LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
Department of Information Technology

Highlight Removal in Spectral Images

Master's Thesis

The topic of the Thesis has been confirmed by the Department Council of the Department of Information Technology on 22 March, 2006

Supervisor: Ph.D. Vladimir Bochko
Examiner: Diana Kalenova

Author: Ivan Shishkarev
Korpikunnaankatu 5 B 14
53850, Lappeenranta, Finland
Phone: +358 407395297
E-mail: ivan.shishkarev@lut.fi

Lappeenranta, 18.05.2006

ABSTRACT

Lappeenranta University of Technology
Department of Information Technology
Author: Ivan Shishkarev
Thesis Title: **Highlight Removal in Spectral Images**
Thesis for the Degree of Master of Science in Information Technology, 2006
55 pages, 27 figures, 1 table
Examiner: Diana Kalenova
Supervisor: Vladimir Bochko, Ph.D
Keywords: *color constancy, spectral images, highlight removal*

In many applications it is very important to reduce influence of a light source to see a real color of objects. Especially this technique is very useful for E-museums, telemedicine, E-shops and E-money. In this study the highlight removal technique is proposed and developed.

An overview for a general color image understanding is given in the thesis and based on it the different highlight removal techniques were analyzed, and the new method on highlight removal was proposed based on the Dichromatic Reflection Model which characterizes a spectral data for highlight and body reflection. The proposed method involves different methodologies for performing – such as principal component analyzing and data classification techniques.

The attempt to invent the fast working algorithm performing well and having a small computational time was made successfully. The experiments on highlight removal were made according to the proposed method, and the results of the algorithm working demonstrated desirable performance. Also the suggestion for the future improvement of the algorithm is presented.

TIIVISTELMÄ

Lappeenrannan Teknillinen Yliopisto
Tietotekniikan osasto
Tekija: Ivan Shishkarev
Tutkielman nimi**:** **Kirkkaiden heijastusten poistaminen spektrikuvista**
Diplomityö
2006
55 sivua, 27 kuvaa ja 1 taulukkoa
Tarkastaja: Diana Kalenova
Ohjaaja: TkT Vladimir Bochko
Hakusanat: värin muuttumattomuus, spektrikuvat, kirkkaan heijastuksen poistaminen

Monissa sovelluksissa on hyvin tärkeää vähentää valolähteen vaikutusta kohteen oikean värin havainnoimiseksi. Tämä on tarpeen mm. virtuaalisissa museoissa, telelääketieteessä, verkkokaupassa ja verkkorahassa. Tässä tutkielmassa on kehitetty tekniikkaa kirkkaiden heijastusten poistoon spektrikuvista.

Työ sisältää katsauksen yleisen värillisen kuvan ymmärtämiseen, mihin perustuen analysoitiin erilaisia kirkkaiden heijastusten poisto–tekniikoita. Työssä kehitettiin uusi kirkkaiden heijastusten poisto–menetelmä, joka perustuu dikromaattiseen heijastus–malliin, joka kuvaa spektrisen datan objektin omaan väriin ja valaisevan valon väriin perustuen. Ehdotettu kirkkaiden heijastusten poisto–menetelmä hyödyntää erilaisia olemassaolevia menetelmiä, kuten pääkomponenttimenetelmää ja tiedon luokittelu–menetelmää.

Yritys kehittää nopeasti toimiva algoritmi, joka myös suoriutuu tehtävästä hyvin, on onnistunut. Kokeet toteutettiin ehdotetun menetelmän mukaisesti ja toimivalla algoritmilla saatiin halutut lopputulokset. Edelleen työ sisältää ehdotuksia esitetyn algoritmin parantamiseksi.

Acknowledgments

I would like to thank my supervisor Vladimir Botchko for providing with a supervision of my thesis, for involving me in image processing world.

I want to thank Lappeenranta University of Technology for giving me a possibility to get education in Finland, for funding me here that has been a matter thing to be involved in educational process entirely without digressing for finance question.

There were people around me who have helped me with my troubles arising during being and studying here. Thank you everybody. Especially great thank for our elder comrades Albert, Alexander and Diana providing with consultations for different questions.

I wish to thank Arseni for starting the translation of my abstract into Finnish, and Arto Kaarna for completely finishing it and finding exactly correspondences for used terms.

Table of Contents

1. INTRODUCTION…… 4

2. PHYSICS OF COLOR…… 6

2.1 Light and color …….. 6

2.2 Color perceptive system……. 7

2.3 Color representation…… 9

2.3.1 Tristimulus color space……... 9

2.3.2 RGB model…… 10

2.3.3 XYZ model…… 12

2.3.4 CIELAB color model…… 14

2.3.5 Multi-spectral images…… 15

3. COLOR IMAGE UNDERSTANDING…… 16

3.1 Dichromatic reflection model……... 16

3.2 Object shape and spectral variation…….. 17

3.3 Illumination geometry and spectral histogram shape…… 19

3.4 Highlight analysis and removal techniques …….. 22

3.4.1 Using the dichromatic reflection model for color image understanding…… 22

3.4.2 Highlight analysis according to Tong and Funt…… 23

3.4.3 Highlight analysis using three color images……. 24

4. LATENT VARIABLE MODELS……. 26

4.1 Principal component analysis…… 26

4.2 Probabilistic principal component analysis……... 29

4.3 Mixture of PPCA…… 31

5. DATA CLASSIFICATION TECHNIQUES……. 34

5.1 Varieties of classification…….. 34

5.2 K-mean clustering…… 36

6. EXPERIMENT…….. 39

6.1 Algorithm……. 39

6.2 Results of the experiment……. 44

7. **CONCLUSION**…….. 53

8. **REFERENCES**…….. 54

Abbreviations

DRM	Dichromatic Reflection Model
KNN-algorithm	k-nearest neighbor algorithm
PCA	Principal Component Analysis
PPCA	Probabilistic Principal Component Analysis
RGB	Red, Green, Blue

1. Introduction

In many applications it is very important to reduce influence of a light source to see a real color of objects. Especially this technique is very useful for E-museums, telemedicine, E-shops and E-money. In this study the highlight removal technique is proposed and developed.

The question of highlight removal has been in touching from different points of view. In general it is a part of color constancy problem which is leading to skip the illumination influence on the surface of objects in the scene. It can also be understood as computation of perceived surface color. In this thesis highlight removal will be resolved by applying to spectral images which are presented as a high dimensional data.

Highlights put to inconvenience not only for humans observing images, but more matter – for computer vision algorithms. They are looking as a part of surface of objects on the image scenes but actually they are caused by lighting that change in position and appearance under different viewing conditions. Some problems could be appeared due to it – result of a correspondence analysis in stereo images may be falsified significantly, recognition and image segmentation errors. Because of these undesirable effects of highlights on image analysis, there have been several previous works on its removal.

The one of the works on color images by Klinker, Kanade et all [9] provided with a theory of the Dichromatic Reflection Model. This physical aspect of color images was used by applying to highlight removal in different approach – Schluns, Koshan [14], Tong and Funt [17] and others.

There are some methods which are just able to detect specular reflection but the generation of a matte image is not possible, and as it should be highlight removal is not possible too.

Ideally highlight removal should be performed globally on whole image – for every object on the scene, and this task is separated into local highlight removal tasks by clustering whole color image into certain regions and performing highlight removal applying to every cluster. In the thesis we will work with physically segmented color regions to demonstrate usability of proposed method.

The most of the previous works were performed in a 3-D RGB-space. In the thesis the spectral images will be objects for analysis, so the data dimension will be higher. Bochko and Parkkinen worked with spectral images in their work on highlight removal [4] and my thesis is logical continuation of their approach. The method proposed by them is well productive and executable, but time demanding due to the time consuming of the one inner part of the approach – the needed for time resources KNN algorithm for mapping highlight pixels onto the body cluster was user there. In the thesis performance of the method is going to be improved. In my study the proposed highlight removal technique improves the results of the method [4] by accelerating the algorithm work and using better mapping algorithm.

And so the task of the research is removal a highlight area from an image by producing a simulated color of an object instead of a highlight. For this task the algorithm based on machine learning is proposed.

Several techniques will be used to rich result we intend. Therefore some background will be given at the beginning of the work.

2. Physics of Color

2.1 Light and Color

We live in the world of electromagnetic waves and every time we are in touch with them. The one of the characteristics of the wave is the wavelength measured by units of the length. The spectrum of waves is quite wide. It is presented with waves from the shortest (minor quotes of meter) to the longest (entire meters) wavelengths as shown in Fig. 1. Human senses can perceive a part of them, and more narrow band can be seen and sensed by the human eye – it is so-called visible spectrum. The human eye can detect waves with wavelengths from 360 nm (blue color) to 760 nm (red color).

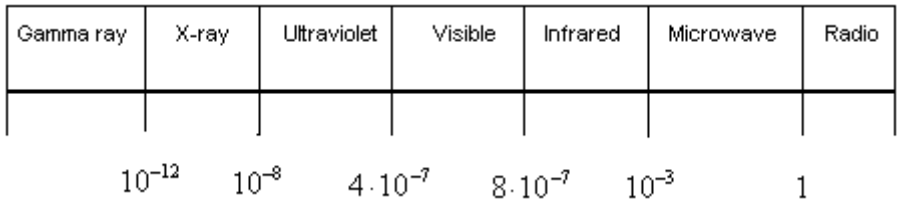

Figure 1. Spectrum of wavelengths

The definition of the word "color" itself provides some interesting challenges and difficulties. There are several terms in the nature and science which are understandable, but anyway they have to be described in strong definitions in order to provide possibilities to work with them correctly. While most of us know what color is, it is an interesting challenge to try to write a definition.

Let to take the next – color is the attribute of visual perception of any combination of chromatic and achromatic content. This attribute can be described by chromatic color name such as yellow, orange, brown, red, pink, green, blue etc. or by achromatic color names such as white, gray, black etc., and qualified by bright, dim, light, dark or by combination of such names. Terms mentioned before can be described in the next way. Achromatic color is perceived color devoid of hue, and chromatic color is a perceived color which possesses a hue [7].

Color is the main characteristic by which colored rays of light are distinguished. Every color can be varied depending on saturation or purity. Saturation indicates visible brightness or intensity of a color. Brightness is a quality of a color. Generally it depends on the quantity of light rays reflected from the surface of current color, that

is equal to brightness of a color with respect to others colors with current illumination.

In physical context brightness is amount of a light energy outgoing from light per second. Absolute brightness depends on its capability for reflecting, brightness and strength of illumination. But in case of human visual system we should refer that human eyes have capability of adaptation, and usual we do not grasp absolute brightness of a surface. And as example here – we can not imagine that some black surface in daylight can be brighter than some light surface in a moon light, but actually it is in physical context.

Albert Munsell created a presentation of 3-D body demonstrating well-ordered arrangement of colors in three-dimensional space by color, brightness and saturation – Fig. 2. Colors are arranged around a core which consists from achromatic colors: black, gradations of grey and white. Colors are clearing from bottom to top according to extent of their light reflection. Brightness of colors is rising in proportion to moving of the core and less mixing with a grey color.

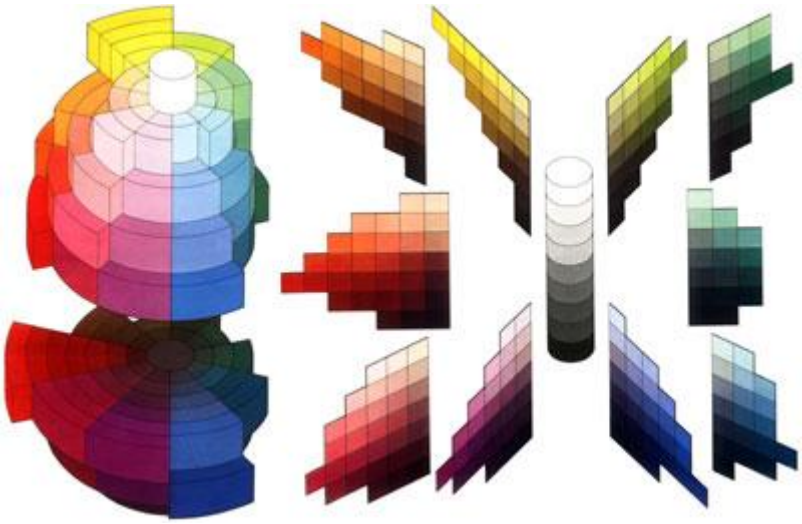

Figure 2. Color Diagram by Albert Munsell

2.2 Color perceptive system

Light falls on the object surface and most parts of a light are absorbed by a surface, but some its part of definite spectrum is reflected and it is perceived as a color by visual system. When we are talking about color we have in mind color perception by some sensors – either by human eyes or by robot visual sensors. Let us take to

consideration the construction of human eye to have presentation on the way of color perception.

The light goes through a pupil and reaches a retina which consists of two types of photoreceptors: rods and cones. The important distinction between them is in visual function. Rods serve for vision at low luminance level (less than 1 cd/m^2), while cones serve for vision at higher luminance level. Thus the transition from rod to cone vision is mechanism that allows our visual system to function over a wide range of luminance levels.

Rods and cones also differ substantially in their spectral sensitivities. There are three types of cone receptors with peak spectral responsiveness spaced throughout the visual spectrum as shown on Fig. 3, and only one type of rod receptor with a peak spectral responsiveness at approximately 510 nm. Since there is only one type of rod, the rod system is incapable of color vision. This can be observed by viewing a normally colorful scene at very low luminance. The three types of cones clearly serve for color vision and they are named as *L*, *M*, and *S* cones - long-wavelength, middle-wavelength, short-wavelength, or other designation *RGB* (red, green and blue sensitivities). These photoreceptors absorb light in their respective regions and send a signal to the rest of the visual system.

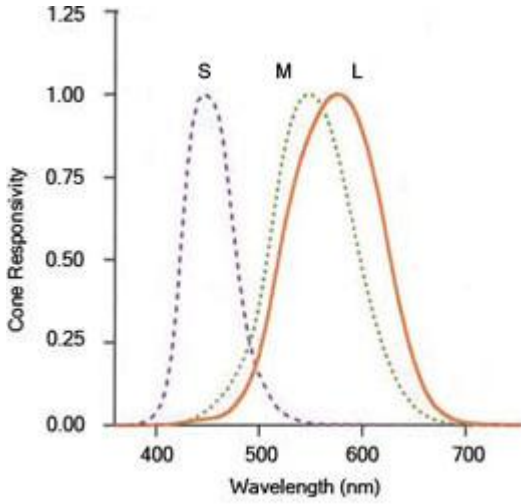

Figure 3. Cones responsiveness [7]

And as it could be seen spectral responsiveness of the three cone types broadly overlap. It is one of the differences from color separation responsiveness that are usually built into physical imaging systems. Such sensitivities, typically incorporated

in imaging systems for practical reasons, are the fundamental reason that accurate color reproduction is often difficult, if not possible, to achieve [7].

2.3 Color representation

A color model is an abstract mathematical model describing the way colors can be represented as tuples of numbers, typically as three or four values or color components. When this model is associated with a precise description of how these components are interpreted (viewing conditions, etc.), the resulting set of colors is called a color space.

There are several models of color representation. Based on the number of primaries used for color description, two groups of them can be considered – tristimulus models and model describing color through its spectral distribution.

2.3.1 Tristimulus color space

This space can be pictured as a region in a 3-D Euclidian space if one identifies the *x*, *y* and *z* axes with the stimuli for the long-wavelength (*L*), medium-wavelength (*M*) and short-wavelength (*S*) receptors – Fig.4.

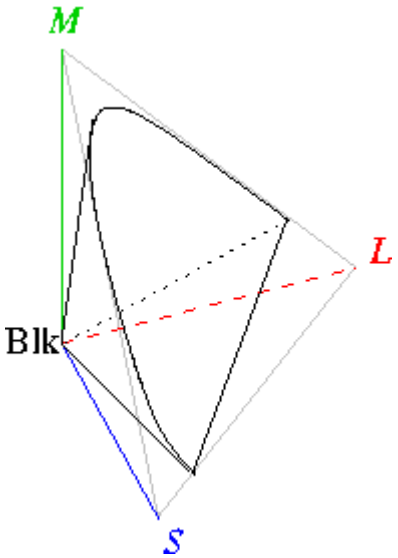

Figure 4. Tristimulus color space [18]

The origin $(S, M, L) = (0, 0, 0)$ corresponds to black. White has no definite position in this diagram, rather it is defined accordingly to the color temperature – it is determined by comparing its hue with a theoretical, heated black-body-radiator. The human color-space is a horseshoe-shaped cone as it is seen on the picture. The most saturated colors are located at the outer rim of the region, with brighter colors farther

removed from the origin. As far as the responses of the receptors in the eye are concerned, there is no such thing as “brown” or “gray” light – these names refer to orange and white with an intensity that is lower than the light from surrounding areas.

The human tristimulus space has the property that additive mixing of colors corresponds to the adding of vectors in this space. That makes it easy to describe possible colors that can be constructed from the red, green, and blue primaries in a computer display, for example.

2.3.2 RGB model

There is a type of color models named additive model which involves light emitted directly from a source or illuminant of some sort. By leading to the tristimulus model described before, the concept of this system is that color matches can be specified in terms of the amounts of three additive primaries required to visually matching a stimulus. This is illustrated by the equation (1):

$$C \equiv R(R) + G(G) + B(B). \quad (1)$$

The way of reading this equation is that a color C is matched by R units of the R primary, G units of the G primary, and B units of the B primary.

The RGB model is this type of additive models in which three primaries “red”, “green” and “blue” are combined in various ways to reproduce other colors. The RGB model itself does not define what is meant by “red”, “green” and “blue”, and results of mixing them are not exact unless the exact spectral make-up of the primaries is defined. And so the spectral values for tristimulus values have to be defined, and the Grassman’s law of additivity and proportionality to sum tristimulus values for each spectral component has to be taken. The spectral tristimulus values are obtained by matching a unit amount of power at each wavelength with an additive mixture of three primaries. There is demonstration of the set of spectral tristimulus values for monochromatic primaries at 435.6 nm (B), 546.1 nm (G) and 700.0 nm (R) in Fig.5

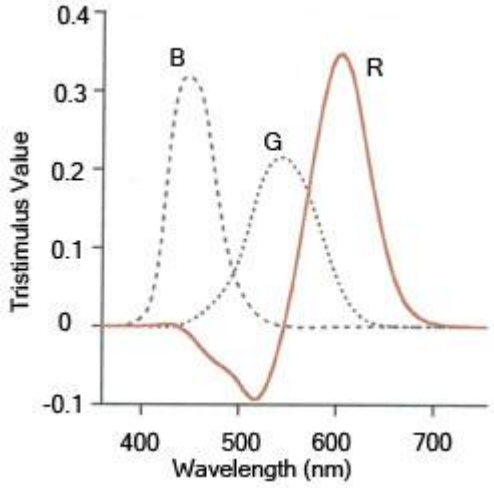

Figure 5. Color-matching function [7]

It is noticeable that some of the spectral tristimulus values are negative. This implies the addition of a negative amount of power into the match. This is because some wavelengths are too saturated to be matched by the particular primaries – they are out of gamut. Negative tristimulus values are obtained by adding the primary to the monochromatic light to desaturate the light and bring it within the gamut of the primaries.

By considering any given stimulus spectral power as an additive mixture of various amounts of monochromatic stimuli, one can obtain the tristimulus values for a stimulus by multiplying the color-matching functions by the amount of energy in the stimulus at each wavelength (Grassman's proportionality) and then integrating across the spectrum (Grassman's additivity). Thus Eq. 2 demonstrates generalization for calculating the tristimulus values of a stimulus with spectral power distribution, $\Phi(\lambda)$, and $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ are the color-matching functions:

$$\begin{aligned} \mathrm{R} &= \int_{\lambda} \phi(\lambda)\bar{r}(\lambda)d\lambda \\ \mathrm{G} &= \int_{\lambda} \phi(\lambda)\bar{g}(\lambda)d\lambda \\ \mathrm{B} &= \int_{\lambda} \phi(\lambda)\bar{b}(\lambda)d\lambda \end{aligned} \tag{2}$$

The RGB model has reputation due to its simplicity, but there are and some disadvantages due to negative part for some spectrum for R primary. Therefore the next XYZ model was introduced to overcome existing disadvantages [7].

2.3.3 XYZ model

It is possible to derive a linear transform (3×3 matrix transformation) to convert tristimulus values from one set of primaries to another. This transformation also applies to the color-matching function. The result of transformation for color-matching function is shown on Fig.6:

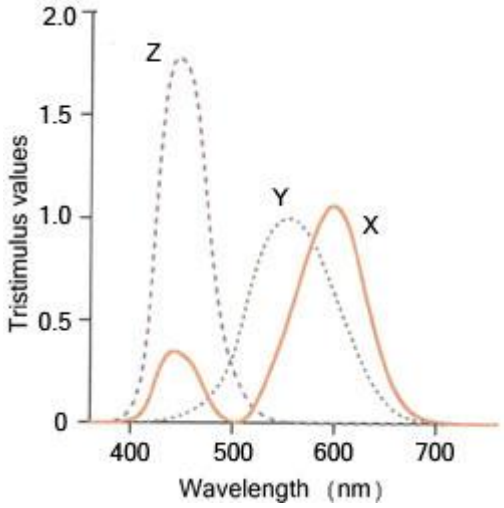

Figure 6. Color matching function for XYZ model [7]

The transformation was intended to eliminate the negative values in color-matching function. The negative values were removed by selecting primaries that could be used to match all physically realizable color stimuli.

XYZ tristimulus values for colored stimuli are calculated in the sane fashion as the RGB tristimulus values described previously. The general equations are given in Eq. 2, where $\Phi(\lambda)$ is the spectral power distribution of the stimulus, $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the color-matching functions, k is a normalizing constant.

$$X = k\int_{\lambda} \phi(\lambda)\bar{x}(\lambda)d\lambda$$

$$Y = k\int_{\lambda} \phi(\lambda)\bar{y}(\lambda)d\lambda \qquad (2)$$

$$Z = k\int_{\lambda} \phi(\lambda)\bar{z}(\lambda)d\lambda$$

The spectral power distribution of the stimulus is defined in different ways for various types of stimuli. For self-luminous stimuli (e.g. light sources) $\Phi(\lambda)$ is typically spectral radiance or a relative spectral power distribution. For reflective materials $\Phi(\lambda)$ is defined as the product of the spectral reflectance factor of the material, $R(\lambda)$, and the relative spectral power distribution of the light source, $S(\lambda)$ - that is $R(\lambda)S(\lambda)$. For transmitting materials $\Phi(\lambda)$ is defined as the product of the spectral transmittance of the material, $T(\lambda)$, and the relative spectral power distribution of the light source, $S(\lambda)$ - that is $T(\lambda)S(\lambda)$.

The normalization constant, k, is defined differently for relative and absolute colorimetry. In absolute colorimetry k is 683 lumens/watt. For relative colorimetry k is defined by Eq. 3:

$$k = \frac{100}{\int_{\lambda} S(\lambda)\bar{y}(\lambda)d\lambda} \qquad (3)$$

The directly linear transformation from RGB model to XYZ is presented by the next Eq.4:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.18 & 0.81 & 0.01 \\ 0 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad (4)$$

Because of linear dependency of models, the XYZ color system has the same disadvantages as RGB model has, for example device dependency – when colors having the same amounts of primaries appear differently on different devices. The problems of uniformity and device dependency were solved in CIELAB model [7].

2.3.4 CIELAB color model

CIE L*a*b* (CIELAB) model is the most complete color model used conventionally to describe all the colors visible to the human eye. It was developed for this specific purpose. This space extends tristimulus colorimetry to 3-D space with dimensions that approximately correlate with the perceived lightness, chroma, and hue of a stimulus. The main its aim is to provide uniform practices for the measurement of color differences, something that can not be done reliably in tristimulus or chromaticity spaces.

The three parameters in the model represent a lightness of the color (**L***, L* = 0 yields black and L* = 100 indicates white), its position between magenta and green (**a***, negative values indicate green while positive values indicates magenta) and its position between yellow and blue (**b***, negative values indicate blue and positive values indicates yellow). Color space described by this model is shown in Fig. 7:

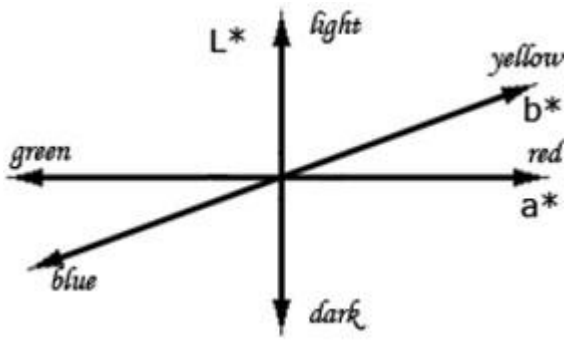

Figure 7. Three-dimensional representation of the CIELAB model

The LAB color model was created to serve as device independent, absolute model to be used as a reference. The components of this color space are defined by Eq. 5:

$$L^* = 116 \cdot (Y / Y_n)^{1/3} - 16$$

$$a^* = 500 \cdot \left[(X / X_n)^{1/3} - (Y / Y_n)^{1/3} \right]$$

$$b^* = 200 \cdot \left[(Y / Y_n)^{1/3} - (Z / Z_n)^{1/3} \right] \quad (5)$$

$$C^*_{ab} = \sqrt{(a^{*2} + b^{*2})}$$

$$h_{ab} = \tan^{-1}(b^* / a^*)$$

Here X, Y and Z are tristimulus values of the stimulus and X_n, Y_n and Z_n are the tristimulus values of the reference white. As it is noticeable tristimulus values are normalized to the white.

CIELAB has become almost universally used for color specifications, but it is still has some disadvantages [7].

2.3.5 Multi-Spectral Images

We will work with multi-spectral images which are presented by spectral color model. A multi-spectral image is a collection of several monochrome images of the same scene, each of them is taken with a different sensor which is set to definite frequency. The idea of spectral color model is to approximate color distribution. The accuracy of color description depends on the amount of primitives – with increasing the number of components the accuracy is raising but not in the linear way as shown in Fig.8. Until primitives are non-correlated the function of dependency accuracy on dimensionality is linear, otherwise non-linearity is proportional to correlation.

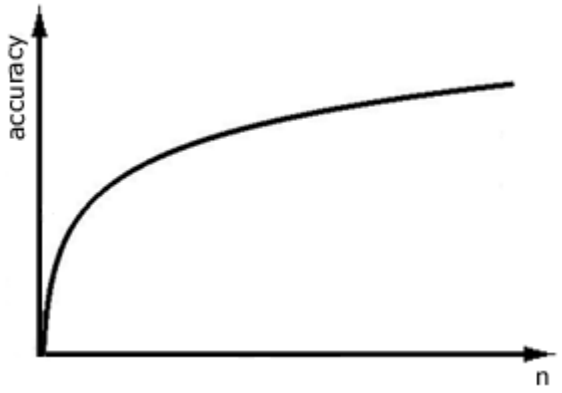

Figure 8. Graph dependency accuracy on dimensionality

For forming multi-spectral image sensors usually take several images from spectral band in the visual and non-visual range. The mentioned above RGB color model is a particular case of spectral image with three components.

The disadvantage of multi-spectral images is in high requirements to computational time and significantly increasing of memory using.

3. Color Image Understanding

3.1 Dichromatic Reflection Model

There were some works on image understanding which used intrinsic models of physical processes in the scene to analyze intensity or color variations in the image [1], [2], [4], [14], [17]. The Dichromatic Reflection Model is one of these models which were used for color image processing. It was used to generate image segmentation, along with description of object and highlight colors and intrinsic reflection images, showing how shading and highlight reflection vary in the image.

The color image analysis system alternates between generating hypothesis about the scene from the image data and verifying whether the hypothesis fit the image. The hypotheses relate object color, shading, highlights and camera limitations to the color variations in local image areas. They are used to segment images and to separate them into intrinsic images – the one of results of it is showing the scene without highlight. In this way, the algorithm is driven by a physical model of light reflection to incrementally identify local and global properties of the scene and to use them in interpreting and segmenting pixels in the images.

The Dichromatic Reflection Model describes the light, $L(\lambda, i, e, g)$, which is reflected from a point on a dielectric, nonuniform material as a mixture of the light $L_s(\lambda, i, e, g)$ reflected at the material surface and the light $L_b(\lambda, i, e, g)$ reflected from the material body as shown on Fig. 9.

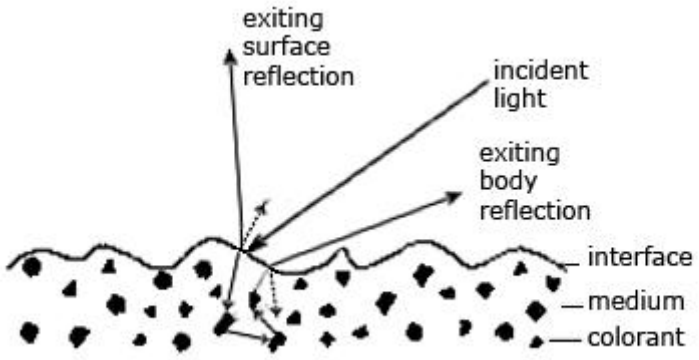

Figure 9. Light reflection from dielectric materials [9]

The parameters i, e and g describe the angles of the incident and emitted light and the phase angle, λ is the wavelength parameter. L_s is called the surface reflection

component – it generally has approximately the same spectral power distribution as the illumination and appears as a highlight or as gloss on the object. L_b is called the body reflection component. It provides the characteristic object color and exhibits the properties of object shading. Eq. 6 demonstrates the light describing:

$$L(\lambda, i, e, g) = L_s(\lambda, i, e, g) + L_b(\lambda, i, e, g) \tag{6}$$

The model separates the spectral reflection properties of L_s and L_b from their geometric reflection properties, modeling them as product of spectral power distribution, $c_s(\lambda)$ and $c_b(\lambda)$, and geometric scale factors, $m_s(i, e, g)$ and $m_b(i, e, g)$, which describe the intensity of the reflected light. And so the Dichromatic Reflection Model equation is Eq.7:

$$L(\lambda, i, e, g) = m_s(i, e, g) \cdot c_s(\lambda) + m_b(i, e, g) \cdot c_b(\lambda) \tag{7}$$

The model describes the light that is reflected from an object point as a mixture of two distinct spectral power distributions, $c_s(\lambda)$ and $c_b(\lambda)$, each of which is scaled according to the geometric reflection properties of surface and body reflection.

3.2 Object Shape and Spectral Variation

The Dichromatic Reflection Model itself describes the spectral properties separately for every single pixel on an object. Klinker, Shafer and Kanade presented a new upgraded mode of DRM [9].

It was observed that the light reflected from all points on an object uses the same two spectral vectors and these two factors are thus constant over an object. And that a dense spectral cluster in the dichromatic plane is formed by the light mixtures – the shape of this cluster is closely related to the shape of an object. Fig.10 illustrates a cylinder under light:

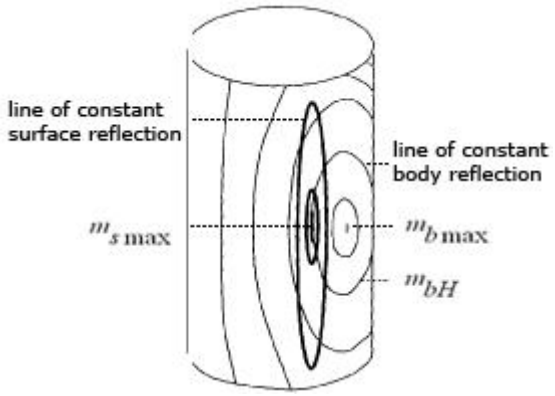

Figure 10. Surface and body reflection from a cylindrical object [9]

The curves show the loci of constant body reflection (perspective viewing and illumination geometry and Lambertian body reflection were assumed). And also the loci of constant surface reflection are shown. Parameter $m_s(i,e,g)$ decreases sharply around the object point with maximal surface reflection $m_{s\max}$. In fact this area is called highlight points. The remaining object points are matte points. Fig. 11 shows corresponding spectral histogram in the dichromatic plane and demonstrates the forming two linear clusters in the histogram.

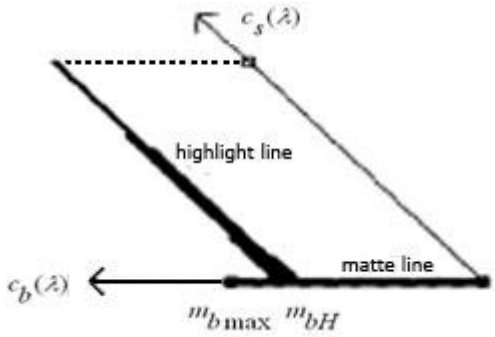

Figure 11. Spectral cluster on the dichromatic plane

Light reflection at matte points is primarily determined by the body reflection process. Additionally matte points may contain some surface reflection as the result of light diffusion at a rough material surface, but this can be neglected.

The observed light at matte points depends mainly on $c_b(\lambda)$, scaled by $m_b(i,e,g)$ according to the geometrical relationship between the local surface normal of the object and the viewing and illumination directions. The matte points form a matte line in the dichromatic plane in the direction of the body reflection vector $c_b(\lambda)$.

Highlight points exhibit both body reflection and surface reflection. Since $m_s(i,e,g)$ is much more sensitive to a small change in the photometric angles than $m_b(i,e,g)$, the body reflection component is generally approximated constant in a highlight area. And so according to the Dichromatic Reflection Model Eq. 7 obtains a constant value $m_{bH} \cdot c_b(\lambda)$ and all spectral variation within highlights comes from varying amounts of $m_s(i,e,g)$. So the new view of dependency of a light inside of highlight area is:

$$L \sim m_s(i,e,g) \tag{8}$$

The highlight points form a highlight line in the dichromatic plane in the direction of the surface reflection vector $c_s(\lambda)$. The line departs from the matte line at position $m_{bH} c_b(\lambda)$. Because of the small variation of the body reflection component over the highlight, the highlight cluster looks like slim, skewed wedge.

The combined spectral cluster of matte and highlight points looks like a skewed T. The skewing angle of the T depends on the spectral difference between the body and surface reflection vectors while the position of the highlight line depends on the illumination geometry.

3.3 Illumination geometry and Spectral Histogram Shape

There is a close relationship between the illumination geometry and the amounts of body and surface reflection. It has an influence on the shape of the clusters in the spectral histogram (in Fig.11). Fig.12 demonstrates dispositions of object, light source and observer:

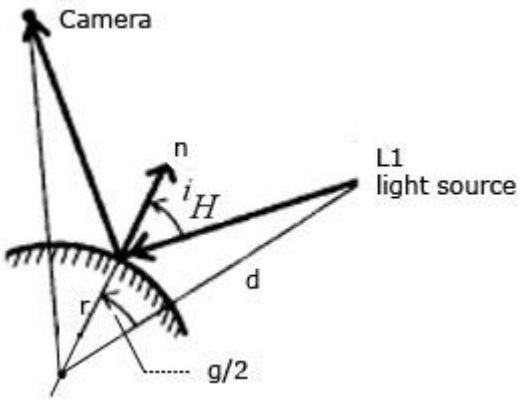

Figure 12. Dispositions of object, light source and observer [9]

The amount of body reflection under highlight depends on the phase angle g between illumination and viewing direction - if g is small, that means close disposition of a camera to the light source, the incidence direction of the light at the highlight is close to the surface normal and amount of body reflection is very high. The highlight line in the dichromatic plane then starts near the tip of the matte line, and skewed T becomes a skewed L. With the increasing the phase angle g the highlight moves away from the area with maximal body reflection and the amount of body reflection under the highlight decreases. The highlight line moves accordingly away from the tip of the matte line. The next formulas show these assumptions in a strong way.

Using the law of sines and the law of cosines, i_H can be described as:

$$i_H = \arcsin \frac{d \cdot \sin(g/2)}{\sqrt{d^2 + r^2 - 2 \cdot d \cdot r \cdot \cos(g/2)}}, \tag{9}$$

where i_H itself is the illumination angle at the highlight, g is a phase angle between illumination and viewing direction, r - the radius of our assumpted to be spherical object, d - the distance from the object center to the light source.

Assuming Lambertian body reflection, $m_{bH} = \cos(i_H)$, the underlying body reflection component, m_{bH}, at the highlight is given by:

$$m_{bH} = \cos(i_H) = \sqrt{1 - \frac{d^2 \sin^2(g/2)}{d^2 + r^2 - 2 \cdot d \cdot r \cos(g/2)}}. \tag{10}$$

According to these equations, m_{bH} approaches 1 when g goes to 0^o, confirming that the spectral cluster looks like a skewed L when camera and source are close.

The next thing has to be taken into attention – globally maximal body reflection $m_{b\max}$ is not always visible from the camera. When it is occluded, the matte line in the spectral histogram does not extend entirely to $m_{b\max} c_b(\lambda)$ but only to the point representing the brightness visible point - $m_{bMaxVisible} c_b(\lambda)$. The next Fig.12 shows positions for the camera and the light source with which the maximum body reflection is still visible:

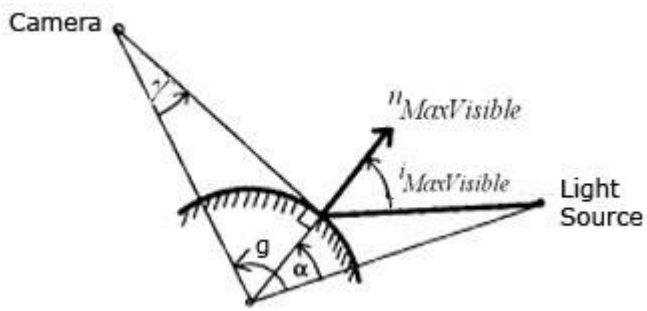

Figure 12. Visible points from a camera [9]

Then the angle between the illumination vector and the surface normal at the local maximum $i_{MaxVisible}$ is described analogously:

$$i_{MaxVisible} = \arcsin \frac{d \sin \alpha}{\sqrt{d^2 + r^2 - 2dr \cos \alpha}} \tag{11}$$

The amount of body reflection at that point $m_{bMaxVisible}$ is given similarly:

$$m_{bMaxVisible} = \cos(i_{MaxVisible}) = \sqrt{1 - \frac{d^2 \sin^2 \alpha}{d^2 + r^2 - 2dr \cos \alpha}} \tag{12}$$

And then the starting point of the highlight line on the matte line can now be described relative to the length of the matte line. The ratio q of body reflection at the highlight in relation to the maximally visible amount of body reflection is:

$$q = \frac{m_{bH}}{m_{bMaxVisible}} \tag{13}$$

Fig. 12 shows variations of q from g and d:

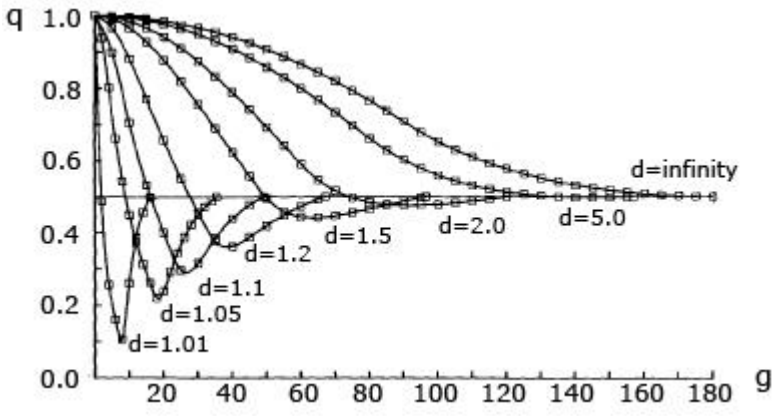

Figure 12 [9].

Each curve in the figure shows for a different fixed d how q depends on g. And in all curves q approaches 0.5 as g approaches $g_{\max}$. With increasing d minimum of q function becomes less and q function becomes a monotonically decreasing function with the point of inflexion [9].

Later we will closely work with matte cluster and highlight cluster to reach the goal we are intending, but next subsection demonstrates directly application of the Dichromatic Reflection Model for highlight removal.

3.4 Highlight analysis and removal techniques

3.4.1 Using the Dichromatic Reflection Model for Color Image Understanding

Klinker, Shafer and Kanade used Dichromatic Reflection Model on relative work for highlight removal [9]. Their algorithm involved segmentation of whole image for small "windows" – small local areas of the image – with which the procedure of merging or growing was performed. This combining was made based on belonging these single "windows" to concrete type of cluster on dichromatic plane – either to matte region or to highlight cluster. This pixels merging was made step by step – including the neighboring pixels or areas, sided with initial guessed regions, skewed Ts of which are the same as in starting area.

Thus the algorithm looks in a bottom-up process for color clusters from local image areas that exhibit the characteristic features of the body and surface reflection process. When it finds a "promising" clusters in an image area, the algorithm generates a hypothesis that describes the object color or highlight color in the image area and determines the shading and highlight components of every pixel in the area. The algorithm then applies the new hypothesis to the image, using a region growing approach to determine the precise extent of the image area to which the hypothesis applies. The resulting intrinsic images and the hypotheses together instantiate the concepts of shading and highlights of the Dichromatic Reflection Model, describing the specific reflection process that occur in this part of the scene.

This physical knowledge from all hypotheses is used to incrementally adapt the steps of the image analysis. Their algorithm performs this generate-and-test analysis in several stages, each of which is related to a particular aspect of the Dichromatic Reflection Model. The steps of the algorithm are the next:

1. Computation an initial, rough description of color variation in local image areas.
2. Generation of the hypotheses on matte color clusters and exploiting them for image segmentation.
3. Extension of the matte hypotheses into hypotheses on skewed Ts in dichromatic planes and resegmentation of the image, exploiting these hypotheses.
4. Analyzing of the effects of blooming and color clipping.
5. Exploiting of the hypotheses to split the pixels into their reflection components.

This structure exploits a physical model of light reflection to incrementally identify local and global properties of the scene – such as object and illumination colors [9].

Suggested Dichromatic Reflection Model and approach based on it for highlight detecting showed quite applied results and became a basis for an order of other approaches – for example for highlight analysis by Tong and Funt.

3.4.2 Highlight analysis according to Tong and Funt

Tong and Funt suggested the next technique - instead of computing vector c_s and c_b and the scaling factors m_s and m_b for every segmented region, they combine the information about the regions and initially estimate the vector c_s, and due to the Dichromatic Reflection Model the color of the interface reflection c_s is identical to the color of illumination – thus identical for all segmented regions. After a coarse segmentation of the image the corresponding dichromatic planes are computed for all single pixels – and then by intersection of all dichromatic planes the vector c_s is obtained. It is a characteristic of a surface reflection – thus a characteristic of a highlight area.

As a finding the intersection of the planes, they find the line which is most parallel to all the planes by calculation the normal to each plane and fitting by a least-squares method for getting the line closest to the perpendicular to all the plane normals.

This idea of computing the vector c_s for the entire image only once is rather efficient, but the single dichromatic planes have to be separated and the corresponding normals have to be calculated when this technique is applied to a color image. And to remind the field of application of this approach it should be mentioned that all these exploits are applied to whole image – means to all objects on the scene at the time.

Also for time efficiency this costly analysis in the three-dimensional color space can be mapped onto two-dimensional case by going from RGB-space to for example YUV-space given by Eq. 14 and by working only with normalized u and v coordinates.

$$(Y,U,V) = (R,G,B)\begin{pmatrix} 1/3 & 1/2 & -1/2\sqrt{3} \\ 1/3 & 0 & 1/\sqrt{3} \\ 1/3 & -1/2 & -1/2\sqrt{3} \end{pmatrix} \tag{14}$$

And it is useful to take to attention that surfaces of ideal matte materials correspond to exact one point in the uv-space. Therefore the dichromatic mate cluster corresponds to one point too, but dichromatic highlight clusters form line segments in the uv-space [17].

3.4.3 Highlight analysis using three color images

There was used a technique named “spectral differencing” by Lee and Bajcsy [21] for highlight analysis. The idea here is to take a color image under different viewer conditions but with a constant illumination direction, and then by spectral differencing to define a group of highlight pixels. The objective of this technique is to detect highlight but not to remove it. And it is assumed that the illumination color and color of the objects are different.

The idea of the method is to detecting points for a viewer. For this so-called minimal spectral differences images are calculated (MSD-images). Let C_α and C_β be two color images of an object scene taken under different viewer directions. And $MSD(C_\alpha \longleftarrow\!\!\!\!- C_\beta)$ denotes the MSD-image calculated from C_β to C_α. The color value of a pixel in the MSD-image $MSD(C_\alpha \longleftarrow\!\!\!\!- C_\beta)$ is defined as the minimal value of all spectral differences between a pixel in image C_α and all pixels in image C_β. Every MSD-value which exceeds a predefined threshold belongs to a potential

specular reflection pixel. The threshold itself depends on the noise introduces by the imaging sensor. Actually pixels representing an object point which is invisible in some of the images are also understandable classified as highlights if the MSD-value is larger than the predefined threshold.

The similar technique is used by Schluns and Koschan in [14] – the three images are captured from a fixed camera position with three different illumination directions.

Thus there were mentioned approaches to highlight detecting from physical point of view – we will take it to attention during of working with our problem.

4. Latent Variable Models

We will work with techniques decreasing the dimensionality of the initial data. This is because originally we have a multi-component image, but the dichromatic plane is two-dimensional. This chapter provides with a survey on such methods and here we will choose the approach suitable for our case.

The models for analyzing principal components are usually used for visualization of high-dimensional data transformed to low-dimensional case with intend for the decreasing of the dimensionality and at the same time for minimization possible losses of the quality of the data. We are going to use these techniques later in our work for a little bit different goals but nevertheless saving the mean of these approaches.

There are the next varieties of the methods for dimensionality decreasing – linear and non-linear approaches [12]. The second are built on the standard PCA in applications where the data demonstrates the nonlinear behavior. Non-linear PCA methods may be categorized into the next groups:

- generative models
- Generative Topographic Mapping
- a mixture of probabilistic principal component analyzers
- Bayesian nonlinear factor analysis

The data we are going to be in touch with has the non-linear nature - therefore the non-linear latent variable model should be used for our intention. Mixture of probabilistic principal component analyzer was chosen as suitable for our goal.

4.1 Principal Component Analysis

Principal Component Analysis (PCA) is the most commonly used feature extraction and visualization technique in practice [12]. It is fast and easy to compute, and nevertheless it demonstrates quite usable results by retaining the most matter information.

Suppose that we are going to map a dataset of vectors x^n where $n = 1, ..., N$ in a space $V = R^d$ to vectors z^n in a space $U = R^M$, which presents a subspace of V. We can

choose an orthonormal basis $u_1,...,u_M$ for U and extend it to an orthonormal basis $u_1,...,u_d$ for V. And so by using the basis a vector x can be represented as:

$$x = \sum_{i=1}^{M} x_i u_i + \sum_{i=M+1}^{d} x_i u_i \tag{15}$$

Now suppose that we project to the M-dimensional space spanned by the first M vectors:

$$z = \sum_{i=1}^{M} x_i u_i + \sum_{i=M+1}^{d} b_i u_i \tag{16}$$

where the b_i are constants. The coefficients b_i and vector u_i are chosen so that the projected vectors z^n best approximate x^n. The quality of the approximation is measured by the reconstruction sum of squares error between the two vectors:

$$E = \frac{1}{2}\sum_{n=1}^{N} \left\| x^n - z^n \right\|^2 = \frac{1}{2}\sum_{n=1}^{N}\sum_{i=M+1}^{d}\sum_{j=M+1}^{d} (x_i^n - b_i)(x_j^n - b_j) u_i^T u_j =$$

since $u_i^T u_j = \delta_{ij}$, where δ_{ij} is the Kroneker delta

$$= \frac{1}{2}\sum_{n=1}^{N}\sum_{i=M+1}^{d} (x_i^n - b_i)^2 \tag{17}$$

Setting the derivative of E with respect to b_i to zero, the next is got:

$$b_i = \frac{1}{N}\sum_{n=1}^{N} x_i^n \tag{18}$$

which is the i-th coordinate of the mean vector $\bar{x}$ with respect to the coordinate system $u_1,...,u_d$. And so the error term can be written:

$$E = \frac{1}{2}\sum_{i=M+1}^{d}\sum_{n=1}^{N}\left\{u_i^T (x^n - \bar{x})^T\right\}\left\{(x^n - \bar{x}) u_i\right\} = \frac{1}{2}\sum_{i=M+1}^{d} u_i^T \Sigma u_i \tag{19}$$

where Σ is the covariance matrix of the data. And it can be shown that the stationary points of E with respect to the vectors u_i occur at the eigenvectors of Σ, so that $\Sigma u_i = \lambda_i u_i$. So the final view for residual error function is:

$$E = \frac{1}{2}\sum_{i=M+1}^{d} \lambda_i \tag{20}$$

Hence the minimal error E is achieved by choosing the $d-M$ smallest eigenvalues, and the initial data is projected onto the space spanned by the eigenvectors corresponding to the largest M eigenvalues. These eigenvectors are called the first M principal components.

There is no general techniques for deciding how many principal components should be used to present the data adequately with minimization of possible losses, although some estimation on it was proposed by Thomas P. Minka in [11] – this estimation guarantees picking the correct dimensionality but needs for complex computations. But still a useful heuristic to get number of principal components is to choose a fraction of the variance to be retained by computing:

$$\frac{\sum_{i=1}^{M}\lambda_i}{\sum_{i=1}^{d}\lambda_i} \tag{21}$$

And so for getting principal components from *n*-dimensional vector x the following should be performed. At first the data centralization has to be made:

$$x \longleftarrow x-\mu\,, \tag{22}$$

where μ is the mean of the dataset. It is illustrated in Fig.13:

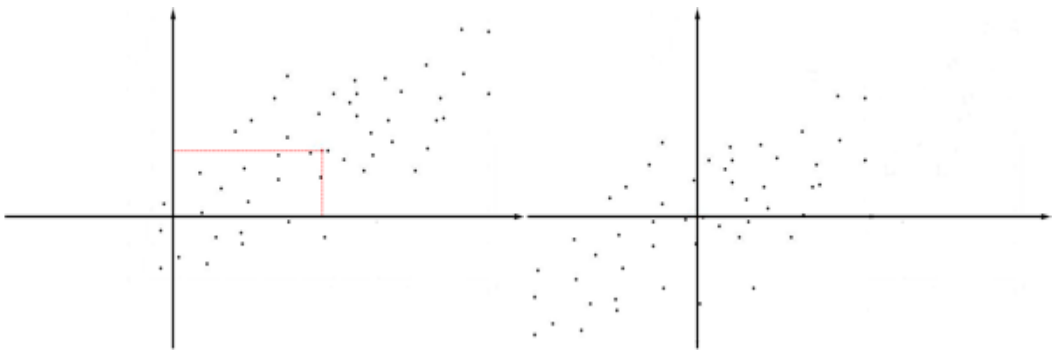

Figure 13. Data centralization

Then the calculation of the covariance matrix C of new shifted data:

$$C = E(x \cdot x^T)\,, \tag{23}$$

where E denotes a mathematical expectation operator. PCA involves a decomposition of a covariance matrix as follows:

$$C = U\Lambda U^T\,, \tag{24}$$

where $U = (e_1, ..., e_n)$, $e_i = (e_{i1}, ..., e_{in})^T$ are the eigenvectors and $\Lambda = diag(\lambda_1, ..., \lambda_n)$, where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ are the eigenvalues of C sorted in the decreasing order.

And so the first k principal components are:

$$y_k = U_k^T x, \tag{25}$$

where Eq. 20 defines the residual error [12].

4.2 Probabilistic Principal Component Analysis

The motivation for the definition of PCA given before is in terms of minimizing the squared reconstruction error defined in Eq.17 between the projected and original vectors [12]. The disadvantage of this approach is that it does not define a generative model: there is no density model $p(x|z)$ - means there is no principled interpretation of the error function *E*. A density model offers several advantages:

- The definition of a likelihood allows to compare this model with other density models in a quantitative way.
- If PPCA is used to model class-conditional densities, then posterior probabilities of class membership may be calculated.
- A single PPCA model may be extended to a mixture of PPCA models.
- Bayesian inference methods may be applied if a suitable prior is chosen.

Classical PCA is made into a density model by using a latent variable approach, derived from standard factor analysis, in which the data x is generated by a linear combination of a number of hidden variables z:

$$x = Wz + \mu + \varepsilon \tag{26}$$

where z has a zero mean, unit isotropic variance, Gaussian distribution $N(0, I)$, μ is a constant whose maximum likelihood estimator is the data mean, and ε is an x-independent noise process. The latent variable space has dimension q, which is usually chosen to be less than the data dimension d so that the model is a more economical description of the data. The noise model is $\varepsilon \sim N(0, \Psi)$, with Ψ diagonal. The mode for x is then also normal $N(\mu, L)$, with $L = WW^T + \Psi$. Since Ψ is diagonal, the observed variables x are conditionally independent given the

values of the latent variables z. So all conditional dependence is captured in the latent distribution over z, while ε represents the independent noise. Unlike PCA there is no analytic solution for W and Ψ - so an iterative algorithm must be used to compute them.

The change needed to make this into a probabilistic model for PCA is very small. In the PCA model there is a systematic component in the data plus an independent error term for each variable with common variance. This is captured by assuming a noise model with an isotropic variance $\Psi = \sigma^2 I$. Then the probability model for PPCA can be written as a combination of the conditional distribution:

$$p(x \mid z) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left\{-\frac{\|x - Wz - \mu\|^2}{2\sigma^2}\right\} \tag{27}$$

and the latent variable distribution:

$$p(z) = \frac{1}{(2\pi)^{q/2}} \exp\left\{-\frac{z^T z}{2}\right\} \tag{28}$$

Then the distribution of the observed data, which is also Gaussian, is:

$$x \sim N(\mu, C) \tag{29}$$

where $C = WW^T + \sigma^2 I$. Thus this model represents the data as a "pancake" consisting of a linear subspace surrounded by equal noise in all directions.

To fit this model to data, the log-likelihood as an error measure is used:

$$L = \sum_{n=1}^{N} \log p(x_n) = -\frac{N}{2}\left\{d \log(2\pi) + \log|C| + (C^{-1}S)^T\right\}, \tag{30}$$

where

$$S = \frac{1}{N}\sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T \tag{31}$$

is the sample covariance matrix of the observed data, provided that μ is set to its maximum likelihood estimate, which is the sample mean. Estimates of W and σ^2 can be obtained by an iterative maximization of L using an EM algorithm. However, it is also possible to find an analytic solution for the maximum likelihood estimate:

$$W_{ML} = U_q(\Lambda_q - \sigma^2 I)^{1/2} R, \tag{32}$$

where the q column vectors in the $d \times q$ matrix U_q are the principal q eigenvectors of S, with corresponding eigenvalues $\lambda_1, ..., \lambda_q$ in the $q \times q$ diagonal matrix Λ_q, and R is an arbitrary $q \times q$ orthogonal matrix (rotation matrix). The latent space is a projection of the original space onto the principal subspace of the data. The effect of the matrix R is simply to choose different orthogonal axes for the principal subspace.

If the maximum-likelihood solution for W is used, then the maximum likelihood estimator for σ^2 is given by:

$$\sigma^2_{ML} = \frac{1}{d-q} \sum_{j=q+1}^{d} \lambda_i, \tag{33}$$

which can be interpreted as the variance lost in the projection to the latent space, averaged over the dimensions left out.

We can also take a probabilistic view of projecting the observed data onto the latent space by using the posterior distribution of the latent variables z given the observed data x:

$$p(z \mid x) \sim N(M^{-1}W^T(x-\mu), \sigma^2 M^{-1}), \tag{34}$$

where the $q \times q$ matrix $M = W^T W + \sigma^2 I$. It is usually more convenient to map x to a single point in latent space rather than a complete distribution. If consider the mean of the posterior distribution $M^{-1}W^T(x-\mu)$ and when $\sigma^2 \longrightarrow 0$, $M^{-1} \longrightarrow (W^T W)^{-1}$ than Eq.(34) represents an orthogonal projection equivalent to standard PCA. In practice, with $\sigma^2 > 0$, the latent projection becomes skewed towards to origin. Because of this, the reconstruction $W_{ML}\langle z_n \mid x_n \rangle + \mu$ is not an orthogonal projection of x_n, and does not minimize the squared reconstruction error as in PCA Eq.17.

4.3 Mixture of PPCA

Because PCA defines only a linear projection of data, it is rather limited technique. One way around this is to use a global non-linear method [12]. It is possible to model a complex non-linear structure by a collection of local linear models. The attraction of

this is that each model is simpler to understand and usual easy to fit. The mixture of PPCA model is appropriate when the data is approximately piece-wise linear.

To fit a collection of linear models two steps are required: first, a partition of the data into regions; second, fitting models by estimating the principal components within each region. The question of how to combine these two steps is not trivial – because the partition depends on the data model and the principal components depend on the data partition.

A major advantage of developing a probabilistic model for PCA is that we can formalize the idea of a collection of models as a mixture of PPCA. The fact that the PPCA is a Gaussian model makes it clear that it is possible to train such a model in a maximum likelihood framework using an Expectation-Maximization algorithm (EM) – which is an iterative likelihood technique. This gives us the principled way of tracking the two steps, which correspond to E-step and the M-step respectively [12].

The log-likelihood of the dataset is given by:

$$L = \sum_{n=1}^{N} \ln \sum_{j=1}^{M} \pi_i \, p(x_n \mid j), \tag{35}$$

where each $p(x_n \mid j)$ is a PPCA model and π_j is the mixing coefficient. Each component is associated with a mean vector μ_j, projection matrix W_j and noise model σ_j^2.

The EM algorithm for this model has a very similar form to that for other Gaussian mixture models. In the E-step, we compute the responsibility R_{nj} of component j for generating data point x_n:

$$R_{nj} = P(j \mid x_n) = \frac{p(x_n \mid j)\pi_j}{p(x_n)}, \tag{36}$$

In the M-step we re-estimate the parameter of each component. The equation for the means and mixing coefficients are identical for other Gaussian mixture models:

$$\pi_j^{(m+1)} = \frac{1}{N} \sum_{n=1}^{N} P^{(m)}(j \mid x_n), \tag{37}$$

$$\mu_j^{(m+1)} = \frac{\sum_{n=1}^{N} P^{(m)}(j \mid s_n) x_n}{\sum_{n=1}^{N} P^{(m)}(j \mid x_n)}. \qquad (38)$$

To re-estimate the covariance structure, we apply PPCA to S_j, the covariance matrix computed for data weighted by the responsibility of the j-th components:

$$S_j = \frac{1}{\pi_j^{(m+1)} N} \sum_{n=1}^{N} P^{(m)}(j \mid x_n)(x_n - \mu_j^{(m+1)})(x_n - \mu_j^{(m+1)})^T . \qquad (39)$$

To perform clusterization for our nonlinear data into piece-wise linear groups in order to apply the PPCA separately for every cluster the next will be used. At first the K-mean method is applied to find means for our clusters, and then computation of likelihoods for samples is made based on results of K-mean clustering. And by taking the maximum likelihood the membership of each sample with respect to definite clusters is evaluated. The next chapter is presented in order to provide with a theory on classification techniques.

5. Data classification techniques

The one of the techniques for object classification will be utilized to solve our task. The tested image region has a nonlinear global data structure. Therefore we are interested in techniques which can incorporate the nonlinearity of the data. We will need it for using the mixture PPCA applying to our nonlinear data to form the set of linear clusters. Below there is a general description on pattern classification systems.

When we are working with some objects from some field, the convenient description of the patterns has to be used to provide a comfortable processing of them. The objects can be images or signals or any type of measurements and for every type there should be clear definite presentation of the object. The numeric presentation is generally used – when the individuals are described by specific features represented in numbers, and so the single patterns are presented as a vector of the required dimensionality. To provide with a processing of the patterns the set of operations including comparison of the objects is determined. Of course there are some specific features of the objects processing depending on the field they are, but in general it does not matter from what area our objects are – we need just their complete and exactly presentation describing the individuals..

The question of the completeness of the data is individual and separate for discussion – we do not need to consider it due to possession by the concrete data set, describing the objects we are working with – the multi-spectral images.

5.1 Varieties of classification

The two types of approaches are involved for pattern classification – supervised and unsupervised. In a case of supervised learning the training data is known and the goal of classification is to find a mapping from a feature space into a class space with a minimal cost. The training samples used here to design a classifier are labeled by their category membership, and procedures that used labeled samples are said to be supervised. In a case of unsupervised learning classes are not known for training data and the objective is to cluster patterns into groups with minimal within group differences and maximal between groups differences. Thus is initially there is a collection of samples without possession by information about their categories.

The need in a supervised learning is clear – the set of various objects is distributed between some set of specific classes (patterns), unsupervised learning serves for

finding groups (clusters) of similar patterns. Interest in unsupervised versus supervised learning for a classification goal is that collecting and labeling a large set of sample patterns can be much costly. Also their cooperation is often used – training with a large amount of unlabeled data, and after then use supervision to label the groupings found. This could be applied for large data mining application, where the contents of a large database are not known beforehand. And additionally unsupervised methods can be used to find features that will be used then for categorization.

The basic steps in building a system for clustering are:

- proximity measure selection
- clustering criteria selection
- clustering algorithm selection

And the result of clustering depends on all these parts. Sometimes it happens that there is no such thing as the correct result. The difficulty of clustering is shown by the next Eq.40 which expresses the Stirling number:

$$S(N,m)=\frac{1}{m!}\sum_{i=0}^{m}(-1)^{m-1}\binom{m}{i}i^{N}, \tag{40}$$

where N is whole number of vectors which are going to be classified into m groups, $\binom{n}{k}$ is a binomial coefficient which expresses the number of ways of picking k unordered outcomes from n possibilities: $\binom{n}{k}\equiv\frac{n!}{(n-k)!k!}$. And as it is noticeable $S(N,2)=2^{N-1}-1$ - it shows the number of possible clustering of N objects into two groups.

Categories of clustering algorithms are:

- Sequential – all feature vectors are presented once or a few times, the result here usually depends on the order of vectors.
- Hierarchical algorithm – agglomerative algorithms produce clusters by merging existing clusters, divisive algorithms split existing clusters into parts.

- Cost function optimization: typically iterative optimization schemes. Can be based on hard clustering, probability theory, fuzzy set theory.
- Some other approaches: branch and bound algorithms, genetic clustering algorithms, stochastic relaxation, competitive learning.

5.2 K-mean clustering

In mixture PPCA which was used in our work to take the principal components from the data we used, the k-mean algorithm to find clusters of the initial data was used.

The efficiency for clustering algorithm is measured by cost function - the best clustering approach minimizes it. Here the cost function is:

$$J(\theta,U)=\sum_{i=1}^{N}\sum_{j=1}^{m}u_{ij}\left\|x_i-\theta_j\right\|^2, \qquad (55)$$

where $\theta=(\theta_1...\theta_m)$ is a matrix containing cluster centers, U is a matrix of u_{ij} containing value 1 if sample x_i belongs to cluster j and 0 otherwise. And so the cost function measures the sum of squared Euclidian distance from each sample to the nearest cluster, and in the best case of clustering this sum should be minimized.

Simply speaking k-means clustering is an algorithm to classify or group objects into K number of groups. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid.

The algorithm of it is presented below:

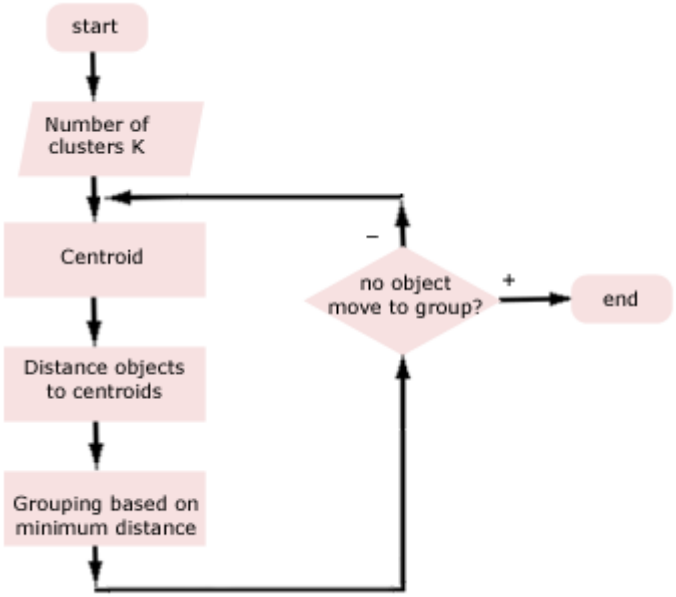

Figure 14. K-mean algorithm

The exactly its description is given below.

1. Initialize θ_j ($j = 1...m$) randomly ($m = K$ - number of clusters)
2. Repeat
3. For $i = 1$ To N
4. find nearest cluster j for sample i: $b(i) = \arg\min_j \left\| x_i - \theta_j \right\|$
5. End For
6. For $j = 1$ To m
7. Update θ_j as the mean of samples where $b(i) = j$
8. End For
9. Until no changes in any of θ_j between two successive iterations

As a result we have K centers (means) which define K clusters. Then likelihood for every sample is calculated, and by taking the maximum likelihood the membership of the sample is defined. And so every sample is marked as belonging to a definite cluster.

Like in other approaches there are some weaknesses of the K-mean algorithm:

- when the numbers of data are not so many, initial grouping will determine the clusters significantly
- we never know the real cluster, using the same data, if it is inputted in a different order may produce different cluster if the number of data is a few
- sensitive to initial conditions – different initial conditions may produce different results of cluster, the algorithm may be trapped in the local optimum
- weakness of arithmetic mean is not robust for outliers – very far data from the centroid may pull the centroid away from the real one
- the result is circular cluster shape because based on distance

By applying to our data with which we worked to cluster them for our needs K-mean algorithm is quite suitable due to well-recognized groups of samples and a lot amount of well-identified patterns.

All these mentioned before methodologies were used to solve the problem of highlight removal. The next section provides with a describing of the work.

6. Experiment

There was made similar work on highlight removal in spectral images by Bochko and Parkkinen – [4]. The techniques proposed by them demonstrated well results, but their algorithm needed for a lot of time for executing. The idea of their algorithm was to use the K-nearest-neighbor algorithm to find replacing for highlight pixels with body-reflection pixels and this procedure was so much time consuming. The goal of presented here algorithm is a highlight removal and additional demanding is to decrease the time needed for this work performing.

6.1 Algorithm

The objects we are working with are the high quality spectral images of 81-dimensionality. They are presented as 3-D matrix:

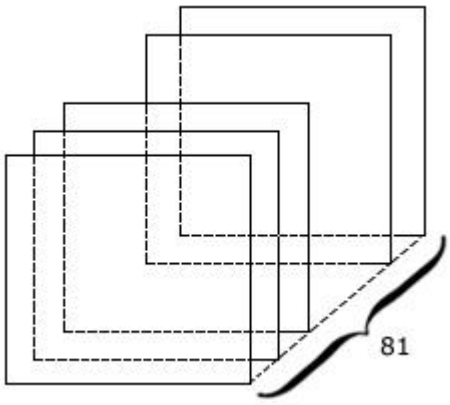

Figure 15. Representation of spectral image

The set of objects are presented on scenes of images with a natural highlight on their surfaces caused by the light illuminating them. The feature of the algorithm is that it works on the concrete manual marked area of the part of an object with a highlight, and the surface of the object should possesses by the monotonous color for the chosen area. It could be said that it is a weakness of it but nevertheless it works properly in this case and in a future it could be demand on the general case for highlight removal processing for the all objects on the scene at once.

According to the Dichromatic Reflection Model from Chapter 3 the data we are working consists from the body reflection and the surface reflection regions. This surface reflection field appears as highlight area which we are going to remove. And

so there are two groups of pixels should be taken in attention to be processed. It is clear seen on the Fig.16:

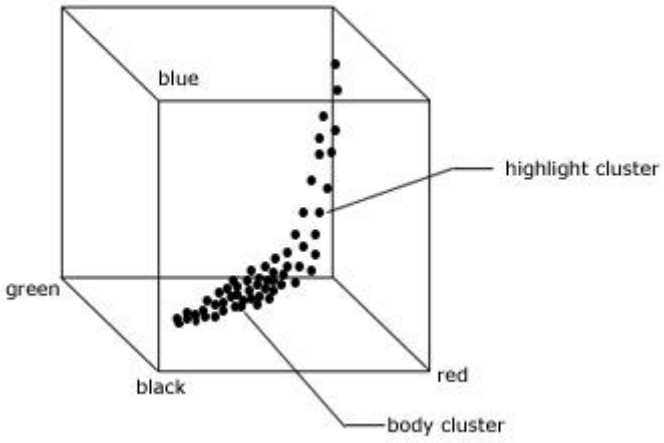

Figure 16. Body and highlight pixels disposition

In general the algorithm is the next - discovering of the chosen area for finding two clusters (highlight pixels and body reflection pixels) is performed and then replacing the highlight pixels with necessary body pixels is made. Below there is more precisely description.

For the initial data we applied a mask to get more bounded area for processing – in order to choose a concrete object or a part of the object with a highlight from an image. The mask is presented as a matrix of size equal to image size which consists from 1 and 0 respectively for pixels from interesting area and other pixels which are out of our exploration area. After getting data by mask we received vector $x = (x_1, x_2, ..., x_d)^T$ where $d = 81$, and each x_i is a spectral component – vector of corresponding dimensionality.

Then standard PCA described in Chapter 4 is applied to our 81-D data to decrease the dimensionality to 2-D, and then we work on highlight removal exactly in 2-D space. So following to the PCA technique, for the vector x covariance matrix C is calculated as in Eq.23, then by involving a decomposition of the covariance matrix as in Eq.24 the matrix U is received which consists from eigenvectors of C which are arranged respectively to corresponding eigen values ordered in a decreasing way. And so mapping of our 81-D vector x onto 2-D case is presented as:

$$y = U^T \cdot x, \tag{56}$$

where due to our needs for 2-D case matrix U is presented only by two first eigenvectors. This matrix U will be used later for reconstructing data to the initial dimensionality.

The next step is to use the mixture of PPCA described in Chapter 4 for our non-linear 2-D data which possess by piece-wise linearity. As was described in Chapter 4 PPCA technique uses the latent variable model where our variable y is presented by:

$$y = Wz + \mu + \varepsilon\,, \tag{57}$$

where W is a transfer matrix, y is a latent variable (principal component) whose distribution is Gaussian with a zero mean, μ is a mean vector and ε is an y-independent noise variable.

The mixture model of PPCA uses a linear combination of an M number of PPCA distributions $P(x|\pi,\mu,C)$ where C is a covariance matrix $C = WW^T + \sigma^2 I$.

The mixture distribution for $M = 2$ components corresponding to highlight and body components is as follows:

$$P(x|\pi,\mu,C) = \sum_{i=1}^{2} \pi_i N(x|\mu_i, C_i)\,, \tag{58}$$

where π_i is a mixing coefficient. The log-likelihood of the equation is maximized using an expectation-maximization algorithm to fit the model to the data. The EM algorithm computes the responsibility of each component to generate the data point and revalues the mixing coefficient, the means and the covariance structure.

The K-mean algorithm described before in Chapter 5 was used for clusterization needs in the mixture of PPCA approach. Due to a big amount of samples and their clear separateness the clustering of the data by this algorithm was executed quite well. The highlight and body pixels were well recognized from each other and there were not some obstacles at a time of clustering. As result of clustering the two clusters are received, and to define what cluster is referred to either body or highlight the powerful of each multitude is calculated. Then the cluster with a maximum amount of samples is the body reflection cluster, and the second is the highlight.

And so as a result we have two clusters of data in 2-D space and corresponding to them basises for 1-D spaces, where the basis for every 1-D subspace is eigenvector of

corresponding covariance matrix which was got during the processing of the mixture PPCA for 2-D data. Now the task is to map samples from highlight cluster onto the body cluster – let us say, to find correspondences for highlight pixels among body reflection pixels. The way of this mapping is made in the next way – the sample from highlight cluster is mapped onto body cluster along its basis vector up to intersection with a basis vector of a body cluster. And then at the time of restoration of 2-D highlight samples up to 81-D space these intersection points, which are actually the coordinates of the highlight pixels in basis of body cluster, are used instead of the initial highlight points.

And so we are going to find a cross point between a line for which the directing vector is the basis 2-D vector for the body cluster and the means of it belongs to this line, and for the all lines for which the directing vector is the basis 2-D vector for the highlight cluster and the samples of this cluster belong to the line in series.

The well-known view for a line is:

$$y = ax + b \tag{59}$$

The equation for the line for which two points (x_1, y_1) and (x_2, y_2) belongs in 2-D space is:

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_2} \tag{60}$$

For the two lines $y = a_1 x + b_1$ and $y = a_2 x + b_2$ the intersection point $(x_{cross,} y_{cross})$ is calculated in a simple way. Due to equality of y-coordinates:

$$a_1 x_{cross} + b_1 = a_2 x_{cross} + b_2, \tag{61}$$

then:

$$x_{cross} = \frac{b_2 - b_1}{a_1 - a_2}, \tag{62}$$

and so:

$$y_{cross} = \frac{b_2 - b_1}{a_1 - a_2} \cdot a_1 + b_1 \tag{63}$$

For the body cluster the first eigenvector of corresponding covariance matrix, which comes as the basis vector for 1-D space, is $e_b = (e_{xb}, e_{yb})$, the mean of the cluster is $\mu_b = (\mu_{xb}, \mu_{yb})$ - and so the points $A1 = (\mu_{xb}, \mu_{yb})$ and $B1 = (\mu_{xb} + e_{xb}, \mu_{yb} + e_{yb})$ define the line for which we are going to find the cross points with other lines. By using Eq. 60 the general view for it could be written:

$$y = \frac{e_{yb}}{e_{xb}} x + (\mu_{yb} - \frac{e_{yb}\mu_{xb}}{e_{xb}}) \tag{64}$$

The directing vector for other lines is the basis 2-D vector for the highlight cluster which is $e_h = (e_{xh}, e_{yh})$, and the samples from highlight cluster are the points belonging to line in series - $p_i = (p_{xi}, p_{yi})$, $i = 1...N$, N - number of samples in the highlight cluster. And so the points $A2 = (p_{xi}, p_{yi})$ and $B2 = (p_{xi} + e_{xh}, p_{yi} + e_{yh})$ define the other lines. And again following by Eq. 60 the general view for these lines is:

$$y = \frac{e_{yh}}{e_{xh}} x + (p_{yi} - \frac{e_{yh} p_{xi}}{e_{xh}}), \tag{65}$$

where $i = 1...N$, N - number of samples in the highlight cluster (highlight pixels).

And so the cross points are defined following to Eq. 62-63:

$$x_{cross,i} = \frac{p_{yi} - \frac{e_{yh} p_{xi}}{e_{xh}} - \mu_{yb} + \frac{e_{yb}\mu_{xb}}{e_{xb}}}{\frac{e_{yb}}{e_{xb}} - \frac{e_{yh}}{e_{xh}}}, \tag{66}$$

$$y_{cross,i} = \frac{p_{yi} - \frac{e_{yh} p_{xi}}{e_{xh}} - \mu_{yb} + \frac{e_{yb}\mu_{xb}}{e_{xb}}}{1 - \frac{e_{yh} e_{xb}}{e_{xh} e_{yb}}} + \mu_{yb} - \frac{e_{yb}\mu_{xb}}{e_{xb}}, \tag{67}$$

where $i = 1...N$, N - number of highlight pixels.

And as it was mentioned, these points of intersection will be used instead of original 2-D samples from the highlight cluster for restoration data from current 2-D case up to initial 81-D space.

There is below step by step precisely algorithm description. It was implemented using Matlab and a Netlab toolbox [13].

Algorithm. The highlight removal algorithm

Input: the observed 81-D vector **x**

Do:

1. Compress data by using PCA technique – go down from 81-D case onto 2-D space, mapping 81-D data onto 2 principal components, which are the two first eigenvectors of corresponding covariance matrix.
2. We are going to apply the mixture of PPCA to 2-D data to get the basis 2-D vectors for 1-D space. For this fitting the mixture model to the data: $P(x \mid \pi, \mu, C) = \sum_{i=1}^{2} \pi_i N(x \mid \mu_i, C_i)$ has to be made.
3. The K-mean algorithm is used to find means of clusters - highlight and body reflection clusters.
4. By supposing our data to be presented as Gaussian mixture model train it with Expectation-Maximization algorithm calculating likelihoods of samples. Defining membership of every sample by choosing the maximum likelihood.
5. Computation of eigendecomposition of C_b and C_h to get the first their eigenvectors which are the basis vectors for body and highlight cluster.
6. Mapping each highlight data point onto the basis vector for body cluster as lines intersection in 2-D space.
7. Data restoration back to 81-D case with substitution highlight points with a points which were got as mapping highlight points onto the basis vector for the body cluster.

Reproduce: the spectral image with a synthetic part instead of a highlight region in a suitable color system.

6.2 Results of the experiment

The experiments were made on a spectral image with *cherries* and a *pen* which surfaces possess by a highlight. All images are acquired in the visible range of

wavelengths from 380 nm to 780 nm and the spectral components are taken at 5 nm apart. Thus as was mentioned the spectral dimension consists of 81 components.

In the first experiment we test the part of the image *pen*. The image itself is shown on Fig.17:

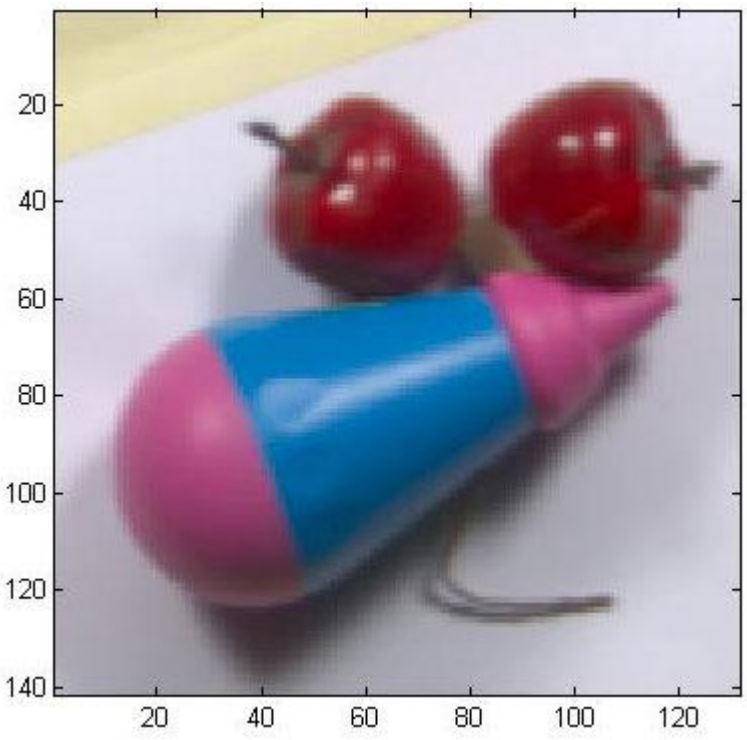

Figure 17. The original image

The mask for the pen is:

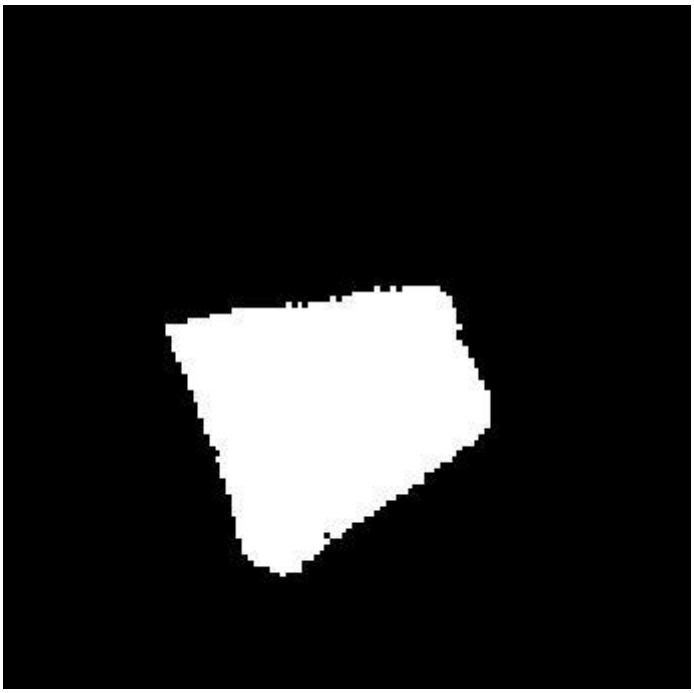

Figure 18. The mask for the *pen* object

After applying PCA and decreasing dimensionality down to 2-D space and clustering the data two body (blue) and highlight (yellow) clusters are clearly seen in Fig.18:

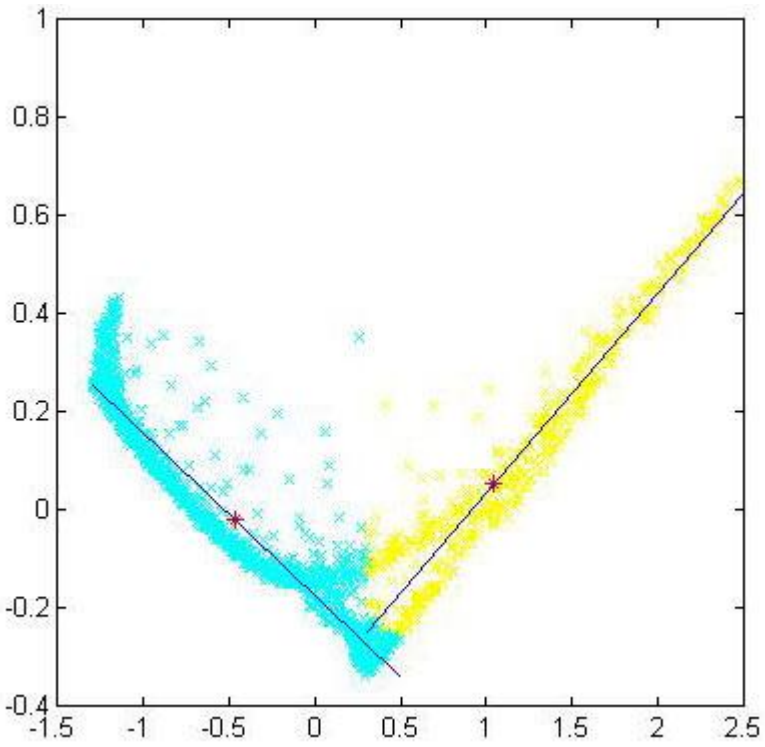

Figure 19. Body and highlight clusters for the pen

The means of every cluster are shown by red asterixes.

And then mapping of highlight points onto basis vector for the body cluster is performed:

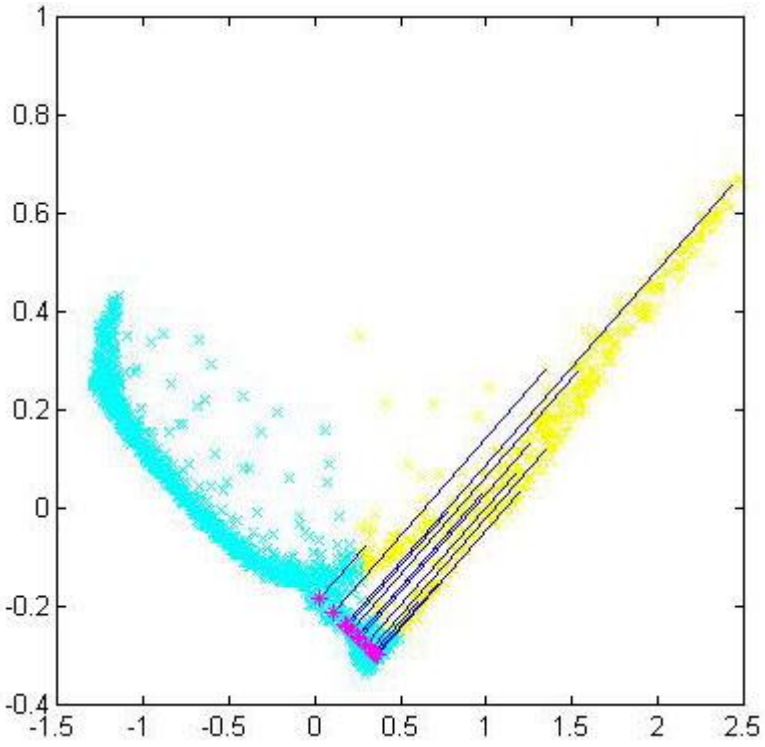

Figure 20. Mapping of the highlight points onto the body cluster

Intersection points are shown as magenta asterixes which will be used for restoration data back to 81-D space instead of initial highlight points (yellow crosses). At a time of restoration these highlight points will be replaced by the body reflection points.

And so as a result we receive the image where the highlight from the pen is removed:

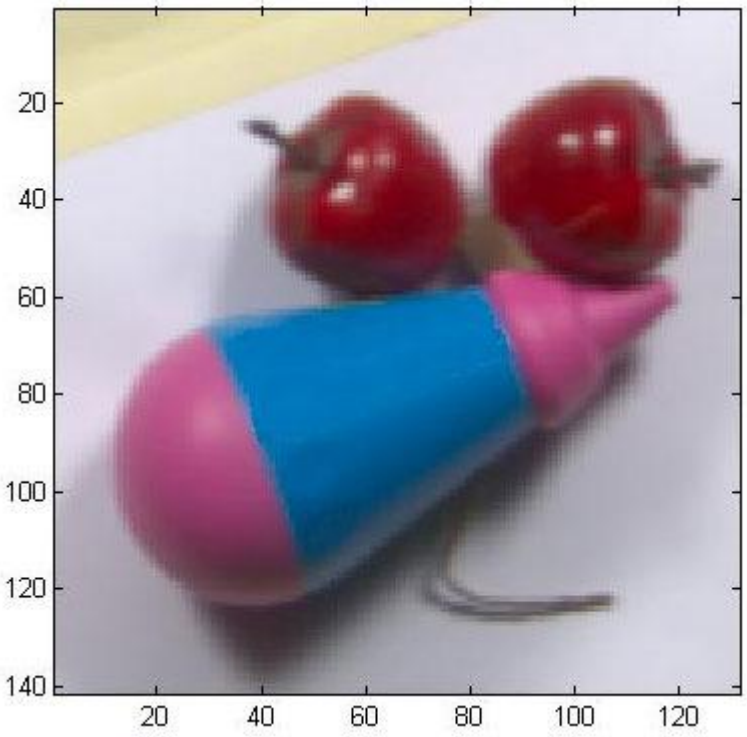

Figure 21. The image with removed highlight from the pen

We can also see histogram of *r-g* chromaticities for the pixel points under the mask for the *pen* object before highlight removal (left image) and after algorithm processing (right image):

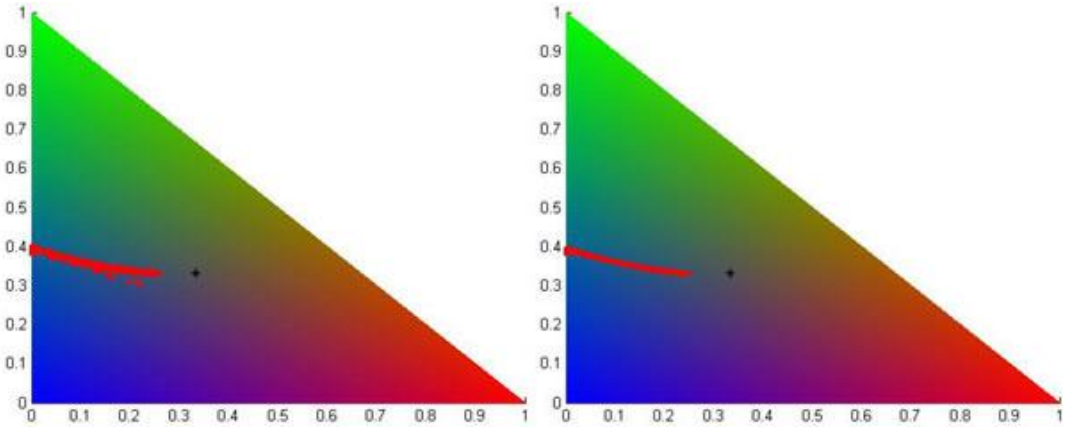

Figure 22. Histogram of *r-g* chromaticities for the pen

The similar work was done for a *cherry* objects. Again the initial image is shown in Fig.17. The mask for the object *cherries* is:

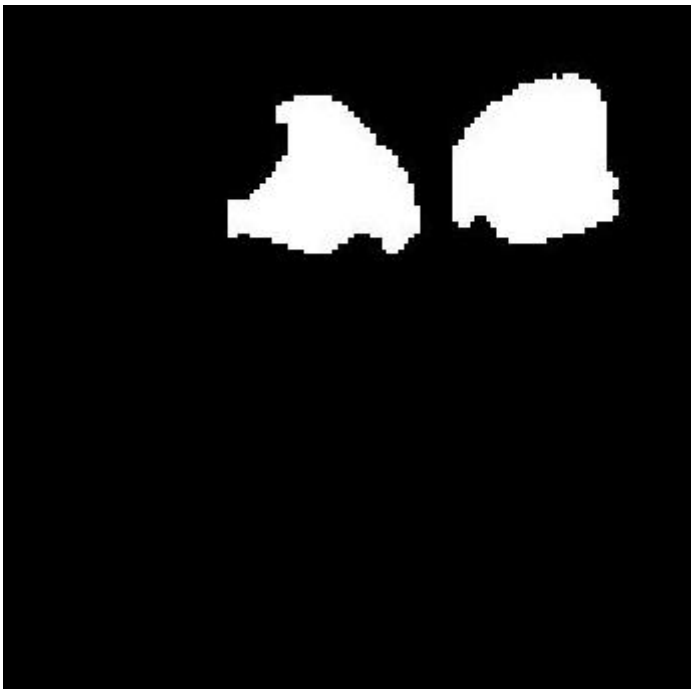

Figure 23. The mask for the object *cherries*

After applying PCA and decreasing dimensionality down to 2-D space, and clusterization data for body (blue crosses) and highlight (yellow crosses) clusters the result is:

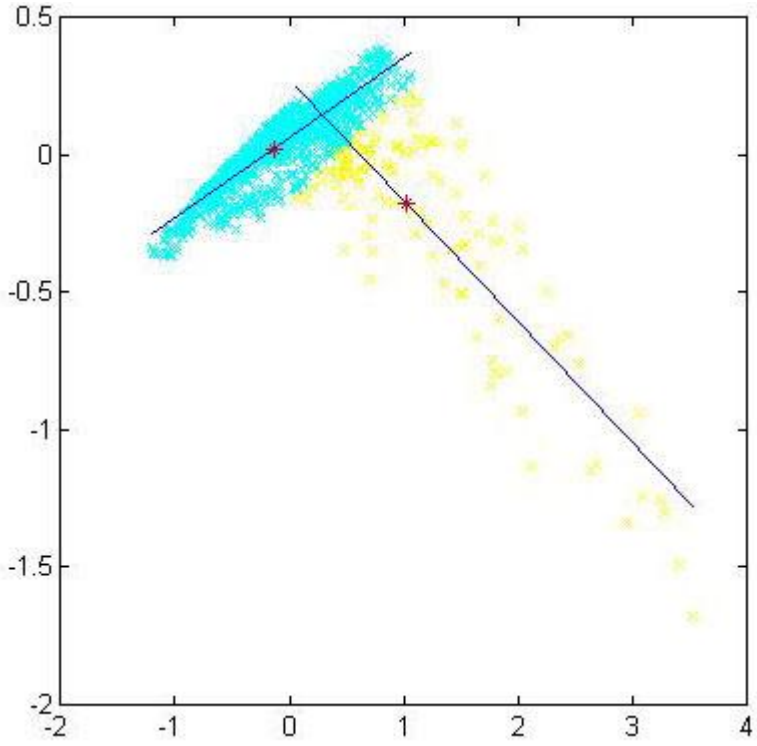

Figure 24. Body and highlight clusters for the cherries

The means of the clusters are marked as red asterixes. And similarly mapping of highlight points onto basis vector for the body cluster is performed:

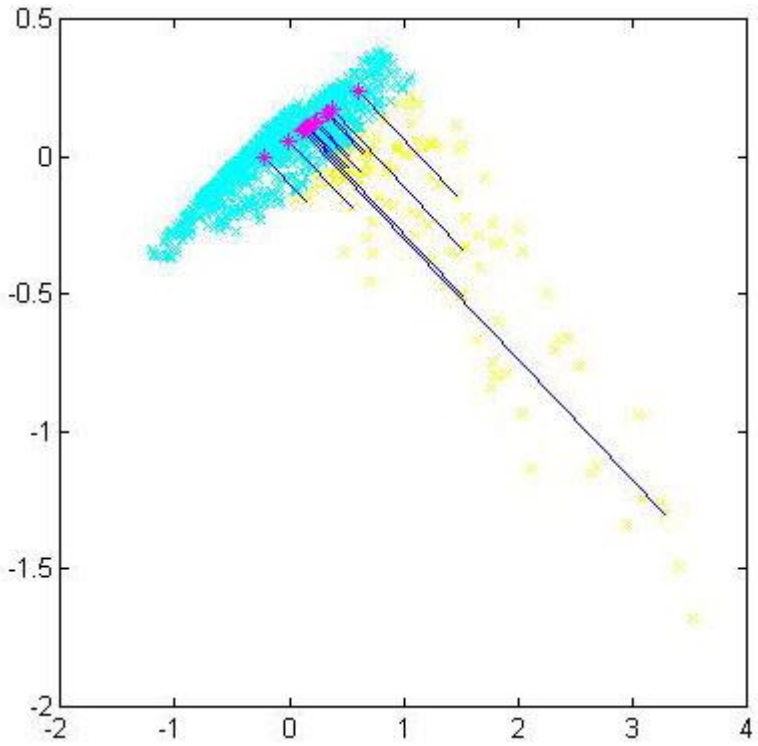

Figure 25. Mapping of the highlight points onto the body cluster

And at a time of data restoration back to 81-D space these intersection points will be used instead of initial highlight points. And so the result of algorithm working for a *cherry* object is:

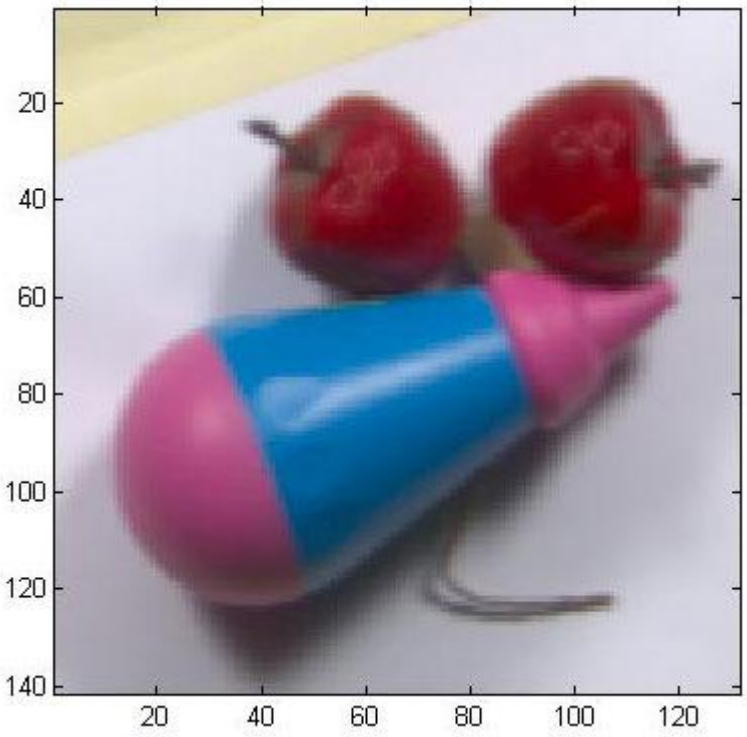

Figure 26. The image with removed highlight from the cherries

Again we can see histogram of *r-g* chromaticities for the pixel points under the mask for the *cherries* object before highlight removal (left image) and after algorithm processing (right image):

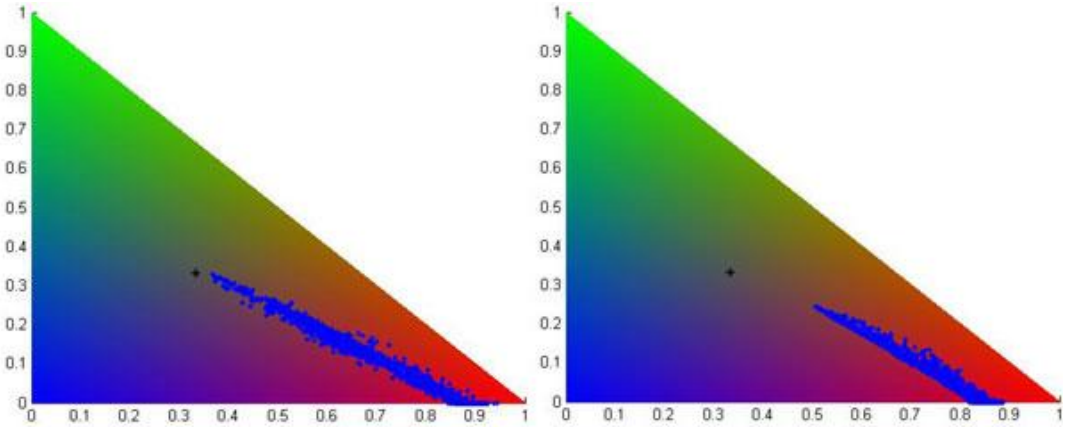

Figure 27. Histogram of *r-g* chromaticities for the cherries

It could be noticed some inaccurateness for the *cherries* image. Deal is that the initial highlight for the object was not so monotonous – due to presence of a little transparency layer onto fruit surface (fruit wax) – therefore it puts some disturbances. But in general the algorithm shows quite well results and the short time of execution. Below there is a table with a time of productivity for the algorithm. The experiment

was made with using workstation with CPU Intel Pentium 3 500 MHz and 256 Mb RAM.

Object	Execution time, sec
Pen	2,57
Cherries	2,013

Table 1. Productivity of the algorithm

7. Conclusion

In this thesis the algorithm on highlight removal in spectral images was proposed. It is based on machine learning and ideally it was intended for the automated work in a spectral space, but on this stage of work it still needs manual involving – for taking the data under a mask. For the artificial chosen area the algorithm efficiency separates data into body-reflection and highlight clusters and performs projection of highlight points to generate necessary color pixels for highlight removal.

In general the algorithm works quite well – the visual quality of the results is relatively good. Actually for the highlight removal there is no the standard measure for quality of an algorithm processing.

In future this algorithm could be extended for the common case – for highlight removal for the whole scene on an image – for the all objects on the scene at a time, without hand-segmenting for the regions to which the algorithm should be applied. For this goal, algorithms for pattern recognition should be improved in order to provide with a possibility for separation objects on the scene of an image. Naturally there are some obstacles for it because ideally the work should be performed for objects with not a simple color structure of their surfaces – therefore the segmenting algorithm should possesses by the possibility to pick out the regions of the object with a monotonous color covering which is disturbed with a natural caused highlight.

In general the result obtained in the thesis can be applied for machine vision and also additionally the performance could be improved in future work by taking in attention generalization of the algorithm performance up to more general case mentioned above.

References

[1] Hans J. Andersen, Moritz Storring, Classifying Body and Surface Reflections using Expectation-Maximization, In: The PICS Conference, An International Technical Conference on The Science and Systems of Digital Photography, including the Fifth International Symposium on Multispectral Color Science, PICS 2003, May 13, Rochester, NY, USA. pp. 441-446. ISSN/ISBN: 0-89208-245-3

[2] V.Bochko, Y.Miyake, Highlight Removal in Endoscope Images, Department of Information Technology, Lappeenranta University of Technology, Lappeenranta, Finland, accepted for CGIV 2006 conference

[3] V.Bochko, J.Parkkinen, A Spectral Color Analysis and Synthesis Technique, Department of Information Technology, Lappeenranta University of Technology, Lappeenranta, Finland

[4] V.Bochko, J.Parkkinen, Highlight Analysis Using a Mixture Model of Probabilistic PCA, WSEAS Trans. On Systems, Issue 1, Vol. 4, 2005, pp. 55-59

[5] V.Bochko, J.Parkkinen, Principal Component Analysis Using Approximated Principal Components, Research Report 90, Department of Information Technology, Lappeenranta University of Technology, Lappeenranta, 2004

[6] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern classification, New York, Wiley, 2001

[7] M.D. Fairchild, Color Appearance Models, Adison Wesley, USA, 1998, ISBN: 0-201-63464-3

[8] R.C. Gonzalez, R.E. Woods, Digital Image Processing. Prentice Hall, 2002, ISBN: 0-20-118075-8

[9] Gudrin J. Klinker, Steven A. Shafer, Takeo Kanade, A Physical Approach to Color Image Understanding, Cambridge Research Lab, Digital Equipment Corporation, Cambridge, International Journal of Computer Vision 4, 1990, pp:7-38

[10] Matlab Tutorial http://www.mathworks.com

[11] Thomas P. Minka, Automatic choice of dimensionality for PCA, MIT Media Laboratory, Vision and Modeling Group, Cambridge, December 29, 2000

[12] Ian T. Nabney, Netlab. Algorithms for Pattern Recognition, Springer, 2003, ISBN: 1-85233-440-1

[13] Netlab Toolbox http://www.ncrg.aston.ac.uk/netlab

[14] Karsten Schluns, Andreas Koshan, Global and Local Highlight Analysis in Color Images, in Proc. 1st International Conference on Color in Graphics and Image Processing CGIP'2000, Saint-Etienne, France, pp. 300-304, October 2000

[15] Ping Tan, Stephen Lin, Long Quan, Heung-Yeung Shum, Highlight Removal by Illumination-Constrained Inpainting, Microsoft Research, In Proc. International Conference on Computer Vision, Nice, France, Oct. 2003

[16] Sergios Theodoridis, Konstantinos Koutroumbas, Pattern recognition, London, Academic Press, 2003

[17] F. Tong, B.V. Funt, Specularity Removal for Shape from Shading, Proceedings of the Vision Interface Conference 1988, pp. 98-103, Edmonton, Canada, 1988

[18] Online encyclopedia www.wikipedia.org

[19] M.D'Zmura, P.Lennie, Mechanisms of color constancy, Journal of the Optical Society of America A (JOSA-A) 3(10):1662-1672, October, 1986

[20] http://mathworld.wolfram.com

[21] S.W.Lee, R.Bajcsy, Detection of specularity using color and multiple views, Proceedings of the 2-nd European Conference on Computer Vision, pp.99-114, Santa Margherita Ligure, Italy, May 1992

[22] Daniel Malacara, Color Vision and Colorimetry Theory and Applications, SPIE PRESS, ISBN: 0-8194-4228-3

[23] Peter G.J.Barten, Contrast Sensitivity of the Human Eye and Its Effects on Image Quality, SPIE PRESS, ISBN: 0-8194-3496-5

[24] Gaurav Sharma, Digital Color Imaging Handbook, CRC PRESS, ISBN: 0-8493-0900-X

Appendix

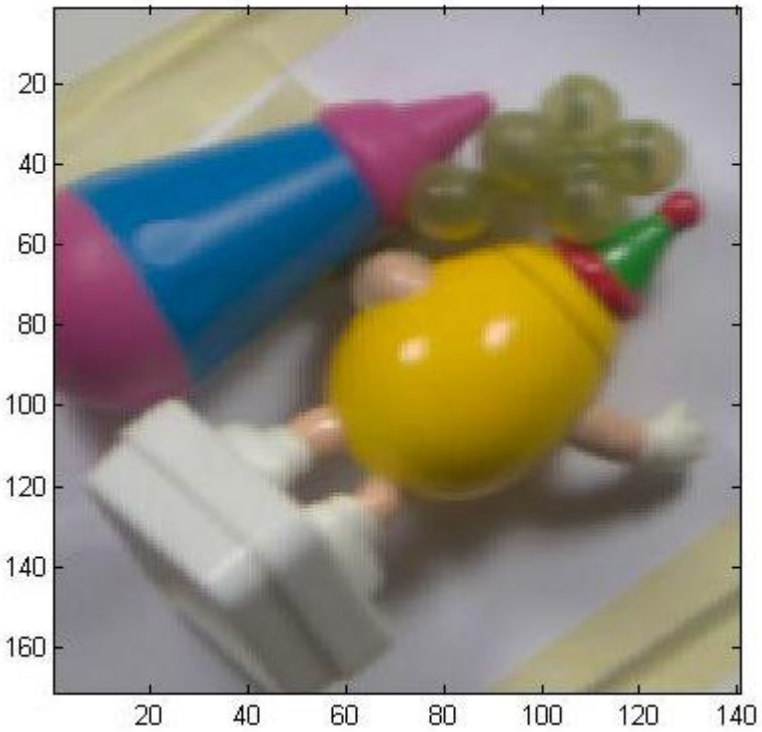

Figure 1. The initial image

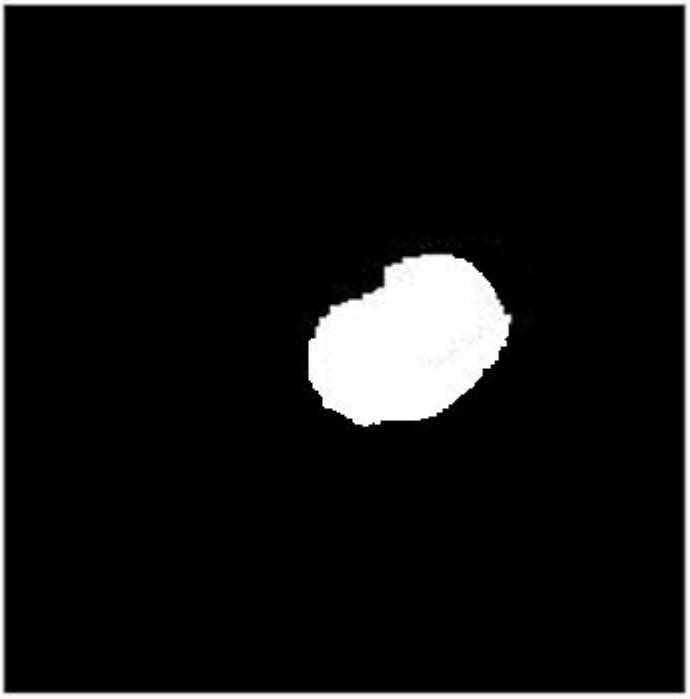

Figure 2. Mask for the *yellow man* object

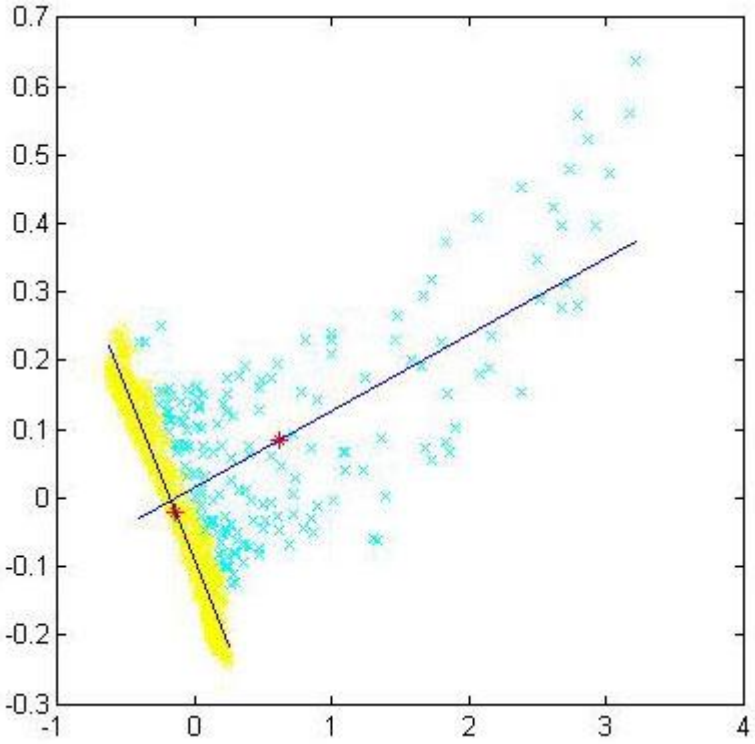

Figure 3. Body and highlight clusters for the *yellow man* object

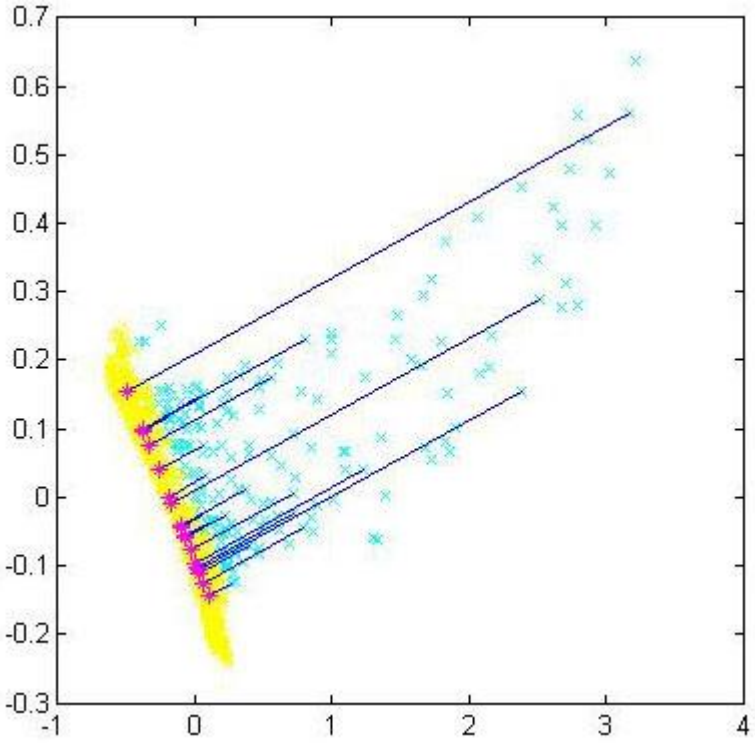

Figure 4. Mapping of the highlight points onto the body cluster

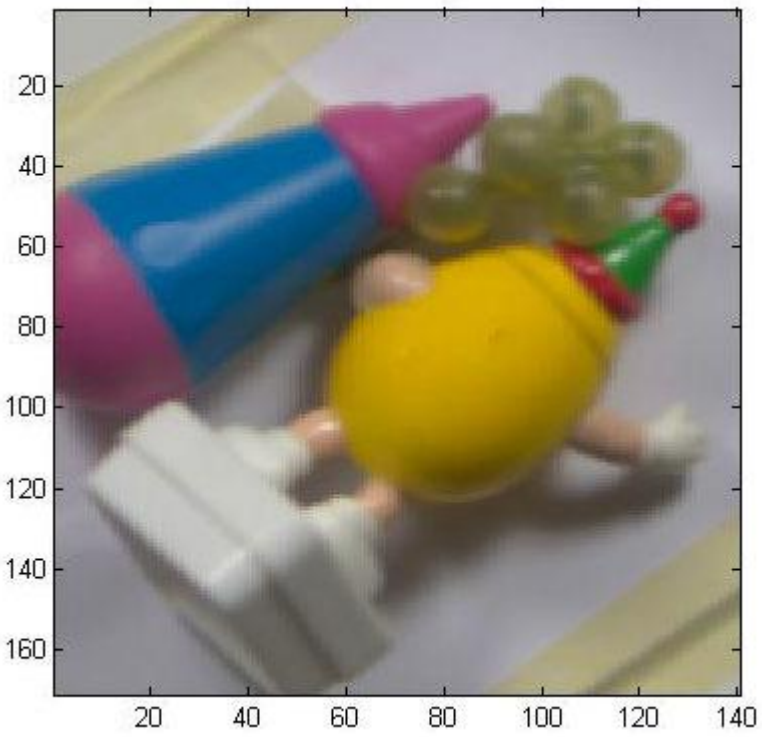

Figure 5. The object after highlight removal