Analysis of spectral images of skin for medical application

Odenna Sagizbaeva

26.11.2004

University of Joensuu
Computer Science
Pro gradu

Preface

I would like to thank **Professor Jussi Parkkinen** for the interesting subject to research and also for his supervising of my research work in the **University of Joensuu**. His technical comments and suggestions helped me many times during the work on the thesis. Further, I would like to say thanks to the people from the **Color group** of the **University of Joensuu**, who advised me in my first steps into the Color science.

The great thanks belong to my **mother** and **sister**, they have been supporting me all the time and I would never have finished the thesis without theirs countenance.

Finally I would like to thank all my **friends**, who have been surrounding me during my studies.

Thank you all.

Announcement:

I announce that I worked up on the thesis by myself under the supervising of the Professor Jussi Parkkinen. I included all the references, which provided information to me.

Joensuu, 2004-11-26 ……..……..

Odenna Sagizbaeva

Abstract

Nowadays color images of objects are widely used in different scientific fields. Studying of the color images helps to make creations, which are impossible without them. There are many works, which made research of human's skin color images for the color reproduction of color film and color television systems; it is known also that some works appeal to skin in order to make skin recognition. At the thesis the multispectral images of human skin for medical application are studied. The analysis of spectral images of skin using Independent component analysis is made. New methods for conversion of spectral images, which are studied by the ICA method, are obtained. Using the FastICA algorithm independent components of skin are obtained, and analysis of them is made.

Keywords:

Analysis of the spectral image of skin, Independent component analysis of skin, FastICA algorithm, independent components of skin.

Аннотация

В наше время цветные изображения широко используются в различных областях науки. Благодаря цветным изображениям зачастую становится возможно сделать открытия, которые были бы невозможны без них. Цветные изображения поверхности кожи человека также широко исследуются в различных областях науки, например таких как репродукция цветных фотографий, цветное телевидение или в задачах распознавания кожи человека. Данная работа посвящена изучению мультиспектральных изображений поверхности кожи человека для медицинских приложений. В течение работы над дипломом был проведени анализ спектральных изображений кожи человека при помощи метода Независимых компонент. Были получены методы конверсии спектральных изображений, к которым был в дальнейшем применен статистический метод Анализ независимых компонент. При использовании алгоритма FastICA, в дипломной работе получены и проанализированы независимые компоненты кожи человека.

Ключевые слова:

Анализ мультиспектрального изображения поверхности кожи, Анализ независимых компонент кожи, FastICA алгоритм, независимые компоненты кожи.

Contents

Nomenclature

CIE	Commission Internationale de Eclairage
D65	artificial daylight with correlated temperature 6504K
FastICA	Fast Independent Component Analysis
FP	FastICA with deflation decorrelation approach
Fpsym	FastICA with symmetric orthogonalization decorrelation approach
FPsymth	FastICA using tanh nonlinearity with symmetric ortogonalization
ICA	Independent Component Analysis
ICs	Independent Components
ML	Maximum likelihood
MV	Mixture Variables
Nm	Nanometers or billionths of a meter
PCA	Principal Component Analysis
RGB	Green, Red, Blue
UV	ultraviolet wavelengths

Variables and constants:

a_{ij}	elements of a mixing matrix
A	unknown mixing matrix
C_X	covariance matrix of the X , $C = E\{XX^T\}$
$H(X)$	entropy of a random variable x
I	unit matrix
$H(X)$, $J(x)$	entropy and negentropy of the random variable x
m_X	sample mean vector of X

$P(i, j, k)$	third dimensional matrix, multispectral image of skin
$Q(c, d, k)$	submatrix of matrix $P(i, j, k)$
$R(u, v, k)$	resized matrix from $P(i, j, k)$
R_X	correlation matrix of the random vector X
s_i	random independent components
x_i	random mixture variable (Also: a coordinate)
$X^{'}$	centered an observable random vector
X_{gauss}	Gaussian random vector
$X(t, k)$	matrix of mixture variables
Y	white vector

Functions:

$E\{\}$	mathematical expectation
$f()$	conversion function
V	linear transformation to white vector, whitening matrix
δ_{ij}	variance of i, j

Other notation:

$\angle$	angle
Oij	coordinate plane
$\overline{1, n}$	from 1 to n
$\leftarrow$	substitution
$\propto$	proportional to

1 Introduction

Skin color studying may be considered as important problem for diagnostics of skin diseases. With the recent progress of spectral imaging, computer graphics and some modern computational applications, the diagnostics of the skin diseases on the base of mathematical modeling of the skin becomes increasingly possible and correct on early stages.

There are many medical and industrial fields, for which accurate skin models and statistical results may be useful. As it is known, in dermatology, skin models can be used to develop methods for computer assisted diagnosis of skin disorders. In the pharmaceutical industry quantification is quite useful when applied to measuring healing progress; such measurements can be used to evaluate and compare treatments and can serve as an early indicator of the success or failure of a particular treatment course. Consumer products and cosmetic industries can use computational skin representations to substantiate claims of appearance changes [27].

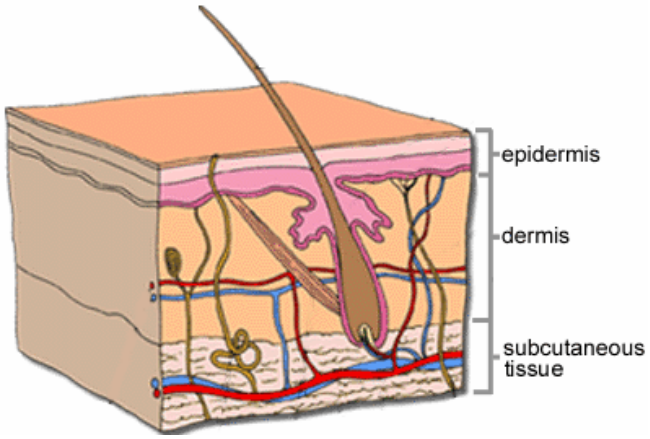


Figure 1. Skin layers.

Human skin is the turbid media with multi-layered structure [62]. Various pigments such as melanin, hemoglobin, bilirubin, etc. are contained in the media and influence on the skin color. Some techniques allow studying each pigment, skin component, separately. The Independent component analysis (ICA) is a technique that extracts the original signals from mixtures of many independent sources without priori information on the sources and the process of mixture [60].

There are several applications of the method in medicine and other scientific fields. In the thesis the ICA model is applied to skin color images, which were obtained in the Color laboratory of the

University of Joensuu. Skin images can be considered as multispectral images, i.e. images with number of spectral channels exceeding the three. In the thesis such type images are studied.

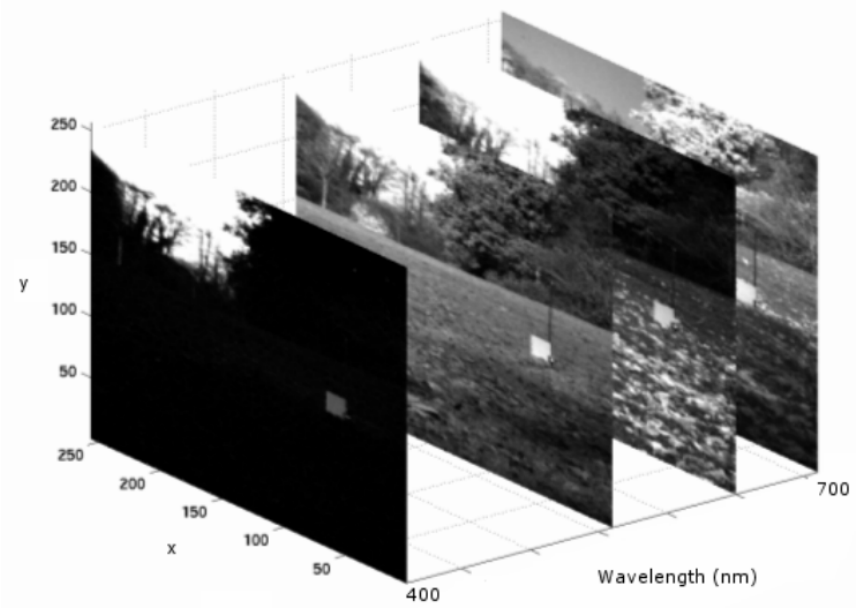


Figure 2. Example of spectral image [36].

For the convenience of readers let us to give a summary of the thesis. The thesis “Analysis of spectral images of skin for medical application” consists of five chapters of research work, the oddments, references and appendices.

Chapter 2 represents a theoretical part; it introduces a reader to historical facts of appearing of radiology and other methods of diagnostics by human body’s images study. In the chapter it is also shown that new technologies allow studying of diseases deeper and diagnose more accurate. The importance of skin cover is considered there, and examples of the skin diseases affecting skin color and structure are given. The skin reflectance properties are discussed in the chapter, the skin model based on analysis of [17] is also presented. The pigments defining the skin color are determined in order to acquaint a reader with the facts of skin optics.

Chapter 3 describes the ICA method, which is studied during the research work. The history of ICA and its medical applications are discussed; definitions, restrictions and assumptions of the model are presented in the chapter. In order to make algorithms of the method simpler the assumption that both mixture variables and the independent components have zero mean is maid. The observed variables are centered before applying of the ICA. An advantage of the whitening is a transformation of a mixing matrix into an orthogonal matrix, and whitening simplifies the problem,

because helps to estimate half of parameters. The preprocessing, which includes centering and whitening of the observed variables, is considered in the section, and the necessity of the preprocessing is clarified there. In conclusion of the section, the FastICA algorithms are discussed and presented.

Chapter 4 is an introduction in Spectral imaging; there are many definitions of different types of spectral images, illuminations in the chapter. Also in the section, an application of the linear interpolation for conversion of radiance type images to reflectance type images is shown. Two algorithms of obtaining the matrix of mixture variables are offered, and an implementation of them is presented in the fourth chapter.

Chapter 5 represents tests description and analysis of results, obtained during the research. In the chapter some statistical approaches are used in order to prove one or another characteristic of independent components. Quantity of independent components, to be observed in the thesis, was chosen in the section. The FastICA algorithm with symmetric orthogonalization was established as the best estimation approach for the component obtaining. At the end of the chapter the separation of three independent components of skin color image is made. Each component was presented as a spectrum, and separated colors were shown in RGB format.

Discussion of the work results and conclusions of the thesis are presented in the sixth chapter.

There are brief summaries after every chapter. In the thesis the cross-references are used. During the work at the thesis more than 60 articles and books were read and used, the most important of them are presented in the thesis. Appendices include illustrations, data, algorithms, and results of the tests using during the work on the thesis. The thesis was written in accordance with reference book on dissertation’s writing [66].

2 Spectral Skin Images in Medical Diagnostics

The old Latin proverb says: “Diagnosis cetra — ullae therapiae fundamentum” (Certain diagnosis forms the basis of medical treatment).

Medicine often appeals to diagnostics of diseases by researching of human body organs’ and whole organism’s images. For that there are many different methods and equipment. Contemporary technology allows researching almost of all human organs and systems, all anatomical formations, which sizes are bigger than microscopical ones. There are many skin diseases and some of them affect skin color and structure, what makes possible diagnostics of them at the basic stages.

2.1 Methods of Medical Diagnostics Imaging the Human Internal Organs

During many ages doctors tried to solve the problem: improve recognizing of human diseases. Need for method, which would give them to look into the human’s body without injuries, was immense. Many years ago information about normal anatomy and anatomical pathology of men was based on dead bodies’ study. With the advent of computer technologies more possibilities for human internals’ images analysis appeared. New era came with X-ray photography. Earlier, during analysis, a radiologist bet on his vision. As known, human eyes are not very sharp. Then new methods, which allow taking of internals’ images, appeared, among them are tomography, ultrasonography, nuclear magnetic resonance tomography and others. And now there are many systems, intensifying images, displaying them on a monitor or keeping them on the hard disk and transferring the data. Also digital monitors of new diagnostics systems allow getting of an image any studying part of body in wanted scale and aspect, what helps to diagnose a problem quickly and qualitatively. Application of these methods and X-ray to medical science results in Radiology science, which concerns of imaging of the human body on a screen, paper, film, and then decoding of the images, makes the determining of organism’s condition possible. The history of radiology written above is derived by analyzing information from [19, 57 and 59].

Some methods of diagnostic by using image analysis is presented in the thesis, the first is the X-ray photography.

X-ray is a form of electromagnetic radiation, just like visible light. In a health care setting, X-rays are emitted by a machine as individual "particles" (photons) that pass through the body and then get detected by a sensitive photographic film. X-ray allows finding of bone and muscle fractures, because they have dense structure that blocks most of the photons and appears white on developed film.

Another method, which uses the ultrasound waves for obtaining images, is an ultrasonography. Ultrasound is sound waves, which frequency higher than 20 kHz. This method allows displaying two dimensional images of internal organs, estimating theirs' size, shape and structure. Using the ultrasound doctor can estimate amount of lesion of an organ without laboratory tests.

The following method of body organs' photography is close to X-ray diagnostic. Fluorography is based on photographing of shadow image by mounting a camera in front of the fluorescent screen. It may be used for comprehensive body examination especially for gastrointestinal and locomotor's systems including pelvis, chest, spine, worm etc.

Developing of X-ray research allowed decreasing of the exposure time and improving of the image quality. But radiation diagnostics gives only sections of organs images on any depth, what presents the whole object, but not a visualization of it. Computers made this problem solvable because of 3-D reconstruction of the patient's body. So computers allow recognizing of more diseases than human eyes.

In 70's a new method of diagnostics based on photographing whole human body or its parts was developed. Computer processes images of sections and display results on the screen. This method is widely used in diagnostics of brain-growth and soft tissues in the organism. Hence the computer tomography can be added to newest methods of diagnostic.

The following picture shows an example of images, which are made by using X-ray. The X-ray on the left picture 3.a shows a normal heart. The heart on the right X-ray, picture3.b, is enlarged [26].

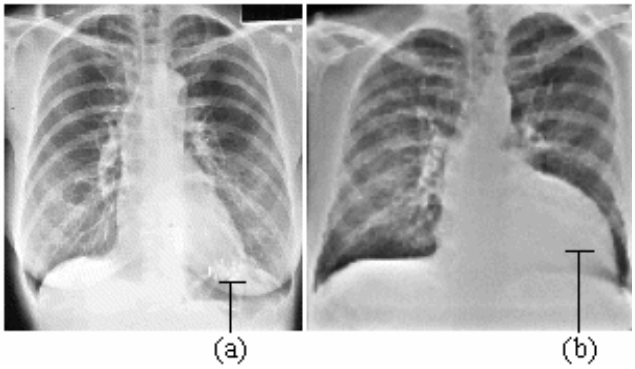


Figure 3. X-ray images of a heart [50].

2.2 Diseases Affecting Skin Structure and Color

Skin is a natural protective body covering of a human, an outer surface of a body, site of the sense of touch. The skin thermoregulates an organism, executes secretor and exchange functions. The skin consist of two layers: an outer epidermis and an inner sensitive dermis composed mostly of connective tissue. Also the skin helps organism to be resistant to dust, wind, humidity, sudden changes of temperature, chemical irritants, touch and rubbing.

While skin is the most accessible organ to research, the diagnostics of skin presents difficulties because of many clinical forms (about 2000) and rare diseases.

There are many diseases affecting structure and color of skin [26]. Some skin diseases are determined by modifications of the skin. Among them is furuncle that is an infection of a hair follicle, a painful sore with a hard pus-filled core. Furuncles may occur in the hair follicles anywhere on the body, but they are most common on the face, neck, armpit, buttocks, and thighs. Another one is eczema that is characterized by the presence of redness and itching, an eruption of small vesicles, and the discharge of a watery exudation, then crusty and thickened skin. Very often an eruption appears on hands, feet, shank and less on a body. A flowing of the eczema distinguishes by lingering. A virus infection, herpes can affect on skin structure. The most common symptoms of herpes are blister or clusters of painful blister-like sores. After the penetration the virus is kept safe in the human organism for whole life as latent infection, which may remain in sensory nerve cells. Under the influence of some diseases herpes can appear again. Harmless, at first glance, changes of skin can be a demonstration of serious illness. There some skin diseases are considered in the thesis.

Some of them become apparent on early stages by changing skin color and morphology. Exactly these changes allow to a doctor to discover disease.

Pigmentation is interesting by changing the color of skin to different colors. Pigmentation is formed during first months and years of life and depends not so much on pigmented cells number as on theirs dynamic abilities. Young people with white skin have proportional distributed pigmentation. At an early age a blood supply is good. The skin of seniors is become thicker and a blood supply is become worse. Moreover, dry old skin admits the light deeper inside, what is a reason of the sallow color of the skin. The distribution of the pigment also is broken, different size chloasmas are appeared in the skin. Also very often an illness affected the skin color.

An allergy influences on skin color depending on allergens. Causes of an allergy are food, drugs, bacteria and many other factors. An allergy is a response, such as rash, hives, itching, asthma, hay fever, etc., to an allergen.

And on the contrary, a vitiligo is a rare skin disease consisting in the development of smooth, milk-white spots upon various parts of the body. Cause of the vitiligo is incapacity of some parts of skin for making a pigment. The illness is almost incurable.

Dermatitis is not uncommon disease, it is a skin disorder and inflammation, producing a rash, sore, cracked skin, and skin becomes itchy and may develop blisters. Doctors discern also a contact dermatitis, which appears due to direct contact with an irritating substance, such as berries or chemical goods. The skin inflammation appears immediately, and an affected area of skin corresponds to a contact area.

Often, a fire, an electric, a radiation or chemical effect or intense heat have affect on human skin. Theirs actions cause a born disease. The gravity of injury depends on temperature, duration of an effect, area of injury and location of the burn.

Chronic skin disease psoriasis, characterized by dry red patches covered with scales, silvery, flaky surface. The primary activity leading to psoriasis occurs in the epidermis, on the scalp and ears and genitalia and the skin over bony prominence. The skin reacts on the infection by growing very fast, trying to "grow" the infection off the skin. The skin also does not mature normally.

Seborrhea produces a disruption of the oil gland working, which is characterized by a skin fat secretion. Seborrhea can appear on any part of skin, where the oil gland is much: scalp, face, back. The skin is thickened and looks dirty gray; pores are quite often with blackheads.

These and other skin diseases can be found from Medical Encyclopedias [26]. Also it is necessary to notice that not only skin diseases have affect on skin color and structure, diseases of many body organs and general state of human health have an influence on skin condition.

2.3 Color of Human Skin

Skin color is one of the most conspicuous ways, in which humans are varying, and which is widely used to define human races. Men have adapted over a long period of time to local conditions, and now we can divide them into races on the basis of geographical distribution. People, living in the tropic zone usually have dark skin and, contradictory, Northern Europe people have a fair skin.

The appearance of human skin is governed by the nature of the surface, by the epidermis, dermis, subcutaneous tissues and pigmentation in the hair shaft and follicle. Keratin lamellae are produced on the skin surface, melanin is produced beneath this horny layer, and yellow carotenoids are deposited deeper in the dermis. The number of blood vessels, the extent, to which they are filled, oxygen and pigmentation in the blood also influence on outward color of the skin. Skin color is an indicator of human well-being. For example, pallor color of the skin is caused by a diminished blood flow, by blood, which has low oxygen-carrying capacity and by fluid beneath the skin layers. On the contrary, redness is caused by too much oxyhaemoglobin, by increased blood flow, or by high temperature, as heat is dissipated from the surface capillaries. An appearance of the yellowness of the skin is explained by high serum carotene levels, yellow-brown or black patches on the skin by exposure to the sun, dispersion or certain specific diseases. A description of these and other kinds of pigmentation as cyanosis (blue appearance), chlorosis (greensickness) and other, which are caused by some changes in human organism and theirs effect on skin color, can be found in [31].

Skin coloration in humans is adaptive and labile. Because of this, skin coloration is of no value in determining phylogenetic relationships among modern human groups.

The major pigment is, however, melanin, and the skin color of all races can be related to the size, aggregation properties, and speed of response of the melanocytes, the cells, which produce the pigment melanin. Cotton *et al.* [17] claim that “The color of the skin is primary due two factors: melanin in the skin and the presence of blood.” In [56] the dependence of skin color from pigments of skin is clarified.

The chromophores, which are considered in next subchapter, are a chemical group that gives a color to a skin and determines the reflectivity of the skin at various wavelengths [67]. In the thesis only several chromophores, which are listed by authors in [2, 3 and 4] are presented.

2.3.1 Formation of Skin Reflectance

There are many techniques to calculate reflectance spectra of the human skin. To understand the development of coloring within the skin it is necessary to understand that skin is a complex structure, which has two structurally different layers (epidermis and dermis) with different scattering, refracting and absorption properties. Different cells structure and blood distribution, amount of chromophores and water content in the skin layers affect on their optical properties. The skin reflectance spectra simulation based on Monte Carlo method can be found from [49]. The model of color formation within human skin based on Kubelka-Munk theory is presented in [2] and [17], and the following figure presents the model of human skin, which is based on the theory.

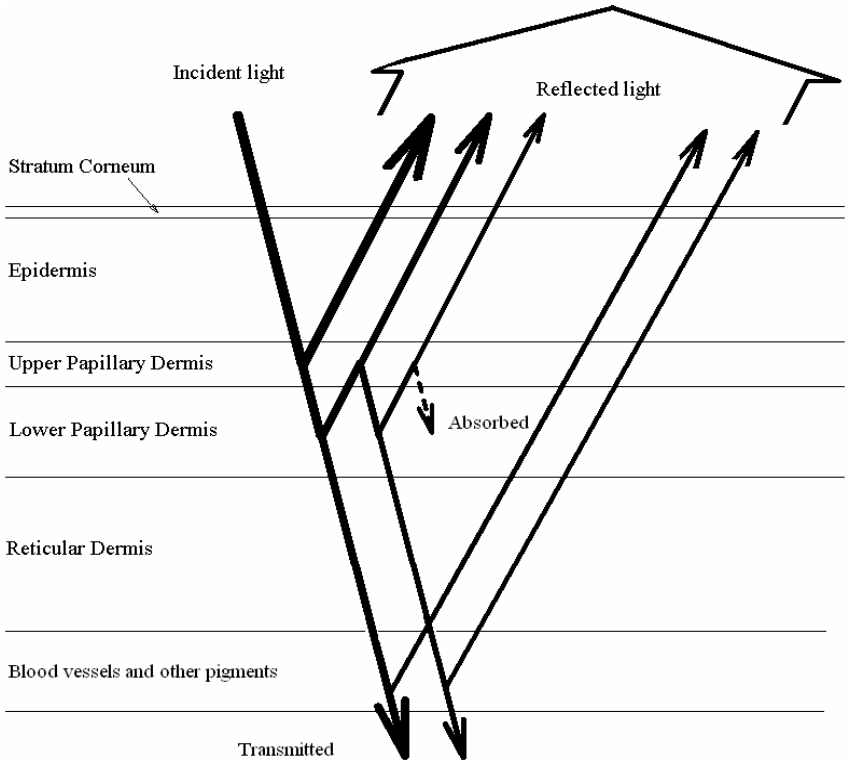


Figure 4. Model of the human skin.

At first, the light enters the epidermis. Melanin pigment absorbs short wavelength more strongly than long one, and resulting light entering the dermis will have lost a blue component, depending on the amount of the pigment. The epidermis contains varying amounts of melanin and keratinized epithelium cells. The surface reflection of the human skin takes place at the epidermis surface. The reflection is approximately 5% of incident radiation in the range 350-3000 Nm independent of the lighting wavelength [17].

Most of the light incoming to a dermis, other than that absorbed, is reflected. The dermis contrasts strongly in structure that the epidermis, it has different optical properties because of constructing from a densely fibrous collection of collagen fibers and blood vessels. The blood born pigments hemoglobin, oxy-hemoglobin, beta-carotene and bilirubin are the major absorbers. The optical properties of the dermis are basically the same for all human races [15].

The light then passes back through the epidermis, where further melanin absorption takes place before being reflected through the stratum cornea.

Melanin.

Melanin accounts for most of the variation in the visual appearance of human skin. Melanin comes in two types: phaeomelanin (red to yellow) and eumelanin (dark brown to black). A number of theories have been advanced to account for man's color via melanin pigmentation. These include its usefulness as camouflage, as an aggressive social signal, as resistance to sunburn and cancer, and as optimization of vitamin D_3 biosynthesis. Jablonski *et al*. [38] demonstrate that the relationship between skin pigmentation in indigenous human populations and latitude is traceable to the strong correlation between skin color and UV radiation. They also present evidence, supporting the theory that variations in melanin pigmentation of human skin are adaptive and that they represent adaptation for the regulation of the effects of UV radiation on deep strata of the integument. Melanin also can be defined as an optical and chemical filter. The penetration of light wavelengths is reduced by using melanin. In [38] a relation between skin pigmentation and vitamin D_3 synthesis is found. Vitamin D_3 is necessary for human growth, calcium absorption and skeletal development. According to Jablonski *et al*., humans with dark skin do not suffer from deficiencies of the vitamin D_3, that can be attributed to lack of sunlight, even they live whole year in a zone of low UV radiation.

Melanin reflectance spectrum in the visible range is monotonically increasing with wavelength, and maximum absorption is happened in the UV range. [5].

Keratin

Keratin is a fibrous protein that occurs in the outer layer of the skin (epidermis). It is formed inside the keratinocytes. Its fibers cause light scattering, but they are too thin to have a critical effect in skin reflectance [3].

Carotene

This dermis chromophore can be also present in the blood. Carotenes are yellow or orange-red fat-soluble pigments in plants. It presence in humans depends on food consumption. Its absorption peaks at 480 Nm, but it has a very weak effect on the overall skin color [3].

Collagen

Collagen is a fibrous protein that can be found in the dermis. Its fibers are the primary source of light scattering in the dermis, determining the depths to which these wavelengths penetrate the dermis, but authors of [3] and [5] do not mention any impact on the reflectance spectrum of the skin [3].

Hemoglobin

The dermis is deeply permeated with blood vessels, which contain hemoglobin (Hb). Hemoglobin is a protein containing in the red blood cells. Hemoglobin has a unique absorption spectrum with characteristic absorption bands at 420 Nm and in the 545 Nm – 575 Nm ranges [3].

The figure 5 shows the hemoglobin absorption spectrum versus skin reflectance, which is measured in [3, 5].

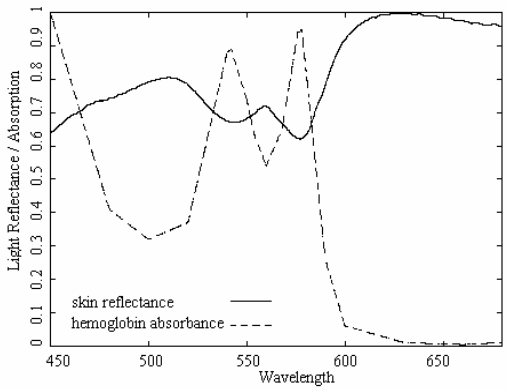


Figure 5. Skin reflectance versus hemoglobin absorption [3].

Bilirubin

Bilirubin is yellow orange-yellow pigment, the waste product that results from the breakdown of hemoglobin molecules from worn out red blood cells. Excessive levels of bilirubin stain the fatty tissues in the skin to the yellow. Along with the blood chromophores hemoglobin and oxy-hemoglobin, bilirubin determines dermal absorption of wavelengths longer than 320 Nm and profoundly modifies skin colors.

2.3.2 Differences in Human Skin Reflectance

Angelopoulou in [3] describes a modeling of skin reflectance data densely sampled over the entire visible spectrum. She measured the hands skin reflectance of different races' volunteers. Two different samples for each subject were taken: back of hands and palms. A light was falling approximately on the center of a hand. For further improvement of a signal she took the average of 10 reflectance measurements for each skin sample. She also kept in her mind that color of reflected light depends on the color of incident light. Thus, the true descriptor of the spectral behavior of a material is the ratio of the light reflected from that material over the light that is incident on that material. Figures 6, 7 are taken from the article and show the skin reflectivity (albedo) of various races, which is the ratio of the amount of electromagnetic radiation reflected by a skin surface to the amount incident upon it, commonly expressed as a percentage. Different races are shown by different colors: Caucasian is shown in red, Asian in green, East Indian in blue and African descent in magenta. All the plots exhibit a gradual increase with respect to wavelength with around 575 Nm.

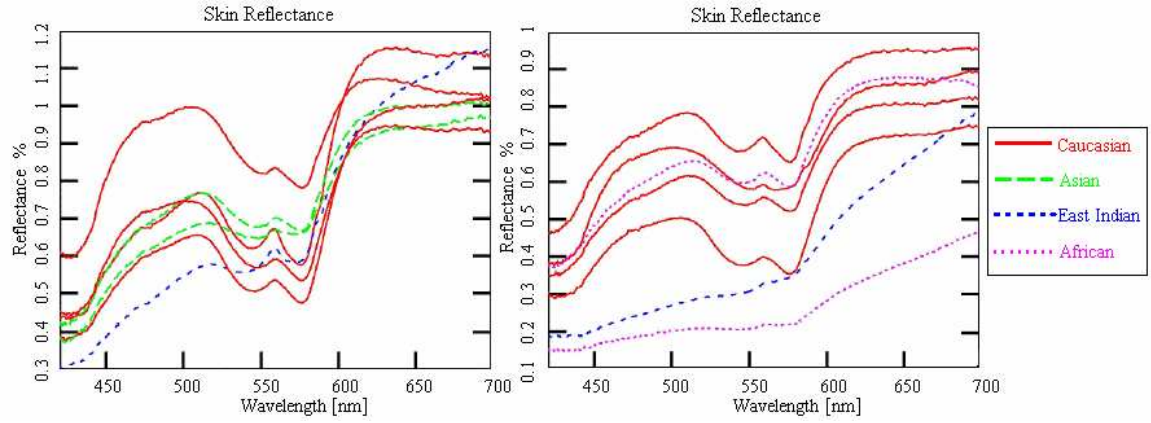


Figure 6. Reflectance spectra of the back of the hand for different races. [2]

In order to test whether skin spectra variations could be attributed to melanin and local skin structures (hair follicles, hair, pores, etc.) they measured the spectra of the palm of the same volunteers. As expected, the palm has a more reddish spectrum than the back of the hand, figure 7. Notice, that the various spectra are much closely clustered and exhibit an almost identical shape.

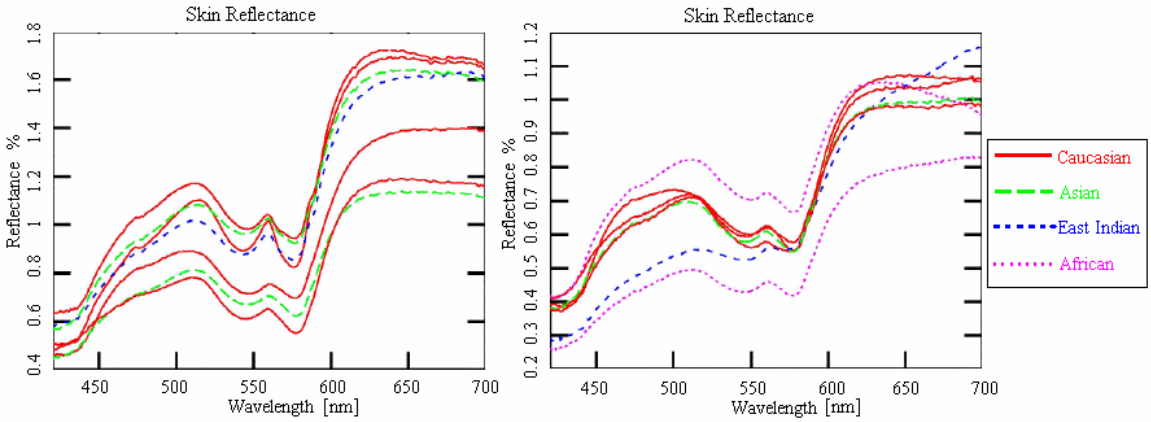


Figure 7. Reflectance spectra of the palm for different race. [3]

The figure 8 shows a subset of the previous curves on the figures 6, 7 after their albedo has been rescaled to vary between 0% and 1%. Angelopoulou *et al*. [4] divided each reflectance distribution by the maximum value of that distribution. So she ignored the effect of the darkness of skin and concentrated on the shape of the spectrum.

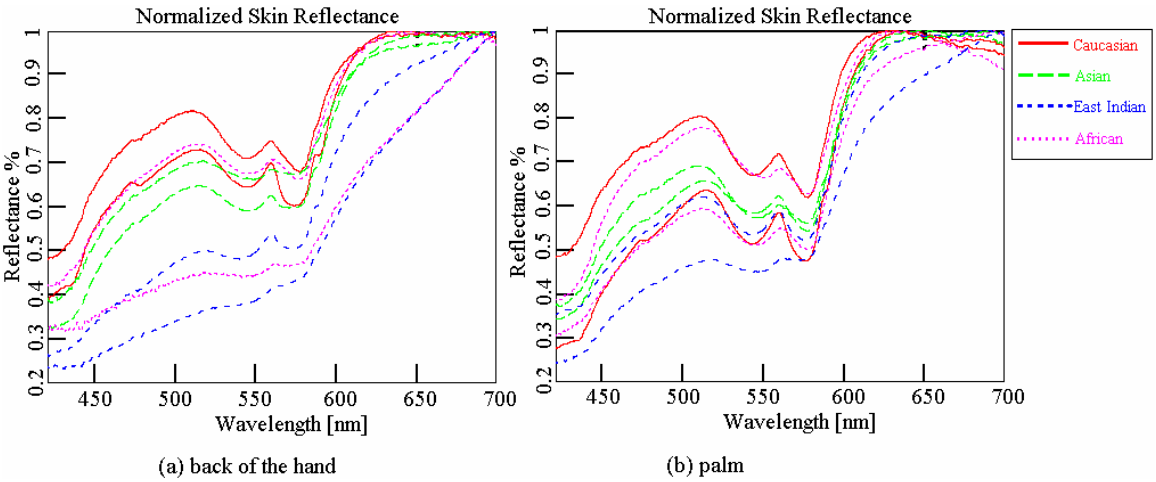


Figure 8. Scaled spectra of the (a) back of the hand, (b) palm [4].

Figures 8.a and 8.b show the rescaled spectra of the palm and the back of hands of the same subjects. Notice, that in the absence of melanin at the palm, the curvatures for the darker skinned subjects are the same as for the rest of curves. The measured spectra demonstrate that there is a specific pattern in the spectral distribution of the color of skin.

2.4 Possible Methods of Skin Analysis

The human skin is examined in different sciences. Researchers study different properties of the skin for applications in different spheres of the life. There are many kinds of methods and approaches to make a skin analysis. New technologies allow studying of skin color and morphology everywhere, not only in laboratories, but also in hospitals, banks and police.

The main role of the skin analysis is in the part of medicine – dermatology. Only by the skin analysis it is possible to make diagnostics of skin diseases. It is the most important part of skin research, but not the only one. Nowadays, at the era of anaplasty, surgeons very often appeal to skin study to improve rhytidectomy and skin implantation.

In criminal identification and security system, and not only there, face recognition systems are employed. There are many approaches to detect human faces. With computers invention most of them are accessible. At the work [30] a statistic approach of the detection of human faces in color nature scene is described. Many methods detect faces by feature matching. In [12] some of the methods are considered and compared. Also an examination of the markings, which are made from

the fingertips, is used wide in criminal identification system. This method of personal identification, which is called dactyloscopy, based on unique physical or behavioral characteristics of every human. Fingerprint examines a shape and detailed characteristics (minutiae) of skin ridges pattern on human hands.

The skin color modeling is widely used for recognizing faces and constructing a realistic face models. Also using reflectance measurements of the human skin it is possible to reconstruct three dimensional images of faces as it is described in [44 and 68]. Interesting works are in speech recognition by using visual information about the corresponding lip movements. For lips recognizing skin color information is used, two colors – skin and lips are marked out in order to improve recognizing [11].

Along with the methods, which are mentioned above, Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are widely used for recognizing faces [7, 8 and 9]. More information about PCA method and comparison of it with ICA can be found from [20], ICA method is considered in the thesis. Also PCA, as applied to skin color, is considered in [35]. Chapter 3 of the thesis represents ICA model description and implementation of it, and chapter 4 describes the application of ICA for spectral images of skin.

Brief Summary of Chapter 2

1. Historical facts of new methods of the diagnostics, which are studied images of body organs, are considered.
2. Some methods of the diagnostics by using image analysis, such as X-ray, tomography, ultrasonography, etc., are presented.
3. Properties of human skin are described, and several diseases affecting on skin color and structure are presented in the paragraph.
4. The model of human skin is constructed, and properties of skin pigments are studied.
5. Applications of skin analysis in different fields are described.

3 Independent Component Analysis

In the thesis spectral images of human skin are analyzed by the Independent Component Analysis (ICA) in order to extract and study information of components in the skin. ICA is a technique that extracts the original signals from mixtures of many independent sources without a priori information on the sources or the process of the mixture. At this chapter a basic conceptions of ICA are considered.

3.1 History of ICA

Hyvärinen *et al*. [32] refer that first mention about ICA technology was made in 1980s by J. Heraut, C. Jutten and B. Ans [6]. The requirement for the method appeared with coming problems in neurophysiologic setting. Researchers implemented ICA for different scientific fields. ICA has been applied to problems of array processing, communication, and medical signal processing and speech analysis. In the field of color image processing, Inoue *et al*. [36] proposed a technique to separate each pigment from compound color images. Useful works with applying the ICA are back-propagation, Hopfield networks [29] and Kohonen's Self-Organizing Map, also for higher-order spectral analysis [41]. The increased interest to the model took place from 1990s with the fixed-point or FastICA algorithm, which allows applying the method to large-scale problems because of its computational efficiency. There are many works, which compare the ICA with other methods, for example, Comon [16] discusses about the ICA, its effectiveness and applications of the model. Due to its generality the ICA model is widely used in many different areas, but in the thesis only medical applications of ICA are discussed.

In [46] authors employ the method for heart rate variability analyzing. Hyvärinen *et al.* [32] report that ICA is valid for measuring electric activity in the brain. In addition to the results they claim that ICA has been applied to other brain imaging and biometrical signals as well: functional magnetic resonance images, optical imaging, which means directly photographing the surface of the brain after making a hole in the skull, and for removal of artifacts from cardiographic (heart) signals and magnetoneugraphic signals. Kiselgov [40] describes the application of ICA method in software development for medical instrument-making industry.

Of course, there another applications of the model in the medicine are existed, may be some of them are on the creation and development stage as this thesis, or unknown to the author of the thesis.

3.2 Description of the ICA Model

The following definitions represent the basic ICA model. The definitions are derived from the book "Independent Component Analysis" [32, subchapters 1.3.1 and 7.2.1].

Definition 1. *Let consider* n *random variables* $x_1, x_2, \ldots, x_n$. *The observed variables are modeled as linear combinations of* n *random variables* $s_1, s_2, \ldots, s_n$ *:*

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{in}s_n, \text{ for all } i = 1, \ldots, n,$$

where the $a_{ij}, i, j = 1, \ldots, n$ *are some real coefficients.*

By definition, the s_i *are statistically mutually independent.*

The mixing model can be rewritten in vector-matrix notation:

Definition 2. *Let us consider random vectors* $X = (x_1, x_2, \ldots, x_n)^T$, $S = (s_1, s_2, \ldots, s_n)^T$, *then the mixing model is*

$$X = AS \qquad (1)$$

where A *is some unknown matrix with elements* a_{ij}.

The independent components (ICs) s_i also cannot be directly observed. Independent Component Analysis consists of estimating both the matrix A and the ICs s_i, when only the x_i variables can be observed. For simplicity, there are no noise terms in the model and the number of independent

components s_i is equal to the number of observed variables. Mathematical background and knowledge in statistics about random variables, distributions and independence can be got from any statistical literature or [32].

There are some restrictions in the ICA method, which should hold for ICs and mixing matrix.

- ICA assumes that ICs are statistically independent.
- The ICs must have no Gaussian distributions.
- For simplicity, the unknown mixing matrix is assumed to be a square matrix. (Mixing matrix is invertible).

The following ambiguities such as impossibilities of determining of ICs' variances and the order are taken place in the ICA model. The reason for both ambiguities is that the ICs and mixing matrix are unknown, but in many cases these facts are insignificant.

3.2.1 Preprocessing for ICA

Before applying of the ICA method, it is often useful to make some preprocessing. In this subchapter the centering of the observable variables and then whitening of them are demonstrated.

In order to make algorithms simpler, the assumption that both mixture variables and the independent components have zero mean is maid. The observed variables are centered before applying of the ICA. The necessity of the centering is clarified in [32, Subchapter 7.2.4]. The following definition is deduced from [32, Subchapters 2.2.4 and 7.2.4.].

Definition 3. *If X is the observed random vector, so a centered vector with zero-mean can be found as*

$$X' = X - m_x ,$$

where $m_x = \frac{1}{K}\sum_{j=1}^{K} x_j$ is a sample mean and K is a number of samples $x_1, x_2, \ldots, x_K$.

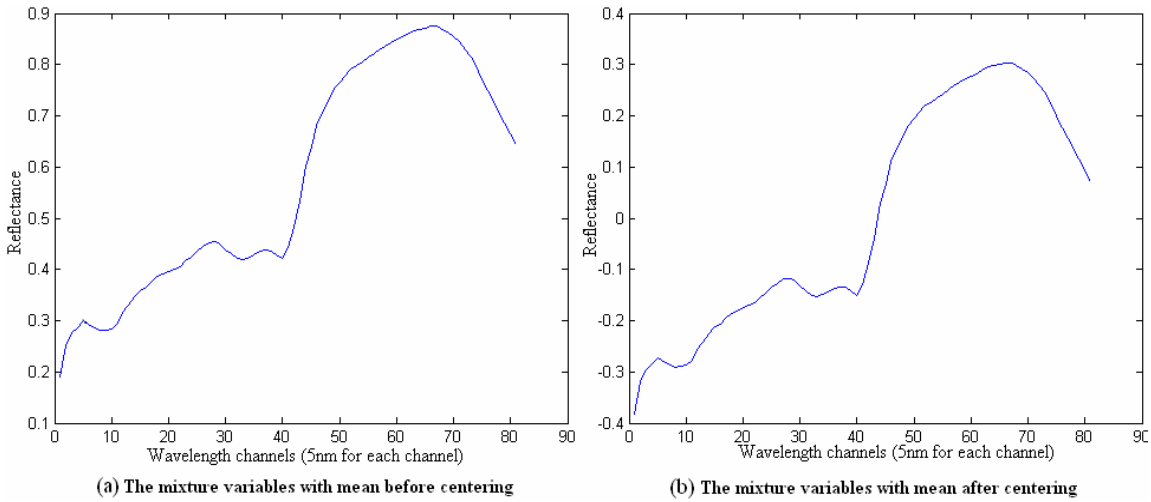


Figure 9. The mixture variables with mean (a) before and (b) after centering.

The figure 9 shows the centering process of the mixture variables, which are retrieved from the spectral images of a face skin (forehead).

Another useful preprocessing technique for ICA is whitening. The term “white” comes from the fact that power spectrum of white noise is constant over all frequencies, somewhat like the spectrum of white light contains all colors [32]. For convenience, in the definition 3.2.4 the designations of the previous one are used.

Definition 4. *A zero-mean random vector* $X' = (x_1', ..., x_n')^T$ *is said to be white, if its elements are uncorrelated and have unit variances* $E\{x_i', x_j'\} = \delta_{ij}$.

Definition 5*. The random vector is called white, if it satisfies the following conditions:*

$$m_X = 0, \quad R_X = C_X = I,$$

where m_X *is a mean vector,* R_X, C_X *are correlation and covariance matrices of the random vector X, respectively.*

Let us consider the random vector $X' = (x_1', ..., x_n')^T$. The task is to find a linear transformation V into white vector Y:

$$Y = V X^{'} \qquad (2)$$

Solution can be found in terms of Principal Component Analysis' (PCA) expansions. Let $E = (e_1, ..., e_n)$ be a matrix, whose columns are the unit norm eigenvectors of the covariance matrix $C_{X^{'}} = E\{X^{'} X^{'T}\}$. That can be computed from a sample of the vectors $X^{'}$ either directly or by one of the on-line PCA learning rules [32, Chapter 6]. Let $D = diag(d_1, ..., d_n)$ be a diagonal matrix of the eigenvalues of C. Then a linear whitening transform is given by

$$V = D^{-1/2} E^T, \quad d_i > 0.$$

Then $C_{X^{'}} = EDE^T$, where E is an orthogonal matrix, i.e. $E^T E = EE^T = I$.

$$E\{YY^T\} = VE\{X^{'} X^{'T}\}V^T = D^{-1/2} E^T EDE^T ED^{-1/2} = I$$

From this it follows that Y is white. The linear operator V is by no means the only unique whitening matrix. The matrix $ED^{-1/2}E^T$ is a whitening matrix too, which is obtained by multiplying V from the left by the orthogonal matrix E. Another method to perform the whitening is on-line learning rules, but they are not considered in the thesis (see [32, Subchapter 6.4]).

Whitening transforms the mixing matrix into a new one $\tilde{A}$. Using equations (1) and (2) we got $y = VAs = \tilde{A}s$. Whitening (uncorrelatedness) is weaker than independence, and is not in itself sufficient of estimation of the ICA model. To consider this, let us use an example from [32]. U is an orthogonal transformation of y: $z = Uy$. Due to orthogonality of U there follows:

$$E\{zz^T\} = E\{Uzz^T U^T\} = UIU^T = I.$$

Since z could be any orthogonal transformation of y, then the whitening gives the ICs only up to an orthogonal transformation. This is not sufficient in most applications. An advantage of the whitening is that the new mixing matrix is orthogonal. The whitening reduces the number of parameters to be estimated. Instead of having to estimate the n^2 parameters that are elements of the original matrix A, we only need to estimate the new, orthogonal mixing matrix $\tilde{A}$. An orthogonal matrix contains $n(n-1)/2$ degrees of freedom. An orthogonal matrix contains only about half of the number of parameters of an arbitrary matrix. That means the whitening reduces the complexity of the problem [32].

3.2.2 FastICA Algorithms

To perform the estimation of ICA the FastICA algorithm, as a computationally highly efficient method is used [33]. It uses a fixed-point iteration scheme that has been found in independent experiments as very fast method for ICA. Another advantage of the FastICA algorithm is that it provides a general-purpose data analysis method that can be used both in an exploratory fashion and for estimation of sources. For more information about this algorithm see [23].

In the thesis only FastICA method is studied. There are three variations of it: using kurtosis, with deflation (FP) or with symmetric orthogonalization (FPsym), and using the tanh nonlinearity with symmetric ortogonalization (FPsymth). A fast fixed-point algorithm using kurtosis, a classic measure of nongaussianity, is treated in [32], and implemented by Leppajarvi *et al*. in [43] in order to find filters for color separation.

The FastICA principle is based on maximization of nongaussianity. As it is mentioned above the main assumption of the ICA method is that the ICs must have no Gaussian distributions. The algorithm of finding of the independent components is based on Central limit theorem, which asserts that in certain cases the sum of independently distributed random variables tends to Gaussian distribution as the number of items is increasing. So, the task of ICA is to find such ICs, whose distributions are maximally far from Gaussian (normal) distribution. For practical measure of nongaussianity some estimation methods are derived. As it was proven in [32], a Gaussian variable has larger entropy among all random variables of unit variance. This means that entropy could be used as a measure of nongaussianity.

Definition 6. *A measure, that is zero for a Gaussian variable and always nonnegative, is called negentropy. Negentropy J is defined as*

$$J(x) = H(X_{gauss}) - H(X),$$

where X_{gauss} *is a Gaussian random vector of the same covariance matrix as* X, *and* H *is the entropy.*

Using the fixed-point algorithm, a fast method for maximizing negentropy can be found. The FastICA for unit, which estimates only one independent component, finds a unit vector w such,

that projection $w^T Y$, where Y is a whitened random vector and $E\{(w^T Y)^2\} = \|w\|^2 = 1$, maximizes nongaussianity. Nongaussianity is measured by the approximation of negentropy J, given in [32, Subchapter 8.3] as:

$$J(y) \propto [E\{G(y)\} - E\{G(\theta)\}]^2.$$

For practically any nonquadratic function G, and θ is a Gaussian variable with zero mean and unit variance, the variable y is assumed to have zero mean and unit variance.

For estimating of several ICs two FastICA algorithms with deflationary orthogonalization and symmetric orthogonalization are obtained in [32]. The property of the methods that vectors w_i, corresponding to different independent components are orthogonal in the whitened space, so in the whitened space

$$E\{(w_i^T Y)(w_j^T Y)\} = w_i^T w_j$$

and, therefore, uncorrelatedness is equivalent to orthogonality. This property is confirmed because of orthogonality of the mixing matrix after whitening. From the last described, it follows that w_i are the rows of inverse of the mixing matrix A^{-1} are equal to the columns of mixing matrix according to the property of orthogonal matrix: $A^{-1} = A^T$.

At first, in FastICA algorithm nonlinearity g, which is the derivative of nonquadratic function G, is chosen. In the thesis the nonlinearities presenting in [32] are used, these are:

$$\begin{aligned} g_1(y) &= \tanh(a_1 y), \\ g_2(y) &= y e^{\frac{-y^2}{2}}, \\ g_3(y) &= y^3, \end{aligned} \qquad (3)$$

where $1 \le a_1 \le 2$ is some suitable constant, often $a_1 = 1$.

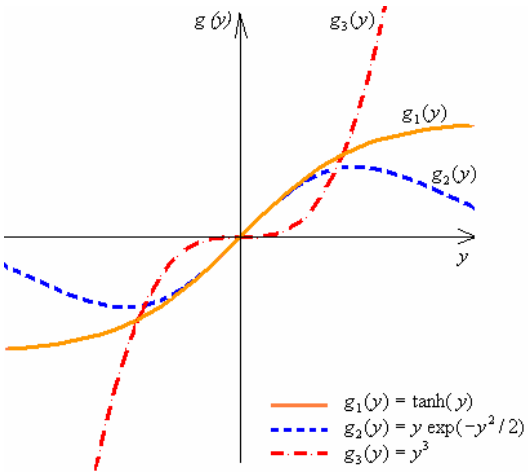


Figure 10. The robust nonlinearities.

Figure 10 shows graphics of three nonlinearities. The $\tanh$ function nonlinearity is used in the Maximum Likelihood Estimation method. In order to estimate several independent components, in [32] one-unit algorithm running several times is used. And after iteration, in order to prevent vectors $w_1,...,w_n$ from converging to the same maxima, orthogonalization of the vectors is made. There two approaches to estimate several ICs are offered in [32]: estimation of the ICs one by one and in parallel. More deep information about the method can be found in the book.

The tables 1 and 2 describe shortly the methods for achieving decorrelation.

Table 1. FastICA with Deflationary Orthogonalization [32].

1.	Center the data.
2.	Whiten the data to give Y.
3.	Choose number k of ICs to estimate.
4.	Set $i \leftarrow 1$.
5.	Choose an initial value of unit norm for w_i, e.g. randomly.
6.	Let $w_i \leftarrow E\{Yg(w_i^T Y)\} - E\{g'(w_i^T Y)\}w$, where g is one of nonlinearities.
7.	Do orthogonalization: $w_i \leftarrow w_i - \sum_{j=1}^{i-1}(w_i^T w_j)w_j$.
8.	Let $w_i \leftarrow w_i / \|w_i\|$.
9.	If w_i has not converged, go back to step 6.
10.	Set $i \leftarrow i+1$. If $i \le k$, go back to step 5.

Table 2. FastICA with Symmetric Orthogonalization [32].

1. 1-3 steps of the algorithm above.
2. Choose initial values for the w_i, $i = \overline{1,k}$, each of unit norm. Orthogonalize the matrix W as $w_i \leftarrow w_i - \sum_{j=1}^{i-1} (w_i^T w_j) w_j$.
3. For every i let $w_i \leftarrow E\{Yg(w_i^T Y)\} - E\{g'(w_i^T Y)\} w$
4. Do a symmetric orthogonalization of the matrix $W = (w_1, ..., w_k)^T$ by $W \leftarrow (WW^T)^{-1/2} W$, or by the iterative algorithm:
1) Let $W \leftarrow W / \|W\|$.
2) Let $W \leftarrow \frac{3}{2} W - \frac{1}{2} WW^T W.$
3) If WW^T is not close enough to identity, go back to step 2.
5. If not converged, go back to step 3.

FastICA by Maximum Likelihood Estimation.

A popular approach for estimating of independent components is maximum likelihood (ML) estimation. Information about the method can be found in any book of Statistics or in [32]. Likelihood can be maximized by a fixed-point algorithm fast and reliably.

Hyvärinen *et al.* [32] obtained the basic iteration of FastICA as:

$$B \leftarrow B + diag(\alpha_i)[diag(\beta_i) + E\{g(H)H^T\}]B,$$

where $\alpha_i = -1/(\beta_i + E\{g'(y_i)\})$, $\beta_i = -E\{y_i g(y_i)\}$, and $H = BX$, this process is without whitening. Then after every step, the matrix B must be projected on the set of whitening matrices. This can be accomplished by the classic method involving matrix square roots,

$$B \leftarrow (BCB^T)^{-1/2} B,$$

where $C = E\{XX^T\}$ is the correlation matrix. And nonlinear function g is the tanh function. In practice is useful using of the whitened vectors, which can be obtained by applying of the whitening matrix $WY = WVX$, which implies $B = WV$ and $H = WY$.

In FastICA, convergence speed is optimized by the choice of the matrices $diag(\alpha_i)$ and $diag(\beta_i)$. These two matrices give an optimal step to be used in the algorithm. Shortly with the whitening the algorithm can be written as:

Table 3. FastICA with Maximum Likelihood Estimation.

1. Center the data to make theirs mean zero. Whiten them to get Y. Compute the correlation matrix $C = E\{YY^T\}$.
2. Choose an initial (e.g., random) matrix W.
3. Compute $H = WY$, and for $i = \overline{1,n}$,
   $\beta_i = -E\{y_i g(y_i)\}, \alpha_i = -1/(\beta_i + E\{g'(y_i)\})$.
4. Update the matrix by $W \leftarrow W + diag(\alpha_i)[diag(\beta_i) + E\{g(H)H^T\}]W$.
5. Decorrelate and normalize by $W \leftarrow (WCW^T)^{-1/2}W$.
6. If not converged, go back to step 3.

The algorithm in the table 3 is derived from [32] using whitened vectors.

Brief Summary of Chapter 3

1. The history of Independent Component Analysis is written. Some medical applications, which use the ICA model, are presented.
2. The main definitions and restrictions of the ICA method are adduced.
3. Centering and whitening preprocessing operations are described in the paragraph.
4. The FastICA algorithm as main algorithm for the research work is considered.

4 Analysis of Spectral Image of Skin

4.1 Preparing of Data for ICA Analysis

Extracting qualitative information and spatial distribution of components through analysis of an observed image is required in many fields of image analysis such as remote sensing, medical diagnostic and robot vision [61]. The process of spectral image retrieving is very important for correct results during a research. Inaccurate analysis, which can involve incorrect diagnosis of a disease, also can depend on the quality of produced spectral images. During the imaging process it is important to take into consideration the light sources and instruments' quality. This chapter describes the process of spectral image retrieval and transformation of them to research data.

4.1.1 Some Necessary Definitions of Color Science

For the last hundred years "images" have been recorded on film either in gray-scale or in color defined by the intensity of red, green and blue (RGB). Spectral imaging provides a significant additional dimension that enables each object to be defined by multiple, even hundreds, of wavelengths. Spectral images are images with high number of spectral channels. In general the number of channels in a spectral image exceeds three, which are found in typical RGB color imaging, and can range to several hundred channels. The possibility to take spectral images is very important for color research. Characteristic features of spectral images in some color imaging applications and medicine are described in [21]. Importance of spectral images in medicine and theirs role in diseases' detection are described in [22].

In a laboratory it is possible to get two kinds of spectral images: reflectance and radiance. Both these type images can be retrieved from each other. Reflectance type images can be received from radiance type images by the separation of the light source, under which the radiance type images are taken. The illumination has important role in the process of transformation. In this case, radiance type images can be defined as pure color images without any outside light. Such kinds of images can be measured under any illuminants that are very important if it is needed to study images in

situations of different light sources. More deeply the transformation is considered below in any book of color science, for example [65].

In order to retrieved a spectral image it is necessary to know properties of light sources, under which the images are obtained. In this subchapter illuminant Daylight, which is commonly used for the spectral images retrieving, is presented. Also as it is mentioned above, the illumination performs the important role for the transformation from reflectance type to radiance type and vice versa.

An object that emits light or radiant energy, to which the human eye is sensitive, is a light source. The emission of a light source can be described by the relative amount of energy emitted at each wavelength in the visible spectrum, thus defining the source as an illuminant. An illuminant is a simulation of a light source. It is a mathematical representation of a theoretical real light source. The light source affects the perception of a color. Both natural daylight and artificial simulated daylight are used for visually examining the color difference.

To define the artificial light sources the standard illuminants are established by the Commission Internationale de I'Eclairage (CIE). CIE standard illuminants have spectral characteristics similar to natural light sources, i.e. they represent an aim spectral power distribution of a theoretical real light source and can be reproduced in a laboratory. In 1963 "D" series of illuminants were proposed to the CIE. The D illuminants are the daylight illuminants, defined from 300 Nm to 830 Nm, across the ultraviolet (UV), visible and near-infrared (IR) wavelengths. So they represent daylight more completely and accurately than do illuminants B and C.

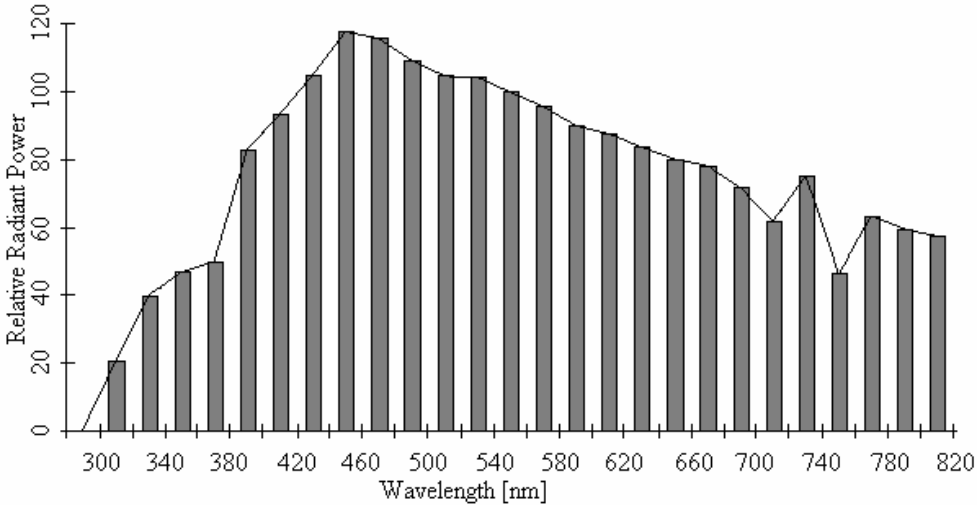


Figure 11. Relative spectral power distribution curve for natural daylight.

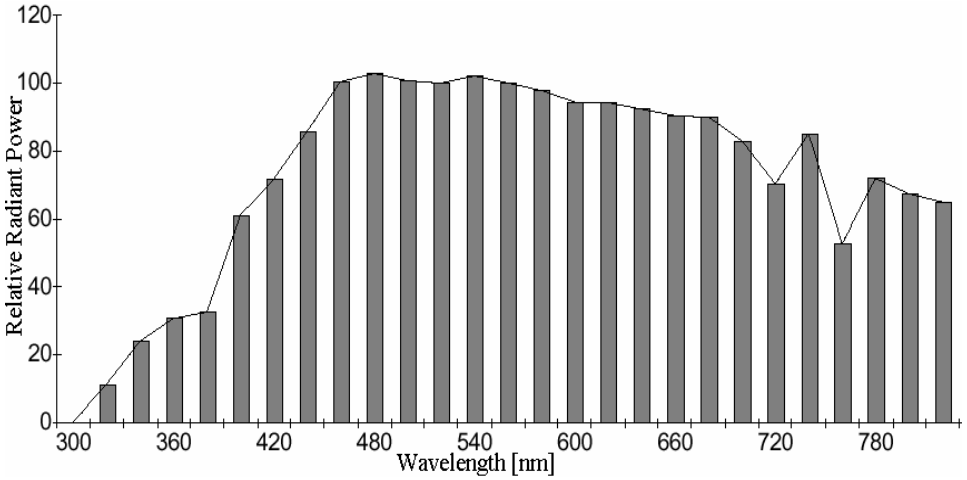


Figure 12. Relative spectral distribution curve for illuminant D65.

D65 is the most commonly required for visual evaluations and color measurements illuminant, it is a mathematical representation of noon sky daylight. D65 has a correlated color temperature of 6504K.

4.1.2 Retrieving of Spectral Images for ICA Analysis

The multispectral images of human skin (hands, faces), which were produced in the Color laboratory of University of Joensuu by using the ImSpector V8 spectrograph, are used for the research work in the thesis. The following picture describes the position of the camera during the process of images' obtaining.

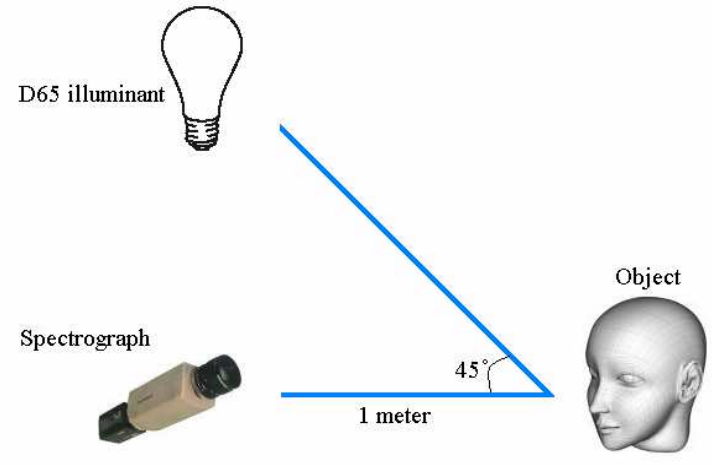


Figure 13. The process of spectral image retrieving.

Spectral images, which are using in the thesis, are two types of images: reflectance and radiance type images. The spectral images of the faces are reflectance type images with wavelength range from 380 Nm to 780 Nm and with 81 components or wavelength channels in each image (with 5 Nm intervals for each channel). The radiance type images are spectral images of different parts of the hands. These skin images are measured with 5 Nm intervals, the starting wavelength is 380 Nm, so in this case the wavelength range is 380-775 Nm, and there are 80 wavelength channels in each image. Information about these images can be found in the tables B.1 and B.2 of the Appendix B.

For radiance type images a transformation of them to reflectance type is made. An interpolation of the Daylight illumination spectrum range (from [18]) to the image data spectrum range (380-775 Nm with 5 Nm intervals) is made in order to transform to reflectance type image, and each radiance spectrum is divided with that daylight spectrum. As the Daylight illumination the illuminant D65 (artificial daylight) was used, the spectrum range of the illuminant is from 376 Nm to 776 Nm with 2 Nm intervals. The process of interpolation is presented in the figure 14.

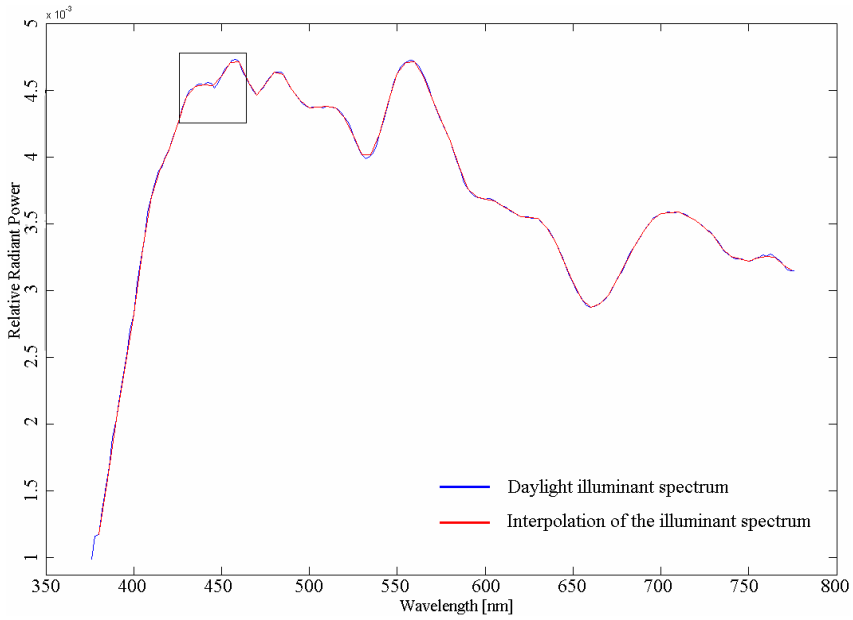


Figure 14. Interpolation of the Daylight illumination spectrum.

Definition 7. *[64] Linear interpolation is a method to find approximately a function* $f(x)$*, which is based on substitution of* $f(x)$ *by linear function* $L(x) = a\,(x - x_1) + b$*, where the parameters* a *and* b *are chosen such, that values of* $L(x)$ *and* $f(x)$ *are equal in the given points* x_1 *and* x_2*. The only function, which satisfies to this, is*

$$L(x) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1) + f(x_1),$$

which is approximated to $f(x)$ *in the segment* $[x_1, x_2]$ *with inaccuracy*

$$f(x) - L(x) = \frac{f''(\zeta)}{2}(x - x_1)(x - x_2), \quad \zeta \in [x_1, x_2].$$

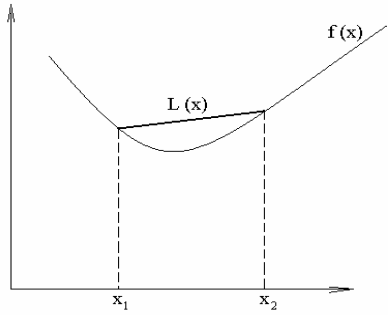


Figure 15. Definition of the linear interpolation.

The figure 16 shows the detail of the spectrum interpolation, which is marked up by the square frame on the previous figure.

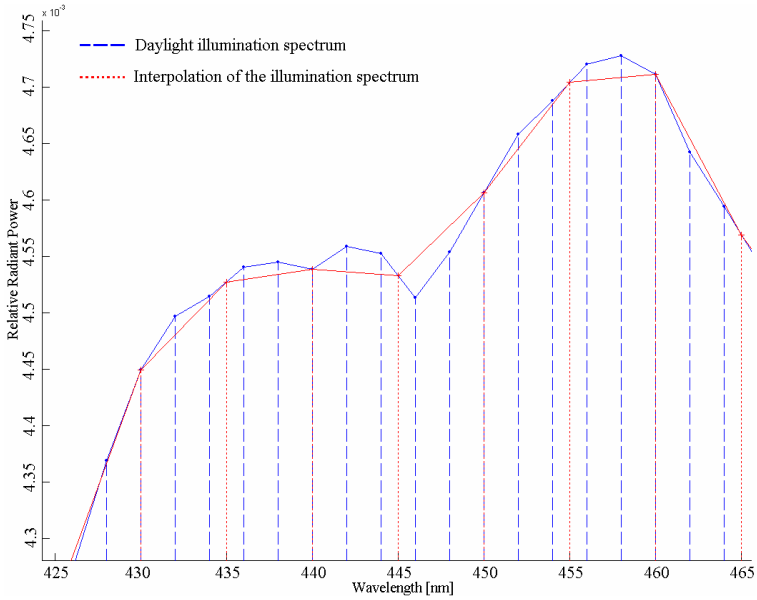


Figure 16. Detail of the interpolation of the Daylight illumination spectrum.

Let us consider a function, which describes the daylight illumination spectrum in the segment $[434, 436]$, the linear function $L(x)$ approximating the function of daylight spectrum is:

$$L(x) = \frac{f(436) - f(434)}{436 - 434}(x - 434) + f(434) = 0{,}01(x + 18),$$

and in the point $x = 435$, the $L(435) = 4{,}53$.

This transformation gets the estimates for reflectance spectra. The level of the spectra is not enough correct one, but that does not matter because in order to apply the ICA method the centering of all data are made, and they have zero-mean before the analysis.

Mathematical definitions of different interpolation methods can be found in [34] and can be applied by using the Matlab program [48].

The spectral images of human skin, which are analyzed in the thesis, allow retrieving of the mixture variables. And they are observed by using the ICA method in order to find independent components. The transformation of the spectral images to mixture variables is needed in order to use them as input data for ICA algorithm. Next paragraph describes methods, which are used for that conversion.

4.2 Retrieving of Mixture Variables

In this subchapter the methods to retrieve the matrix of mixture variables (MV matrices) $X(t,k)$ are developed. The mixture variables or random variables are observed by the ICA method in order to find independent components (see definition 1 of ICA model).

Methods of the transformation from the multispectral images to two dimensional matrices are described in the thesis. The matrices, which are retrieving by the methods, are not centered matrices of mixture variables and can be used as an input data for ICA algorithm, which estimates the independent components from the given multidimensional signals.

Let consider the three dimensional matrix $P(i, j, k)$, where i, j, k are dimensions of the matrix. Our task is to find the function $f(i, j)$ in order to convert the matrix P to matrix $X(f(i, j), k)$, where $f(i, j)$ and k are dimensions of the matrix X, and $f(i, j)$ depends on dimensions i, j.

In mathematical symbols it will be:

for any $i, j, k, f(i, j) \in \overline{1,n}$:

$$f : P(i, j, k) \to X(f(i, j), k)$$

The conversion function can be found as a multiplication $(i \cdot j)$, so the problems looks like:

for any $i, j, k, f(i, j) \in \overline{1,n}$:

$f : P(i, j, k) \to X(f(i, j), k)$, where $f(i, j) = i \cdot j$.

But in this case we can confront with another problem such as big time and space complexity of the ICA algorithm. In this situation it is better to work with not whole image but select some part of it. It is necessary to decrease the size of the matrix of mixtures in order to avoid the problems, which are mentioned above.

The thesis offers two different conversion methods for retrieving of the matrix of mixture variables: Part of image method and Threshold method. Both these methods provide the finding of matrices of mixture variables with the sizes, which satisfy a researcher.

4.2.1 Part of Image Method

As is mentioned above the first method for MV matrix retrieving is a Part of image method. The main property of the method that a submatrix of three dimensional input matrix is extracted. And the conversion method is applied not to whole spectral image but to the submatrix.

Definition 8. *[47] Submatrix of a rectangular* $i-by-j$ *matrix* P *is a matrix* $Q(c,d)$, *where* $1 \le c \le i$, $1 \le d \le j$, *and the submatrix* Q *is developed by elements at the intersection of the fixed* c *rows and* d *columns in the matrix with preservation of the previous order.*

The Definition 8 works only for two dimensional matrices, for three dimensional matrices the Definition 9 and 10 is developed. In the definition 9, the rules (in general case) of choosing the points that determine the 3-D submatrix is described.

Definition 9. $\exists$ *3-D matrix* $P(i,j,k)$, *where* $i,j,k \in \overline{1,n}$. *A submatrix* Q *of the matrix* P *is defined by three points* (i_1, j_1, k_1), (i_2, j_2, k_2), (i_3, j_3, k_3), *for which formula*

$$k_1 = k_2 \vee k_2 \neq k_3 \wedge (i_3 = i_1 \vee i_3 = i_2) \wedge (j_3 = j_2 \vee j_3 = j_1) = True$$

or

$$i_1 = i_3 \wedge i_3 \neq i_2 \wedge (k_2 = k_1 \vee k_2 = k_3) \wedge (j_2 = j_3 \vee j_2 = j_1) = True$$

or

$$j_3 = j_2 \wedge j_2 \neq j_1 \wedge (i_1 = i_3 \vee i_1 = i_2) \wedge (k_1 = k_2 \cup k_1 = k_3) = True.$$

To apply the Part of image algorithm the definition 10 is needed. The algorithm is based on the selection of two points on the front plane Oij as on the figure 17, the second point is given to set the size of submatrix from the beginning. The definition 10 defines the submatrix Q by the given one point; it is a special case of the definition 9.

Definition 10. *∃ 3-D matrix* $P(i, j, k)$, *where* $i, j, k \in \overline{1, n}$. *Let consider a point* $\varphi(i_\varphi, j_\varphi, k_\varphi) \in P(i, j, k)$, *i.e.* $i_\varphi \le i,\ j_\varphi \le j, k_\varphi \le k$. *Let us construct a submatrix* $Q(c, d, e)$ *through the point* φ, *which is supposed to be down-left-front point of submatrix* Q *and* c, d, e *satisfy:*

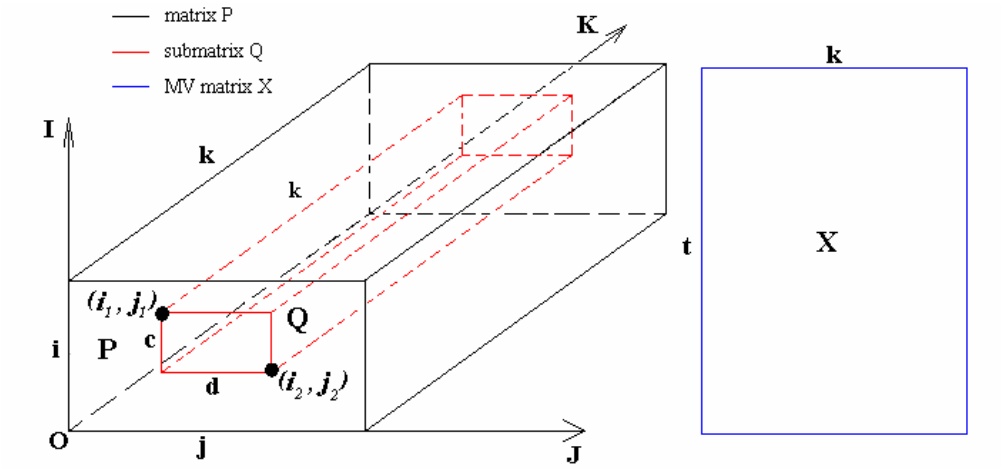
$$1 \le c \le i - i_\varphi, 1 \le d \le j - j_\varphi, 1 \le e \le k - k_\varphi.$$

In our case the submatrix Q of the matrix P is used like it is shown on the figure 17.a. At the thesis the submatrix has the third dimension same as the matrix has.

The size of the submatrix is defined by two input coordinates (i_1, j_1) and (i_2, j_2), which are set by a researcher on a spectral image.

It is very important that an order of elements of the spectral image is kept in a new submatrix.

At the figure 17 the Part of image method is sketched out.



(a) Matrix P(i, j, k) and submatrix Q (c, d, k) **(b)** Matrix of mixture variables X(t, k)

Figure 17. The Part of Image algorithm.

Below there is a definition of a not centered MV matrix, retrieved by the Part of Image method.

The not centered MV matrix is found from the submatrix Q, for that the third dimension vectors are taken from the submatrix in order from left to right, from up to down and put into the MV matrix as row vectors one by one in the bottom.

Definition 11. *$\exists$ 3-D matrix $P(i, j, k)$, where $i, j, k \in \overline{1,n}$. Let develop 3-D submatrix $Q(c, d, k)$, where $c \le i$, $d \le j$ and $c, d \in \overline{1,n}$. The function f makes the conversion from the submatrix Q to a matrix of mixtures $X(t, k)$, where $t = c \cdot d$, i.e.*

$$f : Q(c, d, k) \to X(t, k), \text{ where } t = c \cdot d .$$

And $\forall c_q \in \overline{1,c}$, $\forall d_q \in \overline{1,d}$, $\forall k_q \in \overline{1,k}$ the element of the MV matrix $x_{t_x, k_q} = b_{c_q, d_q, k_q}$, where $t_x = (c_q - 1) \cdot d + d_q$.

Let consider the following example as an illustration of the definition.

Example 1. $\exists$ input 3-D matrix, $P(i, j, k)$, where $i = 4$, $j = 6$, $k = 2$. The submatrix, taken from the matrix P is $Q(c, d, k)$, where $c = 2, d = 3, k = 2$. The MV matrix $X(t, k)$, where $t = 6, k = 2$, is shown on the figure 18.

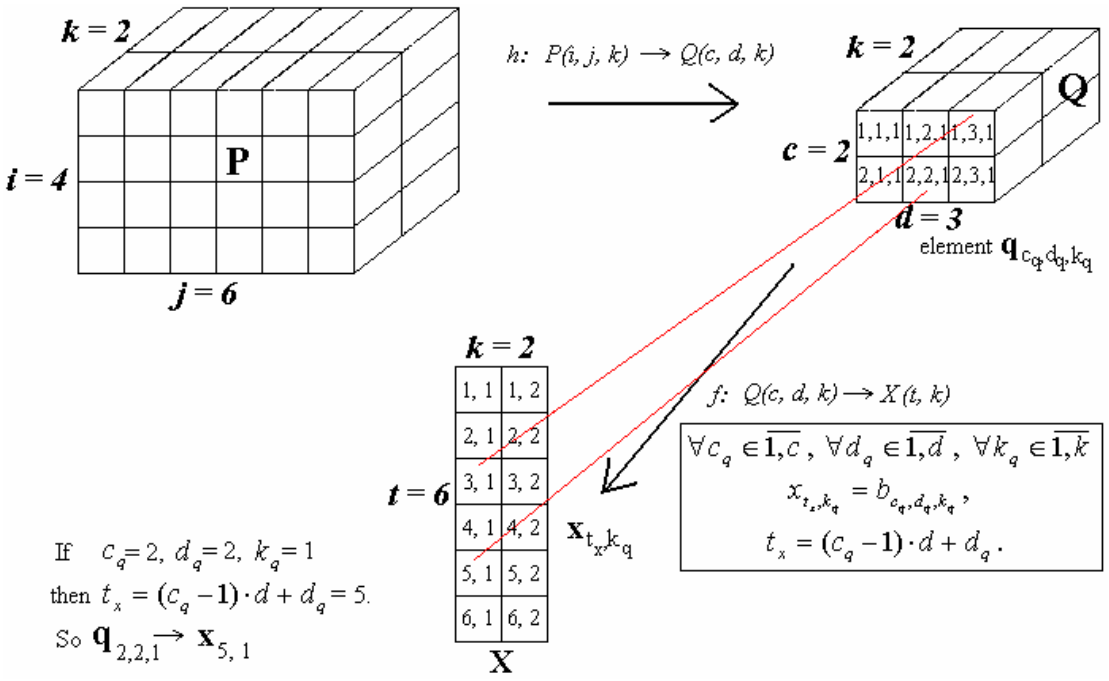


Figure 18. Working of the Part of Image algorithm.

The figure 18 shows that the Part of image algorithm works for 3-D matrix, which is a spectral image of skin in our case, whose two first dimensions are height and width, accordingly, and the third dimension is spectrum's components.

4.2.2 Threshold Method

The Threshold algorithm represents a different to the Part of image method for retrieving of matrices of mixture variables. The algorithm is constructed on the basis of subsampling method and threshold method. The subsampling method is described below. The threshold means a value, which determines size of the MV matrix; it is a sum of reflectance values of one pixel for all specters. Only those points, which value is higher than the threshold value, will be processed by the method. And points, whose values lower than the threshold value will be omitted. The definition 13 numerically determines the threshold.

As it was mentioned above the ICA algorithm has big time and space complexity. Therefore, if the MV matrix is going to be large, then the input spectral image can be resized in order to decrease the number of mixture vectors in the MV matrix, by using the Threshold algorithm. The resized matrix is developed by the subsampling method, where a square window is used. Information about this method can be found in lecture notes [24]. An algorithm of finding of the resized matrix is processed by the Threshold method and described below.

Definition 12. *$\exists$ 3-D matrix $P(i, j, k)$, where $i, j, k \in \overline{1, n}$ and $\exists$ size of window W, then a subsampling algorithm h such as:*

$$h : P(i, j, k) \to R(u, v, k), \text{ where } u \in \overline{1, i},\ v \in \overline{1, j} \text{ and } u = \frac{i}{W},\ \ v = \frac{j}{W}$$

can be found.

And $\forall u_r = \overline{1, u}, v_r = \overline{1, v}, k_p = \overline{1, k}$ the element of the resized R is defined as:

$$r_{u_r, v_r, k_p} = p_{i_p, j_p, k_p}, \text{ where } i_p = (u_r - 1) \cdot W + 1, j_p = (v_r - 1) \cdot W + 1.$$

The figure 19 shows a work of the subsampling algorithm.

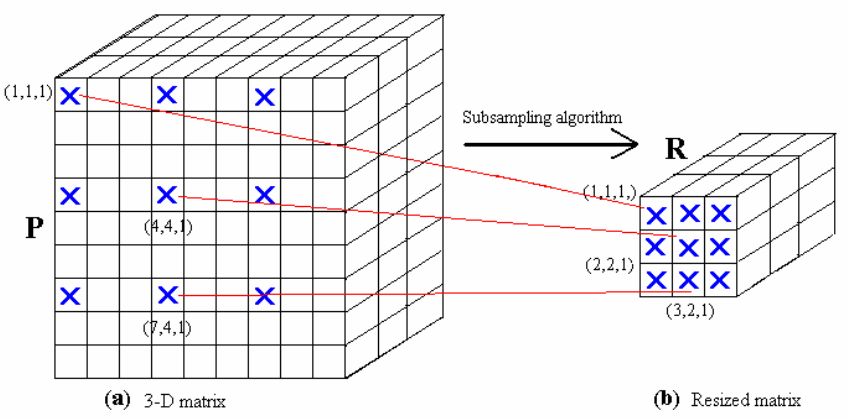


Figure 19. Subsampling method.

The definition 13 describes the Threshold algorithm.

Definition 13. *Let* $R(u,v,k)$ *is a resized matrix, where* $u,v \in \overline{1,n}$. $\exists$ *points* $u_T \in \{1,...,u\}$ *and* $v_T \in \{1,...,v\}$, $(u_T, v_T) \in R(u,v,1)$. *Then*

- *the threshold value T is* $T = (\sum_{\xi=1}^{k} r_{u_T, v_T, \xi})$.
- *the conversion function* $f : R(u,v,k) \to X(t,k)$, *where* $t \le u \cdot v$, *is defined:*

$\exists Q = 0$, *then* $\forall u_r \in \{1, ..., u\}$, $\forall v_r \in \{1, ..., v\}$: *if* $\sum_{\xi=1}^{k} r_{u_r, v_r, \xi} \ge T$ *then* $Q = succ(Q)$.

And $\forall k_r \in \{1, ..., k\}$ *an element* x *of the MV matrix* X *is defined as* $x_{Q,k_r} = r_{u_r, v_r, k_r}$.

Definition 14. For any $Q \in Z$ $\exists succ()$ function, this returns next in order value, such as:

$$succ(Q) = Q + 1.$$

Let consider an example of the Threshold algorithm's work.

Example 2. ∃ 3-D matrix $P(i, j, k)$, where $i = 6,\ j = 4, k = 2$ and ∃ the size of window $W = 2$ and the threshold value $T = 2$. The resized matrix $R(u, v, k)$, where $u = 3, v = 2$, and MV matrix $X(t, k)$, where $t = 3, k = 2$, are shown on the figure 20.

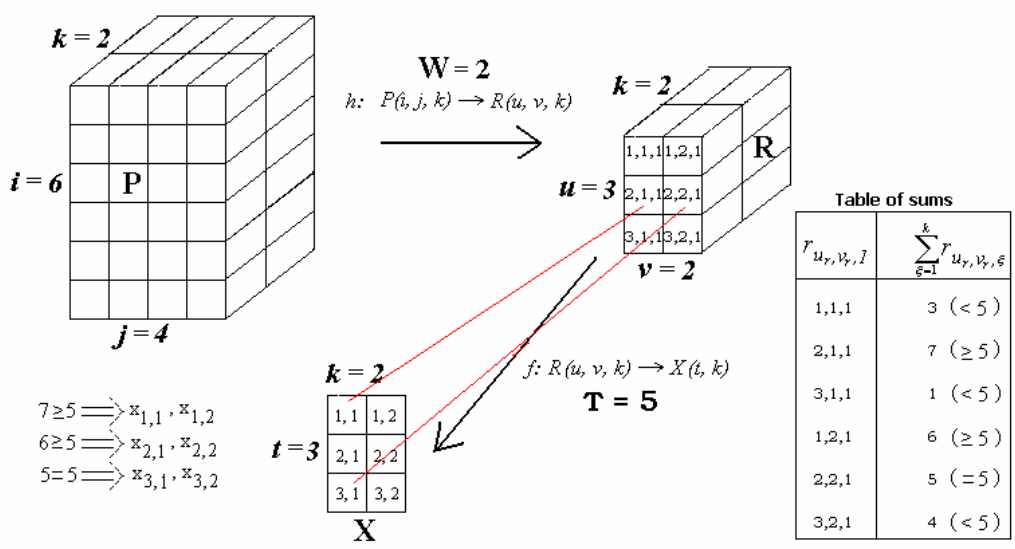


Figure 20. The Threshold algorithm.

Note. The realization of the Part of image and Threshold methods in practice, and visualization of them using Matlab program, can be found in [54, 55]. Both these methods were implemented in ICA_analysis_GUI program (see Appendix C), and results of test are presented in the IT-project [54, 55]. The mathematical package Matlab has wide application in different scientific fields, especially in medicine [39]; it is very useful to work with images, which are actually represented matrices. The application allows making calculations easier because of included set of mathematical operation and formulas.

Brief Summary of Chapter 4

1. Definitions of spectral images and daylight illuminants are presented.
2. The process of retrieving of the multispectral images of human skin and conversion of the radiance type images to the reflectance type images are described.
3. Two algorithms of retrieving of the matrix of mixture variables: Part of image and Threshold, are obtained and implemented.

5 Independent Component Analysis for Skin

As it was mentioned above, spectral images of human faces and hands are tested at the thesis. For spectral images of faces, sample areas were selected on the faces by using Part of image method. There are six regions considered: forehead, bridge of nose, eyelid, cheeks and chin (cheek-bone). The following figure represents an example of the sample areas selection.

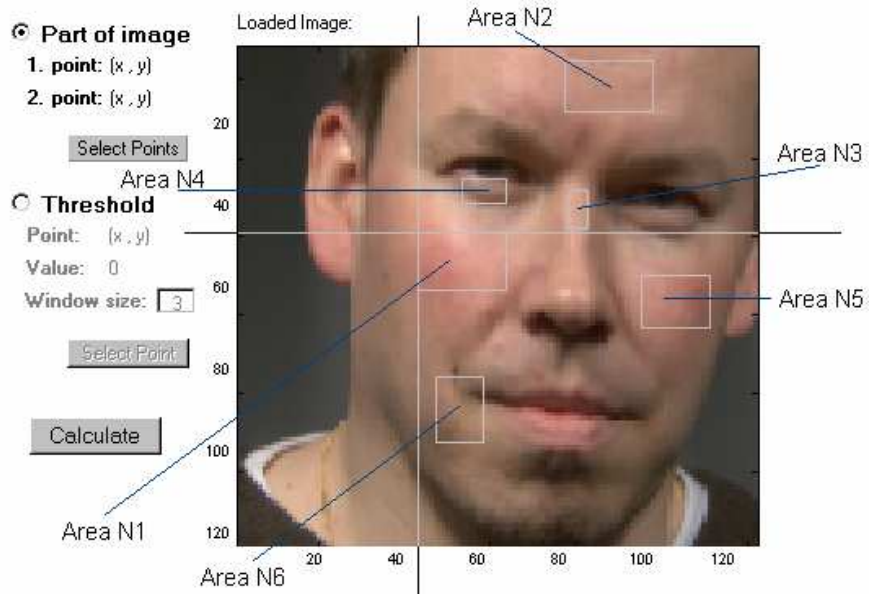


Figure 21. Sample areas of tested faces.

For hands no special regions are selected, so both Part of image and Threshold method could be chosen. “Truncate” sometimes is needed to remove noises appearing during the image retrieval.

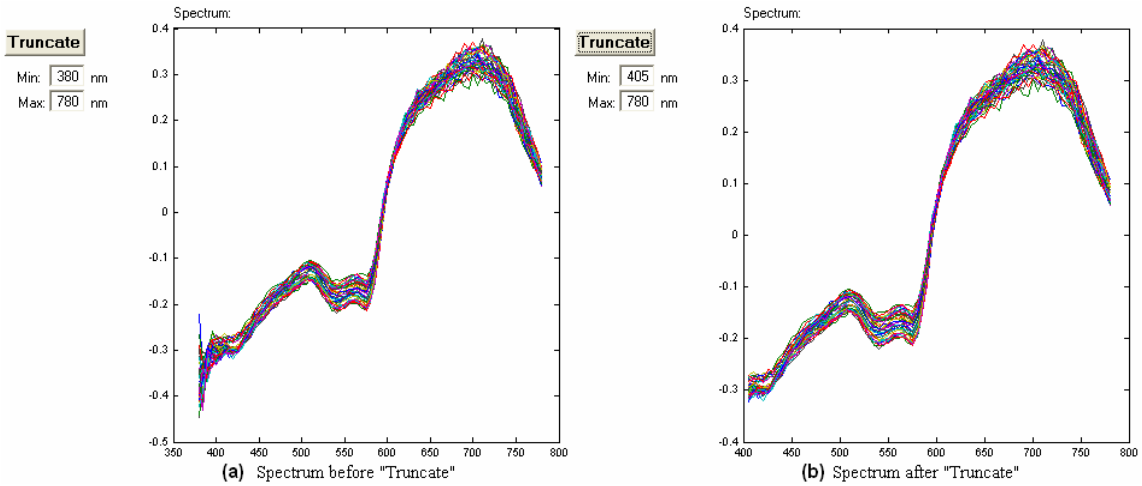


Figure 22. Using of "Truncate" operation, the spectrum is cut off at 405 nm.

As it is seen from the figure 22.b, cutting of the noise can affect the color spectra and change the color parameters of the image. But in our case, when only components of skin are studied, this problem can be omitted, and that is why it is allowed to cut big enough part of spectra in order to save from noises.

5.1 Characteristics of Independent Components

At this chapter the independent components of skin are analyzed. At first, criterions if IC's satisfy the main restriction are verified. Then the number of sufficient IC's for skin research and the method to obtain them are defined. In conclusion the IC's are analyzed and discussion of them is made.

5.1.1 Checking of the Main Restrictions

After applying of the FastICA algorithm to the mixture variables, independent components of skin were obtain. First of all, it is needed to check the main restrictions of the ICA method, which are described in the Subchapter 3.2. Results of the tests and testing data are presented in Appendix D.

Mutual independence of the IC's is main restriction of the ICA method. There are many statistical methods to check if components are independence or not, for example spectral test, t-test or Pearson's test. But most of them are presumed to work with big number of data, what is impossible

in our case. Therefore for the checking of independence of components the Matlab's Statistics Toolbox [48] was applied. Using the cross-tabulation function for several vectors at significance level $\alpha = 0.05$, the null hypothesis that observed components are independent is accepted.

Let us make use of [42] to prove nongaussianity of obtained IC's. The Kolmogorov-Smirnov empirical test is used to reject the hypothesis that empirical distribution function of observed IC's is equal to Gaussian (Normal) distribution.

Let consider IC's – vectors in n-D cube. The norms of the vectors were calculated:

$$\left|s_i\right| = \left\| \sqrt{(\xi_{i_1}^2 + ... + \xi_{i_n}^2)} \right\| ,$$

and ranged in non-descent variation series: $s_1 \le s_2 \le ... \le s_i \le ... \le s_n$. Divide n-D cube o k sub-cubes and count rates of outcomes fallen in each sub-cube:

$$\nu_1 = \frac{m_1}{n}, ..., \nu_i = \frac{m_i}{n}, ..., \nu_k = \frac{m_k}{n},$$

where m_i - amount of outcomes fallen in i-th sub-cube. And then plot the empirical distribution function.

Theorem 1. (A.N. Kolmogorov) *[69]. Probability* $\Phi_n(\lambda)$ *of the inequality*

$$D_n = \sup_{-\infty < x < +\infty} \left|F_n(x) - F_N(x)\right| < \frac{\lambda}{\sqrt{n}},$$

with $n \to +\infty$ *tend to a limit*

$$\Phi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2},$$

(uniformly relative to λ*, whatever the function* $F_N(x)$ *is).*

The Kolmogorov's theorem gives a possibility to define a goodness measure of empirical distribution function $F_n(x)$ with theoretical distribution function $F_N(x)$.

K–S-test is implemented for obtained IC's using Matlab and statistical tables in [51]. At significance level $\alpha = 0.05$ the hypothesis $H_0 : F_{empirical} = F_N$ is rejected. An example of K–S-test applying is illustrated in Appendix D.

5.1.2 Choosing of Number of Independent Components

The problem, which is arisen during the tests, is how many independent components should be estimated? It is clear that this problem does not arise if the same number of components as the dimension of the data is estimated. However, this may not always be a good idea, and explanation of this supposition can be found in [32, Subchapter 13.3]. Authors accomplish a task by choosing the minimum number of components that explain the data well enough, containing, for example, 90% of the variance. And confirm that often, the dimension is actually chosen by trial and error with no theoretical guidelines. It is possible also to reduce dimension by using PCA preprocessing, but this method is avoided in the thesis, and in case of need if necessary, can be found in [32, Chapter 6]. Hyvärinen *et al.* [32] refer to some methods for choosing the quantity of components, such as Information – theoretic, Bayesian, and other criteria, which are considered in detail in [13, 14, and 50].

The following figure shows that for skin color spectra a rather small number of IC's are sufficient, because the shape of the skin spectra is the almost same for all samples. Of course, as is easy to see from the figure 23 that the similarity of the original and reconstructed spectral is increasing with the number of components. The figure 23 shows us the reconstruction of the skin spectra from the different number of independent components. And as in case of PCA, not many components are needed to present the skin spectra. On the figure the results of the tests of three samples of second face area are shown: from top to bottom the following are sketched out: original skin spectra, retrieved from the spectral image by Part of image method; reconstruction of the spectra from 2, 4, 5, 6, 12, 33, 65 independent components; and from left to right – results for different samples.

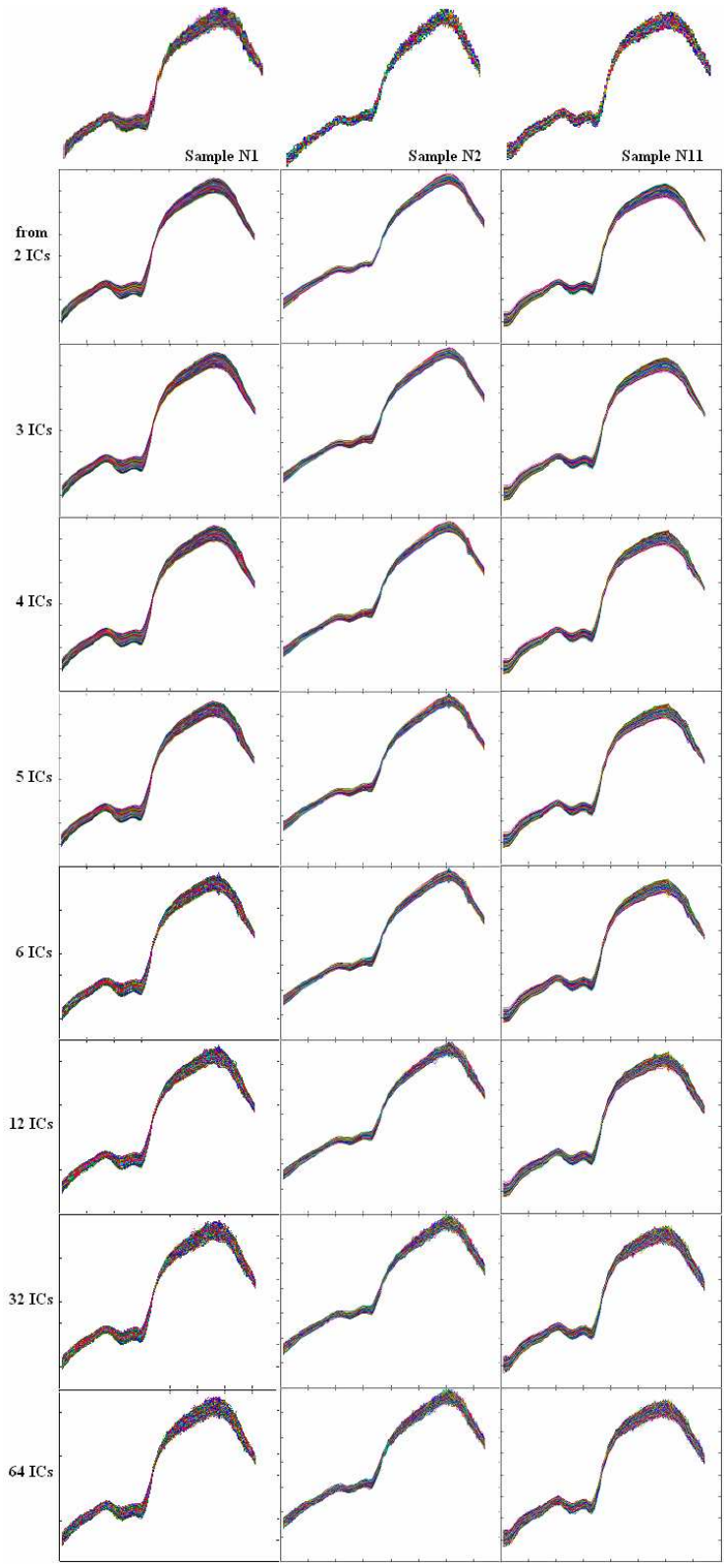


Figure 23. Comparing of original skin spectra with reconstruction from 2, 3, 4, 5, 6, 12, 32, 64 components.

The dependency of number of components and similarity of reconstruction to original spectra also was proved numerically. Let us consider three same samples as on the previous figure. The figure 24 shows that with the increasing of the number of independent components the difference between reconstruction and original is reduced. To calculate difference, the absolute value of sum of deviation squares is used, which is characterized the divergence of values in each spectrum channels from channel's mean value. Mean of all divergences in all channels gives a difference.

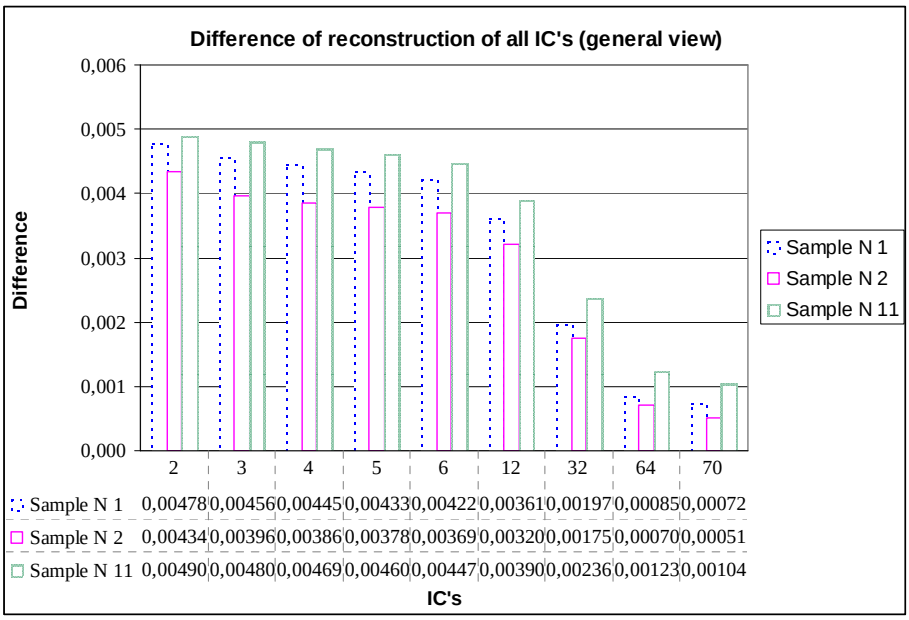


Figure 24. Dependency between reconstruction difference and number of components.

But as it was mentioned above it is not necessary to consider big number of independent components for the image of skin. For that let us consider only significant independent components, which are quite enough to describe the skin spectra. Let $\delta = 0.01$ is an error limit. It means that if some independent components with small value will be excluded the total error will not be more than 0.01 of reflectance value. This limit is presented small enough value, which is imperceptible to a human eye, this value was chosen on the basis of works [45, 52 and 69]. The figure 25 represents difference between original mixture variables and reconstructed ones from the significant independent components. For the purposes of illustration the same samples and number of components were tested as for the figure 23 and 24.

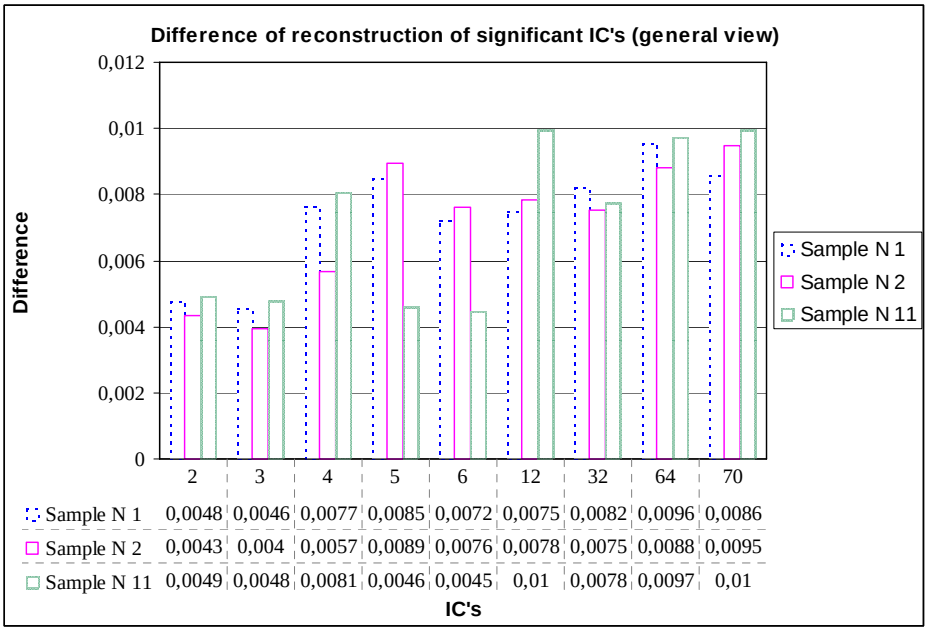


Figure 25. Behavior of the difference after reconstruction by the significant IC's.

Table 4. Number of significant IC's with error limit δ = 0.01

	Number of all retrieved independent components								
	2	3	4	5	6	12	32	64	all
Sample	**Number of significant independent components**								
N 1	2	3	2	4	4	11	30	57	63
N 2	2	3	3	3	5	10	30	52	61
N 11	2	3	3	5	6	9	30	58	64

In the table 4 the numbers of significant components with error limit δ = 0.01 is shown. As it is seen from the figure 25 and the table, the smallest and stabile difference is guarantee with small number of independent components, for example three independent components (further this number of IC's is used for the tests). This number of components will be considered in the thesis below. Hyvärinen *et al*. [32, Chapter 14] also give a notice of error for increasing number of components. In detail the results of tests with different number of components are represented in Appendix D.

The figure 26 represents difference between original spectra of skin (down – center) and reconstruction of it by the significant number of independent components. On the figure the results of testing of one sample of the second face area are shown, the figures with other samples of the

second face area can be found in Appendix D. The list of samples, which are tested in the thesis, is presented in Appendix A.

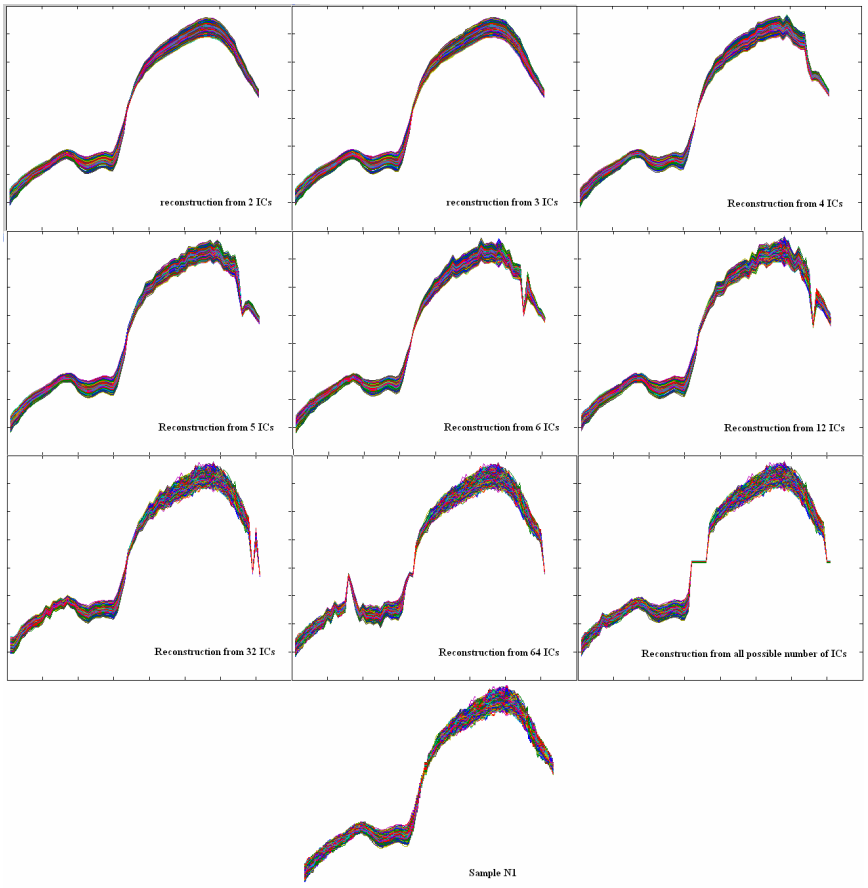


Figure 26. Sample N1. Reconstruction by the significant IC's versus original spectra.

5.1.3 Choosing of Estimation Approaches

In the preceding chapter 3 several different estimation principles and algorithms for ICA were introduced. In this chapter an overview of these methods and chose of the better for our research method is provided. In [32] the connection between all estimation principles is shown, also on comparing of the nonlinearities they confirm that $G(y) = \tanh(a_1 y)$, where $1 \le a_1 \le 2$ is a constant, is a good general-purpose non quadratic function. During the test the main choices should be between nonlinearities and decorrelation methods. Three different statistical criteria for estimation of the ICA model, including likelihood and nongaussianity measures are tested in the thesis. Maximization of nongaussianity is achieved two FastICA algorithms with deflationary

orthogonalization and symmetric orthogonalization. And using of [32], it became known that the version of the fixed-point algorithm (FP), which uses symmetric orthogonalization and $\tanh$ nonlinearity (FPsymth) performs as well as the maximum likelihood (ML) algorithm.

Let us consider 10 samples from the area N2 of the face (forehead). The number of IC's, which are obtained by the methods, are 2, 3, 4 because of chosen number of components in the previous chapter. Results of the testing of the methods are presented in the tables in the Appendix E. Methods, which are tested in the thesis, are named accordingly:

M_1, M_2, M_3 — FastICA with Deflationary Orthogonalization and $g(y) = y^3$, $g(y) = \tanh(y)$, $g(y) = y\exp(-y^2/2)$ nonlinearities correspondingly,

M_4, M_5, M_6 — FastICA with Symmetric Orthogonalization and $g(y) = y^3$, $g(y) = y\exp(-y^2/2)$ nonlinearities, and Maximum Likelihood Estimation with $g(y) = \tanh(y)$.

Each method was launched 10 times. It is necessary to note that all ICA algorithms have some stochastic aspects, and can give every time different results.

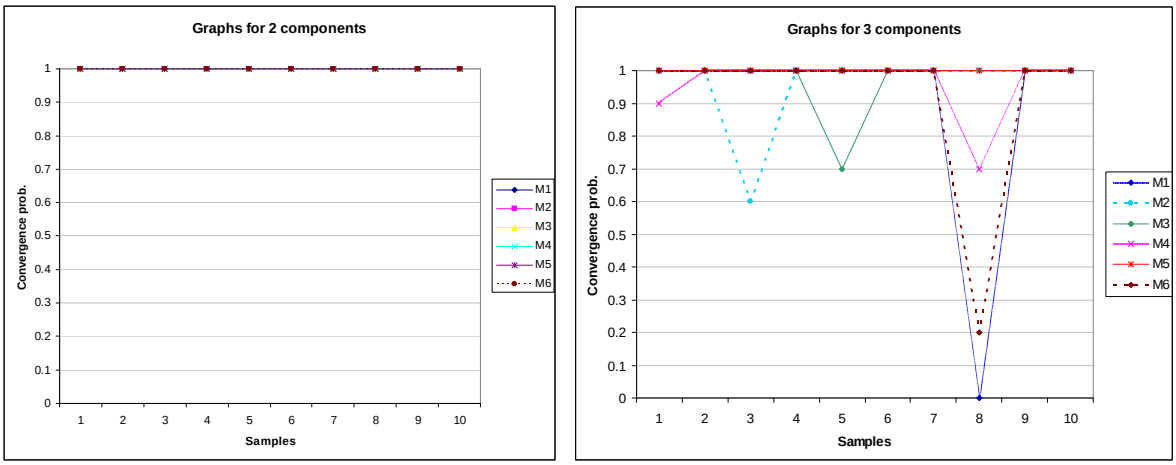


Figure 27. Convergence of the methods M_1-M_6 for 2 and 3 IC's.

As it is seen from the figure 27, all methods are converged successfully for two IC's, and in this case the priority can be given to a faster method, such as Deflation estimation method. The case of three IC's more interesting for study, because such number of components is tested at the thesis. In the figure the superiority of the method M_5 is seen. This allows us to assume that Symmetric estimation approach is the best for the case of three components. In the case of larger number of IC's, this method also gives good results, so this method will be used below.

In the figure 28 all training methods are shown. Here also the confirmation of the correct chosen method is shown. For obviousness, the convergence values were scaled from 0 to -0.3.

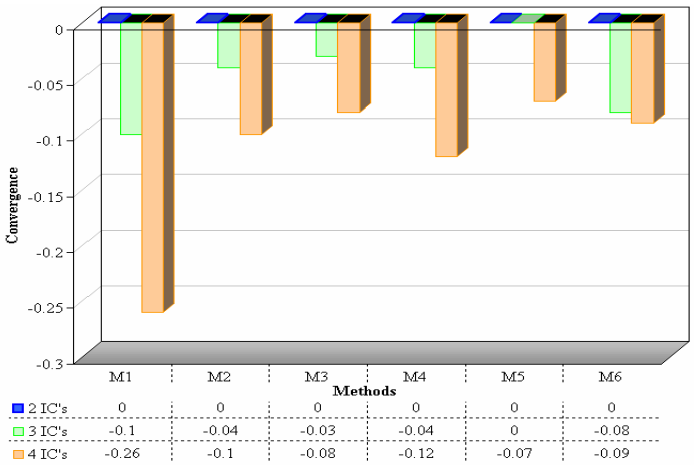


Figure 28. Comparison of methods.

But in general case all methods have no significant difference, what is proved by using a method of Analysis of variance [42]. At significant level $\alpha = 0.05$ the method gives big probability of the type-1 error (error of rejection of a right hypothesis $H_0 : \mu_1 = ... = \mu_6 = \mu$); see ANOVA method in Appendix D.

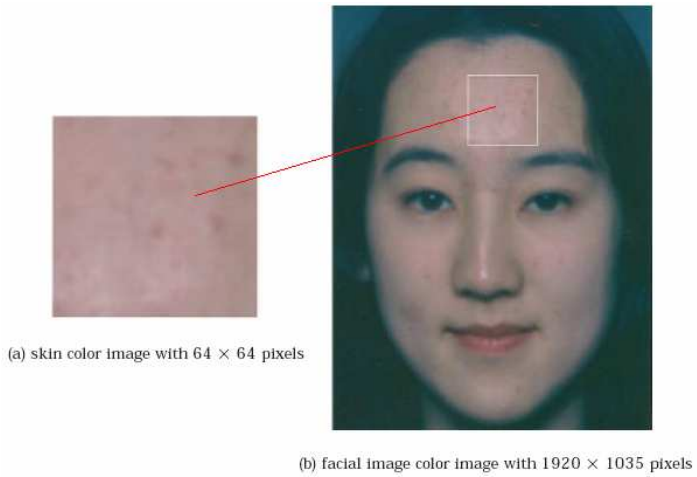
In order to check the greater convergence probability for the methods and nonlinearities a statistical hypothesis was checked for the results in the table D.4. The method of comparison of two binomial distribution's probability [25] to the methods for 3 and 4 IC's was applied. Method was applied to provide best nonlinearity and choose the FastICA algorithms with deflationary or symmetric orthogonalization. As a result of statistical test, the following derivations are made: the nonlinearity $g(y) = y \exp(-y^2/2)$ provided the greater probability of the method convergence, and the symmetrical approach gives better convergence. The description of the method can be found in Appendix 4.

Methods	Advantages	Disadvantages
M_1, M_2, M_3 **(Deflation)**	Deflation is useful in the case, when only a very limited number of components estimated. It has high convergence speed on small numbers of IC's. In contrast to MLE, the method does not depend on reasonable approximation of the densities, and it is less problematic to use it.	Estimation errors in the components that are estimated at first accumulate and increase the error in the later components. Often, when large numbers of components are obtained, the method does not converge.
M_4, M_5, M_6 **(Symmetric and ML)**	The method estimates all independent components in parallel, what allows obtaining large number of components in small number of steps.	Often, when small numbers of components are obtained, the method converges after large number of steps. If the nonlinearity is chosen not correctly, the ML algorithm will not converge properly.

The nonlinearity $g(y) = y\exp(-y^2/2)$ is chosen as the best for the testing, in case of necessity of another nonlinearity the $g(y) = \tanh(y)$ will be chosen, because of wholly satisfactory results and as recommended by the authors of [32].

5.2 IC Separation of Skin Color Image

The separation of three independent components of skin color image is made in the thesis. Such kind of process obtained in [60]. Tsumura *et al.* separated skin color images into two images by ICA in the optical density domain of three color channels. They believed that the images correspond to distributions of melanin and hemoglobin, because the result of separation agreed well with physiological knowledge. As a test material they used RGB color face image and many assumptions were made in the analysis: linearity among the quantities and the observed color signals in the optical density domain, spatial color variation caused by only two pigments, spatial independence of the two pigments, and zero quantity at a certain point of the skin image. The detailed process description can be found in [60]. The figures 29 and 30 show testing color image and final result: two independent components, which are assumed by Tsumura et al. to be caused by hemoglobin and melanin, respectively.



(a) skin color image with 64 × 64 pixels

(b) facial image color image with 1920 × 1035 pixels

Figure 29. Analyzed color images in [60].

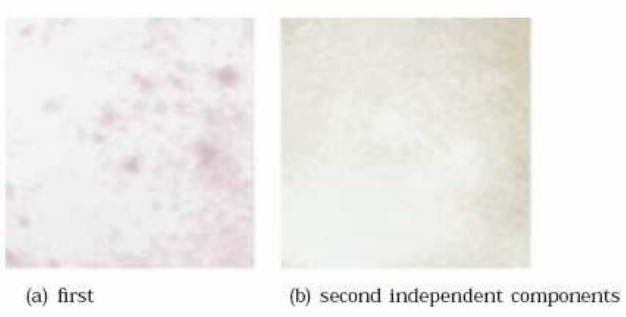


Figure 30. Two separated independent components of the skin color image [60].

In applying this technique to various parts of the body, however, it will be necessary to consider the violation of assumptions from [60] depending on the area of skin image, skin structure, skin condition, and so on. But the proposed techniques could be applied to a spectral reflectance image as well [60].

Let us consider two examples of separation of independent components for the sample spectral images 11 (area 2), and 14. On the figures 31, 33 there are ICs, obtained during the tests. For each separate component inverse operation using equation (1) is made. It is important to notice that mixing matrix A performs a role of a coefficient, which rotate and rescale each new spectrum's plot. The process of color separation is presented on figures 32 and 34. For the visualization of the color images the RGB color images are used. The figures 32.d and 34.d shows the process of obtained colors mixing, which can be compared with original skin color image.

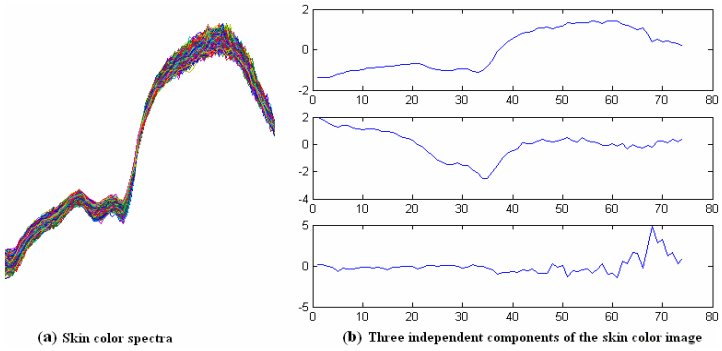


Figure 31. Skin spectra and ICs, obtained during the tests (face).

The independent components obtained during the test are centered as it is required in the subchapter 3.2.1. Therefore they were rescaled from 0 to 1 as and moved along reflectance axis by multiplication on the mixing matrix (see figures 32.a, 34.a). Each component was assign as spectra of (1×1) pixel image with 80 spectrum channels.

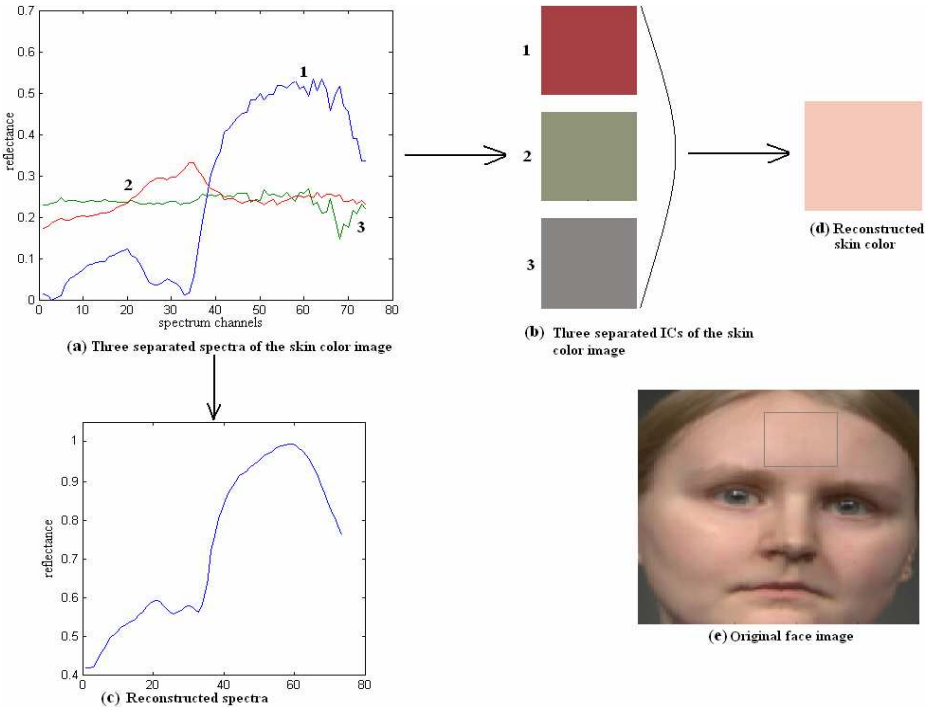


Figure 32. Separation of ICs of the skin color image (face).

On the figures 32.b and 34.b there are three separated colors correspond to three obtained independent components.

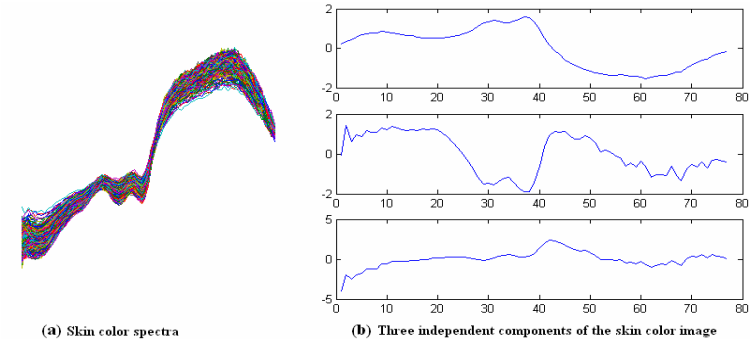


Figure 33. Skin spectra and ICs, obtained during the tests (finger).

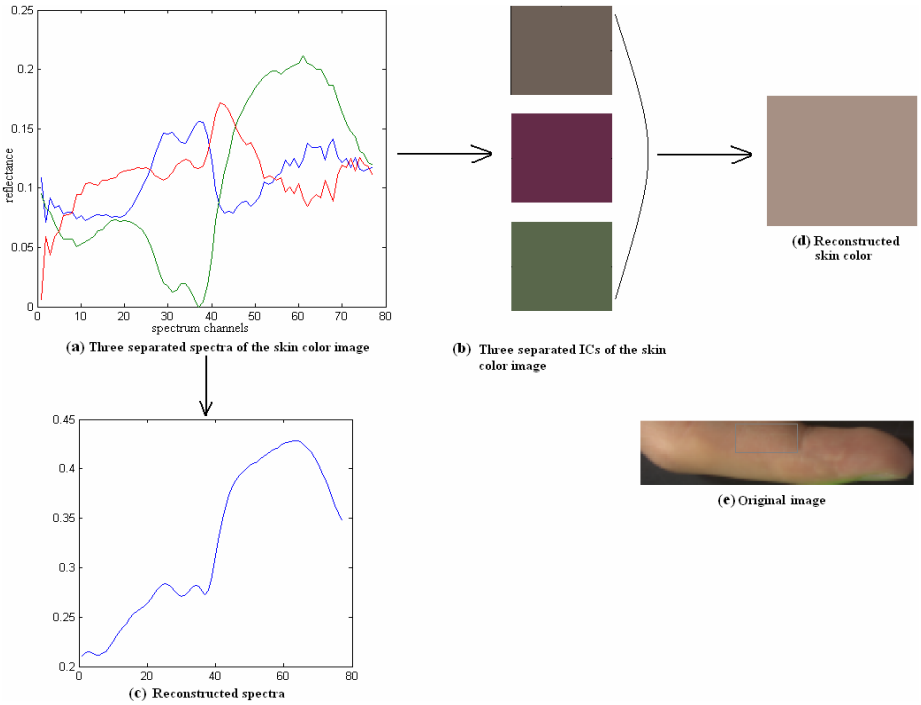


Figure 34. Separation of ICs of the skin color image (finger).

Thus in this section the separation of independent components of skin color spectral image is shown. It is recommended to use small parts of skin to get certain colors by decreasing the errors.

Brief Summary of Chapter 5

1. Independence of the obtained components was confirmed.
2. The hypothesis of Gaussian distribution of the ICs was rejected using *K*–*S*-test.
3. Three independent components as enough number for analysis of skin were chosen.
4. The symmetrical approach with nonlinearity $g(y) = y\exp(-y^2/2)$ was chosen as the most suitable estimation method to obtain ICs. Two statistical methods: ANOVA and method of comparison of two binomial distribution's probability were applied for that.
5. The facial color spectral image was separated into three images by independent component analysis.

6 Discussion

Independent components of skin were estimated in the thesis. During the tests and analysis of results some conclusions were drawn, and they are discussed in this chapter. The potentials of future employment of the work direction are versatile and can profit to medicine. Interest in this topic of the competent scientists also justify important of the further problem development.

6.1 Related Work

As it was mentioned above Tsumura *et al.* [60] have previously proposed a FastICA method to get independent components of skin. The separated components are synthesized to simulate the various facial color images by changing the quantities of the two separated pigments. However, it seems that their method is only applicable in the case of two skin pigments, melanin and hemoglobin, whereas most analysis of skin color images purpose studying of two and more skin components. The reason is that they assume that spatial variation of color in the skin is caused by two pigments. They claimed that melanin and hemoglobin are predominantly found in the epidermal and dermal layer, respectively as on the schematic model of human skin.

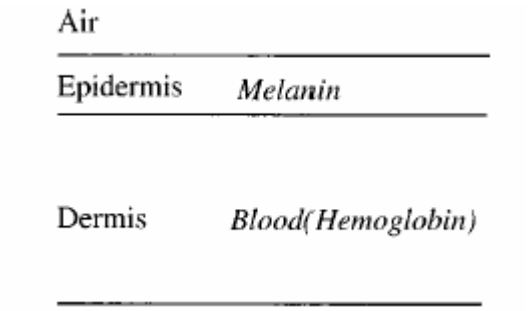


Figure 35. Schematic model of human skin with plane-parallel epidermal and dermal layers [59].

While this assumption is true for the skin model they used, it is not true for the most widely-used cases in medicine, when even the change of the most insignificant pigment or its quantity could be a serious illness indication. Tsumura *et al.* did not consider the algorithm reliability, either. Also in [60] the facial image taken by HDTV camera is used, and each pixel of the color image has three

channels: red, green, and blue. And an idea of the thesis was to apply the Independent component analysis to multispectral color images with, for example, 80 spectral channels.

6.2 Discussion of the Components

During the research work three independent components of skin were obtained. The quantity of the components was chosen to simplify the work and the principle of choice is described in Chapter 5. It was noticed that all tested images have one similar, according to shape, component, which is present in every set of three components. It enables to make a conclusion that every skin color image can be dissimilitude examined, if to fix this component and mixing matrix. This component could be accepted as a main function and others will present a difference among skin color. But it is impossible to realize in practice because of ICA algorithms feature. Every image's components are little bit different from each other, and it is impossible to make a conclusion that one skin differs from another by using the components.

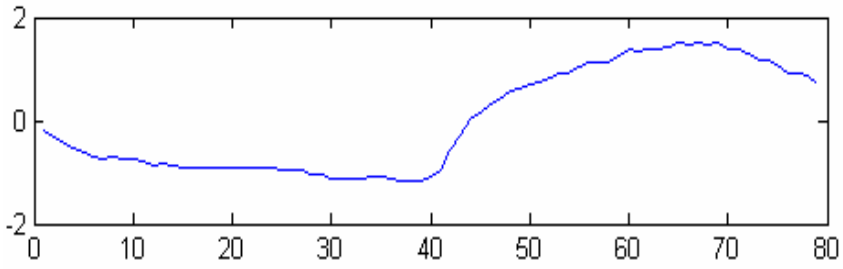


Figure 36. Independent component.

The cause of the problem is in the ICA algorithms' work, which was studied during the tests. Independent component analysis is a general-purpose statistical model, which can be widely used in the analysis of skin diseases. A major problem in application of ICA is that the reliability of the estimated IC's is not known. An ICA method gives a specified number of components, but it is not known which ones are to be taken seriously. As with any statistical method, it is necessary to check the statistical significance of the estimated components. A further problem is that ICA has random elements. It is impossible to fix the mixing matrix, an algorithm works so that every time creates randomly a new one. Somewhat different results were obtained at every run of the program; such kind of work of the algorithm throws often into confusion. As it was mentioned above in Chapter 3, the FastICA algorithm, discussing in the thesis, is based on minimization of gaussianity

(maximization of nongaussianity). The problem is that algorithm is based on methods related to gradient descent, and in case of a high-dimensional signal space, the probability of finding the global minimum may be very small [28]. There are some methods to assess the statistical reliability of obtained independent components. Therefore it is impossible to discuss the validating of the ICs directly in the thesis, but could be discussed in further works.

6.3 Future Directions

For the simplicity of work, as it was proved in the thesis, three components are enough for current study. But in future it will be interesting to consider greater number of components to estimate, and agree them with existent skin pigments that affect skin color. Also it is necessary to make tests with different kinds of human skin. Images, analyzed during the work at the thesis belong to two types of people, one Asiatic and few Scandinavian. Analysis of other types of skin and skin with skin diseases or some peculiar pigmentation will enable to make miscellaneous experiments, to obtain authentic and actual for medicine results.

6.4 Conclusions

I have carried out immense and interesting work in spectral analysis of skin color images. The aim of the research work “Analysis of spectral images of skin for medical application” was to obtain independent components of skin. The extracted components were analyzed to according to main restrictions of the ICA algorithm. During the work it was found that time, space complexity of ICA algorithm is high, and it is hard to work with large parts of skin. Because of this, I have selected small skin areas to be able perform enough tests in order to obtain reasonable amount of data for the analysis. And research work was carried out with small sections. For that two algorithms of partition were studied and implemented. I have developed the methods to convert spectral images into mixture variable and software package for implementation of them. The testing materials were available with the support of the Color laboratory at the University of Joensuu.

7 References

[1] *Accurate Image Manipulation*. Internet WWW-page, URL: http://www.aim-dtp.net (19.11.2004).

[2] Anderson R.R., Parrish J.A. The optics of human skin. *Journal of Investigative Dermatology*, Vol. 77, 1981. — pp. 13-19.

[3] Angelopoulou, E. *The reflectance Spectrum of Human Skin.* Technical Report MS-CIS-99-29. – University of Pennsylvania, December 1999. — 14p. URL: http://www.cs.stevens-tech.edu/~elli/ (19.11.2004).

[4] Angelopoulou, E. Understanding the color of human skin. *Proceedings of the SPIE Conference on Human Vision and Electronic Imaging VI (SPIE) 2001*. SPIE Vol. 4299. SPIE Press, May 2001. — pp. 243-251.

[5] Angelopoulou, E., Molana, R., and Daniilidis, K. Multispectral Skin Color Modeling. *IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2001.* IEEE Computer Society Press, December 2001. — pp. 635-642.

[6] Ans, B., Heraut, J., and Jutten, C. Adaptive neural architectures: detection of primitives. *In Proceedings of COGNITIVA'85.* – Paris, France, 1985. — pp. 593-597.

[7] Barlett, M.S., Lades, H.M., and Sejnowski, T.J. Independent component representations for face recognition. *Proceedings of the SPIE Symposium on Electronic Imaging: Science and Technology;* Conference on Human Vision and Electronic Imaging III. – San Jose, California, January 1998. — 12p.

[8] Barlett, M.S., Sejnowski, T.J. Independent components of face images: A representation for face recognition. *Proceedings of the 4th Annual Symposium on Neural Computation*. – Pasadena, CA, May 17, 1997. — 8p.

[9] Barlett, M.S., Sejnowski, T.J. Viewpoint invariance face recognition using independent component analysis and attractor networks. *Neural Information Processing Systems* – Natural and Synthetic, Vol. 9. – M. Mozer, M. Jordan, & T. Petsche, eds. MIT Press, Cambridge, MA, 1997. — pp.817-823.

[10] Behavioral Sciences Department, Palomar College. Human Biological Adaptability: *An Introduction to Human Responses to Common Environmental Stresses.* Internet WWW-page, URL: http://anthro.palomar.edu/adapt/ (19.11.2004).

[11] Bregler, C., Konig, Y. “Eigenlips” for robust speech recognition. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’94)*, Vol. 2, 1994. — pp.669-672.

[12] Cai, J., Goshtasby, A. Detecting human faces in color images. *Image and Vision Computing*, No. 18, 1999. — pp. 63-75.

[13] Cichocki, A., Karhunen, J., Kasprzak, W., and Vigario, R. Neural networks for blind separation with unknown number of sources. *Neurocomputing*, 22, 1998. — pp.113 – 129.

[14] Cichocki, A., Karhunen, J., Kasprzak, W., and Pajunen, P. On neural blind separation with noise suppression and redundancy reduction. *International Journal of Neural Systems*, Vol. 8(2), 1997. — pp.55-93.

[15] Claridge, E., Cotton, S., Hall, P., and Moncrieff. From colour to tissue histology: Physics-based interpretation of images of pigmented skin lesions. – *Medical Image Analysis*, No. 7, 2003. — pp. 489-502.

[16] Comon, P. Independent component analysis. A new concept? *Signal processing*, No. 36, 1994. — pp. 287-314.

[17] Cotton, S., D'Oyly, and Claridge, E. Developing a predictive model of human skin colouring. *Proceedings of the SPIE Medical Imaging*, Vol. 2708, 1996. — pp. 814-825.

[18] Data base of the artificial lights of the color laboratory of university of Joensuu. Internet WWW-page, URL: http://cs.joensuu.fi/spectral/intra/valot/iso_valotaulu.xls (21.11.2003).

[19] Department of Radiology. Internet WWW-page, URL: http://www.brighamandwomens.org/radiology (19.11.2004).

[20] Draper, B.A., Baek, K., Barlett, M.S., and Beveridge, J.R. Recognizing faces with PCA and ICA. – *Computer Vision and Image Understanding 91*, 2003. — pp. 115-137.

[21] Fairchild, M.D., Rosen, M.R., and Johnson, G.M. Spectral and Metameric Color Imaging. *Technical Report.* – Munsell Color Science Laboratory, August 2001. — 8p. URL: http://www.cis.rit.edu/research/mcsl/research/reports.shtml#SpectralImaging (10.02.2004).

[22] Farkas D.L., Becker, D. Applications of spectral imaging: detection and analysis of human melanoma and its precursors. *Pigment cell research* / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society, No. 14(1), February 2001. — pp.2-8.

[23] FastICA package for Matlab. — Helsinki university of Technology. Home page of the Laboratory of Computer and Information Science. Internet WWW-page, URL: www.cis.hut.fi/projects/ica (10.02.2004).

[24] Fränti, P. *Image Compression.* Lecture notes. – University of Joensuu, Department of Computer Science. URL: www.cs.joensuu.fi/pages/franti/image/notes.html (19.11.2004).

[25] Gmurman, V.E. *Theory of probability and Mathematical statistics. Textbook for institutes of higher education*, 7th stereotyped edition. – Moscow, "High school", 2001. — 479 p., graphics. (Original: Гмурман В.Е. *Теория вероятностей и математическая статистика*. Учебное пособие для ВУЗов, изд. 7-е, стереотипное. – М: Высшая школа, 2001. — 479 с.: ил.)

[26] *Great Medical encyclopedia. In twenty nine volumes.* — V.: 1, 4, 5, 7, 9, 14, 17, 19, 21, 23, 26, 27, 28. – Moscow: Soviet encyclopedia, 1974. — 15912p. (Original: *Большая медицинская энциклопедия. В двадцати девяти томах*. — Том 1, 4, 5, 7, 9, 14, 17, 19, 21, 23, 26, 27, 28. – М: Советская энциклопедия, 1974. — 15912с.)

[27] Gula, O.G., Dana, K.J. Image-based Skin Analysis. *Proceedings of Texture 2002* - The 2nd international workshop on texture analysis and synthesis, June 1, 2002, Copenhagen, Denmark (in conjunction with European Conference on Computer Vision ECCV 2002). — pp. 35-41.

[28] Himberg, J., Hyvarinen, A., and Esposito, F. Validating the independent components of neuroimaging time-series via clustering and visualization. *Neuroimage*, No. 22 (3), 2004. — pp.1214-1222.

[29] Hopfield, J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc. Nat'l Academy of Science*, Vol. 79, 1982. — pp. 2,554-2,558.

[30] Hsieh, I.-S., Fan, K.-C., and Lin, C. A statistic approach to the detection of human faces in color nature scene. *Pattern Recognition*, No. 35, 2002. — pp.1583-1596.

[31] Hutchings, J.B. *Color in plants, animals and man.* – North-Holland, 1998 — 25p.

[32] Hyvarinen, A., Karhunen, A., and Oja, E. *Independent Component Analysis.* – John Wiley & Sons, Inc., New York, USA, 2001. — 482p.

[33] Hyvarinen, A., Oja, E. *Independent Component Analysis: A Tutorial.* – Finland, Helsinki University of Technology, 1999. — 30p.

[34] Ibraghimov, I.E. *Function interpolation methods and theirs applications.* – Moscow: "Nauka", Main editorship of physical and mathematical literature, 1971. — 520p. (Original: Ибрагимов И.И. *Методы интерполяции функций и некоторые их применения.* – М.:"Наука", гл. Редакция физико-математической литературы, 1971. — 520с.)

[35] Imai, F.H., Tsumura, N., Haneshi, H., and Miyake, Y. Principal Component Analysis of Skin Color and Its Application to Colorimetric Color Reproduction on CRT Display and Hardcopy. *Imaging Science and Technology*, No. 40, 1996. — pp. 422-430.

[36] Inoue, T., Fujii, Y., Itoh, K., and Ichioka, Y. Block blind separation of independent spectra in hyperspectral images. International Topical Meeting on Optical Computing: *Technical Digest*, Vol. 1. – The Japan Society of Applied Physics, Tokyo, 1996. — p 40.

[37] Introduction to color research. Slides. Internet WWW-page, URL: http://spectral.joensuu.fi (19.11.2004).

[38] Jablonski, N.G., Chaplin, G. The evolution of human skin coloration. – *Human Evolution* 39, 2000. — pp. 57-106.

[39] Katzman, P., Doyle, L. NIR Tool Gives Burn Patients Better Chance of Strong Recovery. *Scientific Computing & Instrumentation*, The Math Works Inc., August 1, 2002. — pp.20-22.

[40] Kiselgov, E.N. *Independent component analysis and application for statistical independent signals retrieving.* Technology of instrument-making industry 1-2. – Kharkov, 2001. — pp.142-147. (Original: Е.Н. Кисельгов. *Независимый компонентный анализ и его применение к задачам выделения статистически независимых сигналов.* Технология приборостроения 1-2. – Харьков, 2001. — с. 142-147.)

[41] Kohonen, T. The Self-organizing Map. *Proceedings of the IEEE*, Vol. 78, No. 9, September 1990. — pp.1464-1480.

[42] Korolyuk, M.S., Portenko, N.I., Skorohod, A.V., and Turbin, A.F. *Hand-book in theory of probability and mathematical statistics*. – Moscow, Nauka. The main editorial office of physical mathematical literature, 1985. — 640 p. (Original: Королюк В.С., Портенко Н.И.,

Скороход А.В., Турбин А.Ф. *Справочник по теории вероятностей и математической статистике*. – М.: Наука. Главная редакция физико-математической литературы, 1985. — 640 с.)

[43] Leppajarvi, S., Parkkinen, J. Independent Filters for Color Separation. – *Research Report*, No. 63, Department of Information Technology, Lappeenranta University of Technology, 1999. — 14p.

[44] Marschner, S.R., Westin, S.H., Lafortune, E.P.F., Torrance, K.E., and Greenberg, D.P. Reflectance Measurements of Human Skin. *Technical report*, No. 2, PCG-99-2, Program of Computer Graphics, Cornell University, January 1999. — 12p.

[45] Martinkauppi, B. *Face colour under varying illumination - analysis and applications*. Department of Electrical and Information Engineering, University of Oulu. URL: http://herkules.oulu.fi/isbn9514267885/html/index.html (19.11.2004).

[46] Martynenko, A. Spectral Independent Component Analysis in Heart Rate Variability Nonlinear Dynamic. – Kharkov National University. *Second HRV Congress Materials*, 2002. — 9 slides. URL: http://www.hrvcongress.org/second/russian/first/15.html (19.11.2004).

[47] *Mathematics. Great Encyclopedic Dictionary.* / Main redactor Prohorov, Y.V. — Third publication. – Moscow: Great Russian Encyclopedia, 2000. — 848p. (Original: *Математика. Большой энциклопедический словарь.*/ Гл. Редактор Ю.В. Прохоров. — 3-е изд. – М.: Большая Российская энциклопедия, 2000. —848с.:ил.)

[48] *Matlab Help.* Internet WWW-page, URL: http://www.mathworks.com/access/helpdesk/help/techdoc/matlab.shtml (19.11.2004)

[49] Meglinski, I.V., Matcher, S.J. Computer simulation of the skin reflectance spectra. *Computer Methods and Programs in Biomedicine*, No.70, 2000. — pp. 179-186.

[50] Roberts, S. Independent Component Analysis: Source assessment & separation, a Bayesian approach. *IEE Proceedings – Vision, Image & Signal Processing*, 145, 1998. — pp. 149-154.

[51] Romanovskiy, V.I. *Mathematical statistics*. – Tashkent: Academy of Sciences of Uzbek SSR, Vol. 2, 1963. — 794 с. (Original: Романовский В.И. *Математическая статистика*. – Ташкент: Изд-во Академии наук Узбекской ССР, Т.2, 1963. — 794 с.)

[52] Romney, A.K., Indow, T. *Estimating physical reflectance spectra from human color-matching experiments*. URL: http://hypatia.ss.uci.edu/imbs/personnel/romney/cmfcolor.pdf (14.05.2004).

[53] Ryer, A. *Light Measurement Handbook*. – International Light Inc., 1998. URL: http://www.intl-light.com/handbook (03.03.2004).

[54] Sagizbaeva, O. Implementation description. *IT – project*. – Finland, University of Joensuu, Department of Computer Science, 2004.

[55] Sagizbaeva, O. User Manual. *IT – project*. – Finland, University of Joensuu, Department of Computer Science, 2004.

[56] *Skin color. Human Pigmentation and adaptation*. Internet WWW-page, URL: http://www.as.ua.edu/ant/bindon/ant475/Skincolor/SKINCOLOR.PDF (15.05.2004).

[57] The Department of Radiology. Weill Medical College of Cornell University. Internet WWW-page, URL: http://www.nycornell.org/radiology (19.11.2004).

[58] The Franklin Institute Online. Internet WWW-page, URL: http://sln.fi.edu/biosci2/monitor (19.11.2004).

[59] The Medical Radiography Home Page. Internet WWW-page, URL: http://home.earthlink.net/~terrass/radiography/history.html (19.11.2004).

[60] Tsumura, N., Haneishi, H., and Miyake, Y. Independent component analysis of skin color image. – *Optical Society of America*, Vol. 16, No. 9, 1999. — pp. 2169-2176.

[61] Tsumura, N., Haneishi, H., and Miyake, Y. Independent Component Analysis of Spectral Absorbance Image in Human Skin. – *Optical Review*, Vol. 7, No. 6, 2000. — pp. 479-482.

[62] Tsumura, N., Kawabuchi, M., Haneishi, H., and Miyake, Y. Mapping pigmentation in human skin by multivisible-spectral imaging by inverse optical scattering technique. *Imaging Science and Technology (IS&T)*/ Society for Information Display (SID's) 8th Color Imaging Conference, Color Science, Systems and Applications, 2000. — pp.81-84.

[63] Vadzinskiy, R.N. *Probability distribution hand-book*. – St. Petersburg, Nauka, 2001. — 295 p., 116 graphics. (Original:Вадзинский Р.Н. *Справочник по вероятностным распределениям*. – СПб.: Наука, 2001. — 294 с.: ил. 116.)

[64] Volkov, E.A. *Numerical methods: A tutorial for institutes of higher education*, 2nd edition, revised. – Moscow, Nauka. The main editorial office of physical mathematical literature,

1987. — 248 p. (Original: Волков Е.А. *Численные методы: Учебное пособие для ВУЗов*, 2-е изд., испр. – М.: Наука. Гл. Ред. Физ.-мат. Лит., 1987. — 248 л.)

[65] Wyszecki, G., Stiles, W.S. *Color science: concepts and methods, quantitative data and formulae*, 2nd edition. – New York, Wiley, 2000. — 950p.

[66] Yevseev, V.S. T*raining and defense of dissertation: Information methodical manual.* – St. Petersburg, "Polytechnics", 1991. — 304 p.: graphics. (Original: Евсеев В.С. *Подготовка и защита диссертации: Справочно–методическое пособие*. – СПб.: Политехника, 1991. — 304 с.: ил.)

[67] Young, A.R. Chromophores in human skin. *Physics in Medicine and Biology*, No. 42, May 1997. — pp. 789-802.

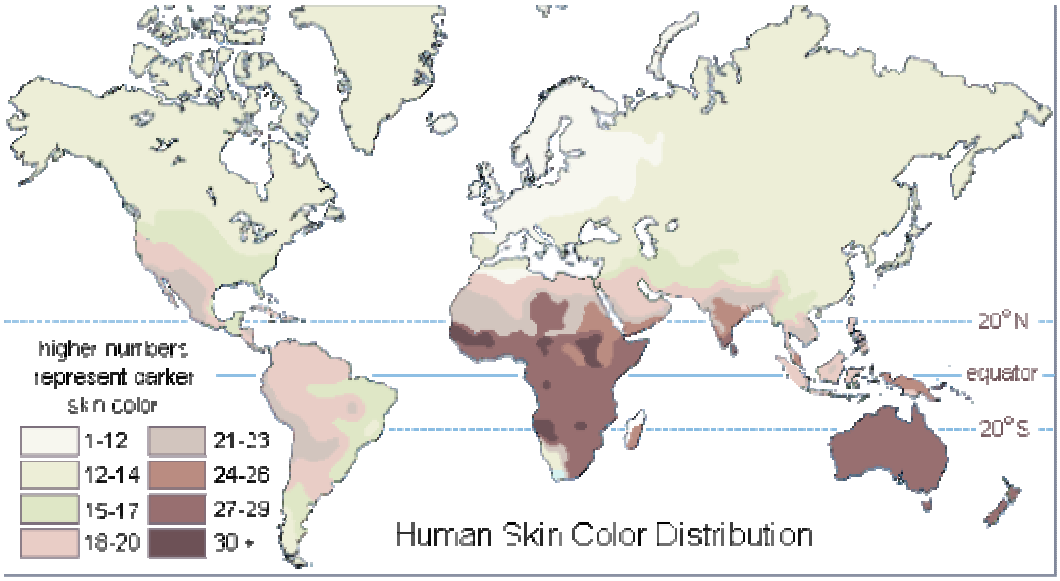
[68] Zhang, Y., Prakash, E.C., and Sung, E. Constructing a Realistic Face Model of an Individual for Expression Animation. *International Journal of Information Technology*, Vol. 8, No. 2, September 2002. — pp.10-25

[69] Resource for watercolor painters on the Internet. Internet WWW-page, URL: http://ttt.upv.es/~gbenet/teoria%20del%20color/water_color/wcolor.html (19.11.2004).

Appendices

A. Skin Color Distribution Around the World

Before the mass human migrations of the last 500 years, dark skin color was mostly concentrated near the equator and light color progressively increased further away, as illustrated in the map below. In fact, the majority of dark pigmented people lived within 20° of the equator. Most of the lighter pigmented people lived in the northern hemisphere north of 20° latitude.



Such a non-random distribution pattern of human skin color was predicted by Wilhelm Gloger, a 19th century naturalist. In 1833, he observed that heavily pigmented animals are to be found mostly in hot climates where there is intense sunshine. Conversely, those in cold climates closer to the poles commonly have light pigmentation. The relative intensity of solar radiation is largely responsible for this distribution pattern [10].

B. Testing Material

B.1 Spectral Images Used in the Thesis

Altogether skin sample images were taken in color lab, using ImSpector V8 multispectral camera (spectrograph) and under the "Daylight 65" (D65) illumination produced by the Light Booth.

In the table B.1 the radiance type spectral images are presented. Recorded spectrum range is from 380 Nm to 775 Nm (5 Nm for each channel).

Table B.1. Radiance type spectral images of skin.

File name	Description	Size	Volunteer
Simage_1.mat	backside of hand, where contains two mosquito bites (size 100*160*80)	10MB	Chinese
Simage_2.mat	backside of hand (size 300*160*80)	30MB	Finnish
Simage_3.mat	middle finger with a cut wound near the first joint, where the green color shown in RGB image was caused by light reflection of a green paper nearby (size 40*160*80)	4MB	Chinese
Simage_4.mat	little finger (size 50*160*80)	5MB	Finnish
Simage_5.mat	Palm (size 50*160*80)	10MB	Russian
daylight.mat	illumination spectrum (size 100*160*80)	1KB	D65

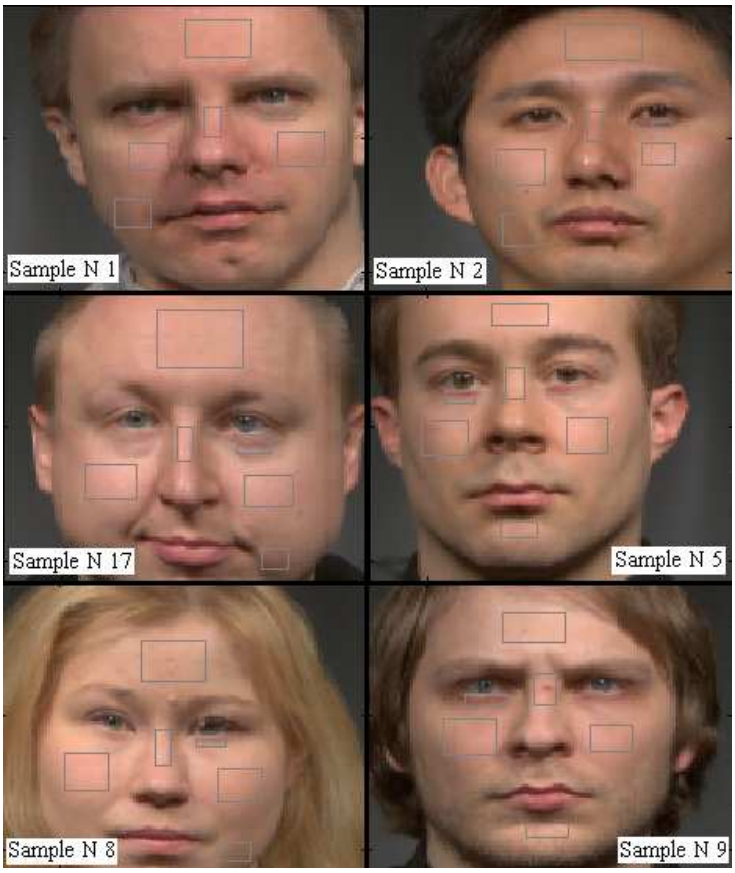
In the table B.2 the reflectance type spectral face images are presented. Recorded spectrum range is from 380 Nm to 780 Nm (5 Nm for each channel). The all images' size is 130*125*80.

Table B.2. Reflectance type multispectral face images.

File name	Description	Size	Volunteer
anssi.mat	face with the reddening at the nose area	10,119Mb	Finnish
hiroaki.mat	face	10,119Mb	Japanese
jouni.mat	rosy-cheeked face	10,119Mb	Finnish
juha.mat	face	10,119Mb	Finnish
kimmo.mat	face	10,119Mb	Finnish
lari.mat	face	10,119Mb	Finnish
laura.mat	face	10,119Mb	Finnish
markku.mat	face	10,119Mb	Finnish
oili.mat	face	10,119Mb	Finnish
pekka.mat	face with the pimple on the nose	10,119Mb	Finnish
petri.mat	face, containing birthmarks	10,119Mb	Finnish
tuija.mat	face	10,119Mb	Finnish

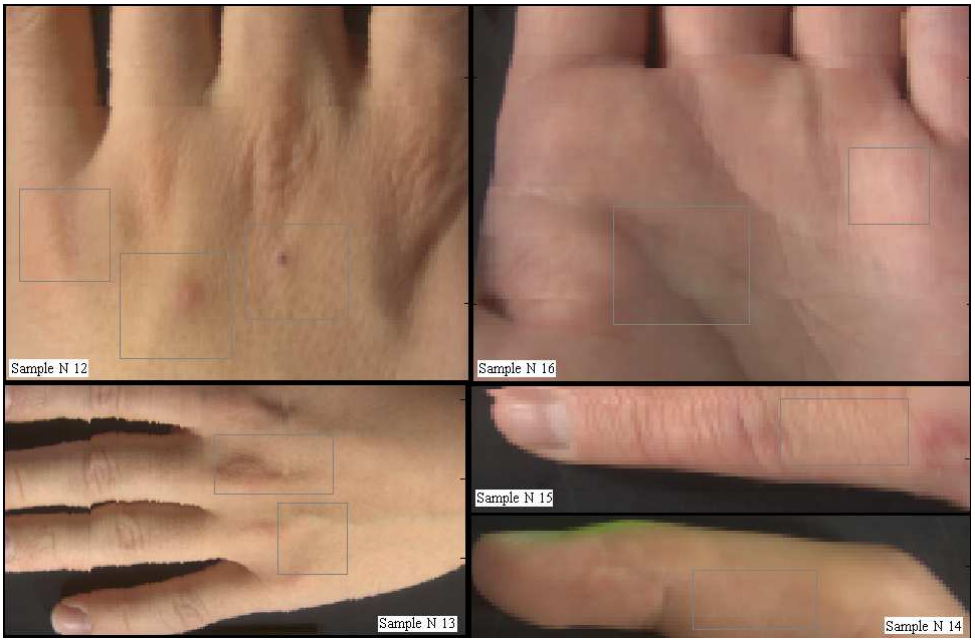
B.2 Catalogue of Tested Images

Here are the images, which were tested during the work on the thesis. The grey squares mean the areas selected for the analysis.



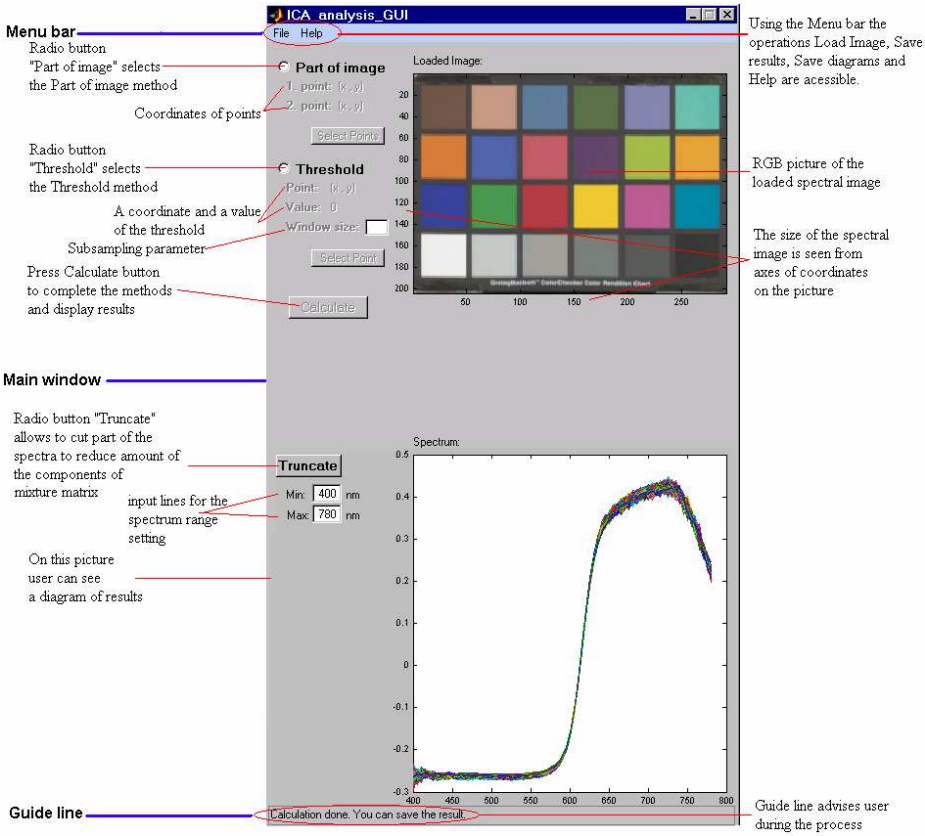
Picture B.1. Sample images (part 1).

Picture B.2. Sample images (part 2).

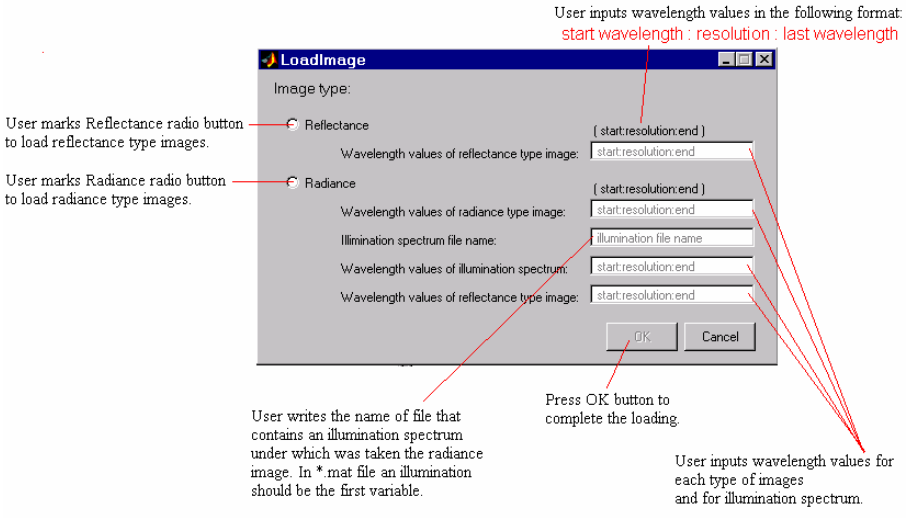


Picture B.3. Sample images (part 3).

C. ICA_analysis_GUI Program. Interface.



Picture C.1. Main window of the program ICA_analysis_GUI.



Picture C.2. LoadImage dialog window.

D. Preprocessing Operations

D.1 Reconstruction of Independent Components

Here the additional results, which were retrieved during the work on the thesis, are presented. Figures D.1, D.2 show the difference between original skin spectra (the diagram in the down-centre of the figures) and reconstruction of it by the significant independent components. The reconstruction of the significant IC's are shown on the figures from left to right and from up to bottom. During the test in the same order as in the figure 2, 3, 4, 5, 6, 12, 32, 64 and all independent components were obtained. As example three different samples from the list of samples were tested. Here the two samples are shown, the third one is considered in the chapter 5. Also, in detail, the process of reconstruction is described in the subchapter 5.2.

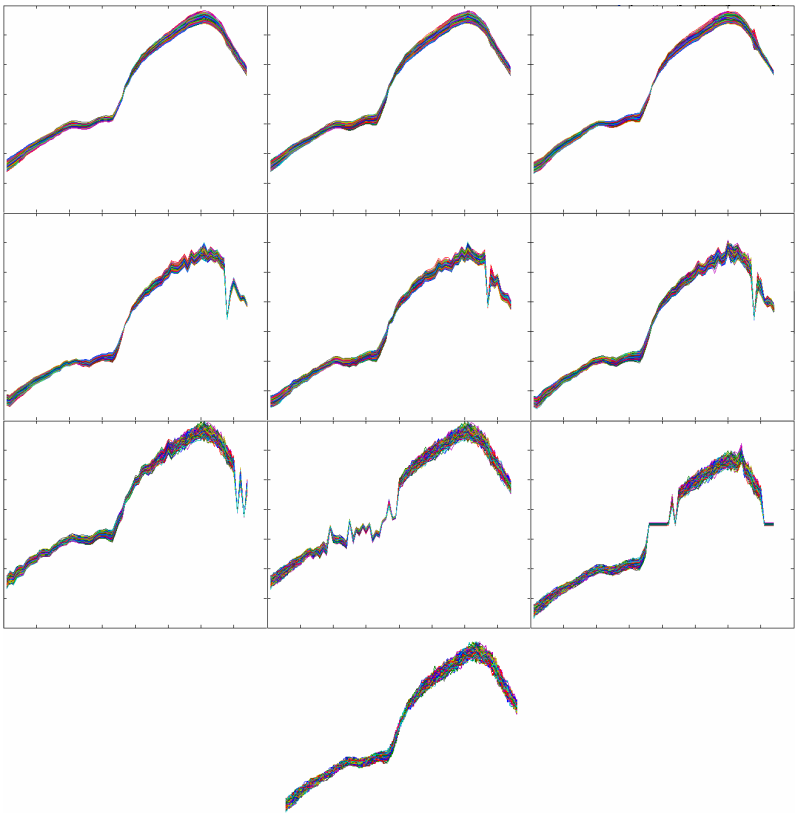


Figure D.1. Reconstructions by the significant IC's versus original skin spectra: area 2, sample 2.

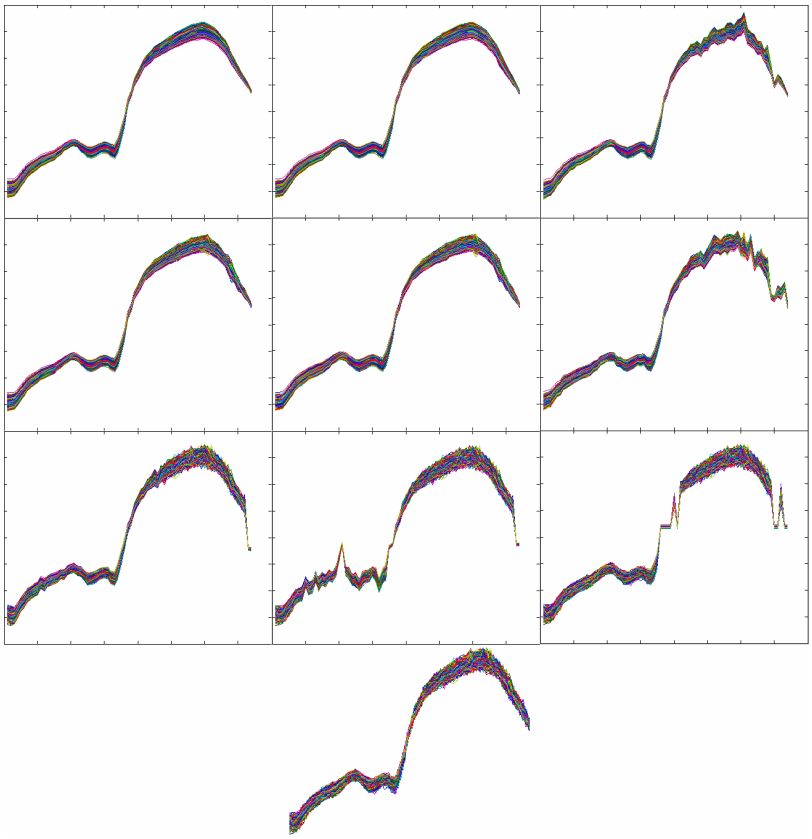


Figure D.2. Reconstruction of the significant IC's versus original skin spectra: area 2, sample 11.

In the tables D.1.–D.3. there are numerical characteristics of the tests are shown. For the test three different samples were considered from the list of samples (see Appendix A). The samples represent part of skin, which is indicated as area N2 (forehead). The differences of reconstruction by all independent components and significant IC's are presented in the tables. For the test the following numbers of components were obtained: 2, 3, 4, 5, 6, 12, 32, 64 and all (depends on how many channel were truncated in order to cut noise). For the differences the minimal, maximum and average values are shown; also a standard deviation was calculated for difference range, which gives us a notion of difference behavior.

Table D.1. Testing of the Sample N1 of the area N2.

IC's	Difference of reconstruction of all IC's				Error limit, δ	Difference of reconstruction of significant IC's				Number of significant IC's
	mean	max	min	standard deviation		mean	max	min	standard deviation	
2	0,00478	0,00977	0,00217	0,00215	0,01	0,00478	0,00977	0,00217	0,00215	2
3	0,00456	0,00919	0,00215	0,00208	0,01	0,00456	0,00919	0,00215	0,00208	3
4	0,00445	0,00891	0,00215	0,00193	0,01	0,00765	0,03443	0,00279	0,00549	2
5	0,00433	0,00881	0,00215	0,00185	0,01	0,00848	0,09301	0,00280	0,01146	4
6	0,00422	0,00748	0,00213	0,00175	0,01	0,00721	0,06225	0,00301	0,00746	4
12	0,00361	0,00632	0,00150	0,00135	0,01	0,00748	0,13255	0,00152	0,01548	11
32	0,00197	0,00328	0,00086	0,00075	0,01	0,00822	0,14661	0,00091	0,02143	30
64	0,00085	0,00148	0,00072	0,00019	0,01	0,00956	0,10721	0,00107	0,02163	57
70	0,00072	0,00072	0,00072	0,00000	0,01	0,00859	0,10123	0,00237	0,02029	63

Table D.2. Testing of the Sample N2 of the area N2.

IC's	Difference of reconstruction of all IC's				Error limit, δ	Difference of reconstruction of significant IC's				Number of significant IC's
	mean	max	min	standard deviation		mean	max	min	standard deviation	
2	0,00434	0,00856	0,00213	0,00168	0,01	0,00434	0,00856	0,00213	0,00168	2
3	0,00396	0,00854	0,00181	0,00184	0,01	0,00396	0,00854	0,00181	0,00184	3
4	0,00386	0,00785	0,00181	0,00174	0,01	0,00568	0,00971	0,00187	0,00173	3
5	0,00378	0,00690	0,00181	0,00164	0,01	0,00894	0,13988	0,00258	0,01633	3
6	0,00369	0,00688	0,00178	0,00157	0,01	0,00763	0,12067	0,00214	0,01412	5
12	0,00320	0,00544	0,00159	0,00115	0,01	0,00782	0,13875	0,00191	0,01598	10
32	0,00175	0,00314	0,00067	0,00072	0,01	0,00751	0,13777	0,00075	0,02060	30
64	0,00070	0,00140	0,00051	0,00025	0,01	0,00881	0,10102	0,00056	0,01822	52
73	0,00051	0,00051	0,00051	0,00000	0,01	0,00949	0,09241	0,00115	0,02062	61

Table D.3. Testing of the Sample N11 of the area N2.

IC's	Difference of reconstruction of all IC's				Error limit, δ	Difference of reconstruction of significant IC's				Number of significant IC's
	mean	max	min	standard deviation		mean	max	min	standard deviation	
2	0,00490	0,00955	0,00240	0,00189	0,01	0,0049	0,00955	0,00240	0,00189	2
3	0,00480	0,00869	0,00239	0,00175	0,01	0,0048	0,00869	0,00239	0,00175	3
4	0,00469	0,00848	0,00239	0,00164	0,01	0,00808	0,05708	0,00270	0,00811	3
5	0,00460	0,00758	0,00239	0,00154	0,01	0,0046	0,00758	0,00239	0,00154	5
6	0,00447	0,00758	0,00239	0,00153	0,01	0,00447	0,00758	0,00239	0,00153	6
12	0,00390	0,00635	0,00231	0,00121	0,01	0,00996	0,07345	0,00236	0,01289	9
32	0,00236	0,00373	0,00116	0,00075	0,01	0,00776	0,13894	0,00116	0,02041	30
64	0,00123	0,00183	0,00106	0,00022	0,01	0,00974	0,12040	0,00112	0,02208	58
73	0,00104	0,00104	0,00104	0,00000	0,01	0,00995	0,12027	0,00117	0,02602	64

D.2 Choosing of the Method

Table D.4. Methods' convergences.

Sample N1	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	10 / 10	9 / 10	10 / 10	10 / 10
4	10 / 10	5 / 10	7 / 10	10 / 10	10 / 10	10 / 10

Sample N2	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
4	10 /10	10 /10	10 / 10	10 / 10	10 / 10	10 / 10

Sample N3	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	6 / 10	10 / 10	10 / 10	10 / 10	10 / 10
4	6 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10

Sample N4	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
4	1 / 10	9 / 10	10 / 10	8 / 10	10 / 10	10 / 10

Sample N5	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	7 / 10	10 / 10	10 / 10	10 / 10
4	8 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10

Sample N6	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
4	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10

Sample N7	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
4	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10

Sample N8	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	0 / 10	10 / 10	10 / 10	7 / 10	10 / 10	2 / 10
4	9 / 10	8 / 10	6 / 10	0 / 10	3 / 10	1 / 10

Sample N9	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
4	10 / 10	10 /10	10 /10	10 / 10	10 / 10	10 /10

Sample N10	Convergence / Number of launchings					
Number of IC's	M1	M2	M3	M4	M5	M6
2	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
3	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10
4	0 / 10	8 / 10	9 / 10	10 / 10	10 / 10	10 / 10

D.3 K–S-test

Here is the statistical *K–S* test is presented. As example, testing of sample N1 from area N5 is shown. Nongaussianity of 3, 30 and all possible 77 IC's was proved. Each figure includes information about test parameters. Cumulative distribution functions of observed IC's and Normal distribution function (Gaussian) were plotted. The Gaussian curve was plotted on the basis of formulas from [62]. A hypothesis that observed distribution function is Gaussian was rejected with significance value 0.05. The theoretical explanation of the test is shortly given in Chapter 5.

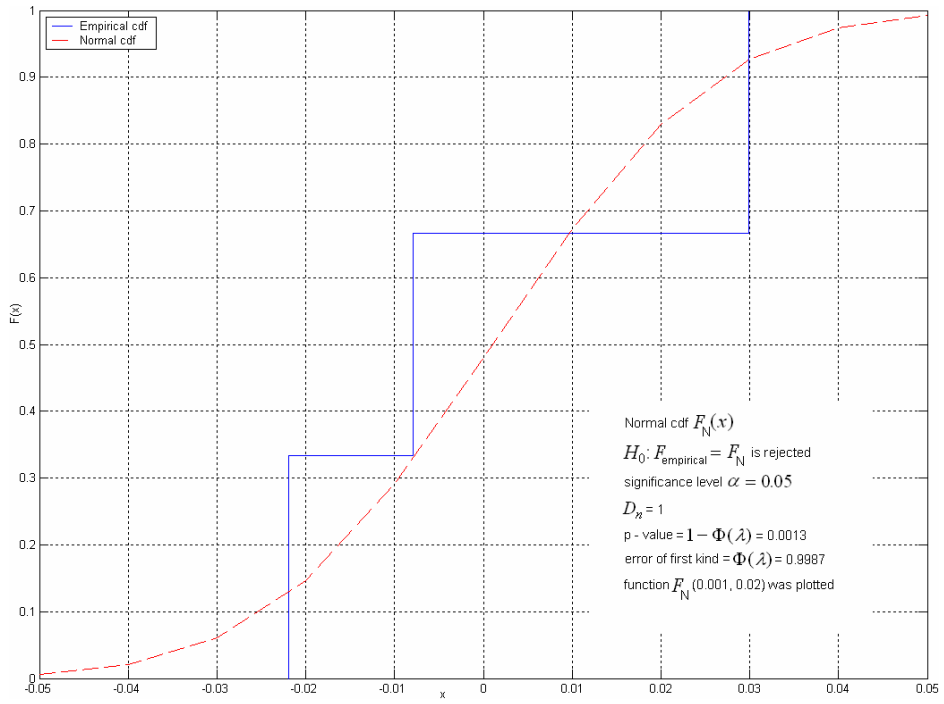


Figure D.3. K–S-test with 3 IC's.

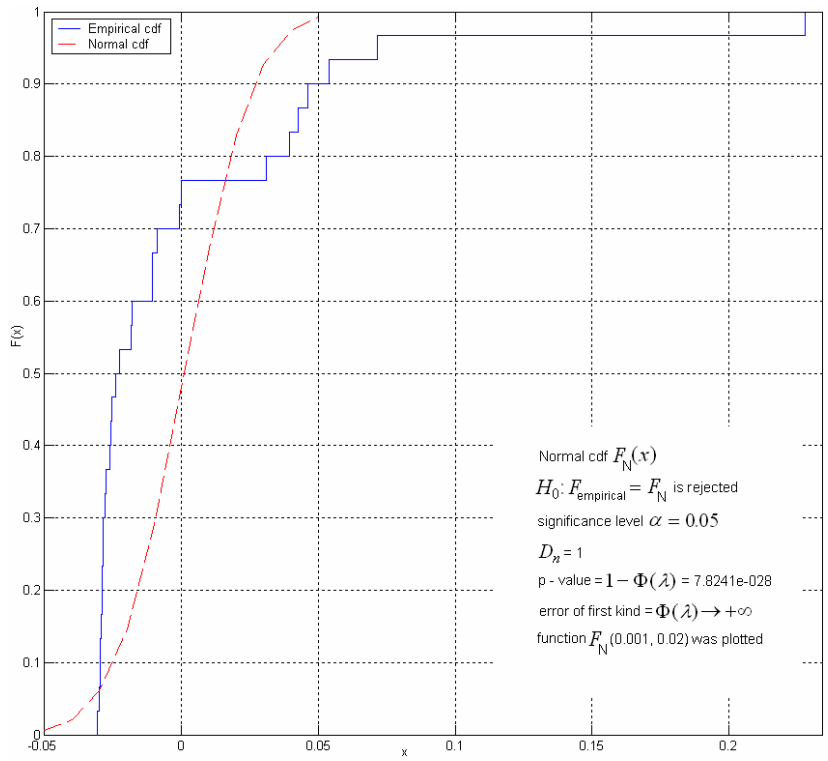


Figure D.4. K–S-test with 30 IC's

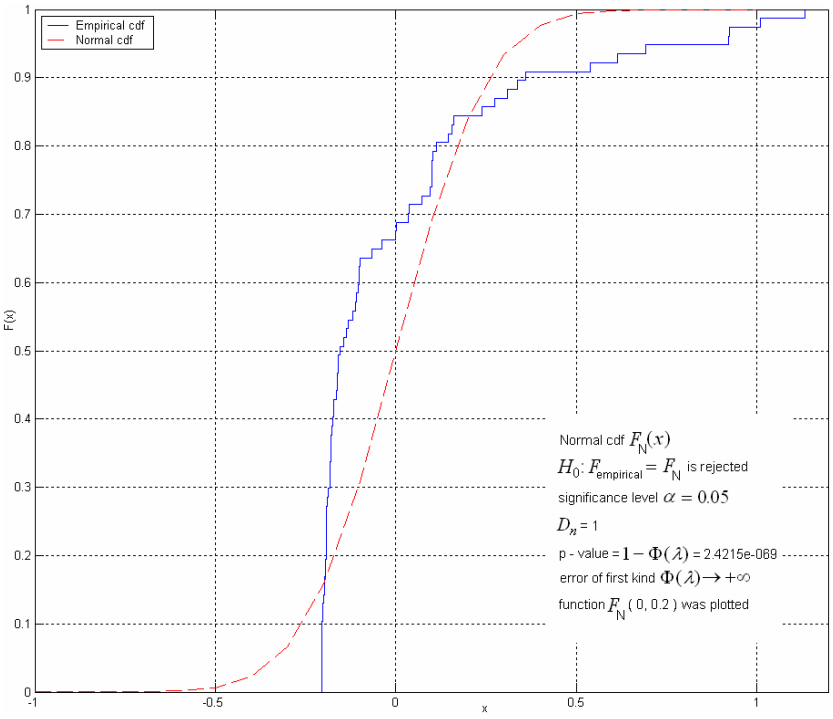


Figure D.5. K–S-test with 77 IC's

D.4 ANOVA Test

To compare more than two groups the Analysis of variance (ANOVA) method is used. The test uses a test statistic called the F statistics, if the calculated value F_{calc} is greater than the tabled value of F_{crit}, the null hypothesis that says, difference between the sample means are not significantly different from each other, is rejected. Shortly, the algorithm can be written as follow:

Table D.5. ANOVA algorithm.

1. Record data in columns.
2. Compute total sum of squares: $SS_T = \sum x^2 - \frac{(\sum x_T)^2}{N}$, where N–a number of observation in all groups
3. Compute between and then within groups sum of squares: $SS_b = \sum \frac{(\sum x)^2}{n} - \frac{(\sum x_T)^2}{N}$ and $SS_W = SS_T - SS_b$, where n–a number of observations in a group.
4. Calculate the degrees of freedom: between groups $df_b =$ number of groups -1, for the total $df_T =$ number of subjects -1, and within groups $df_W = df_T - df_b$.
5. Compute the mean squares for each: $MS_b = \frac{SS_b}{df_b}$, $MS_W = \frac{SS_W}{df_W}$.
6. Compute value for F : $F_{calc} = MS_b / MS_W$.
7. Find the critical value of F in the statistic table: critical F_{α, df_b, df_W}.
8. Determine whether F is significant: if $F_{calc} > F_{crit}$ the null hypothesis is rejected.

The advantage of the ANOVA test is that it tests the difference between the means of two or more groups. It can be used with any number of data sets, recorded from any process. The data sets need not be equal in size. Data sets suitable for ANOVA can be as small as three or four numbers, to infinitely large sets of numbers. The following tables are the results of ANOVA test implemented in Excel program. Test was applied to 3, 4 and all components. The main important values are F_{calc} and F_{crit}, theirs' comparison gives the test result. As it is seen from the tables D.6 – D.11, $F_{calc} < F_{crit} = F_{0.05,5,54}$, so the hypothesis $H_0 : \mu_1 = ... = \mu_6 = \mu$ can not be rejected at significant level $\alpha = 0.05$. The other values in the tables are calculated by the algorithm. A p -value is a type-

1 error, explaining in the Paragraph 5.1, shows that the probability of rejecting of the right hypothesis is high for all three cases.

1. For 3 ICs

Table D.6. SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
M1	10	90	9	10
M2	10	96	9.6	1.6
M3	10	97	9.7	0.9
M4	10	96	9.6	0.933333
M5	10	100	10	0
M6	10	92	9.2	6.4

Table D.7. ANOVA

Source of Variation	*SS*	*df*	*MS*	*F calc*	*p-value*	*F crit*
Between groups	6.483333	5	1.296667	**0.392269**	0.851994	**2.386066**
Within groups	178.5	54	3.305556			
Total	184.9833	59				

2. For 4 ICs

Table D.8. SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
M1	10	74	7.4	14.93333
M2	10	90	9	2.666667
M3	10	92	9.2	2.177778
M4	10	88	8.8	9.955556
M5	10	93	9.3	4.9
M6	10	91	9.1	8.1

Table D.9. ANOVA

Source of Variation	*SS*	*df*	*MS*	*F calc*	*p-value*	*F crit*
Between groups	25	5	5	**0.702028**	0.624334	**2.386066**
Within groups	384.6	54	7.122222			
Total	409.6	59				

3. For all

Table D.10. SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
M1	10	164	16.4	21.37778
M2	10	186	18.6	3.377778
M3	10	189	18.9	2.544444
M4	10	184	18.4	16.48889
M5	10	193	19.3	4.9
M6	10	183	18.3	28.9

Table D.11. ANOVA

Source of Variation	*SS*	*df*	*MS*	*F calc*	*p-value*	*F crit*
Between groups	50.68333	5	10.13667	**0.783875**	0.56579	**2.386066**
Within groups	698.3	54	12.93148			
Total	748.9833	59				

D.5 Probability Comparison Method of Two Binomial Distributions

Let perform $n = 100$ tests in every populations, method's *Convergence* is observed $m(M_i)$ times, therefore the relative frequency [41] of occurrence of event in each population:

$$\varpi_i(Convergence) = m_i / n .$$

At significance level $\alpha = 0.05$, accepted in medicine, the hypothesis $H_0 : p_i = p_j = p$ that probabilities of convergence for both methods are equal is checked. The competing hypothesizes are $H_1 : p_i > p_j$ and $H_1 : p_i < p_j$.

Rule 1. At the competing hypothesizes $H_1 : p_i > p_j$ the following criterions are carried out:

By the data, competing hypothesis is right-side, that means, the critical point must be found using Laplace table and equation: $\Phi(u_{critical}) = (1 - 2\alpha)/2$.

- If $U_{observed} < u_{critical}$ the hypothesis H_0 is accepted.

- If $U_{observed} > u_{critical}$ the hypothesis H_0 is rejected.

Rule 2. At the competing hypothesizes $H_1 : p_i < p_j$ the following criterions are carried out:

The critical point must be found as in the Rule 1, and then border of the left-side critical area is supposed to be $u_{critical}{}' = -u_{critical}$.

- If $U_{observed} > -u_{critical}$ the hypothesis H_0 is accepted.

- If $U_{observed} < -u_{critical}$ the hypothesis H_0 is rejected.

The criterion to check of null hypothesis is

$$U = \frac{M_i / n_i - M_j / n_j}{\sqrt{p(1-p)(1/n_i + 1/n_j)}},$$

where the unknown probability can be changed by most likelihood estimate:

$$p^* = (m_i + m_j)/(n_i + n_j)$$