accessed on 01/04/2005.
31. Principe C. J., Dongxin X. and Wang, C. *Generalized Oja's rule for linear discriminant analysis with Fisher Criterion*, Computational NeuroEngineering Laboratory, University of Florida. Viewed at <http://www.cnel.ufl.edu/files/1023120412.pdf>, accessed on 24/02/2005.

BIBLIOGRAPHY

32. http://staff.science.uva.nl/~jverbeek/teaching/mlpr/slides/6.Linear_Discriminant04-05.pdf, accessed on 05/04/2005.
33. <http://cmp.felk.cvut.cz/~xfrancv/stprtool/stprtool.pdf>, very useful for use Statistical Pattern Recognition Toolbox, accessed on 05/04/2005.
34. Eblenkamp M. et al. [2004]. "Umbilical Cord Stromal Cells", *Orthopäde*, Vol. 33, Nr. 12, pp.1338-1345.

7. ACKNOWLEDGMENTS

I dedicate this work to the community and *Studentinnenheim St. Michael, München* for a very precious support during my stay in München. Thank you *sister Adelgart!*

I would like to express my sincere gratitude and appreciation:

To *Dr. Otto Likar* who generously sponsored my 10-month stay in München with the scholarship of Otto and Karla Likar's foundation.

To *Slovene scientific foundation* for a financial injection when it was the most needed.

Next fold of acknowledgements goes to people who helped me with their valuable advices considering the work itself:

To *Dr. Markus Eblenkamp* for his welcome to his young team and for the supervision of my work.

To *Dipl. Ing. Matthias Pigerl* for his friendship and for being the best working colleague. I am grateful for his patience and willingness to explain me some things 10 times if needed.

To *Uschi Hopfner, Sepp Hintermaier* and *Dr. Joachim Aigner* for their help in a biological lab.

To *Prof. Dr. Brigitte Forster* for her valuable help in solving some abstract mathematical problems.

To *Prof. Dr. Erich Wintermantel* and *the whole team* at the Zentral Institut für Medizintechnik for making my stay possible and enriched with friendly atmosphere.

To *Dr. Maria Athellogou* for help at the work with the Cellenger software.

To *Brian Lambert* for the extensive language revision.

To *Prof. Dr. Stane Kovačič* for his mentorship and his contribution of helpful advices which encouraged me to continue my work in the right way.

To *Prof. Dr. Janko Drnovšek* and his team for being so hospitable to welcome me and gave me the courage to continue my post gradual studies.

To *Dr. Jure Skvarč Dipl. Ing.* for valuable discussions and tips about the topic.

And last but not least special thanks go to *my home colleagues* who helped me out from situations I thought them to be impossible.