FACE COLOUR UNDER VARYING ILLUMINATION - ANALYSIS AND APPLICATIONS

BIRGITTA MARTINKAUPPI

Department of Electrical and
Information Engineering and
Infotech Oulu,
University of Oulu

OULU 2002

BIRGITTA MARTINKAUPPI

FACE COLOUR UNDER VARYING ILLUMINATION - ANALYSIS AND APPLICATIONS

Martinkauppi, Birgitta, Face colour under varying illumination - analysis and applications

Department of Electrical and Information Engineering and Infotech Oulu, University of Oulu, P.O.Box 4500, FIN-90014 University of Oulu, Finland

Oulu, Finland

2002

Abstract

The colours of objects perceived by a colour camera are dependent on the illumination conditions. For example, when the prevailing illumination condition does not correspond to the one used in the white balancing of the camera, the object colours can change their appearance due to the lack of colour constancy capabilities. Many methods for colour constancy have been suggested but so far their performance has been inadequate. Faces are common and important objects encountered in many applications. Therefore, this thesis is dedicated to studying face colours and their robust use under real world illumination conditions. The main thesis statement is "knowledge about an object's colour, like skin colour changes under different illumination conditions, can be used to develop more robust techniques against illumination changes".

Many face databases exist, and in some cases they contain colour images and even videos. However, from the point of view of this thesis these databases have several limitations: unavailability of spectral data related to image acquisition, undefined illumination conditions of the acquisition, and if illumination change is present it often means only change in illumination direction. To overcome these limitations, two databases, a Physics-Based Face Database and a Face Video Database were created. In addition to the images, the Physics-Based Face Database consists of spectral data part including skin reflectances, channel responsivities of the camera and spectral power distribution of the illumination. The images of faces are taken under four known light sources with different white balancing illumination conditions for over 100 persons. In addition to videos, the Face Video Database has spectral reflectances of skin for selected persons and images taken with the same measurement arrangement as in the Physics-Based Face Database. The images and videos are taken with several cameras.

The databases were used to gather information about skin chromaticities and to provide test material. The skin RGB from images were converted to different colour spaces and the result showed that the normalized colour coordinate was among the most usable colour spaces for skin chromaticity modelling. None of the colour spaces could eliminate the colour shifts in chromaticity. The obtained chromaticity constraint can be implemented as an adaptive skin colour modelling part of face tracking algorithms, like histogram backprojection or mean shift. The performances of these adaptive algorithms were superior compared to those using a fixed skin colour model or model adaptation based on spatial pixel selection. Of course, there are cases when the colour cue is not enough alone and use of other cues like motion or edge data would improve the result. It was also demonstrated that the skin colour model can be used to segment faces and the segmentation results depend on the background due to the method used. Also an application for colour correction using principal component analysis and a simplified dichromatic reflection model was shown to improve colour quality of seriously clipped images. The results of tracking, segmentation and colour correction experiments using the collected data validate the thesis statement.

Keywords: image colour analysis, machine vision, computer vision, skin colour, varying lighting conditions, colour camera

Acknowledgements

List of symbols

Greek Letters

α	weight for refreshing model histogram
δ	spectral reflectance of skin
Δ	difference
ε	basis function
η	spectral sensitivity or spectral response
Θ	imaging geometry like photometric angles
λ	wavelength
μ	mean
ρ	spectral reflectance of the sample

Abbreviations

bmp	bitmap
CCD	Charge Coupled Device
CIE	Commission Internationale de l'Éclairage
CS	colour signal
D	dimension (1D, 2D or 3D)
DIN	Deutsches Institut für Normung
DR	dichromatic reflection
ICA	independent component analysis
IR	infrared
K	Kelvin (unit for colour temperature)
MA	moving average
NCC	Normalized Colour Coordinates
NCS	Natural Colour System

nm   nanometer
PCA   principal component analysis
RGB   red, green and blue pixel values
SCE   spectral component excluded
SCI   spectral component included
SOM   self-organizing map
SPD   spectral power distribution of illumination
SVD   singular value decomposition
UO   University of Oulu
WWW   World Wide Web

List of original publications

I   Marszalec E, Martinkauppi JB & Pietikäinen M (1997) Evaluation of the per-
        formance of color camera for measuring small color differences. SPIE 3208 In-
        telligent Robots and Computer Vision XVI: Algorithms, Techniques, Active
        Vision, and Materials Handling, 348-359.
II   Marszalec E, Martinkauppi JB, Soriano MN & Pietikäinen M (2000) Physics-
        based face database for color research. Journal of Electronic Imaging 9(1): 32-
        38.
III   Soriano MN, Marszalec E, Martinkauppi JB & Pietikäinen M (1999) Making
        saturated facial images useful again. SPIE 3826 EUROPTO Conference on Po-
        larization and Color Techniques in Industrial Inspection, 113-121.
IV   Soriano MN, Martinkauppi JB, Huovinen S & Laaksonen MH (2002) Adaptive
        skin color modeling using the skin locus for selecting training pixels. Pattern
        Recognition, in press.
V   Martinkauppi JB & Soriano MN (2001), Basis functions of the color signal of
        skin under different illuminants. Proc. 3rd International Conference on Multi-
        spectral Color Science MCS'01, Joensuu, Finland, 21-24.
VI   Martinkauppi JB, Soriano MN & Laaksonen MH (2001) Behavior of skin color
        under varying illumination seen by different cameras at different color spaces.
        SPIE 4301 Machine Vision in Industrial Inspection IX, 102-113.
VII   Martinkauppi JB, Soriano MN, Huovinen S & Laaksonen MH (2002) Face
        video database. Proc. 1st European Conference on Color in Graphics, Imaging
        and Vision (CGIV'2002), Poitiers, France, 380-383.
VIII   Martinkauppi JB, Sangi P, Soriano MN, Pietikäinen M, Huovinen S & Laakso-
        nen MH (2001) Illumination-invariant face tracking with mean shift and skin
        locus. Proc. IEEE International Workshop on Cues in Communication (Cues
        2001), Kauai, Hawaii, 44-49.

The author participated in the research and writing of Papers I-IV and was mainly respon-
sible for the practical arrangements and measurements done for the database created in Pa-
per II and the measurements for Paper I. She wrote Papers V-VII while the other authors
gave their useful comments. In the Papers V-VI, she was responsible for the research made.
Papers VI-VIII are based on her ideas, while Paper V was based on Dr. Soriano's idea. For
Paper VII, all the authors participated in the practical research. In Paper VIII, the present

author applied the chromaticity constraint technique for the mean shift algorithm which was implemented by Mr. Sangi. While she mainly performed the writing and experiments, also the co-authors Prof. Pietikäinen and Mr. Sangi participated in the writing process. The other authors gave once again useful comments.

Contents

Appendix 1: Transforms from RGB to other colour spaces
Appendix 2: Visualization of skin chromaticities at different colour spaces
Appendix 3: Mean shift algorithm
Errata
Original papers

1 Introduction

1.1 Background

Colour cameras, video cameras and their applications have become increasingly popular among professionals and amateurs alike. Still, many colour related problems have not yet vanished, like problems of a colour camera keeping stable colour appearance for an object or producing similar colour appearances as the human vision system. To make the situation more difficult, different colour cameras do not necessarily produce the same colour appearances for the same scene under the same imaging conditions. One of the main reasons for different appearances is in the first stage of image formation: spectral sensitivities of the sensors diverge from those of the human eye and from the other cameras. Of course, there are cameras with responses similar to the human eyes, but at least so far they are rarely used today.

One of the remarkable things in the human vision system is its ability to disregard the effects of widely varying illumination, automatically. This ability aids in keeping the object's colour appearance stable, and it is often erroneously called colour constancy which is only approximately true. In the literature, it has been claimed to be both a high level brain process (which contains, among other things, a memory for some colours, and adjustment for lighting level) or a low level process. The details behind the colour constancy mechanism are still under research, although many theories and studies have been suggested, but they are beyond the scope of this thesis.

Unfortunately, colour cameras themselves do not have this kind of "built-in" mechanism against illumination dependency. They cannot separate changes in an object's reflectance from changes in illumination over the object. The proper white balancing or white calibration of the camera to the prevailing light source does not guarantee any other colour than the "white" calibration object having the same colour appearance in images taken under different light sources. The problem worsens when the illumination changes from the calibrated cases: distortion can appear in objects' colours (both in intensity and in chromaticity) due to the illumination variation and the properties of cameras, like limited dynamic range.

Problems caused by illumination in colour imaging are handled in general in four different manners: 1) preventing changes by controlling illumination or ignoring information taken under changed condition, 2) using a process which disregards illumination, 3) adapting to the changes or 4) combining the second and the third to improve robustness. The first possibility is inadequate in many applications because it is impossible to control illumination in many real world situations and ignoring information may lead to a loss of essential data. The second option is to use illumination invariant (or robust) features or colour correction, in other words, colour constancy for cameras. Illumination invariance here means invariance / robustness towards lighting with different spectra and intensity although in some cases it has been used to stand for invariance to the direction of a light source. The goal of colour correction is usually correction of chromaticities back to the original values, while the invariant features try to present colour information independent of lighting conditions. A massive number of papers have been published in this area, but still for machine vision applications their performance is not necessarily enough. For example, some colour cameras do have an automatic colour correction method like the grey world algorithm (Buchsbaum 1980) and these methods can produce satisfactory results at least for a human observer as long as the assumptions and constraints imposed by the methods are valid. But in many scenes, the results are poor even for human evaluation and it is very easy to show that these algorithms fail. In fact, the correction can lead to unstable colour appearance and wrongly corrected colours. Almost all correction algorithms except Retinex (Land 1977, Land 1986, and Land & McCann 1971) work under one global illumination change whereas in practise, local changes are common. There have been anyway suggested methods for correcting nonuniform intensity (Chang & Reid 1996, Powell *et al.* 1999) but this can be also cancelled by using only chromaticities. Illumination invariant features can be pixel based or region based but they are not successful either for the same reasons as the colour correction algorithms. In an extreme case, these features are obtained by quantization to a few possible colour values (Redfield & Harris 2000). This causes poor discrimination capability and is therefore useful only in a couple of applications. In general, once the illumination has changed and sensor readings obtained, it is impossible to reconstruct the ideal values due to information losses introduced by the change. The third option is investigated in this thesis whereas the fourth option will be hopefully studied in the future.

In this thesis, the adaptive schema are studied with colour images or frames of human faces and facial skin colour because a practical solution for realistic illumination problems is being sought for machine vision purposes. Also a colour correction schema for facial colours is presented under severe information loss due to clipping. Faces are selected as the study target since they are common and important objects in videos and images. But what is skin colour? Although the answer to this question might seem trivial - perceived colour appearance of skin - a closer look at it reveals an interesting dependence on the perceiver. The human perceiver usually sees the skin colour as quite constant and stable over a wide range of illumination conditions. The skin chromaticities observed are few and are located in a limited region in the chromaticity space. In fact, humans can easily notice even a small deviation from these chromaticities and therefore it is important to have a high quality representation of skin colour (Harwood 1976, Satyanarayana & Dalal 1996 and Lee & Ha 1997). On the other hand, uncalibrated cameras can produce a rainbow colour appearance for skin under illumination conditions varying between sunset / sunrise and daylight because of the lack of a colour constancy ability. The possible skin chromaticities for the cam-

era cover a large region in a chromaticity space. This skin chromaticity region can be reduced drastically by white balancing the camera properly each time for the prevailing illumination. Although often unspecified in the literature, in this thesis, the skin colour refers to all possible perceivable chromaticities of skin. The term skin tone is used to refer to cases with a smaller skin chromaticity area and shades generally associated with proper skin colour by humans.

1.2 The scope and contributions of the thesis

The main statement of this thesis is: knowledge of an object's colour, like skin colour changes under different illumination conditions, can be used to develop more robust techniques against illumination changes. Faces were selected as the objects to be studied because they are common and important in very many applications. To prove the statement, this thesis employs three different phases: 1) collecting facial skin data under different illumination conditions, 2) analysing the data, and 3) applying the obtained knowledge. The following list shows the novel contributions and their support for the main statement:

* A method for evaluating colour camera performance (Paper I) is developed for studying metamerism in human and camera vision systems and is used in evaluating cameras. This information can also be used as a criterion for selecting a camera or between human vision and device colour spaces.

* A unique Physics-based Face Database (Paper II) is introduced for face related colour research. Its novelty lies in the combination of face images and spectral data related to the formation of those images as well as the procedure for studying the illumination effects on the images. In the procedure, the camera was first white balanced to one of the light sources and then images were taken under this light source and under other light sources with different spectral power distribution. This was repeated for four different light sources. The purpose of the database is to collect knowledge about facial skin colour appearance under known illumination and camera white balancing conditions.

* A novel method for skin colour correction is presented for face images with clipping (Paper III). For this method, the knowledge obtained from the database is shown to be useful for its development.

* Creation of a chromatic constraint which does not only cover different illumination conditions but also takes into account the effect of different camera calibrations (Paper IV). This constraint offers information about possible skin chromaticities perceivable by a colour camera with a certain illumination range and white balancing conditions.

* Use of basis functions obtained from skin colour signals for creating the chromatic constraint (Paper V). This is a spectral based method for obtaining the information about skin chromaticities under different conditions. In addition, it makes it easy to simulate outputs for different cameras.

* A study on how skin colour behaves in different colour spaces and evaluation of their usefulness (Paper VI). The purpose is to analyse different colour spaces for chromatic constraint based applications.

* A novel Face Video Database (Paper VII) which contains videos with drastic colour

changes taken in real environments and face images under known illumination is suggested for the testing and developing of algorithms. The videos have been taken with several different cameras. In addition, the face localizations are available in each frame. The goal is once again data collection.

* Implementing a chromatic constraint as a part of different face tracking methods to make possible adaptive skin colour modelling (Paper IV and Paper VIII). Here it is shown that the knowledge obtained about skin chromaticities can be used to provide robustness against illumination change.

* Visualizing how different skin models can be used for segmenting faces in videos (Paper VIII). This is another example of how the chromaticity constraint provides improved results against illumination change.

1.3 The outline of the thesis

The remaining chapters of this thesis are organized as follows:

In Chapter 2, an overview is given of the properties of human skin and the earlier research related to face and skin analysis.

Chapter 3 presents a basis of image acquisition and its physical background for a CCD camera. Especially issues related to a camera's non-idealities and white balancing conditions are studied in detail. An example of automatic colour correction failure is demonstrated. In addition, the responses and outputs of colour cameras are evaluated and compared to those of the human vision system. There is also a short overview of illumination types and different device-dependent colour spaces.

Next, in Chapter 4, a unique Physics-based Face Database is presented for colour research on faces. The skin reflectances from the database are used to evaluate uniformity of the skin and the general shape of the spectra. Then the database is shown to be useful for developing a method for correcting skin colours in severely clipped images.

Then in Chapter 5, skin chromaticities perceived by a colour camera are studied under challenging illumination and camera white balancing conditions. Based on the available data, two methods are suggested for creating the skin chromaticity constraint. In addition, skin RGB is converted to seventeen different colour spaces which are compared using the behaviour of skin chromaticities.

After this, some applications of obtained knowledge are shown in Chapter 6. The chromatic constraint introduced earlier is used in face tracking under drastic and challenging illumination conditions. For test purposes, a novel Face Video Database is created containing videos and images taken by several cameras. Next, the constraint is applied as an adaptive part of a tracking method called histogram backprojection (Swain & Ballard 1991). The results obtained using chromatic constraint are compared with those obtained using nonadaptive modelling and another adaptive schema with backprojection. Then it is shown that the chromatic constraint is as well applicable to another tracking method called mean shift. Once again, a comparison between different ways of skin modelling is presented, also the use of these modelling methods is investigated for segmenting faces in colour videos.

Finally, conclusions are drawn about the databases created, studies made and applications in Chapter 7.

At the end, three appendices list further details. Appendix 1 contains transforms from RGB to other colour spaces, and visualization of skin chromaticities at these colour spaces are displayed in Appendix 2. The mean shift algorithm is presented in Appendix 3.

2 An overview of colour-based face image and skin analysis

2.1 Some basic concepts in colour theory and spaces

The reflection from a surface can be diffuse ("body"), specular ("interface" or "regular") or a mix of the two (Wyszecki & Stiles 2000). In the diffuse case, the incoming light is scattered by the surface without any regularities. Mirror like interaction with light is called specular reflection. The mixed reflection can be either gloss or retro-reflection.

Because an ordinary visual system describes the spectra only with a few descriptors, different reflectances can obtain the same descriptor values. If two colour samples with different reflectance functions have the same colour appearance (= the same descriptor values) under one viewing condition whereas under another they are discriminated to be separate colours, they are called metameric samples. A common factor causing metamerism is illumination change.

Illumination can be described accurately using spectral power distribution SPD which is its radiant output over a wavelength range. A more rough descriptor of illumination is colour temperature. Colour temperature (Wyszecki & Stiles 2000) relates a light source or an illuminant to an ideal model called a Planckian radiator (also called a blackbody radiator and a full radiator) and illustrates the relationship between the red and blue wavelength areas of the SPD. The Planckian radiator is a thermal radiator (hot body) with a continuous SPD depending only on the temperature of the body material. Colour temperature gives a reasonable good sensation of the "colour" of light: a high colour temperature refers to a more bluish light, while a low colour temperature means a light with more reddish components. It defines uniquely the SPD of a Planckian radiator which presents a light emitted by an ideal blackbody source when heated at this certain temperature.

The Planckian locus is the curve formed by the chromaticities of different Planckian radiators in a colour space. The CIE colour spaces model the colour processing of the human vision system. The basic human colour space is CIE XYZ tristimulus values which can be obtained by an illumination dependent transformation from the linear RGB values of the camera. The CIE xy chromaticity coordinates are obtained from the normalization of X and Y tristimulus values by the sum of all three tristimulus values. The CIE Lab and CIE Luv

spaces were developed to obtain more perceptually uniform space for colour presentation like Farnsworth's uniform-chromaticity-scale UCS. The CIE Luv values can be processed further to obtain CIE SH values which correspond to the saturation and hue of the colour.

Device colour spaces like RGB, HSV, YIQ and NCC rgb describe the colour responses for a device which on the other hand can be very different from those of human space. Their and other colour spaces' formulae can be found in Appendix 1.

2.2 Properties of human skin

From the biological point of view, skin can be described as a layered structure, shown in Fig. 1 (Nienstedt *et al.* 1984). The three main layers are subcutaneous tissue, dermis (cortium), and epidermis. The surface of skin itself can be approximated to be diffusional or matt because the uppermost level of skin is covered with dead cells causing no regular reflection. These dead cells are optically inactive (i.e. no fluorescent). The glossiness of skin can be due to sweat, skin oil or some chemical products covering the surface.

The matt skin colour appearance is influenced by the light filtering capabilities of three main colouring agents: melanin in epidermis, carotene in dermis and subcutaneous fat, and blood capillaries across the dermis. Melanin is a brown pigment and carotene gives an orange tint. Haemoglobin (an element of blood) can produce two different tints: if the haemoglobin is oxygenated (oxyhemoglobing) the tint is reddish or pinkish apart from when it is deoxygenated (reduced haemoglobin) then the tint is bluish.
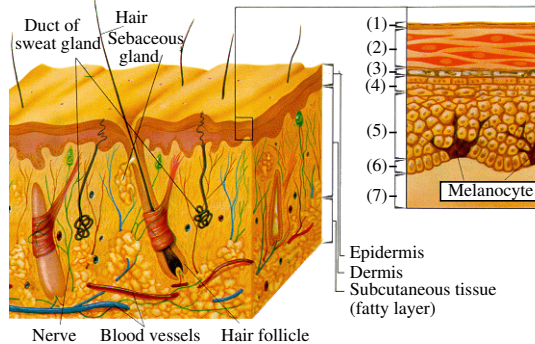
Fig. 1. Structure of the skin. Structure of the epidermis: (1) Keratin, (2) Horny layer, (3) Lucid layer, (4) Granular layer, (5) Spinous layer, (6) Basal layer and (7) Dermis.

The final skin spectra are formed by the interaction between skin and light: light striking skin is transmitted, absorbed, and reflected through the layers. The spectra for human skin generally form a continuous homologous series because of characterization caused by absorption of melanin and haemoglobin (Edwards & Duntley 1939). It has a higher relative reflectance in long wavelengths (orange and red) than in short ones (blue and green). Like most natural objects the skin has spectral variability which are in this case mainly due to amount, density, and distribution of melanin. The skin can be described as an optically inhomogeneous material because under the surface there are colourant particles which interact with light, producing scattering and colouration.

2.3 Skin reflectances, PCA and ICA

Earliest studies on skin reflectances were made by Edwards and Duntley (1939), Buck and Froelich (1948), and Stimson and Fee (1953) according to Wyszecki and Stiles (2000). Recently, Angelopoulou (2001) has made noncontact measurement of skin at different places to separate skin objects from those which have skin coloured appearance. There are some shape differences in skin reflectances between her results and earlier measurements. Many models have also been presented for generating and simulating easily different skin reflectances, for example by Ohtsuki and Healey (1998), and So-Ling and Ling (2001). The reflectances obtained can be used for skin colour simulation as in Störring *et al.* (1999) who computed skin colour appearance under different light sources with one camera calibration. They also compare the calculated skin chromaticities to an average of those obtained from images and find the difference to be reasonable small.

Skin reflectances have been also subjected to principal component analysis, PCA. According to Imai *et al.* (1996) and Nakai *et al.* (1998), skin reflectances can be presented by just three basis functions which correspond to different skin colourants like melanin and carotene. PCA (Moon & Phillips 1998) and independent component analysis ICA (Hyvärinen *et al.* 2001) have also been applied to face images (a comparison between ICA and PCA for colour recognition has been presented by Laamanen *et al.* (2000). The eigenvectors produced by PCA are called eigenfaces and they are applied to face recognition but usually on grey scale images. Soriano *et al.* (1999) extended the eigenface approach to RGB images by applying PCA on each colour channel. They found that the first three eigenfaces contain information about the illumination and camera calibration, and are therefore useful for colour correction. ICA has been applied to colour face images to achieve different components like melanin concentration and to simulate the skin colour appearance with different degrees of components (Tsumura *et al.* 1999). ICA has been also used in medical analysis (Tsumura *et al.* 2001) and cosmetic research (Shimizu *et al.* 2001) of skin images.

2.4 Face databases

Up to date, databases containing many faces have been created employing cameras and some of them contain colour images and even videos. Their main purpose has been to provide material to test and develop face recognition and detection algorithms. If the illumination changes are taken into consideration in these images, it is typically caused by variation in illumination direction, camera viewpoint or different white balancing light sources. Table 1 summarizes the properties of face databases some of which can be even downloaded from the WWW (see links at http://www.ee.oulu.fi/research/imag/color/). Also Yang and Ahuja (2001) and Gong *et al.* (2000) provide information on some of these databases. The physical basis of image formation data like camera responses are not considered with these databases and in some cases not even the type of the camera is mentioned. The illumination conditions are not reported and in some cases illumination changes are in fact caused by changes in illumination direction. Some databases do contain videos, for example from TV,

but they are not taken in real, drastic conditions.

Table 1. Face databases.

Face database	Number of persons	Images	Variables	Other data related to face
MIT (Turk & Pentland 1991)	16 men	27 images per person grey images	1. illumination direction 2. head tilt (orientation) 3. scale	a video sequence of person moving behind a plant
Shimon Edelman's	28 persons	minimum: 60 images per person grey images	1. horizontal illumination level 2. viewpoint 3. face expressions (3)	
CMU test images for face detection	3 datasets; 2 test sets not mentioned	grey images	1. frontal and profile views 2. different backgrounds	ground truths
University of Stirling	not mentioned	1591 colour and grey images	1. illumination (not defined) 2. expression 3. different views and poses	
M2VTS (Pigeon & Vandendrope 1997)	37 and 295	185 and 295 colour images	1. rotation 2. expressions 3. glasses on / off	four video sequences per person (295, head rotation under controlled lighting) and speech data
Yale (Belhumeur *et al.* 1997, Georghiades *et al.* 2001)	two databases, 15 and 10	165 grey images and 5850 grey images	1. facial expressions 2. glasses on / off 3. lighting direction and level	

Table 1. Face databases (continued).

Face database	Number of persons	Images	Variables	Other data related to face
UMIST (Graham & Allinson 1998)	together 20 men and women	564 grey images	1. different poses from profile to frontal view	
Purdue AR University (Martinez & Benavente 1998)	126, 70 men and 56 women	over 4000 colour images frontal view	1. facial expressions 2. occlusions 3. illumination: some images with different direction of yellowish light	
Goudail *et al.* (1996)	116	11600 frontal, grey images	1. pose	two 30 s moving head videos
AT & T (Olivetti) (Samaria & Harter 1994)	40	400 grey images	1. time 2. lighting level 3. facial expressions 4. glasses on / off	
University of Bern	30	450 grey images	1. head position 2. size 3. contrast	
FERET (Phillips *et al.* 2000)	not mentioned	14051 grey images	1. different poses from profiles to frontal view 2. different lighting level 3. facial expressions	ground truths

Table 1. Face databases (continued).

Face database	Number of persons	Images	Variables	Other data related to face
Kodak data set (Loui *et al.* 1998)	not mentioned	colour	1. size 2. pose 3. illumination between images (near white balanced ones)	videos (no big skin tone changes)
The Japanese Female Facial Expression (JAFFE) Database	10 women	213 grey images	1. facial expressions	emotion ratings
PEIPA (Pilot European Image Processing Archive)	two datasets	over 750 colour and grey	1. pose 2. contrast	
Harvard (Hallinan 1995)	10	not mentioned grey	1. illumination direction	
Usenix face dataset	not mentioned	5592	1. variable viewing conditions	
NISTS Special Database 18 (Mugshot Identification Database)	1573: 1495 men and 78 women	3248 grey images	1. poses: frontal and profile 2. size	

2.5 Studies of skin colours at different spaces

Because of increasing interest in faces, there have been studies on behaviour of skin chromaticities at different colour spaces. Many studies have indicated that the skin tones differ mainly in their intensity value while they are very similar in chrominance coordinates, see for example Graf *et al.* (1996), Yang and Waibel (1996), Graf *et al.* (1995), and Hunke and Waibel (1994). Terrillon *et al.* (2000) evaluated both different chrominance spaces and skin colour distribution models. They use a single Gaussian and Gaussian mixtures for modelling skin chromaticity distributions in nine colour spaces (TSL, NCC rgb, CIE xy, CIE SH, HSV, YIQ, YES, CIE Luv and CIE Lab). (For other than CIE colour spaces (Wyszecki & Stiles 2000), see Appendix 1). The images used in the evaluation were taken under slowly

varying illumination conditions under one camera or downloaded from the Internet. This most probably means that their study considered only skin colours obtained under white balanced or near white balanced conditions. According to their research, for a single Gaussian model the best results were obtained in illumination normalized colour spaces, whereas the use of Gaussian mixture models improved results with those colour spaces which do not use illumination normalization. The use of Gaussian mixture in an illumination normalized colour space produced comparable results to a single Gaussian model. They found that skin colour distribution in a space with no illumination normalization is complex shaped. The normalization produced distributions which were simpler to model, confined and more efficient for skin colour segmentation. An interesting observation was made on the behaviour of HSV space: the saturation S is sensitive to skin colour and it took almost all values for a limited hue H range. An illumination normalized colour space, TSL, was developed and then produced better performance. In their paper, they also presented a technique for calculating the threshold based on true positives and true negatives. Later, Terrillon *et al.* (2001) found that NCC rgb and CIE xy were most efficient for skin segmentation and these spaces produced the smallest area for skin chromaticities. They also tested portability of colour spaces between two cameras and concluded that the most portable was CIE xy and then NCC rgb. These two spaces were confirmed again to be best fitting for a single Gaussian colour model and most effective for face detection. NCC rgb had the highest correct face detection rate and correct nonface rejection rate.

Zarit *et al.* (1999) compare five colour spaces for classification of skin pixels in a colour histogram based applications. The colour spaces were CIE Lab, Fleck HS, HSV, Normalized RGB and YCrCb. The colour histogram based methods were based on a look-up table and Bayesian decision theory. Most of the images in their study were downloaded from the Internet, which means that the images do most probably contain very much shifting of chromaticities of skin tones. They found that for the look-up table method, the HS-spaces performed best while the CIE Lab and YCbCr were poorer. For Bayesian decision based classification, the choice of colour space did not matter but the maximum likelihood method produced better results than the maximum a posteriori method.

Three colour spaces, RGB, YUV and HSV, were evaluated for PCA based face recognition by Torres *et al.* (1999). According to them, RGB and luminance Y produced equal recognition rates, but better performance was obtained with SV components and YUV space. However, the skin appearance did not have very many colour shifts between the test image and found match image, and in all images shown faces and other skin objects seem to have skin tone or near skin tones colour appearance.

However, these studies have not considered so much colour shifts from skin tones because they do not address clearly real illumination changes. They do not specify under which camera white balancing and prevailing illumination conditions the images were taken, although this might be difficult for images downloaded from the Internet. It is therefore necessary to make a study about the behaviour of skin colours under defined camera white balancing and prevailing illumination conditions for different colour spaces.

2.6 Colour based detection, localization and tracking of skin

The skin colour is often used as a cue for detecting, localization and tracking targets containing skin, like faces and hands in an image. It is often not enough to separate skin objects from non-skin objects like wood, which can appear to be skin coloured. Therefore, skin is often combined with other cues like motion, texture and edge features, but in this section only the handling of colour is overviewed.

The goal is to divide the pixels of the image into skin coloured and non-skin coloured ones. The simplest methods define skin colour to have a certain range or values in some coordinates of a colour space. This can easily be implemented as a look-up table or as threshold values as in Chai and Ngan (1998). Dai and Nakano (1996) enhanced orange-coloured parts in YIQ space by selecting only a certain range of the I component. Hidai *et al.* (2000) defined an "ideal skin colour" by an average of precaptured face images, and based on the closeness of image pixels to this point they defined skin and non-skin pixels. Additionally, histogram equalization was made to increase robustness against brightness fluctuations. The second approach is to assume that the skin colours have different probability to occur and these probabilities follow a certain distribution which can be learned. Common features for these approaches are thresholds and tunable parameters; also the use of chromaticity coordinates is typical. The amount of skin pixels used for these off-line probability calculations varies greatly in the literature. Hsu *et al.* (2002) suggested colour correction before skin detection in YCbCr space. The colour correction was a version of the white patch method in which transformation coefficients are calculated from the mean of the highest 5 % luminance pixels if their amount exceeds a fixed threshold and the mean is not a skin tone value. However, their correction algorithm does not take into account saturated channels or the possibility of high valued pixels belonging to chromatic colour. After the correction, a nonlinear transformation was applied to chromatic data to obtain a better fit for the elliptical skin colour model. The detection algorithm was tested with quite moderate illumination change and the most demanding cases have a simple, white background.

The selection of threshold(s) has also been made in various ways to exclude those skin colours which occur too rarely. Comaniciu and Ramesh (2000) use a 1D skin colour distribution with mean shift to track faces (see Appendix 3). The object probability distribution was obtained off-line from an image or images taken in an office room. Although they mentioned that images were taken at different times (morning, afternoon and night) it was not clear how big the skin colour changes were. Generally, their test of mean shift tracking seems to be made under quite stable illumination conditions. Schiele and Waibel (1995) have made a face tracker based on only skin colour. They use a probability distribution to intensify the skin coloured region. Although they mention a colour map for most of the possible face-colours, they do not show or specify the chromaticity changes. Not all distributions are calculated off-line; for example Saxe and Foulds (1996) have suggested an on-line iterative method in which after user-initialization, the histogram of the selected area is compared to other histograms of patches.

The common parametric methods are based on Gaussians: unimodal Gaussian density function (Cai & Goshtasby 1999, Kim *et al.* 1998, Yang & Ahuja 1998) or multimodal Gaussian mixtures (Jebara & Pentland 1997, Jebara *et al.* 1998, Yang & Ahuja 1998). The parameters of the former can be estimated using maximum likelihood (Cai & Goshtasby 1999, Kim *et al.* 1998, Yang & Ahuja 1998) whereas the estimation for the latter requires

an Expectation-Maximization (EM) algorithm (Jebara & Pentland 1997, Jebara *et al.* 1998, Yang & Ahuja 1998). An output image which contains a skin probability has also been presented for face detection: Menser and Müller (1999) applied PCA on skin tone probability images obtained from a 2D Gaussian colour model. However, an interesting study has shown that histogram models provide better accuracy and lower computational cost than mixture models for skin detection (Jones & Rehg 2002). In addition, according to Yang & Ahuja (2001) single Gaussian distribution may detect less well the skin regions than a mixture of Gaussians. Additional assumptions, like an homogeneous intensity field over the object, have been made to separate more effectively skin and non-skin objects which have similar chromaticities (Abdel-Mottaleb & Elgammal 1999). Skin colour distributions have been learned also by neural network based approaches. Karlekar and Desai (2000) used a multilayer perceptron to learn skin colour distribution and classify pixels into skin-tone and non-skin tones. A Self Organizing Map (SOM) for labelling skin tones was used by Piirainen *et al.* (2000). It seems that all these different approaches work only in very well behaving illumination conditions; at least they seem to be designed for stable illumination conditions due to static models.

Adaptive approaches have also be suggested in order to cope with changing conditions. One way is to define a range of possible skin colours in which a finer model is found. Sahbi and Boujemaa (2000) collect a coarse skin colour model using neural networks from "a very large population ethnicity" which is used for coarse level skin detection. Later, the areas found are subjected to Gaussian colour modelling for relevant and noisy skin points and the parameters of the models are evaluated using a fuzzy clustering approach. They also assume that skin objects have a homogeneous local colour distribution. Sigal *et al.* (2000) adapted the skin colour histogram using a second order Markov model and feedback from the current segmentation results. They initialized tracking using the model suggested by Jones and Rehg (1999) for Internet images. Bergasa *et al.* (2000) presented a Gaussian skin colour model which is both unsupervised (prototype) and adaptive. They use a prototype cluster for representing human skin and the colour cluster which is closest to the prototype is considered to be skin. However, this limits usability of their approach to quite static illumination conditions. The adaptation of the model is done using a linear combination of previous model parameters. Cho *et al.* (2001) also used a predefined area for HSV skin colours in which a finer area is selected by adjusting several threshold values. They did not consider skin tone shifts because the thresholds for the hue component were fixed. Background areas were eliminated by assuming that their area is small compared to skin regions. Also a cluster analysis was performed to separate dominant background colour vectors from skin coloured ones. The skin coloured vectors were defined to be those which were nearest to predefined values. Approaches with user initialization have also been proposed.

Rasmussen and Hager (1997) have developed a tracking method in which the user gives an initialization region which is subjected to PCA to parametrize an ellipsoidal model. The ellipsoidal model assumes that the object colours can be confined by a simple, point-symmetric cluster. Their tracking method uses a fixed tracking window and based on the target found, the model is once again updated with PCA. However, their targets do not seem to contain any chroma shifts. Tsapatsoulis *et al.* (2001) combine skin colours and shape to template matching. They use an adaptive 2D Gaussian model whose parameters are re-estimated based on the current image. The pixels classified as skin were used for re-estimation of the Gaussian mean value. Schuster (1994) use two colour models: an ellipsoid

model and a mixture density model using RGB values. The mixture density model is obtained as a weighted sum of colour density functions which describe the distribution of colour values. Based on the localized target, colour model parameters are calculated and used for prediction of the parameters in the next frame. He also used a global colour model which contains a priori knowledge about parameters. Shape information was used to make sure that the pixels used for adapting both colour models were part of the object. Yang *et al.* (1998) suggest adapting a Gaussian model using maximum likelihood criteria by modelling it as a combination of the previous Gaussian distributions. Also in this case no big changes in skin colour were shown.

For adaptive tracking, two different spatial constraints have been introduced for selecting the pixel for refreshing the skin colour model. Raja *et al.* (1998) (later also in McKenna *et al.* (1999)) suggested adapting a Gaussian mixture model by a small area inside the localization. The Gaussian mixture model approximates the multi-modal distribution of the object's colours by using a number of suitably weighted Gaussians. They also use a normalized log-likelihood measure to prevent adaptation under tracker failure which seems to be caused by a shift in hue. Another spatial constraint was presented by Yoo and Oh (1999) who used histogram backprojection for face tracking. The purpose of histogram backprojection is to form a greyscale image in which the grey value shows the probability of a colour shade belonging to the object. It is assumed that the blob of high values in the image indicates the presence of the object. The face was assumed to be an ellipse and the pixels inside the located face ellipse were used to update the skin histogram. Also transductive learning has been suggested for skin tracking (Wu & Huang 2000) for a linear subspace of a combination of HSV and RGB spaces. The goal is to transduce the colour classifier so that it works well in the changed conditions. Once again, the main illumination variability seems to be caused by intensity changes.

However, the images and videos used for evaluation of these algorithms so far do not contain very many chromaticity shifts nor a nonuniform illumination colour field. The basic assumption of many methods seems to be that the illumination colour does not vary significantly due to restrictions built in the algorithms. More often the change is in the intensity (due to shadowing for example) or image geometry. It might be that a different choice of colour space would improve results, as was demonstrated by Terrillon *et al.* (2001). An exemption to this is the work done by Störring *et al.* (2001) and Störring *et al.* (1999). They consider skin colour under an illumination colour temperature range of 1500 K-25000 K with one camera calibration condition. Störring *et al.* (1999) named the area of all possible skin chromaticities under the illumination range as a skin locus because the chromaticities followed a Planckian locus. Störring *et al.* (2001) extended the work for mixed illumination (for example cases when there are two light sources causing a nonuniform illumination field over the skin). They concluded that the results for the body reflection chromaticities are the same as in the cases of a single light source. In both papers, they compared the average measured chromaticities to the modelled chromaticity area and found a good match with actual spectral power distributions. Before then also Matas *et al.* (1994) have suggested the use of chromaticity constraints. Unfortunately, their publications have been deprived of details, so further evaluation of their results and constraints is difficult. Another interesting piece of research related to changing illumination conditions was made by Debevec *et al.* (2000) who present a method to acquire the reflectance field of a human face. They use their measurements to render the face under arbitrary illumination

conditions.

Table 2 summarises some colour spaces used for pixel labelling for face based approaches. The most popular approach seems to be NCC rgb.

Table 2. Colour spaces for pixel labelling.

Colour space	Yang & Ahuja 2001	Other works
	Authors	Author
RGB	Jebara & Pentland 1997, Jebara *et al.* 1998, Satoh *et al.* 1999,	Rasmussen & Hager, Yang *et al.* 1998
normalized RGB or NCC rgb	Crowley & Bedrune 1994, Crowley & Berard 1997, Kim *et al.* 1998, Miyake *et al.* 1990, Oliver *et al.* 1997, Qian *et al.* 1998, Starner & Pentland 1996, Sun *et al.* 1998, Yang *et al.* 1998, Yang & Waibel 1996	Bergasa *et al.* 2000, Sahbi & Boujemaa 2000, Schiele & Waibel 1995
HS-based	Kjeldsen & Kender 1996, Saxe & Foulds 1996, Sobottka & Pitas 1996a, Sobottka & Pitas 1996b	Cho *et al.* 2001, Yang *et al.* 1998
YCrCb	Chai & Ngan 1998, Wang & Chang 1997	Hsu *et al.* 2002, Karlekar & Desai 2000, Luo & Eleftheriadis 2000, Menser & Müller 1999
YIQ	Dai & Nakono 1995, Dai & Nakono 1996	
YES	Saber & Tekalp 1998	
CIE XYZ	Chen *et al.* 1995	
CIE LUV	Yang & Ahuja 1998	
ab		Kawato & Ohya 2000a, Kawato & Ohya 2000b
YUV		Abdel-Mottaleb & Elgammal 1999
Farnsworth's UCS		Wu *et al.* 1999

3 Colour image acquisition by a CCD camera

3.1 Overview

Colour signals are the light spectra either from the source or from the interaction between the illuminations' spectra and response properties of materials. CCD colour camera can be described as a filter which transforms continuous colour signals from the limited spectral area to three descriptors ("red", "green", and "blue") values of a limited range. In this sense, colour cameras resembles the human eye; they cannot directly measure the spectra of colour signals because the spectral accuracy is sacrificed for the spatial resolution (Fortner & Meyer 1997). Since the spectral data for a point is described with three values, it is only an approximation of the true, incoming colour signal spectra. Also because of this spectral data compression, colour samples with different reflectances can become metameric, which means, for example, that they appear as two different colours under a certain illumination whereas under a second illumination they cannot be discriminated (Wyszecki & Stiles 2000). According to Fortner and Meyer (1997), there are four reasons why the human eye has only three different cones: 1) there are a limited number of available visual pigments, 2) the increasing number of different cones decreases the light sensitivity of the visual system because a photon can be detected only once, 3) cones need space; and if more different cones are required to form a point, the area needed for seeing a point increases and therefore reduces resolution, and 4) more different cones would mean increasing already the enormous information flow to brain. Cameras are usually monochromatic or colour. There do exist imaging spectrographs to capture more accurately spectral data, but for them, the image forming takes a much longer time due to decreased light sensitivity. This makes them unsuitable for real-time operations and susceptible to environmental changes. Only colour cameras are considered in this thesis. It is important to note that sensor sensitivities vary between colour cameras which makes the descriptors camera dependent. In addition, there are two types of CCD colour cameras: 1CCD and 3CCD colour cameras, depending on the number of CCD elements. The 3CCD cameras have separate CCD detectors for each colour channel, whereas in 1CCD cameras the colours for the output channels are approximated using filters covering the detector. The filters have either stripe or mosaic

layout over the detector and they can produce directly the RGB signals or other colours like cyan, yellow, magenta or white (no colour filter) (Holst 1998). These signals are interpolated to produce the three output colour channels and in the case of filters other than RGB, the channels are converted to RGB colour space. An image taken by a 1CCD camera has poorer spatial resolution and colour reproduction quality than the one taken with a 3CCD camera because of the colour interpolation in 1CCD cameras (Klette *et al.* 1998). 1CCD cameras are susceptible to colour Moire effects which cause colour deviation. On the other hand, the 3CCD cameras are more expensive and need more intense light.

Although in the modelling of colour image formation the main factors are illumination spectral power distribution (SPD), spectral sensitivities of the camera, and surface reflectances, there are many other factors which can have an essential effect: scene and acquisition geometry, surroundings, camera settings, camera type and other nonidealities of the camera. The output of the colour camera is often digitized RGB (Red, Green and Blue). Because the RGB space is redundant, it is often preferred to do further processing in another colour space.

3.2 Illuminants

An essential part of any vision system is electromagnetic radiation from which the range between 400 nm-700 nm, also referred to as visible light or simply light, is studied. In this thesis, the wave effects of light like interference are ignored. This is a reasonably good assumption (Ryer 1998) because the imaging systems used are incoherent and large scale. Light commonly encountered in a real environment can be separated to come from different source types: halogen / tungsten sources (such as incandescent lamps and other Planckian type radiators) and light at sunset or sunrise, fluorescent tubes, and daylight (sun and sky), daylight simulators. A few selected examples of each of these groups are visualized in Fig. 2: the SPD of the Planckian type radiators is smoothest, whereas the fluorescent SPD can be very spiky.

For normal, everyday lighting purposes, the overall impression of the light can be characterized using three classes defined by DIN 5035 (according to (Philips)): warm white (<3000 K), neutral white (3000-5000 K) or cool daylight (>5000 K). However, scientific and industrial applications need more accurate information and designing of the lighting.

To describe the illumination more accurately, colour temperature is used to relate the real illumination to the ideal Planckian radiator and to give an impression of redness (low colour temperature) or blueness (high colour temperature) of the illumination colour. Measurement of the colour temperature can be done relatively easily, quickly and inexpensively with a hand-held instrument (see for example (Broncolor)). It is very often used by professionals for many imaging and machine vision applications for investigating the illumination and its uniformity. Planckian SPD is smooth as shown in Fig. 2c and provides good approximation for tungsten / halogen lamps and sunrise / sunset lighting (Hunt 1987). If the chromaticities of an illumination, like fluorescent and daylight, do not have exact correspondence with those of any blackbody radiator, then a term called the correlated colour temperature is used to show the closest match. The details of procedures obtaining the correlated colour temperature are presented in Wyszecki and Stiles (2000). Later in this thesis,

only the term colour temperature is used assuming that the readers now recognize the difference. The term colour temperature should be used cautiously with fluorescent lamps, and also with accurate scientific calculations, the Planckian approximation of fluorescent lighting is generally not recommend because it can cause severe errors (Holst 1998).
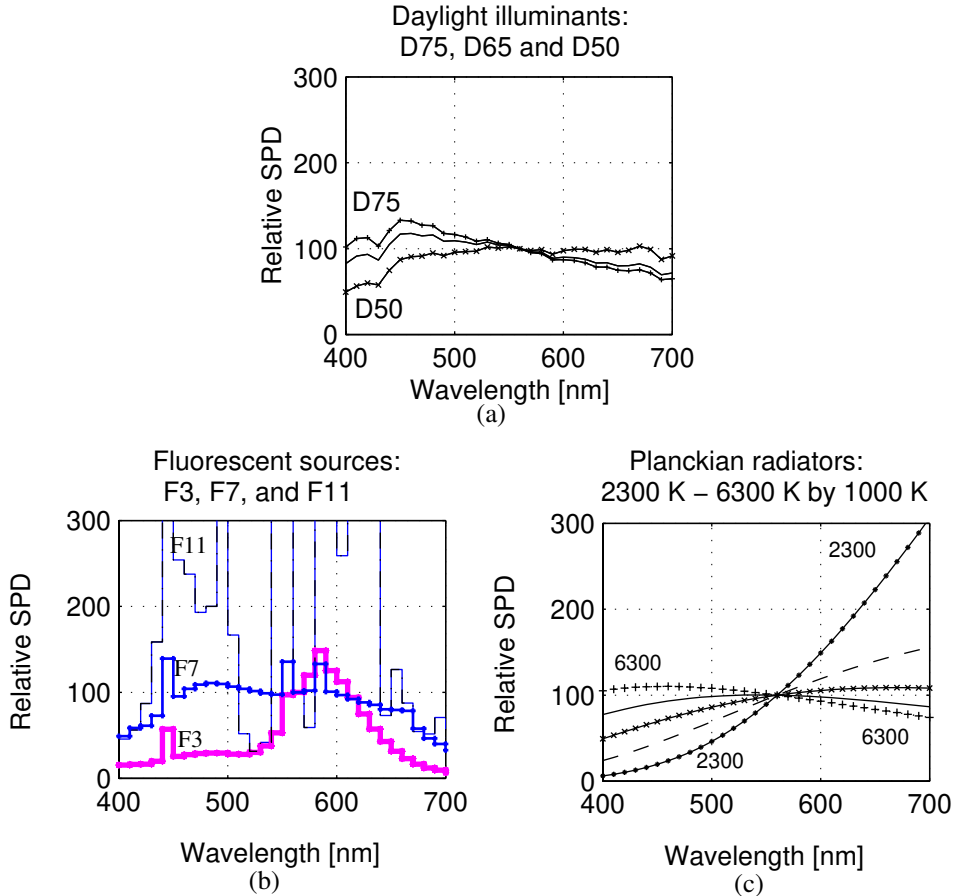
Fig. 2. Examples of different SPDs: (a) CIE standard daylight spectra (Hunt 1987), (b) CIE representative distributions for fluorescent lamps (Hunt 1987), and (c) calculated Planckian radiator spectra (Wyszecki & Stiles 2000). Note: SPD of F11 and of Planckian 2300 K are not shown in their full range.

To obtain better SPD modelling for fluorescent tubes and daylight, the CIE proposes special functions for modelling the daylight SPDs and specific SPD distribution for representing the fluorescent illuminants (Wyszecki & Stiles 2000, Hunt 1987). The colour temperature is used to exclusively define the daylight calculated via CIE daylight functions (Wyszecki & Stiles 2000, Hunt 1987). The daylight SPD for 5000 K (D50), 6500 K (D65) and 7500 K (D75) are presented in Fig. 2a at 400 nm from the lowest curve to highest one, respectively. Fluorescent lamps can be categorized to three different groups (Hunt 1987): normal, broad-band and three-band. A typical lamp in the normal group has high efficiency

but reddish colours are not rendered well (Hunt 1987). Improvements in rendering capability are achieved at the cost of decreasing efficiency; broad-band lamps have the best colour rendering among the fluorescent lamps but the lowest efficiency. In addition, three-band lamps can increase saturation of colours and therefore distort the appearance of colours (Hunt 1987). Fig. 2b shows examples of normal (F3), broadband(F7) and three-band (F11) fluorescent SPDs. The colour rendering index is especially used with fluorescent illuminants (see for example (Philips)). For example, the CIE general Colour Rendering Index is used to compare chromaticities of eight Munsell colours rendered under the light source and the reference source with the same colour temperature (Wyszecki & Stiles 2000). The reference source is Planckian if the colour temperature of the test source is under 5000 K; otherwise it is daylight (Hunt 1987). CIE Publication No. 13.2 (CIE 1974) provides more details on the method. Colour rendering issues are beyond the scope of this thesis.

The most accurate information on a real illumination SPD can be obtained by direct measurement i.e. with a spectroradiometer. The obtained results are rarely useful in general, and they are valid only for the measurement spot at the measurement time. In addition, they usually need a more expensive instrument, a spectroradiometer (like Minolta (1996)), and more time and effort than a plain colour temperature measure. Due to these reasons, and because the real SPD is not very often needed in applications, it is rarely used in practise. The advantage of actual measurement is of course valid data, for example because of changes caused by lamp aging (DeCusatis 1998). For many imaging and colour appearance applications, the SPDs are normalized with respect to some criteria. The normalized SPDs are preferred according to Wyszecki and Stiles (2000). The usual normalization (also recommend by the CIE) divides all SPD wavelength values by the value which is at the wavelength 560 nm and then multiplied by a constant factor of 100:

$$I \; = \; 100 \cdot \frac{I_{original}}{I_{original}(560)}, \tag{1}$$

where I = normalized SPD of the illuminant
$I_{original}$ = SPD of the illuminant.
Nevertheless, there are normalization methods like power normalization (Romero *et al.* 1997), also called Euclidean rule normalization:

$$I \; = \; 100 \cdot \frac{I_{original}}{\sqrt{\sum_{\lambda} I_{original}^2}}, \tag{2}$$

in which the normalization coefficient is the inverse of the total area of the SPD.

In the real world, the illumination is often a mixture between two or more light sources. Although electromagnetic radiation is a vector function with direction, normalization of the SPDs makes it a scalar function. Therefore, the normalized combination of N scalars can be obtained as a weighted sum:

$$I_{new} = \sum_{j=1}^{N} w_j \cdot I_j, \tag{3}$$

where j = the certain illuminant shining on the scene,

I = illuminant SPD, and

w = the degree (or weight) to which the illuminant is effecting the scene.

The sum of the weights is set equal to one and therefore the equation provides normalized mixed illumination SPD.

3.2.1 Responses of the human eye and a colour camera

The human vision system does not have the same response functions as most of colour cameras, as shown in Fig. 3. Therefore it is not a surprise that colours reproduced by the camera differ from those produced by the human vision system (see also Parkkinen & Jääskeläinen 1989). There can be instances where two colours differentiable to a human are not so for a colour camera and vice versa. Generally, the human vision system's colour gamut (area of perceivable colours) is larger than that of a colour camera or a colour scanner (see i.e. Foley *et al.* (1996)).
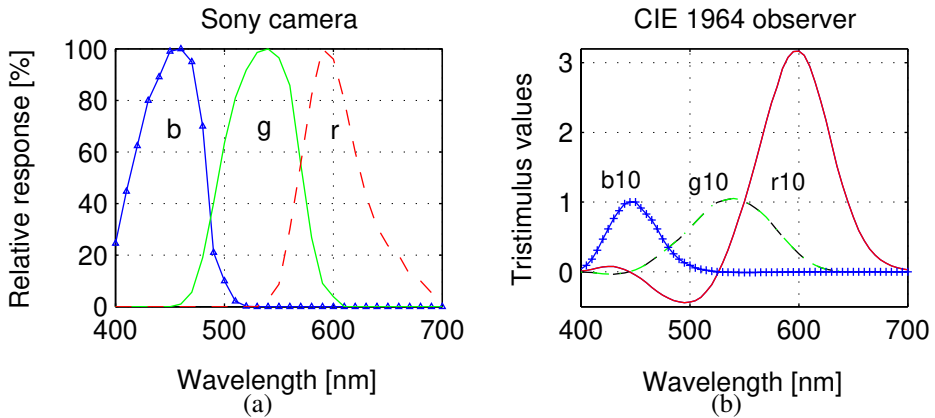

Fig. 3. Human and machine vision systems have different light responses: (a) relative spectral sensitivities of SONY DXC-755P 3CCD colour camera as given by the manufacturer and (b) RGB spectral sensitivities for 1964 Supplementary Observer (Wyszecki & Stiles 2000).

3.2.2 Non-idealities of real colour cameras

In theory, an ideal, analog colour camera has a linear response in an unlimited brightness range for each channel at infinite precision. In practice, all colour cameras have several restrictions which can have drastic effects on the colour appearance.

All real colour cameras have a limited dynamic range which bounds the possible brightness extent still expressible and differentiable to three indicators (RGB). The dynamic range is affected by a variety of factors; for example, integration time and the spectral content of the source (Holst 1998). The values outside this range are clipped and information on their real values is lost (Novak *et al.* 1992). There are two different ways in which clipping can happen: values in one or more channels can saturate to the maximum value ("overclipping") due to extreme brightness or they can go to zero values in channel(s) ("underclipping") due to very low brightness. This clipping can cause distortion in the hue appearance and is a common problem in many videos. One solution is that the user or an automatic gain controller adjusts the camera so that most of the wanted colours stay within the camera's dynamic range. For the user, this can be too cumbersome, demanding, and the quality of the result can vary. In addition, it is difficult to do this uniformly for every scene. The automatic gain controller can produce unstable results, too. For example, if there is a bright object in the scene, the gain adjustment can reduce significantly the range allocated to other scene points. Therefore there is a need for techniques which can tolerate imperfect data caused by clipping.

Another common factor is a nonlinear response of the camera which means that the output value transform is not independent of the input value. It can make the colour appearance dependent on the overall brightness of the channel at a pixel. The nonlinearity is not necessarily unwanted; it can be even a designed property of the camera (Vora *et al.* 1997) like gamma or pre-knee! The purpose of the gamma is to improve image quality reproduced in display devices (Holst 1998). Especially, the gamma in cameras is used to compensate for the nonlinear relationship between output light intensity and input voltage of cathode-ray tubes (Poynton 1996). As in many other machine vision and scientific approaches (Holst 1998), the gamma correction was set to off when possible in this thesis because it can distort colours. For increasing the dynamic range of the camera and protecting its CCD elements from very intense light (Sony 1989), a pre-knee circuit is added to change the linearity of the camera after a certain input signal value (Klette *et al.* 1998 and Holst 1998). In general, many cameras have linear response in the middle brightness range (Lomheim & Kalman 1992) whereas the nonlinearity is present in the both extremes of the brightness range. Its detection can be made with a grey patch or grey intensity scale chart. The linearity of the camera is determined from a graph of a channel which shows the camera's real output versus theoretical, linear output values. Fig. 4 illustrates different responses of a camera channel for different intensities.
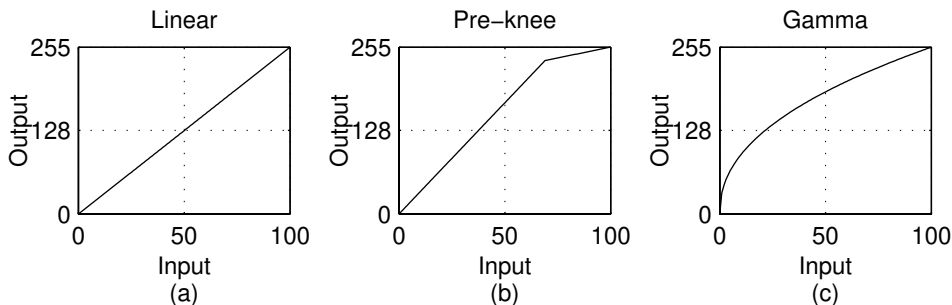
Fig. 4. The response of a camera for intensity in a channel: (a) linear response, (b) non-linear response caused by two different linear regions (pre-knee) and (c) non-linear response by gamma.

Like any other real world measurement devices, also the cameras suffer from noise. There are many noise types in cameras, like quantization noise and pattern noise (due to CCD's dark currents) (Holst 1998). Quantization noise is caused by imperfect analog-digital conversion. Because the values are converted to discrete, finite levels, the reduction of information is evident. Noise varies also between pixels (i.e. pattern noise) and it depends on the channel. The reason for different channel noise is that the CCD's sensitivity is a function of wavelengths. The wavelength area which produces blue values yields a significantly lower response than the one with red value production. In addition, many light sources have low SPD values at the blue end of the spectrum. In white balancing, the channels are scaled differently and therefore also the noise. In some cases, even when the camera's shutter is closed, the output is not a zero valued image. The nonzero values are called black level noise and they have been suggested to be subtracted from the images to get real response of the scene.

In this thesis, it has been assumed, based on experiments, that possible blooming, chromatic aberration or IR-response are not present, or at least, they can be ignored from the machine vision point of view (see Novak *et al.* (1992) for more details on these phenomena).

3.2.3 White balance or white calibration

In many cases, the colour appearances of objects in images are desired to be stable in spite of the prevailing scene illumination. To achieve this, the CCD camera is calibrated to disregard the effects of the illumination either by user-made or automatic adjustment of the camera's channel gains. The goal is usually that a "white" object, which is very often a plate, appears white in the image no matter what is the current illumination condition.

In white calibration or white balance, the gain adjustments are either user induced or automatic. There are three possible options which regulate how a user can influence gains (Sony 1991). The methods are presented shortly in the following sentences. First, the user may select the gain values from a predefined setting, i.e. the camera has a button for indoor

and outdoor conditions. Second, the user indicates to the camera that there is a white object on the scene and the camera itself makes adjustments to the gains. The degree to which the white object should cover the scene depends on the algorithm used by the camera. For example, some cameras make the balancing by calculating the gain values from the pixels with the highest intensities. However, it is the responsibility of the user to make sure that these pixels truly belong to the white object. In the third case, the user adjusts the gains manually and verifies the results. All these options have some serious drawbacks. The first is valid even for the white objects in very limited illumination conditions. The second and the third should be done every time the illumination changes. They are cumbersome and laborious for the user, especially if the illumination is changing constantly.

Calibration made in this way is valid only for the "white" object and for other achromatic colours if the camera is linear. The appearance of chromatic colours is still dependent on the illumination. It can vary in images taken under different illuminants even though the camera was properly calibrated for these illuminants. This is clearly demonstrated in Fig. 5. Although the camera was successfully white balanced at each light source, the skin chromaticities extracted from these images differ. Not even the definition of white is unique; for example, there can be differences between the whites of different objects (Poynton 1996). In this thesis, the cameras were mainly white balanced using one "white" plate as a white reference.

Some colour cameras do have automatic colour correction based on algorithms like the grey world (Buchsbaum 1980). Very often these algorithms make limiting and unrealistic assumptions and constraints on the work, and in many cases it is very easy to make them fail if assumptions (or constraints) are invalid. For example, the grey world algorithm assumes that the average colour is grey (the average values of channels are equal). Fig. 6 demonstrates a colour correction failure produced by the algorithm. But as long as the assumptions of these kinds of algorithms are satisfied, the results from many of these algorithms might be evaluated to be reasonable good by a human observer. There is no guarantee that this automatic correction will not lead to unstable appearances of colours or wrongly corrected colours.

One very constraining assumption of these gain control algorithms is that the illumination is uniform over the scene while in reality the opposite is often encountered. One exception is the Retinex algorithm, but it too has its own drawbacks (Brainard & Wandell 1986). Another underlying and restricting assumption is that the camera is adjusted so that there are no clipping or zero pixel values for chromatic colour. In any case, the colour correction or white balancing offered by these algorithms may not be enough to build reliable machine vision applications. For example, Funt *et al.* (1998) proved that the colour constancy algorithms do not yet provide reliable results for colour indexing presented by Swain and Ballard (1991). But on the other hand, in some cases humans find the colour correction to be satisfactory, especially if no accurate comparison between real scenes and images is made.
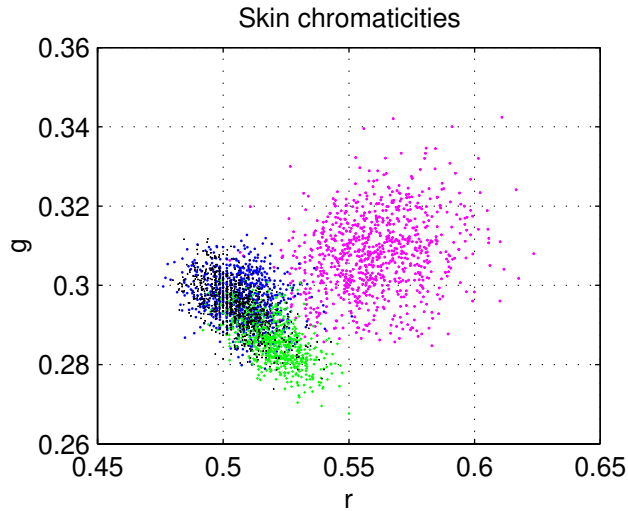
Fig. 5. Skin chromaticities from images taken under different white balancing illumination. The different colours are used to separate the chromaticities obtained from different white balancing cases. The r and g chromaticities shown in the axes are parameters of the normalized colour coordinates obtained by a conversion from skin RGB values.

Fig. 6. The grey world algorithm can fail: (a) original image, and (b) image corrected by the grey world algorithm. The channel average value is set to 100 by the user.

3.3 The RGB response of a camera

In theory, the output of a camera is characterized by three main factors: the spectral reflectance of the object at a point, the spectral power distribution of the prevailing illumination over the point and the spectral sensitivities (or responses) of the camera. The output is nor-

malized against a selected white object. The general equation for the output of a camera channel at a pixel is

$$V_i(x, y) = m_i \sum_{\lambda= 400}^{700} \eta_i(\lambda) \cdot L(\lambda, \Theta), \tag{4}$$

where V_i = the output signal of the ith camera channel,
i = blue, red or green channel,
x,y = pixel location in the image,
m = scaling coefficients,
η = spectral sensitivity or spectral response,
L = the radiance of the incoming light,
λ = wavelength, and
Θ = imaging geometry like photometric angles.

If the light entering to the camera has impinged on some material surface, then the output can be written as

$$V_i(x, y) = m_i \sum_{\lambda= 400}^{700} \eta_i(\lambda) \cdot I(\lambda, \Theta) \cdot R(\lambda, \Theta), \tag{5}$$

where I = spectral power distribution of the illumination, and
R = spectral reflectance of the material surface.

The scaling coefficient can be calculated using the following equation:

$$m_i = \sum_{\lambda= 400}^{700} \eta_i(\lambda) \cdot I_{ref}(\lambda, \Theta) \cdot R_{white}(\lambda, \Theta), \tag{6}$$

where the R_{white} = very often constant and its value is the maximum reflectance, and I_{ref} = the SPD of the illumination used in camera calibration.

Furthermore, according to the Dichromatic Reflection DR model (Shafer 1992), for many materials the reflectance can be divided into two components: interface ("specular") and body ("diffuse") parts. Both of these can be further divided into geometric terms K and the spectral part:

$$R(\lambda, \Theta) = R_{interface}(\lambda, \Theta) + R_{body}(\lambda, \Theta)=$$
$$K_{interface} \cdot r_{interface}(\lambda) + K_{body} \cdot r_{body}(\lambda) \tag{7}$$

The DR model can be used as a good approximation model for light reflection of those materials which are optically inhomogeneous, opaque, covered by an optically inactive surface and are either on a curved or planar surface (Shafer 1992).

After quantization, Eq. 4 is a discrete representation of a transform from a continuous, infinite but limited wavelength area to continuous, 3 dimensional value space. This causes data reduction and loss because the ability to discriminate two colour signals decreases.

From a human vision point of view, the RGB space produced by an ordinary camera is non-uniform and cannot produce all visible colours.

When the SPD of the illumination is the same as it was in the calculation of the scaling coefficient, then the camera is said to be white balanced to this illumination. If it is not the same, the equations can be still used, but normalization of the illumination should be made. Even normalization cannot prevent the problem related to modelling the phenomena encountered in the limited dynamic range or nonuniform illumination.

3.4 Colour spaces

Colour spaces usually either model the human vision system or describe device dependent colour appearances. Although there exist many different colour spaces for human vision, those standardized by the CIE (i.e. XYZ, CIE Lab and CIE Luv, see for example Wyszecki & Stiles 2000) have gained the greatest popularity. These colour spaces are device independent and should produce colour constancy, at least in principle. Among device dependent colour spaces are HSI, NCC rgbI and YIQ (see Appendix 1 for formulae). The different versions of HS-spaces (HSI, HSV, Fleck HS and HSB) are related to the human vision system; they describe the colours in a way that is intuitive to humans.

Usually the output from CCD element is expressed as RGB values or corresponding values. This can be understood as a basic colour space from which the values are converted to the other device colour spaces. The RGB values are redundant and intensity dependent. Therefore, in many device colour spaces the intensity is separated from the chrominances. Use of only chrominance values offers robustness against changes in illumination intensity both in the time and spatial domains. A disadvantage is the loss of information related to different intensity levels of the same chrominance; in other words, for example black, grey and white cannot be separated by using only chromaticity values. It is interesting to note while the intensity may be the most significant feature in segmentation (Ohta *et al.* 1980), it is also the most sensitive component to changes in practical imaging conditions.

The values of device colour spaces can be converted to the corresponding values of a human colour space. For example, this transformation can be made by first selecting representative samples and calculating the transform matrix from them or by using the samples to train a neural network. The transform can be non-linear. The inputs (i.e. RGB) do not necessary have to be a 3x3 matrix; their values can be also obtained using polynomials with different degrees of polynomial. However, the created transform function depends heavily on the illumination conditions under which it was made. Therefore, the transform to human colour space still does not solve the colour constancy problem but alleviates the device dependency problem.

3.5 Evaluation of camera performance

As mentioned earlier, humans might be able to discriminate colours which are indiscriminable for colour camera and vice versa. Those colours which are indiscriminable under one

condition but discriminable under another are called metameric colours. Metamerism can be used to evaluate the camera's ability to handle small colour difference measurements (Paper I).

For human vision, the metamerism of samples can be evaluated using metamerism indices or colour difference formulae. CIE 1976 Lab colour difference is widely used in industry (Pierce & Marcus 1994) and implemented as a Euclidean distance between CIE Lab parameters of two samples:

$$\Delta E_{ab} = \left(\left| L_1 - L_2 \right|^2 + \left| a_1 - a_2 \right|^2 + \left| b_1 - b_2 \right|^2 \right)^{1/2}, \tag{8}$$

where ΔE_{ab} = the CIE 1976 Lab colour difference,
L = lightness,
a = a coordinate indicating location of colour in a greenness-redness axis, and
b = b coordinate indicating location of colour in a blueness-yellowness axis.

For evaluating metamerism, the following general metamerism indices are often used: Bridgeman's index BMAN (Bridgeman & Hudson 1969 according to Choudhury & Chatterjee 1996),

$$BMAN = \left(\sum_{\lambda} \left(\rho_1(\lambda) + \rho_2(\lambda) \right)^2 \right)^{1/2}, \tag{9}$$

and that of Nimeroff *et al.* (N+Y) (Nimeroff & Yurow 1965):

$$(N+Y) = \left(\sum_{\lambda} \left((\bar{x}(\lambda))^2 + (\bar{y}(\lambda))^2 + (\bar{z}(\lambda))^2 \right) (\rho_1(\lambda) + \rho_2(\lambda))^2 \right)^{1/2}, \tag{10}$$

where $x(\lambda)$, $y(\lambda)$ and $z(\lambda)$ = the CIE colour-matching functions for the CIE 1964 Supplementary Standard Colorimetric Observer, and
$\rho(\lambda)$ = spectral reflectance of the sample.

It is interesting to note that these two metamerism indices do not include illumination information in the evaluation. For colour cameras, the Minkowski's distance formula can be used to evaluate the colour difference ΔE_{RGB} in the camera's RGB colour space (Novak & Shafer 1992):

$$\Delta E_{RGB} = \left(\left| R_1 - R_2 \right|^n + \left| G_1 - G_2 \right|^n + \left| B_1 - B_2 \right|^n \right)^{1/n}, \tag{11}$$

where n = power.

Typical values of n are 1 (the sum of absolute difference in values in each band or city block distance), 2 (Euclidean distance) and ∞ (chessboard distance). However, this formula induces a bias against bright colours but it is used because it is the best available.

For the simulated experiments, the sample spectra were obtained from a NCS colour block (NCS 1989) (Natural Colour System block with 1526 samples and measured by a Minolta CM-2002 spectrophotometer), the illuminants were A, D65, and F11, the camera selected was a Temet TVI camera (TVI 1995) with 2 options (8 and 12 bit). For human vision modelling, the CIE 1964 Supplementary Standard Colorimetric observer was used.

In the first experiments, the predictions for human vision are evaluated. Table 3 dis-

plays the number of samples within a certain distance range calculated from metrics presented in Eqs. 8-10. In total, there are 2327150 (1526x1525) possible sample pairs. A colour pair is defined as similar if its value for the general metameric index is in the range of 0-5, or for the CIE Lab difference formula in the range of 0-3. In many colourant industries, a sample pair with a CIE Lab difference in the range 0-1.5 is classified to be metameric (Choudhury & Chatterjee 1996 and Choudhury & Chatterjee 1992) and in the range 1.5-3 similar colours. As can be seen from Table 3, different evaluation methods produce different predictions on the amount of sample pairs for the difference range. This leads us to an obvious conclusion that there is disagreement between these methods on asserting metamerism on sample pairs. CIE Lab formula predicts that the amount varies with prevailing illumination over the samples. For the further study, the effects of metamerism due to illumination are investigated and therefore only the CIE Lab formula is utilized.

Table 3. The values of metameric indices for NCS samples.

Index	number of sample pairs			number of metameric samples
	0-1.5	1.5-3	3-5	
BMAN	3	47	254	303
N+Y	6	53	274	333
$\Delta E_{ab}(A)$	23	382	1426	405
$\Delta E_{ab}(D65)$	15	371	1448	386
$\Delta E_{ab}(F11)$	17	351	1340	368

The results in Table 4 prove that the predicted number of metameric samples is different under different illumination conditions even for human vision. Perfect colour constancy is therefore impossible even for human vision. This implies that for colour cameras, when comparing colour distribution taken under different conditions, precision is limited and robustness is needed.

The different discrimination capabilities of human and colour cameras are demonstrated in the second simulated experiments. The results are shown in Tables 5-7: some colour pairs which are predicted to be metameric or similar for human vision are not necessarily that for the TVI camera. In the Table 5, the sample pairs are arranged into subsets according to their calculated human colour space ΔE_{ab} values for the prevailing illuminant. For these subsets of sample pairs, their minimum (marked as min in the table), maximum (max) and average (avg) colour differences in the camera space are calculated for each illuminant case. The number of metameric pairs for the camera is shown in the rows called pairs. The subsets in the Tables 6 and 7 are constructed by using their colour difference in the camera space (ΔE_{RGB}). The minimum (min) and average (avg) difference values of these subsets

are computed in the human colour space. The amount of sample pairs in each column is shown in the rows marked as a number. At least for ideal, noiseless colour cameras the increase in the bit number will improve discrimination capabilities: the 12-bit option produced an increased discrimination capability for those colours which are strongly metameric for human vision.

Table 4. Number of metameric pairs for CIE Lab difference formulae (NCS samples).

Index	Range	$\Delta E_{ab}(A)$		$\Delta E_{ab}(D65)$		$\Delta E_{ab}(F11)$	
		0-1	1-1.5	0-1	1-1.5	0-1	1-1.5
$\Delta E_{ab}(A)$	0-1	3	0	0	2	1	1
	1-1.5	0	20	1	7	2	9
$\Delta E_{ab}(D65)$	0-1	0	1	1	0	0	0
	1-1.5	2	7	0	14	3	7
$\Delta E_{ab}(F11)$	0-1	1	2	0	3	4	0
	1-1.5	1	9	0	7	0	13

Table 5. Values of colour differences in the Temet TVI RGB camera space for pairs with $\Delta E_{ab} <= 3.0$.

Illuminant		8-bit TVI camera				12-bit TVI camera			
		Range of values of ΔE_{ab}				Range of values of ΔE_{ab}			
		0-1	1-1.5	1.5-2	2-3	0-1	1-1.5	1.5-2	2-3
A	min	2	0	1	1	28.8	10.1	15.8	11.7
	max	12.2	8.4	12.7	23.2	202.5	137.3	192.3	368.2
	avg	5.7	3.7	5.8	7.2	91.7	59.3	91.1	114.7
	pairs	0	1	0	0	0	0	0	0

Table 5. Values of colour differences in the Temet TVI RGB camera space for pairs with $\Delta E_{ab} <= 3.0$ (continued).

Illuminant		8-bit TVI camera				12-bit TVI camera			
		Range of values of ΔE_{ab}				Range of values of ΔE_{ab}			
		0-1	1-1.5	1.5-2	2-3	0-1	1-1.5	1.5-2	2-3
D65	min	2.4	1	1	1	38.6	10.2	11.7	15.3
	max	2.4	7.5	15.8	24.4	38.6	115.2	257.0	386.6
	avg	2.4	3.8	5.6	7.3	38.6	61.4	87.5	115.6
	pairs	0	0	0	0	0	0	0	0
F11	min	1.4	0	1	1	21.8	10.3	15.3	12.6
	max	4.1	8.4	12.0	21.4	59.1	121.6	190.4	341.1
	avg	2.8	4.1	4.9	7.2	45.2	62.4	77.9	115.1
	pairs	0	1	0	0	0	0	0	0

Table 6. Values of colour differences ΔE_{ab} for the similar colour pairs for the 8-bit TVI camera.

Index		Range of values of ΔE_{RGB}					
		0	0-1	1-2	2-3	3-4	4-5
$\Delta E_{ab}(A)$	min	1.1	1.1	0.7	1.0	1.2	1.1
	avg	3.9	3.9	8.9	12.6	18.4	16.7
	number	2	11	39	184	279	511
$\Delta E_{ab}(D65)$	min	5.5	1.2	1.7	0.9	1.2	1.4
	avg	5.5	2.7	3.6	5.0	5.7	6.8
	number	2	11	47	183	282	474
$\Delta E_{ab}(F11)$	min	1.3	1.5	0.7	1.0	0.8	0.8
	avg	1.3	2.7	3.1	4.7	5.7	6.7
	number	1	8	43	183	277	473

Table 7. Values of colour differences ΔE_{ab} for the similar colour pairs for the 12-bit TVI camera.

Index		Range of values of ΔE_{RGB}					
		0-10	10-20	20-30	30-40	40-50	50-70
$\Delta E_{ab}(A)$	min	-	1.1	0.7	1.1	1.0	1.0
	avg	-	2.1	3.0	4.3	4.8	6.1
	number	0	13	30	83	118	459
$\Delta E_{ab}(D65)$	min	5.5	1.2	1.2	0.9	1.2	1.2
	avg	5.5	2.4	3.4	4.3	4.7	6.1
	number	1	11	36	73	126	444
$\Delta E_{ab}(F11)$	min	-	1.1	0.7	1.6	1.0	0.8
	avg	-	2.2	3.1	3.8	5.0	5.9
	number	0	11	33	72	125	436

4 Acquisition of face images by a colour camera

4.1 Overview

Skin is an important object in many machine vision, computer graphics and display applications. Skin colours have been imagined both for professional and leisure purposes, and skin reflectances are sought for more illumination and device independent characterization. Reflectance data is useful for colour correction and simulating the skin colour appearances. Different skin tone appearances have been even simulated using the separate reflectances of skin colourants.

To study facial skin image formation and acquisition under different camera and illumination conditions, we have introduced a unique Physics-based Face Database for research purposes (Paper II). The database has been delivered to several world famous universities and companies (by 8.2.2002 together 43 full or partial copies). The database has been used for skin reflectance analysis and developing a colour correction method for imperfect data.

4.2 The Physics-based Face Database

The Physics-based Face Database is uniquely dedicated to colour issues related to facial skin appearance acquisition and formation under different prevailing illumination and white calibration conditions. It does not contain only face images but also spectral data related to image formation and is therefore called physics-based. The database contains 125 persons who can be roughly classified into three skin groups (Jones *et al.* 1992): dark or Negroid (8 persons), pale or Caucasian (101 persons), and yellowish or Asiatic (16 persons). The factors which make this database unique are: 1) the face images are taken under different illuminants with different camera white calibrations, thus making skin appearance vary both in intensity and in chromaticity, and 2) spectral data - SPDs of illuminants, reflectances of skin and spectral sensitivities of the camera - related to imaging are collected.

The data acquisition from the facial skin has two parts, caption of the face colour image and measurement of its reflectance. The setup for face imaging as shown in Fig. 7 was kept fixed for all subjects. In the image acquisition, the subject sits in front of a large diffuse 80 % grey screen and is approximately 1.5 m meter from a Sony DXC-755P 3CCD colour camera (Sony 1991). Above the camera, a Macbeth SpectraLight II Luminaire (Macbeth 1997) illuminates the scene. It provides four typical real world light sources: a daylight source 6500 K for modelling CIE D65, incandescent A for CIE A 2856 K, horizon sunlight for incandescent 2300 K, and fluorescent TL84 corresponding to fluorescent F11. They are referred later with the letters 'D', 'A', 'H', and 'TL84', respectively. A white plate was adjusted so that it can be reliably set in the centre position of the camera's view when the user needs to calibrate the camera. In the calibration, the camera adjusts its gains using only the brightest pixels. The colour temperature of the prevailing illumination was measured on the white plate by a Broncolor Colormeter (Broncolor). In addition, the SPDs of the light sources have been measured by a Minolta CS-1000 spectroradiometer (Minolta 1996). During imaging sessions, some adjustment were made so that the person's face appears approximately in the middle of the image. The persons were asked to keep his / her face in the same position, with the same facial expression, and to keep their eyes open though many persons find this difficult to do. The obtained images are 24-bit, colour, 428x569 pixels, and in raw bmp-format.
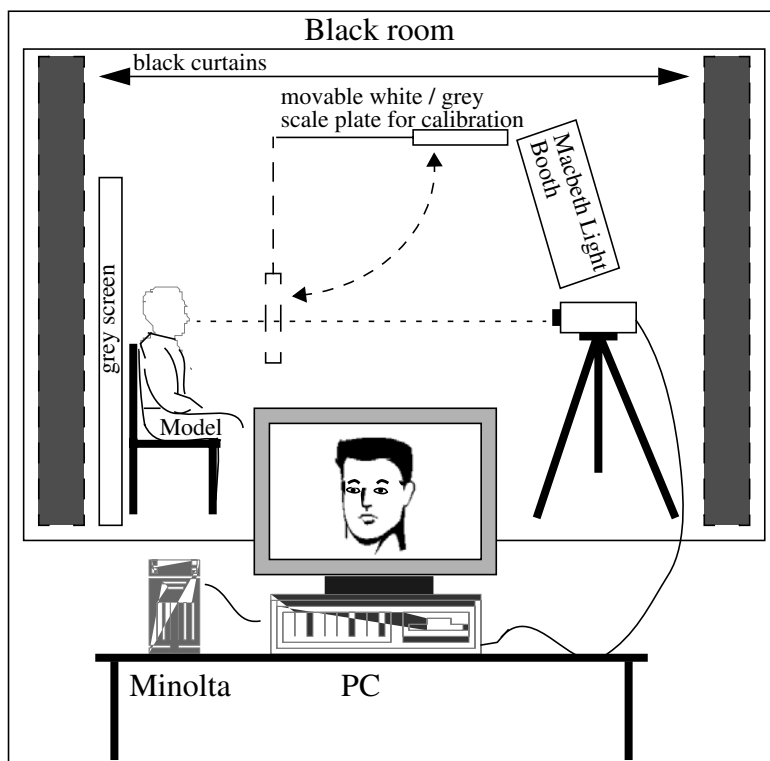
Fig. 7. Experimental setup for the Physics-based Face Database.

The general procedure for imaging with NI different illuminants and NC different camera calibrations is presented in Fig. 8 and it produces NIxNC images. NI and NC were chosen to be four, because SpectraLight had four different illumination options. The procedure was applied to all persons and therefore 16 images for each person were obtained. Additionally another series of 16 images was taken if the person wore spectacles. Each image is marked exclusively with the illumination reference to camera calibration and to the prevailing condition, and the person number. If the person also wore glasses, another identification mark was added. An example of the 16 image series of a person is shown in Fig. 9.

```
Imaging procedure of NI different illuminants

for illuminant_C = 1:1:NC
    1. Switch on the illuminant_C
    2. White balance the colour camera to this illuminant_C
    3. Keep these settings
    for illuminant = 1:1:NI
        3.1. Capture the face image under current illuminant
        3.2. Change to the next illuminant
    end
end
```

Fig. 8. The imaging procedure.

In Fig. 9, the light sources used in white balancing lie on the x axis and the prevailing illuminants are located in the y axis. Illuminants are ordered so that their colour temperature increases when moving from left to right and from top to bottom. The images which are taken under the calibration illumination (canonical conditions) lie on the diagonal axis. Even between them, there are visible colour differences. As can be seen from these images, although the skin colour may appear plausible for the human observer especially in the temporal comparisons, the calibration illumination has an effect on the skin colour appearance. The difference in skin colour appearance is not caused by improper camera white balance but the inability of the balance to eliminate the effect of the illumination. For nondiagonal images which are not taken under white balancing illumination, the skin colour change is even more drastic. As it was mentioned earlier, the higher colour temperature means more bluish content for the illumination whereas the lower one means more reddish content. The influence of this characteristic is evident in the image series: if the colour temperature of the prevailing illumination is higher than the one used in the camera calibration, the colours shift towards blue. In the opposite situation, the shift happens toward red.

Fig. 9. An example from Physics-based Face Database: 16 images of a face.

The measurements of skin reflectances were done with a hand-held, contact spectropho-tometer Minolta CM-2002 (Minolta 1991). For each person, there were measurements of diffusely reflected light (the surface or specular component excluded) on the skin at the wavelength range of 400 nm-700 nm by 10 nm steps. For a group of 20 persons, the skin was also measured with a specular component included option which takes into account also specularly reflected light at the same wavelength range. The measurement positions on the face were the forehead, left cheek, and right cheek, and each result was obtained as an average of three measurements. Fig. 10 displays the spectral measurement results for a person. The facial skin measured is located on a curved surface. The illuminants's SPD and

the camera's spectral sensitivities were obtained from their definitions (confirmed by spectroradiometric measurement) and manufacturers's specifications, respectively.
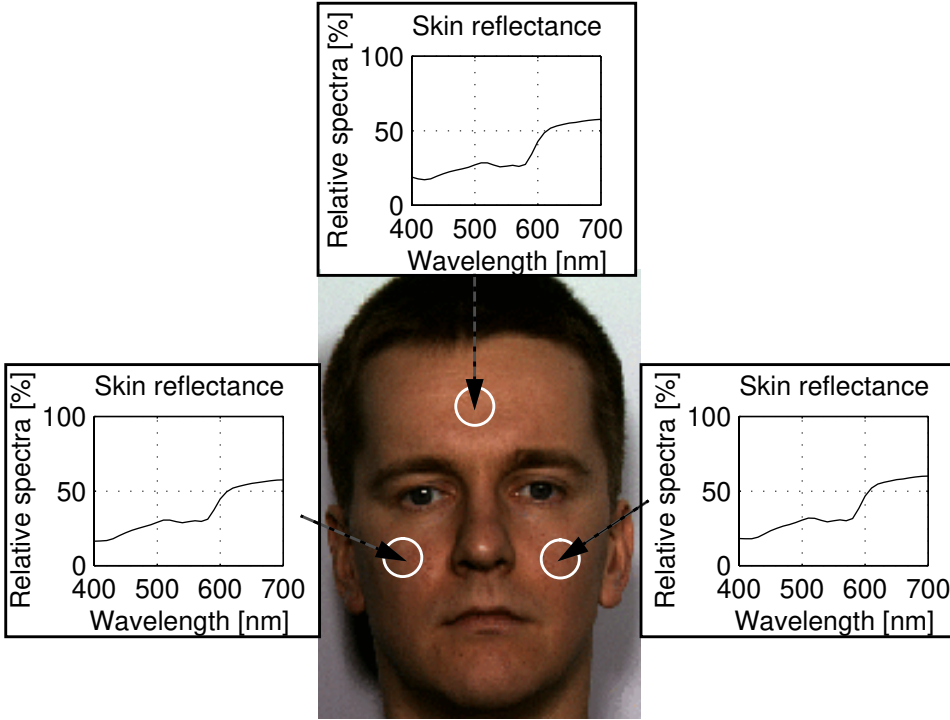

Fig. 10. Skin spectral reflectance measured at three points for a person.

## 4.3 Analysis of spectral characteristics of skin

Skin reflectances are examined with respect to their measurement position, values, and measurement options. The similarity of a person's spectral reflectance at different measurement points is evaluated using the following equation for measuring the overall goodness of the fit:

$$\Delta\delta_{avg(i,k)} = \frac{1}{N}\sum_{1}^{N}|\delta_i - \delta_k|, \tag{12}$$

where i, k = measurement position of the spectral (referred to 1,2,3) and $i \neq k$,
δ = spectral reflectance of skin, and
N = the number of wavelengths.

Fig. 11 shows results after applying Eq. 12 for the spectral reflectances. In 90 % of cases, all reflectance measurements of a face are very close to each other (Fig. 11d). In the rest

of the cases, the cheeks have very similar values (Fig. 11a) but the forehead areas produced slightly smaller values (Fig. 11b and c). Based on Fig. 11, the facial skin is quite uniformly coloured and opaque.
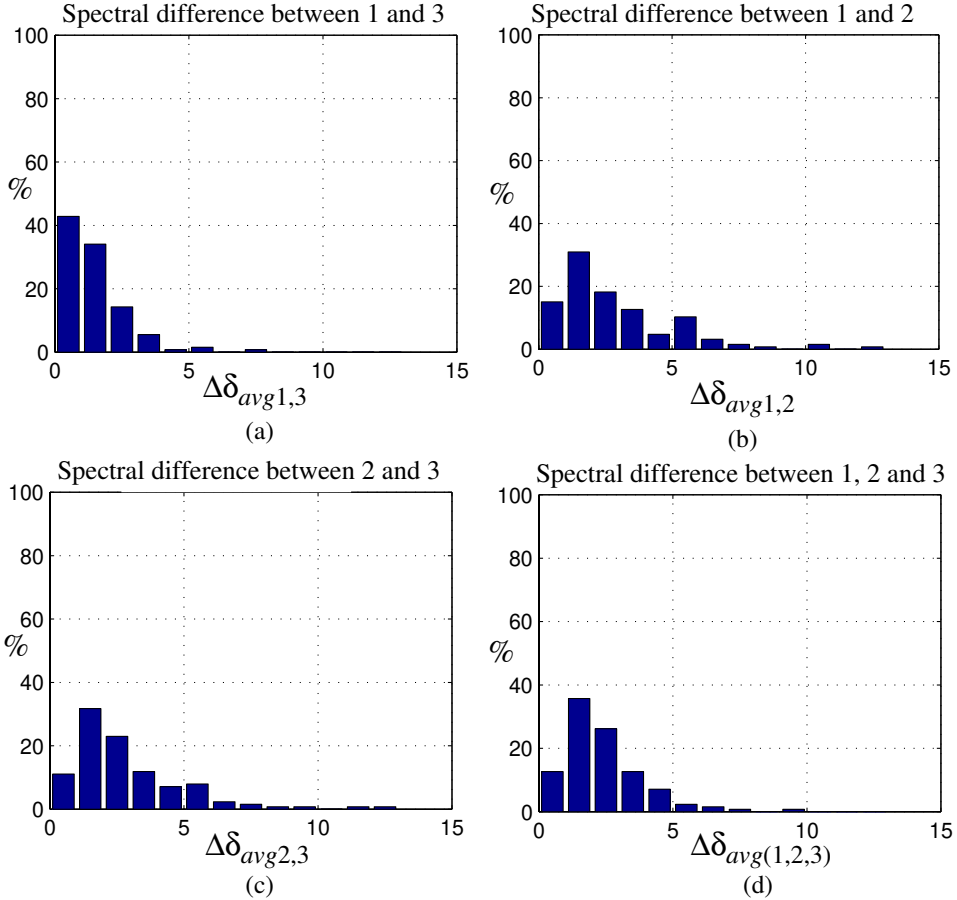
Fig. 11. Histograms of $\Delta\delta_{avgi,k}$ for all skin complexions: (a) for spectral reflectances of cheeks (left cheek = 1, right cheek = 3); (b) for spectral reflectances of a left cheek and a forehead (left cheek = 1, forehead = 2); (c) for a spectral reflectances of a right cheek and a forehead (right cheek = 3, forehead = 2); and (d) for all three spectral curves (left cheek = 1, right cheek = 3), (left cheek = 1, forehead = 2), and (right cheek = 3, forehead= 2).

The total spectral reflectance was measured with the specular component included (SCI option). The diffuse reflectance measured (SCE option) reflectance is slightly lower than that displayed in Fig. 12. The difference in wavelength is approximately 1 %-1.8 % on a scale of reflectance of 100 %. This suggests that skin in the face is mostly a matt finish.

Fig. 13 shows average and extreme skin spectra for different skin groups. The skin spectra between the groups are mainly separated by a bias. Therefore, assuming a reasonably linear camera, all skin chromaticities are very close to each other. Because the spectra are smooth, slowly varying and similar in shape, they can be modelled by a small number of

basis vectors. In addition, these spectral reflectance curves have a close resemblance to those presented by Wyzecki and Stiles (2000).
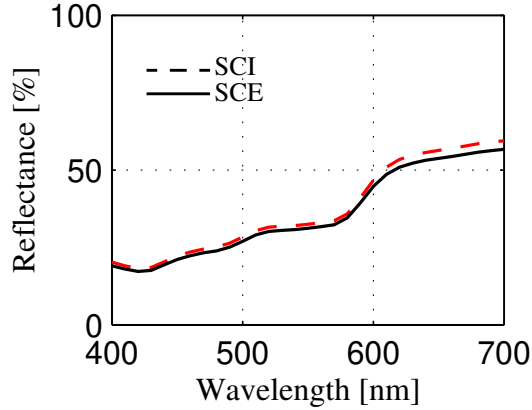

Fig. 12. Spectral reflectance for skin when diffusely reflected light (SCE), and diffusely and specularly reflected light (SCI) are measured.
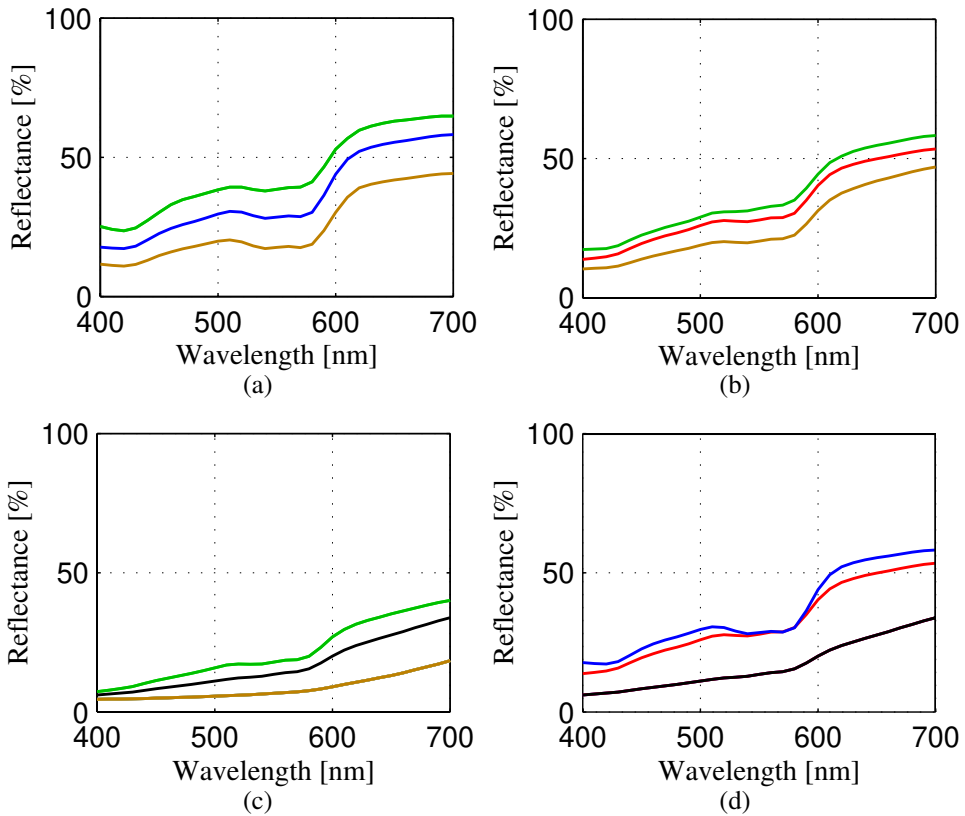

Fig. 13. Spectral reflectance curves (SCE): (a) Caucasian; (b) Asian; and (c) Negroid complexions; (d) average curves for each group.

4.4 Making overclipped facial images useful

The purpose of colour correction is to change the appearance of object colours in the current image to their appearance under another illumination and / or with another camera settings. Because of information loss present in most of the images a perfect correction is rarely possible. This loss can be caused by metamerism, in which colour appearance of objects is discriminable under one light source but not under another, or it can be caused by prevailing illumination which differs from the illumination used in white balancing of the camera. The later one is usually more severe since it can cause bigger colour shifts and clipping. Because clipping of one or more colour channel definitely means inadequate and unsure colour information, they are often disregarded in image analysis. There exists no simple transformation for correcting all these kind of information losses which cannot be always avoided especially with videos. Higher level knowledge and new methods are therefore needed.
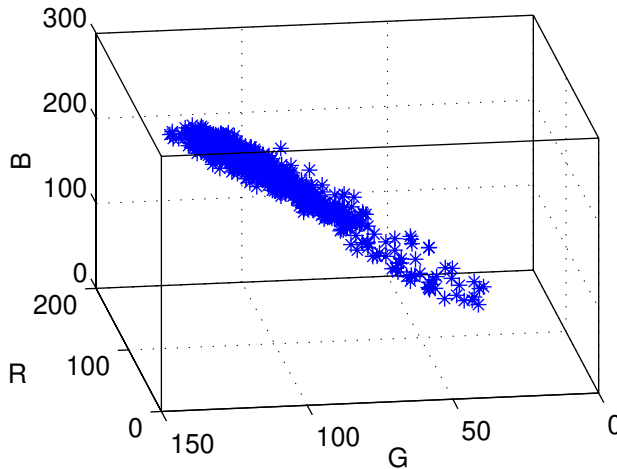
Fig. 14. RGB pixels of an AD image extracted from an unsaturated skin part (nose). AD means that the image is taken under daylight illuminant D when the white balancing of the camera was done under incandescent illuminant A.

Shafer (1992) has shown that there are cases when the colour of a pixel can be approximated by a linear combination of body and surface components. Furthermore, the clipped channel value or values can be calculated at least in theory using knowledge of the object's spectra, camera spectral response functions, SPD of the illumination and image geometry. Even without knowledge of these physical factors, it is still possible to estimate the pixel colour based on the body component if the object is matt and dielectric. The object must be dielectric because dielectric materials have a linear colour histogram for the body component between the minimum and maximum brightness. The colour histogram can be modelled simply by slopes or ratios between channels of unclipped pixels. Klinker (1993) demonstrated the usefulness of these colour histograms for recovering data of clipped pixels. However, the method proposed in Paper III is distinctive in two ways: first, it disregards surface components and assumes the skin to be matt, and second, a scaling is

implemented to preserve shading information. The skin colour can be approximated to be matt and this conclusion is reasonable, as shown by the spectral analysis in Section 4.4. In Fig. 14, the colour histogram of unclipped pixels is displayed and this further verifies that the matt assumption is quite appropriate: the RGB values seem to fall into a straight line (Klinker 1993). If the object is glossy, the RGB values of the object form two lines due to specular reflection (Klinker 1993). For a matt, dielectric object, the average ratios of the body colour can be acquired from unclipped pixels by

$$m_{gr} = mean\left(\frac{G}{R}\right), \;\; and \;\; m_{br} = mean\left(\frac{B}{R}\right), \tag{13}$$

where m = mean value,
gr = ratio between green and red,
br = ratio between blue and red, and
R, G, B = red, green and blue output of the camera.

The R channel was selected to be the divisor because it is often unclipped and therefore usable to recover estimates for G and B values. Of course, the selection of the divisor is case-dependent. The underlying requirement for the ratio method to be valid is that at least one channel is unclipped and a correct ratio is obtainable. Using the mean ratios, G and B values for the overclipped pixels can be approximated:

$$G_{est} = R \cdot m_{gr}, \;\; and \;\; B_{est} = R \cdot m_{br}, \tag{14}$$

where est = estimated value.

After the unclipped values of G and B are recovered, the whole image is scaled by the maximum value of all channels to the dynamic range of the camera. The disadvantage is that by applying this method to correct the image can cause an achromatic appearance for skin because skin is not perfectly matt.

Eigenfaces are face-like appearances having eigenvectors computed from face images and they are used for face recognition in grey scale images (Moon & Phillips 1998, Turk & Pentland 1991). One method of obtaining these eigenfaces is by applying singular value decomposition SVD on a set of F face images N = [N1, N2,...,NF) which produce F eigenvectors n = [n1, n2,...,nf]. Now the original images can be reconstructed at certain accuracy level using only the eigenvectors and their coefficients for the original image:

$$\tilde{N}_i = \sum_{j=1}^{E} c_j \hat{e}_j, \tag{15}$$

where $\tilde{N}$ = new reconstructed image,
E = the number of eigenfaces to be used,
c = the coefficient for the current image,
$\hat{e}$ = the eigenface,
i = the number of the image, and
j = the indicator of the factors related to a certain eigenface.

The eigenface technique is also extendable for colour images. The SVD is just applied to each colour channel separately. The advantage of this extension is better recognition results for complexly-illuminated object (Finlayson *et al.* 1996). Soriano *et al.* (1999) showed

that RGB eigenfaces are useful in colour correction: the coefficients of the first few eigenfaces are replaced by the coefficients obtained from the canonical face images while the facial details are preserved with the higher order eigenfaces. Mathematically, the image Ñ can be then reconstructed with two different sets of coefficients:

$$\tilde{N}_i = \sum_{j=1}^{Hf} ac_j \hat{f}_j + \sum_{j=D+1}^{F} c_j \hat{f}_j \tag{16}$$

where ac = the coefficient obtained from the canonical images,
Hf = the amount of the first few eigenfaces, and
F = total number of eigenfaces used in reconstruction.

In Soriano *et al.* (1999), a neural network was trained for transforming coefficients of the nonideal images closer to the canonical ones.

In the Paper III, the corrected image is compared to the image with the preferred skin colour appearance using a Euclidean distance metric in NCC rg space:

$$E_{rg} = \sqrt{(\bar{r}_o - \bar{r}_c)^2 + (\bar{g}_o - \bar{g}_c)^2} \tag{17}$$

where r, g = mean of NCC rg chromaticities, respectively,
o = original image, and
c = corrected image.

The evaluation was done using only mean values because there was no pixel-to-pixel correspondence between the images. Because of simple transformation from RGB, the comparison of images were done in the NCC rg space but other possibilities like S-CIE Lab (Zhang & Wandell 1996) can be used for evaluation.

The combined technique of RGB eigenfaces and the method of ratios was tested with images of the Physics-based Face Database taken under AD conditions. Because the camera was calibrated to incandescent A but the current illuminant was D65 with high luminous flux at the time, the overclipping was evident without adjustment of the camera aperture. An example of an AD image can be seen in Fig. 15a. In Fig. 15b pixels with one or more channels overclipped are coloured in black and those with no clipped channels in white. Before applying the SVD on the image set of 20 Caucasians, all images are geometrically normalized, cropped and resized. Table 8 shows colour correction results for 20 Caucasians when the correction is done with RGB eigenfaces and the combined technique of RGB eigenfaces and the method of ratios. The combined technique provides much smaller mean error than the use of only RGB eigenfaces.

Table 8. Mean error for 20 corrected face images taken under AD conditions.

Method applied on the images	Average E_{rg}
colour correction with RGB eigenfaces	0.112
colour correction with RGB eigenfaces and method of ratios	0.036

The results are also visualized in Fig. 16. When only RGB eigenfaces are used in colour correction, it makes some previously overclipped pixels appear greenish. If overclipped pixels are adjusted back to their original values, the colour correction works better. As can be seen from Fig. 16, the method of ratios applied before colour correction produces a closer match to the original image. The spatial visual evaluation made by a human between original and corrected images reveals that the presented correction produces acceptable results.

(a)   (b)
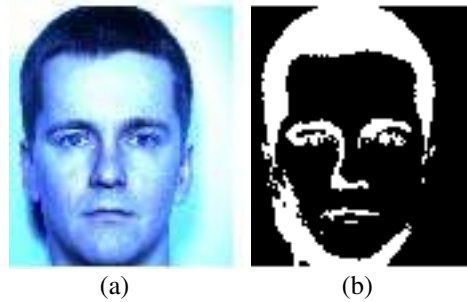
Fig. 15. Testing the colour correction: (a) the original AD image, and (b) the overclipping image in which black means pixels with one or more overclipped channel and white shows those pixels which do not have overclipped values.

(a)   (b)   (c)
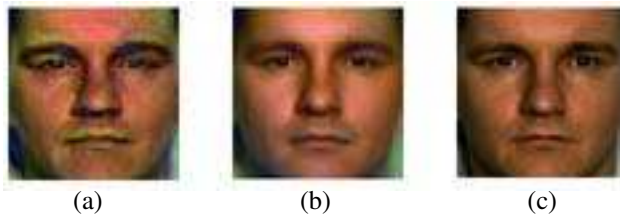
Fig. 16. Colour correction results for (a) RGB eigenfaces, which produce a greenish cast on some pixels, and (b) RGB eigenfaces and method of ratios. Both of these images are geometrically normalized due to the requirements of PCA. (c) is the original image in which both white balancing and the prevailing illumination type is incandescent A.

5 Skin chromaticities seen by a colour camera

The three main factors determining the camera output are the illumination SPD, the camera's channel sensitivities and the reflectance of the object (for more details, see Section 3.3). Another important factor is white balancing or white calibration conditions which are typically done so that a white object will have a white colour appearance. Even proper calibration does not guarantee that colours other than white (and the other achromatic colours if the camera is linear) will appear the same between images taken under different white balancing conditions. Moreover, when the illumination used in white balance is not the same as the prevailing illumination, the appearance of the colours will most likely change in a more drastic way. The bigger the colour temperature difference between these two illuminations, the bigger the shift of chromaticities, which is eventually limited by the properties of the camera. Also the direction of colour temperature change matters. When the prevailing illumination has a higher colour temperature, it makes colours appear more blueish, and when the opposite happens then the colours appear reddish (see Section 4.2).

In practical situations, there can also be other factors causing colour shifting. Nonlinearity of the camera is one factor and it is caused by the input-to-output conversion in which the transformation and its parameters depend on the input signal. The nonlinearity can manifest itself as a changing slope between the input and the output or gamma factor. There are many cameras in which the nonlinearity is deliberately introduced to compensate for nonlinearity in the display (Holst 1998 and Klette *et al.* 1998). In some cameras, the nonlinearity is introduced only in the high and low intensity areas. As an example of this kind of nonlinearity, the Sony DXC-755P has a knee-phenomena for the protection of the CCD elements from intense light (Sony 1989). In the knee-phenomenon, when the intensity of the light is higher than a certain threshold, the slope will become smaller. Another effecting factor is the limited dynamic range of the camera. This manifests itself in clipping of values: overclipping if the values are saturated to the maximum value and underclipping if they are zeros. Both forms of clipping cause serious information loss and in many algorithms pixels with clipping are ignored or it is assumed that there is no clipping in the image. Also the camera settings and controls, background, smearing etc. have their own influence on the formation of the image.

5.1 Basic principles

The purpose of a skin locus is to define an area of possible skin chromaticities perceived by a colour camera under a certain illumination variation range. It should be emphasized that no specific distribution is assumed for the skin colour or skin chromaticities. In this thesis, the chromatic constraint created does not only offer robustness against illumination change, but also independence of the camera white balancing conditions because it consists of subloci obtained from different camera calibration conditions (Paper IV). In any case, it is camera dependent or more accurately, sensor dependent.

Two methods will be presented for gathering this information about skin colour change. The goal of both methods is to find a good model for the chromatic constraint. It will also be shown that the NCC rgb space is an appropriate colour space for the constraint and therefore it is used in both methods. However, there is no reason why some other colour space might not also be applicable for these methods, although the quality of the chromaticity constraint created in these spaces cannot be guaranteed. The first of the methods presented here uses images with skin to create the chromatic constraint (skin locus). The images are taken under such illumination conditions which are thought to be representative to those one encountered in the application. For the second method, the spectral data of the camera, illumination and object need to be known or measured.

5.2 Skin locus from an image series

The chromaticity constraint (called skin locus in NCC space) is formed from a set of images using the method presented in Section 4.2. This kind of image set is shown in Fig. 17 taken by a Nogatech camera (Paper IV). For this image series the selected illumination range varied from the sunset / sunrise conditions to daylight condition. The selection is based on the evaluated robustness needed against the possible illumination changes encountered in the application. It is usually enough to take illuminants from both extremes and a couple from the middle range with suitable colour temperature differences. However, one should be cautious with fluorescent light sources. If a fluorescent light source with strong green components is not included in the locus, then the skin chromaticities perceived under this light source might fall outside the locus. With fluorescent lamps which have a strong red and / or blue component, this is not such a big concern because variations caused by them are usually smaller than those caused by colour temperature changes.

The range of illumination and white balancing conditions in which a skin locus is a valid description of possible skin chromaticities effects its size. If only one white balancing is used, then the obtained locus is smaller than the one which was constructed using many white balancing conditions. The former locus is more specific being only valid for one camera calibration. The later locus is bigger and offers robustness against calibration conditions at the expense of reduced discrimination capabilities between skin and non-skin objects. Therefore, when before making the locus one should consider both the illumination range encountered in the application and the possible issues related to white balancing of the camera.

After the series has been obtained, the skin areas in the images are manually selected

and then extracted as shown in Fig. 18 for the Nogatech camera.

Fig. 17. An image series for the Nogatech camera.

Fig. 18. Extracted facial skin areas.

Then the extracted skin pixels are converted from RGB space to NCC rbgI space or to

some other colour space. In Fig. 19, NCC gb coordinates of skin pixels are visualized. It is possible to model the obtained skin locus with functions describing the boundary area or with a look-up table.
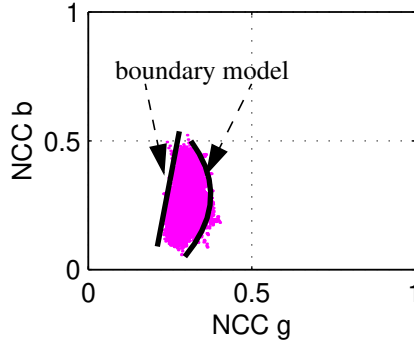

Fig. 19. Skin locus obtained from extracted facial skin areas.

5.3 Skin locus from basis functions

At least in theory, when all spectral information about imaging is available, then the skin locus can be calculated off-line using Eq. 5. This allows us the possibility of very easy and fast simulation of different illumination and white balancing conditions without taking any images. Again, results are sensor-dependent due to the use of the camera's spectral sensitivities. To obtain a more device-independent approach, one can utilize the skin colour signals (Paper V). Generally, the colour signal CS is described as the light which is reflected from a point with a certain reflectance. It can be expressed as a product of the illuminant SPD effecting over the point and the reflectance at that point:

$$CS(\lambda) = \prod_{\lambda = 400}^{700} R(\lambda) \cdot I(\lambda), \tag{18}$$

in which I = SPD of the effective illuminant over the point,
R = reflectance of the object at the point, and
λ = wavelength.
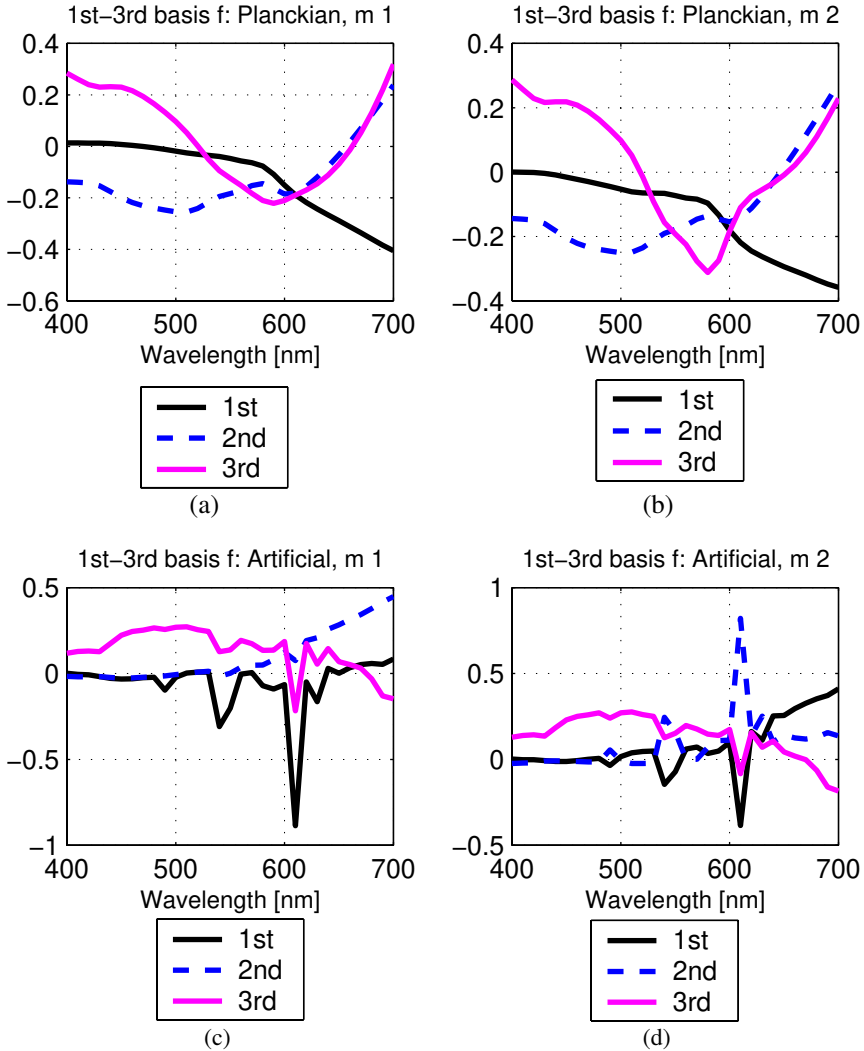
Fig. 20. The three first basis functions of colour signals using a Planckian set with (a) simple scaling normalization and (b) Euclidean normalization, and an artificial set with (c) simple scaling and (d) Euclidean normalization.

Basis functions for skin colour signals are not always available, but they are very simple and quick to calculate by applying principal component analysis (PCA) when illumination and reflectance data are available (i.e. from Physics-based Face Database). In many earlier studies, it has been shown that the obtained basis functions for spectral reflectances of different objects are useful for colour constancy (Maloney & Wandell 1986), colour correction (Lenz *et al.* 1999), segmentation (Hauta-Kasari *et al.* 2000) and colour image synthesis and analysis (Wandell 1987), but here they are used exclusively for skin appearance modelling. Furthermore, according to the author's best knowledge the PCA has usually been

applied separately to illumination SPDs and an object's reflectance functions for data compression and reducing complexity (Maloney 1986, Maloney & Wandell 1986, Parkkinen *et al.* 1989). The inspiration for employing basis functions is that they offer camera-independent information in compressed form about skin colour signals and make it possible to simulate outputs of different cameras easily.

The procedure for obtaining basis functions starts with the choice of illumination to be used. In these experiments, two illumination groups were used, artificial sources and Planckian illuminants. The first group consists of four light sources obtained from the Physics-based Face Database, namely the sources A, Horizon, TL84 and D65. The illuminants used in the Planckian group were calculated using black body radiator formulae (Wyszecki & Stiles 2000) with the colour temperature parameter changing from 2300 K to 6500 K in steps of 100 K. For both of these illumination groups, the results obtained from two commonly used normalization methods, the Euclidean rule and a scaling normalization (Eqs. 1-2 in Section 3.2), are compared and evaluated. When the normalization of illumination is done, the ensemble of skin colour signals is calculated using the illuminant groups and three spectral reflectances of eight persons from each skin type (totalling 72 reflectances). PCA was applied to the zero centered ensemble. The resulting basis functions are displayed in Fig. 20. The basis functions for the Planckian colour signal set are smoother than for the artificial set, but this is to be expected due to their different forms of SPDs. In addition, the illuminant normalization method used effects the shape of the basis functions. Another interesting observation is shown in Fig. 21.
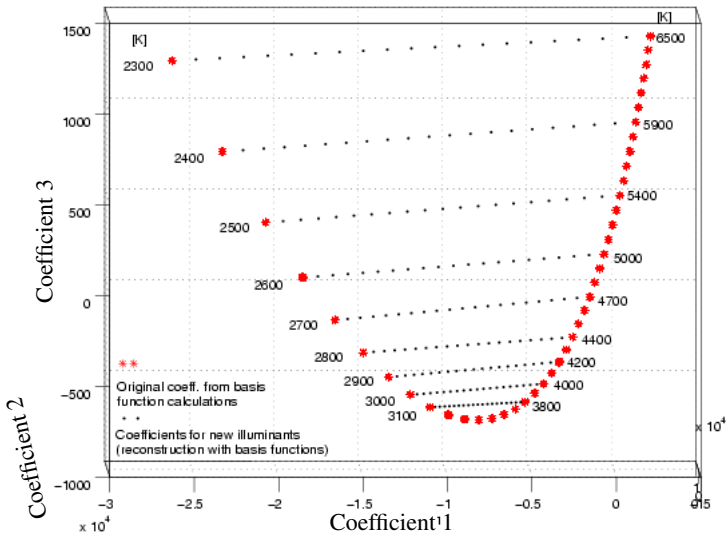
Fig. 21. First three coefficients of basis functions form a quadratic slope when a Planckian set with illumination normalization of scaling is used. The viewing angle is set to maximize the visibility of the slope and due to this the coefficient 2 axis is almost perpendicular. The calculated coefficients of basis functions for one light source situation (referred to as the original and marked with *) and for different mixtures of two light sources (referred as the new illuminant and marked with .) like combinations of Planckians of 2300K and 6500K; and of 2400K and 6000K etc.

The coefficients of the three first functions are shown in Fig. 21 for the Planckian set. The coefficients form a slope in quadratic shape because the SPDs are smooth and similar in shape. This is not true for real illuminants and their slope form is different. The coefficients are more sparse with lower colour temperatures because Planckian SPDs do not change uniformly with colour temperature.

The reconstruction quality of the basis function obtained is tested both for new reflectances and for new illuminants. The function for approximation error E per wavelength was selected so that the error increases rapidly when the difference between the original and reconstructed signal increases. The selected function is

$$E(\lambda) = \sqrt{\frac{(CS - CS_{REC})^2}{(CS)^2}}, \qquad (19)$$

where CS = the original skin colour signal, and
CS_{REC} = the reconstructed signal.

The total mean and standard errors are calculated as the mean approximation error over the wavelength and the number of samples. The reconstruction errors for skin colour signals calculated using 303 reflectances not used in basis function computations are shown in Tables 9 and 10, as well. Due to the division of data between test and training sets, the test set produces a smaller reconstruction error than the training set which was used in the computation of the basis functions. This is caused by reflectances included in the test set: it contains only Caucasian and Asian reflectances, which are more similar to each other than the Negroid reflectances. By comparing these two tables, it can be observed that the Euclidean normalization yields poorer performance than the scaling normalization and it is therefore excluded from the further research.

Table 9. Reconstruction error.

	Number of basis functions used	Error in training set		Error in test set	
		Mean	Std	Mean	Std
Planckian Illuminants	3	0.0301	0.0403	0.0256	0.0169
	5	0.0079	0.0084	0.0072	0.0043
	7	0.0050	0.0062	0.0046	0.0031
	10	0.0018	0.0015	0.0019	0.0014
Artificial sources: A, H, TL84, D65	3	0.1335	0.1294	0.0972	0.0601
	5	0.0548	0.0403	0.0408	0.0222
	7	0.0218	0.0229	0.0177	0.0137
	10	0.0104	0.0079	0.0090	0.0048

Table 10. Reconstruction error with illumination normalized by Euclidean rule.

Number of basis functions used		Error in training set		Error in test set	
		Mean	Std	Mean	Std
Planckian Illuminants	3	0.0305	0.0396	0.0274	0.0238
	5	0.0084	0.0108	0.0074	0.0050
	7	0.0052	0.0075	0.0047	0.0036
	10	0.0019	0.0019	0.0020	0.0015
Artificial sources: A, H, TL84, D65	3	0.1511	0.1457	0.1044	0.0679
	5	0.0475	0.0522	0.0376	0.0348
	7	0.0240	0.0240	0.0169	0.0133
	10	0.0107	0.0095	0.0097	0.0080

Because it is common for real scenes to have more than one illuminant shining on a point of the object, the basis functions are tested for skin signals under mixed illumination conditions. It is easy to show that this can be modelled by using the basis functions calculated earlier only for one illuminant cases. First, let's assume that we have N illuminants giving their contribution to a point with reflectance R. The new relative, effective illuminant I_{new} can be expressed as a weighted sum of different light sources as shown in Section 3.2, Eq. 3. The colour signal with this new illuminant CS_{new},

$$CS_{new} = I_{new} \cdot R = \sum_{j=1}^{N} w_j \cdot I_j \cdot R = \sum_{j=1}^{N} w_j \cdot CS_j, \tag{20}$$

is a weighted sum of colour signals of each individual illuminant. The new colour signal can be reconstructed using basis functions ε and mean μ calculated for the colour signals of each individual illuminant:

$$CS_{new} = \sum_{j=1}^{N} w_j \cdot (coeff_j \cdot \varepsilon + \mu) = C \cdot \varepsilon + \mu, \tag{21}$$

where the combination C is a combination of coefficients and weights,

$$C = \sum_{j=1}^{N} w_j \cdot coeff_j, \tag{22}$$

and *coeff* = the coefficients of basis functions for an individual illuminant.

The reconstruction quality with colour signals using a new illuminant is displayed in Table 11. It shows the reconstruction error with 5 basis functions. The new illuminants are calculated from two, equally weighted "old" illumination SPDs and they can be used to model a mixture of light sources. The purpose is now to simulate cases where two light sources shine on a point in the face equally. The reconstruction error is small, which means that the basis functions can be used to represent new skin colour signals. In fact, the error is smaller than for the average one illuminant case (Table 9) because the illuminant combination smooths out big differences. Moreover, the different degrees of mixtures can be simply simulated by scaling coefficients of the illuminants and summing them as shown in Eq 3. Fig. 21 displays coefficients for a mixture of two and Fig. 22 for three illuminants. The coefficients calculated for mixture colour signals form straight lines.

Table 11. Reconstruction error with illumination normalized by Euclidean rule.

Planckian illuminants			Artificial sources		
The new illuminant	Mean	Std	The new illuminant	Mean	Std
6500 K + 2300 K	0.0061	0.0030	H+TL84	0.0244	0.0155
3000 K + 5000 K	0.0058	0.0027	A+TL84	0.0246	0.0111
4000 K + 5000 K	0.0060	0.0027	D65+TL84	0.0156	0.0098
3000 K + 6000 K	0.0059	0.0027	A+D65	0.0200	0.0106

Next, the basis functions are used to calculate the skin locus: first, the illuminants are chosen for the locus modelling and the number of the basis functions are selected to achieve the needed reconstruction accuracy. Using these two factors, the skin colour signal set is reconstructed. For each output channel, the set is weighted by the camera's channel sensitivity and the sum of the weighted data is calculated. Then the sum is normalized using Eq. 6 with the desired white balancing illuminants. The simulated skin RGB values can be converted to some other colour space like normalized colour coordinates which are then used for further processing. The obtained loci can be seen in Fig. 23. When only three basis functions are used, the obtained locus is somehow smeared as shown in Fig. 23b and resembles the loci obtained with a Winnov camera (see Fig. 24). The results with five basis functions, Fig. 23c are comparable to those obtained using all data in Fig. 23a.
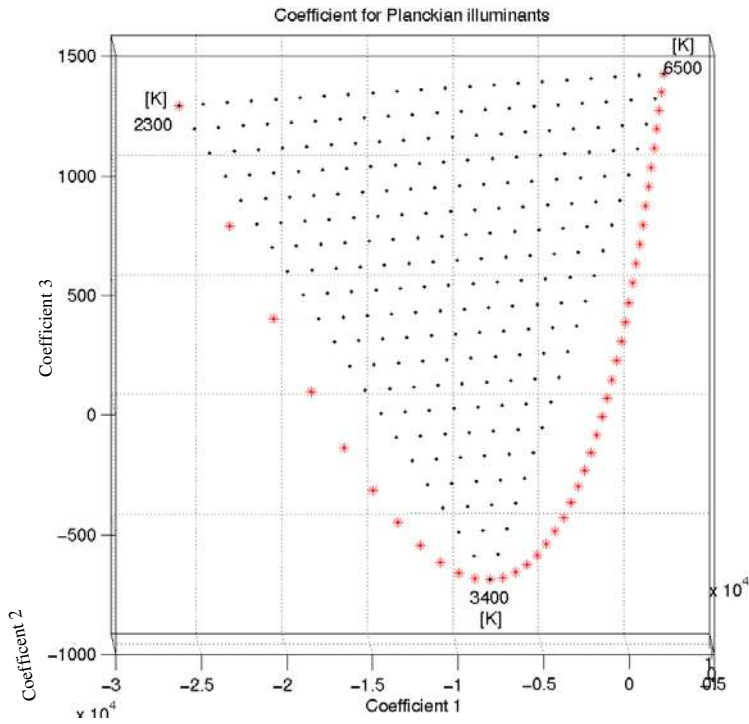
Fig. 22. The mark '*' shows coefficients of basis functions for one light source situation. It is easy to show that the coefficients for mixture illumination can be obtained by a linear combination of the coefficients from one light source case. Here the mark '.' shows calculated coefficients for different combinations of three Planckian illuminants (2300 K, 3400 K and 6500 K) and the linear relationship between the coefficients from one light source case is obvious.
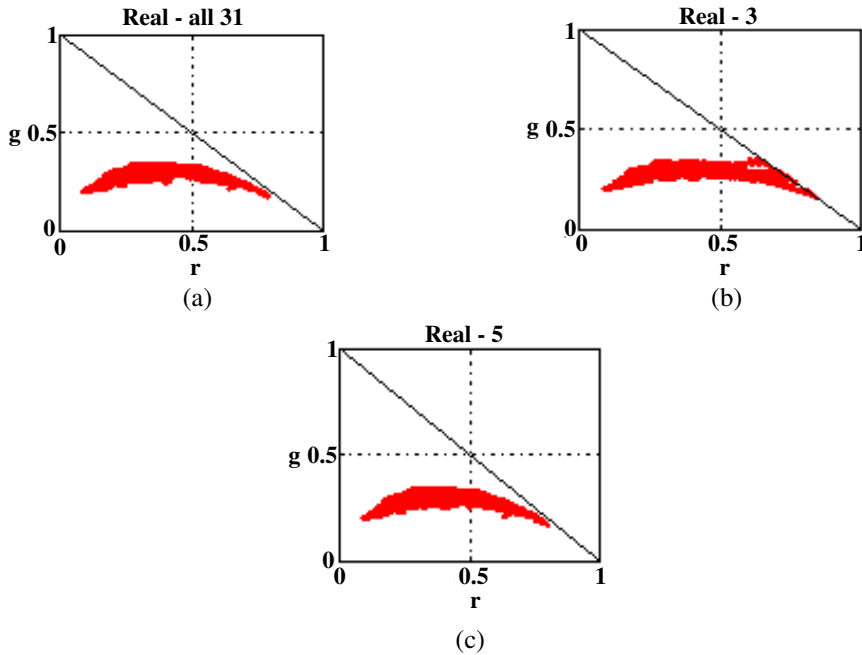
Fig. 23. Computed skin loci for the Sony camera: (a) with all data, (b) with the first three basis functions and (c) with the first five basis functions. The light sources were taken from an artificial group. The r and g chromaticities shown in the axes are parameters of the normalized colour coordinates obtained by a conversion from the calculated RGB values.

5.4 Behavior of skin colour

The chromaticity constraint method based on images is tested in 17 different colour spaces (Paper VI). For obtaining the images, the procedure presented in Section 4.2 was applied to three 1CCD webcameras, Alaris, Nogatech and Winnov cameras. The imaging arrangements were the same as those used with the Sony 3CCD camera except that the images were taken only from 2 persons (one from the yellowish and one from the pale skin group). This produced two 16 image series for each camera. For the Sony camera, ten image series were chosen from the same skin groups totalling twenty series. From the image series taken by these four cameras, skin regions were manually selected and extracted for further processing. The extracted skin RGB values were subjected to colour space conversion and after that quantized to 1 % steps of the whole chromaticity range. However, the colour space conversion does not remove chromaticity shifts or effects caused by noise. It can only remove the redundancy of RGB coordinates by separating intensity and chromaticity data.

These colour spaces are rather device-oriented, because human vision oriented spaces

like XYZ or CIE Lab require an illumination-dependent transform matrix and cannot therefore give much better results.The colour spaces employed can be classified, based on the conversion method, into two groups. The first group's colour spaces are calculated using linear transforms from RGB, and the spaces of the second group are obtained via nonlinear transforms of RGB. The chosen linear transform based colour spaces are: I1I2I3 (Ohta's features), YES, YIQ, YUV, YCrCb (Rec. 601-5 and 709). Among the nonlinear transforms are: NCC rgb, modified rgb, natural logarithm ln-chromaticity, P1P2, l1l2l3, ratios between channels (G/R, B/R, and B/G), HSV, HSL, modified ab, TLS and Yuv. Equations for these transformations are presented in Appendix 1. The RGB data was preprocessed by scaling it between 0 and 1 except for the ln-chromaticity in which case the range was 0-255. Some of the transformations needed further processing of RGB values: 1) pixels with zero intensity (a special handling in NCC rgbI transformation; removed before modified rgb, P1P2, uv and modified ab transformations), 2) pixels with achromatic colours (special handling in the case of HS-spaces, removed before l1l2l3 transformation), 3) pixels with zero green value (removed before modified ab transformation), and 4) one or more channel is zero at a pixel (removed before ln-transformation).

It was noticed that the brightness control or lack of it can have a strong effect on the possible skin chromaticities. If there is no automatic brightness or gain controller, it is possible for one channel to have low values or even underclipping. This can be clearly seen from Fig. 24a for the Winnov camera. The locus is quite big but it can be made smaller by using only values over a certain intensity at each channel as in Fig. 24b. The latter case is that it limits the possible intensity change range while the former reduces discriminability between background and skin objects. However, there are problems with a brightness controller, like lost of colour data due to colour bleaching.
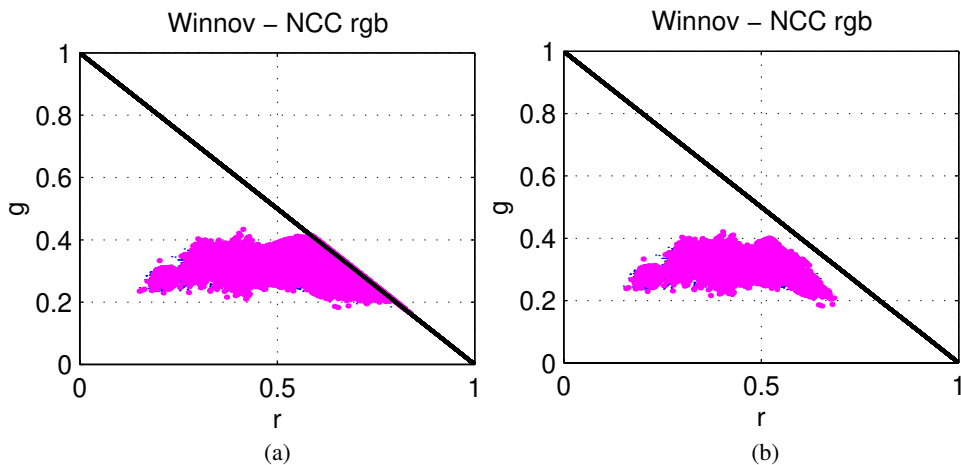
Fig. 24. Winnov locus in NCC rgb from (a) original data and (b) data associated with values higher than 25 in each channel (maximum value 255).

Not all colour spaces give uniform presentation for skin chromaticities. In HS-spaces (HSV, HSL, HSI and HSB) and TSL spaces plotted in Cartesian coordinates, the skin chro-

maticities do not form a uniform cluster, as shown in Fig. 25
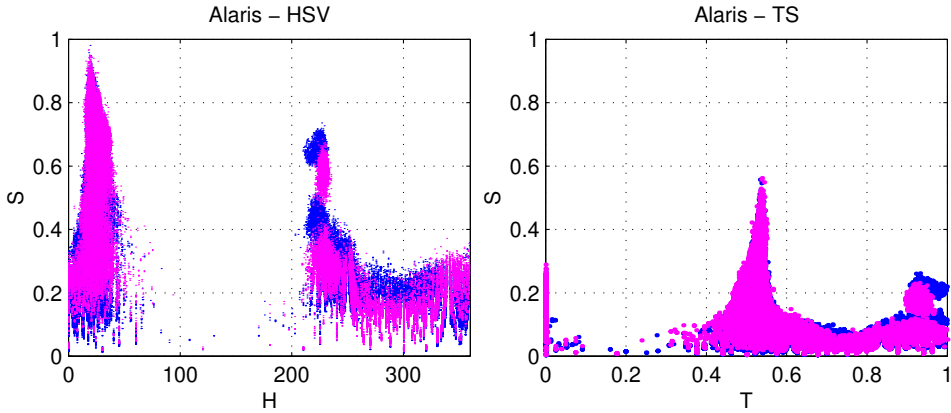
Fig. 25. There was no single cluster for (a) HS-chromaticities in Cartesian coordinates or (b) TS-chromaticities.

If the HS-space is correctly visualized in polar coordinates, this improves uniformity and produces only one cluster (See Fig. 26). Appendix 2 displays more images about skin chromaticities in other colour spaces. As can be observed from these images, the chromaticity constraint can be modelled with straight lines and / or quadratic curves except TSL-space. For HS-spaces, the modelling was considered only in polar coordinates.
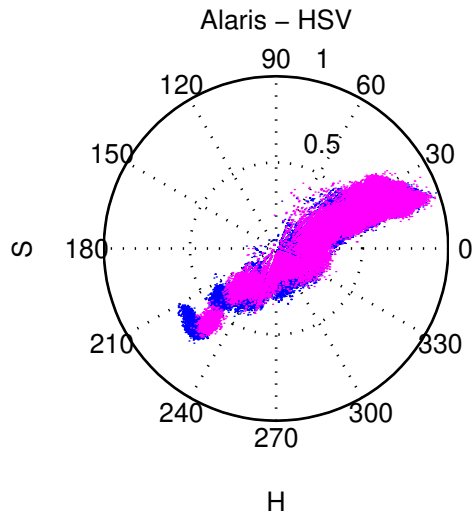
Fig. 26. Skin chromaticities of HS-space plotted in polar coordinates.

Next, it was evaluated how well different skin groups (pale and yellow) overlap in colour spaces. The overlap is defined as the percentage of the number of skin colour bins common to both groups divided by the amount of all skin coloured pixels. It was calculated for

4 calibrated conditions (except Nogatech, where only 3 calibrated cases were used because of limited calibration capabilities) and the results can be seen in Table 12. The same was done also for all 16 cases (Table 13). It can be concluded from these tables, that the overlap between different skin groups was in all cases reasonably high. Table 14 compares the overlap between different cameras and the percentage occupied by the constraint from the total area. As shown in this table, the total overlap between all cameras is much lower than the overlap between brightness (gain) controlled cameras (Alaris and Nogatech) and between cameras without it (Sony and Winnov). However, the sparsity of the colour space values was not evaluated neither was their applicability for modelling mixture illuminations or dichromatic reflection model. It seems that the behaviour of skin chromaticities on a colour space is dependent on the selected camera and the other colour spaces do not produce better results than NCC rgb.

Table 12. Overlapping percentage between different skin groups.

colour space	Calibrated cases in [%]			
	Alaris	Nogatech	Winnov	Sony
TSL	57.1	49.3	65.6	69.6
mod. ab	52.8	49.2	59.8	70.4
HSV	47.2	51.8	69.3	66.8
HSL	49.3	50.5	67.6	69.2
YES	60.9	60.6	72.3	72.9
YIQ	62.5	60.1	73.7	72.2
YUV	58.9	58.4	70.2	74.2
Yuv	53.2	49.7	62.6	72.3
P1P2	56.6	46.6	61.9	71.9
YCbCr1	62.3	61.6	70.8	70.1
YCbCr2	62.4	59.4	73.7	71.8
NCC rg	61.0	47.9	62.9	72.2
NCC rb	63.1	49.1	64.7	73.5
NCC gb	60.7	46.9	61.2	73.2
mod. rg	58.0	54.7	72.3	70.6
mod. rb	58.2	49.8	63.8	70.5
mod. bg	56.3	48.2	62.6	72.9
logchroma	60.9	43.3	58.6	69.2
ratio g/r_b/r	70.6	58.1	69.0	63.9
ratio g/r_b/g	64.3	60.0	65.6	66.4
ratio b/r_b/g	66.7	65.2	68.0	65.1
I1I2I3_I2I3 (Ohta)	62.4	58.9	69.0	71.9
1l1l2l3_1l1l2	58.9	76.9	79.6	71.4
1l1l2l3_1l1l3	67.4	79.4	79.2	77.0
1l1l2l3_l2l3	59.2	70.4	75.8	68.5

Table 13. Overlapping percentage between different skin groups.

colour space	All 16 cases in [%]			
	Alaris	Nogatech	Winnov	Sony
TSL	75.8	67.3	72.7	88.7
mod. ab	68.2	56.1	65.0	78.7
HSV	72.1	69.1	67.2	89.9
HSL	74.3	67.9	69.7	88.0
YES	76.0	65.0	74.3	83.1
YIQ	75.6	64.4	74.2	83.5
YUV	75.2	64.4	72.6	83.5
Yuv	65.6	58.1	69.1	79.6
P1P2	73.8	58.5	76.1	83.0
YCbCr1	76.8	64.6	74.6	82.9
YCbCr2	74.3	64.9	75.9	82.5
NCC rg	74.6	59.1	79.1	82.4
NCC rb	74.5	59.7	79.6	84.8
NCC gb	74.9	59.8	78.4	83.0
mod. rg	67.9	65.4	72.1	83.1
mod. rb	74.5	61.9	75.4	84.2
mod. bg	78.6	60.1	75.9	84.8
logchroma	72.6	54.3	77.6	78.8
ratio g/r_b/r	53.7	63.7	74.6	64.1
ratio g/r_b/g	63.5	64.4	75.3	68.3
ratio b/r_b/g	57.9	67.5	70.7	64.0
I1I2I3_I2I3 (Ohta)	74.6	64.7	73.5	82.9
I1I2I3_I1I2	87.5	95.3	87.9	99.4
I1I2I3_I1I3	89.4	96.4	87.7	98.8
I1I2I3_I2I3	89.3	95.6	85.3	98.8

Table 14. Overlapping between cameras and size of skin locus in the colour space.

colour space	Overlapping [%]			Total size in [%]
	All 4 cameras	Alaris and Nogatech	Sony and Winnov	Sony
TSL	17.7	59.3	53.1	25.8
mod. ab	13.2	53.4	43.0	45.2
HSV	17.2	60.6	47.8	55.6
HSL	17.6	59.6	50.1	44.2
YES	17.6	57.2	32.5	15.8
YIQ	17.6	57.4	32.3	12.7
YUV	14.1	56.3	27.4	14.3
Yuv	14.6	53.2	47.0	28.6
P1P2	17.3	55.5	57.9	13.0
YCbCr1	17.9	56.2	33.4	15.0
YCbCr2	17.7	56.5	32.9	15.5
NCC rg	17.4	55.5	57.4	14.2
NCC rb	18.3	56.9	59.2	14.0
NCC gb	17.9	56.4	58.6	14.0
mod. rg	18.9	55.0	51.9	23.2
mod. rb	20.0	58.7	56.5	25.9
mod. bg	20.6	59.2	64.8	31.0
logchroma	10.2	51.6	52.4	8.1
ratio g/r_b/r	7.3	39.2	21.6	0.0
ratio g/r_b/g	10.5	43.5	32.2	0.0
ratio b/r_b/g	7.7	39.3	25.8	0.0
I1I2I3 (Ohta)	17.3	56.7	31.8	15.8
l1l2l3_l1l2	24.0	83.6	49.4	65.1
l1l2l3_l1l3	30.3	85.8	56.1	65.1
l1l2l3_l2l3	24.7	85.1	50.3	64.5

6 Skin locus in face tracking

The obtained knowledge about the range of skin chromaticities (skin locus) will be shown to be useful in colour based face tracking and segmentation. To envisage this, a Face Video Database under drastically varying illumination conditions has been created. The skin locus is combined with different tracking algorithms and applied to these videos. No cue other than colour is used; obviously adding other cues would improve the overall performance. However, because the colour is the only cue it is much easier to evaluate its performance and usefulness.

6.1 Face Video Database

The Face Video Database (Paper VII) was designed for development, testing, comparison, and verification of algorithms related to face-based applications. Because (at this moment) three cameras (Alaris, Nogatech and Sony) are employed, it is possible to study and compare the performance of algorithms with different cameras.

The database consists of the following data for each camera: face images and information related to their acquisition, face videos and manual localization of faces in the videos. The videos and images were taken with 1CCD cameras (Alaris and Nogatech) and with a 3CCD (Sony) camera. Alaris and Nogatech are both low-cost web cameras with an automatic intensity (gain) level control. When taking videos and images, their automatic colour correction options were turned off after initial white balancing because colour correction can lead to unstable and unpredictable results, and the main interest was in the effect of illumination changes. The Sony DXC-755P does not have automatic gain or colour correction.

When the face images were taken with a camera, the same procedure and same conditions as in the creation of the Physics-based Face Database were applied (See Section 4.2). Table 15 shows a face image series taken by the Nogatech camera with corresponding prevailing and white balancing illumination conditions. The SPDs of the illuminants used and the spectral reflectance of skin for some persons are available. During image acquisition, it was noticed that calibration results were not always very good for the 1CCD cameras due

to their limited capabilities. This can be observed when comparing the uppermost and second rows of Table 15. The white calibration failed to remove totally the reddish cast of illumination H and to make white appear white. Because the white calibration has its own limitations, there is an obvious need for illumination insensitive techniques.

Table 15. Sixteen faces for Nogatech.

	H	A	TL84	D65
H				
A				
TL84				
D65				

The illumination in the face videos is challenging and commonly encountered in practice; the videos are made under both indoor and outdoor illumination conditions. The illumination field over the objects varies in time and in space. The videos have also different initial white balancing settings. The videos have persons with dark, pale and yellowish skin tones. Tables 16 and 17 display a few selected frames from an indoor and outdoor video. In the outdoor video, the person is on the roof of a building. The illumination field over the

face varies from direct sunlight to the cast from the plain sky. The indoor movie shows a person moving in a corridor with an illumination field created by either fluorescent lamps, daylight from the window or both.

Ground truths for face localization are selected manually and they make possible numerical quantification of the results. The localization is made by a box (ground truth bounding box) surrounding the face region in the image. For some selected videos, the face area is defined more accurately by a polygon (a ground truth bounding polygon). Both ground truths are visualized in Fig. 27.

Table 16. Selected frames from an outdoor Nogatech video.

Frame 1	Frame 51	Frame 101
Frame 151	Frame 201	Frame 251
Frame 301	Frame 351	Frame 401
Frame 451	Frame 501	Frame 551

Table 17: Selected frames from an indoor Alaris video.

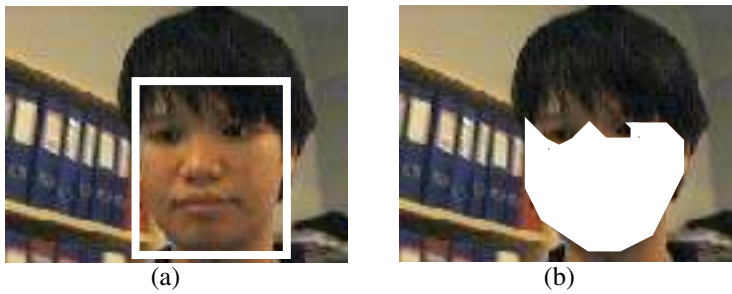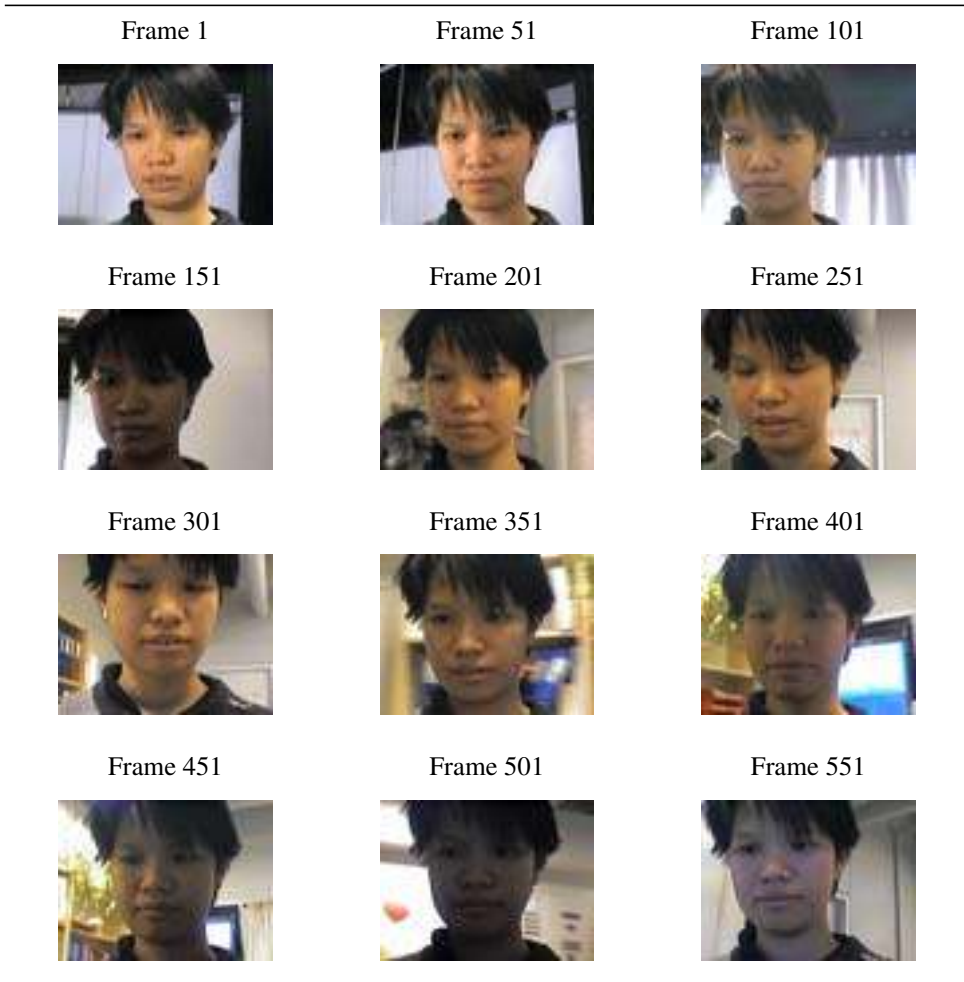| Frame 1	Frame 51	Frame 101
Frame 151	Frame 201	Frame 251
Frame 301	Frame 351	Frame 401
Frame 451	Frame 501	Frame 551

(a)   (b)

Fig. 27. Face localization by (a) a box and (b) a polygon.

6.2 Ratio histogram and histogram backprojection

A histogram can be used in the description of an object. It is useful in detecting and locating the object in an image (Swain & Ballard 1991) or a frame of a video sequence (Comaniciu *et al.* 2000) by forming a model of the object. The colour model of the object in an image or a frame can be made by a common method called a ratio histogram. The ratio histogram R describes the relationship between the histogram M of the selected region (target histogram) and the histogram I of the whole image or R=M/I. This way those objects' pixel chromaticities which appear rarely in the background are emphasized in the characterization of the object. The purpose is to increase discrimination between the background and the object which leads (hopefully) to more reliable object localization.

The calculation of the ratio histogram is the first step of histogram backprojection introduced by Swain and Ballard (1991). Histogram backprojection can also be described as a filtering which leads to the formation of a new grey scale image. The pixel value in the greyscale image is the probability of the ratio histogram defined uniquely by the corresponding chromaticity values of the pixel in the original image. Those pixels whose chromaticity coordinates do not have support from the ratio histogram are attached with value zero. Usually in this grey scale image, there are clustered regions with high values which are then considered as object candidates. A general practice is to apply morphological operations on the grey scale image before object localization to reduce the effects of noise and to connect the components.

6.3 Adaptive ratio histogram

Colour histograms have proved to be very effective and useful in object detection, localization and recognition (Swain & Ballard 1991), but under static or nearly static illumination conditions. For example, in many videos the colour appearance of an object, which is the facial skin in this thesis, is not static in time or in space. To cope with these colour changes, Raja *et al.* (1998) have suggested a dynamic adaptation of the skin colour model. In the spirit of this idea, the ratio histogram of the object can be refreshed for each frame by a moving average (MA) method:

$$R_t = \frac{(1 - \alpha) \cdot R_{t-1} + \alpha \cdot R_{t-2}}{max((1 - \alpha) \cdot R_{t-1} + \alpha \cdot R_{t-2})} \tag{23}$$

where $\alpha = 0.5$, and
t = the frame index.

The MA method provides a smooth transition between frames and reduces the effects of noise. However, when the older colour models are "memorized" in the calculation of the new ratio histogram causing the confidence level of the information as a function of the number of frames used. This means that a high ratio value can be attached to a colour tone no longer belonging to the target due to, for example, a very rapid change in colour appearance.

If the pixels are labelled using just the ratio histogram of the previous frame, this can

cause instability in the colour model and errors; especially with low quality cameras. The sensitive part of updating is the calculation of the new ratio histogram. To calculate the ratio histogram, or more accurately the target histogram M for the current frames, the training pixels are selected using the spatial or chromaticity constraint. The spatial constraint utilizes pixels falling inside a central region (Raja *et al.* 1998) or an oval region (Yoo & Oh 1999). The chromaticity constraint instead filters out obviously non-skin coloured pixels and only data from skin coloured pixels are used in the calculation of the target histogram M.

6.4 Tracking with skin locus: settings and results

Face tracking with adaptive backprojection was implemented with a chromaticity constraint called skin locus and tested with both indoor and outdoor videos to show the applicability (Paper IV). The basic assumption is that the illumination causing skin chromaticity is changing relatively slowly compared to the frame rate of the camera. Henceforth, the skin colour in the previous frame is not totally different that of the current frame. The illumination field over the object can vary not only due to lighting conditions but also due the movement of the object in a nonuniform field. It is also possible that in the case of a nonuniform illumination field over the object, the movement of the camera causes the chromaticities of the object to change due to the different viewing angle. This leads to an additional assumption: between two sequential frames, the difference between the object positions is at most 10 % of the object size in one direction (up, down, left or right). When these two assumptions are both valid, then the skin colour histogram can be used for describing the skin colour appearance in the next frame and the search area is 10 % bigger in one direction than the localized area (maximum size is the size of the frame). Because the search area is not always as big as the whole frame, this increases the speed of the tracking.

   Initialization of the skin model is obtained from a user-defined cut-out region which is assumed to contain facial skin. Each RGB frame of the videos was converted to 2D planes of the NCC rg-chromaticity before making the histogram as described in Section 6.2. The obtained histogram for the object implies the probability of how uniquely a chromaticity pair belongs to the current colour appearance of the skin. The histogram is normalized so that the maximum value is one because a fixed value is used to exclude those chromaticity pairs which appear commonly also in the background. Without normalization, there can be problems if the distribution is wide meaning lower probability values attached to chromaticity coordinate pairs because of a fixed threshold value. Both chromaticity channels are quantized to 65 discrete values in the range of 0-1. There are several advantages for the quantization: faster computation, reduction of noise and small illumination variation, and reasonable presentation of colours. If, for example, a colour is defined with one unit accuracy, there are $256*256*256 = 16777216$ possible colours in RGB space. Although conversion to chromaticity coordinates reduces the total number of colours, it still may overcome the number of the image pixels and lead to a big table with many zero entries. The 65 discrete levels was thought to represent a good balance between accuracy and computation speed but no numerical test was made to verify this (beyond the thesis scope).

   The chromaticity constraint (skin locus) was used as prior knowledge about skin. Fig.

28 shows the skin loci for Nogatech and Alaris cameras. The skin locus was restricted with two quadratic functions. A general formula for classification of chromaticities is presented as a function of NCC g value:

$$M(r, g) = \begin{cases} 1 & if \ (g < g_u) \cdot (g > g_d) \cdot (W > 0.0004), \ and \\ 0 & otherwise, \end{cases} \tag{24}$$

where u = upper bound, d = lower bound, $g_u = J_u r^2 + K_u r + L_u$, $g_d = J_d r^2 + K_d r + L_d$, and $W = (r-0.33)^2 + (g-0.33)^2$. The quadratic coefficients for Nogatech's skin locus are $J_u = -1.377$, $K_u = 1.074$, $L_u = 0.145$, $J_d = -0.776$, $K_d = 0.560$, and $L_d = 0.177$; and for Alaris they are $J_u = -1.842$, $K_u = 1.529$, $L_u = 0.042$, $J_d = -0.728$, $K_d = 0.607$, and $L_d = 0.177$. The modelling used can include either partially or wholly the white point (r = 0.33 and g = 0.33) which causes greyish and whitish pixels to be labelled as skin. To avoid this situation, a circle with a radius of 0.02 around the white point is additionally included in the non-skin set. The range of chromaticity r is restricted to 0.2-0.7.
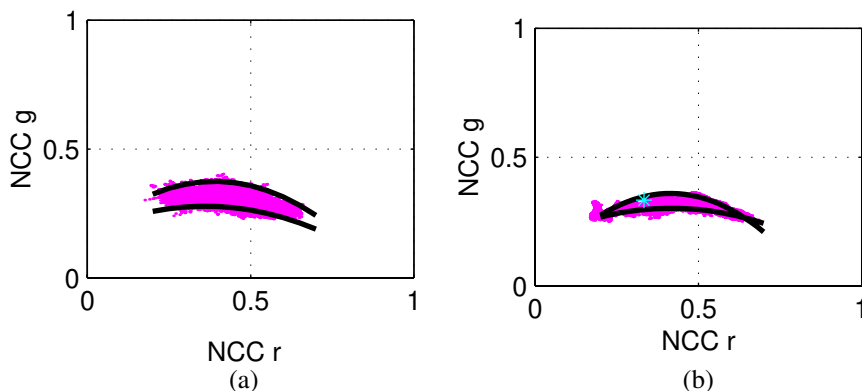
Fig. 28. Skin locus for (a) the Nogatech camera and (b) the Alaris camera. The straight or dashed lines visualize the pair of quadratic functions used to define the upper and lower bound for the skin chromaticity cluster. (b) also shows the white point and the surrounding to be excluded from the image (marked with *).

The adaptive ratio histogram and histogram backprojection are applied for face tracking in a search region. The search region for a face is obtained by expanding proportionally the bounding box which often increases the speed of the algorithm. After backprojection, the obtained grey scale image is thresholded (fixed threshold, 0.33) and then filtered with a majority morphological operator. The size of the majority filter is 3x3 and it essentially sets the middle pixel to one if there are five or more pixels with 1's surrounding it. The purpose of the filtering is to eliminate spurious pixels and connect areas. Next, the connected component analysis is done and the largest component found is assumed to be the face. The result is visualized by drawing a bounding box around the component. The pixels inside the bounding box are used to calculate the face colour model for adapting the ratio histogram. The behaviour of the skin ratio histogram at three frames is observable in Fig. 29. Visualization of the tracking results using a few selected frames is displayed in Fig. 30 for indoor and outdoor videos of the Nogatech camera. The skin locus based adaptive technique is also applicable to other cameras as shown in Fig. 31.

Fig. 29. Behavior of the ratiohistogram. The upper image row shows how the ratiohistogram of a face image is affected by the colour appearance of skin. The skin locus is visualized with two straight lines. The reddish colour in the upper row means high values whereas the blue indicates low values.

(a)

(b)

Fig. 30. Adaptive skin locus based face tracking for video sequences of the Nogatech webcamera: (a) indoor conditions, and (b) outdoor conditions (tracked with a limited skin locus). White bounding boxes indicate localized face whereas magenta boxes show the search area.

Because few images can give only an impression of the behaviour of tracking, the goodness of our localization is expressed with an overlap measure A:

$$A = \frac{A_{GT} \cap A_c}{\sqrt{A_{GT} \times A_c}} \tag{25}$$

where A = the size of the area of the bounding box,
GT = ground truth, and
c = calculated.

If the calculated and ground truth bounding boxes have the same size and totally overlap, the overlap measure A = 1. When reliable colour information is not always available, which means no clusters can be found i.e. due to loss of colours, the tracking is suspended and the error count for the frame is set to 1 (otherwise 0). The old bounding box is kept and used until the clusters are (hopefully) found again. Fig. 32 presents a numerical characterization of tracking with histogram adaptation based on skin locus constraint. A few selected frames for visualization of these results are shown in Fig. 30a.

6.5 Comparison with other tracking methods

Some tracking algorithms use a fixed histogram model for facial colour, and they usually work well if the illumination is unchanged (Paper IV). Often only chromaticity data is used to increase the model's usefulness under varying illumination intensity. As it is shown in Fig. 33, the fixed skin colour model based on chromaticities cannot handle changes in illumination chromaticities. The model obtained from the first frame is valid only for restricted conditions. Fig. 34 shows numerical evaluation for this tracking and clearly demonstrates the failure of the fixed model tracking.

Because of the problems of tracking based on a fixed model and colour constancy, adaptive colour models have been suggested for target tracing (Raja *et al.* 1998). It is obvious that model adaptation without any constraint will fail as it adapts to all colours. For the constraints, two spatial restrictions have been suggested (Raja *et al.* 1998, Yoo & Oh 1999). The pixels for model refreshment are selected either inside the bounding box localization of the face (Raja *et al.* 1998) or oval (elliptical) face localization (Yoo & Oh 1999). Fig. 35 shows some frames from the adaptive tracking using oval constraint. These image frames demonstrate that oval tracking can adapt to something else which is not facial skin (Paper IV)!

Fig. 31. A few selected frames from the skin locus based adaptive face tracking for a video taken by an Alaris webcamera. White boxes visualize localization of the bounding box and magenta boxes show the search area.

Fig. 32. Numerical evaluation of tracking results. The skin locus was used as a constraint on the histogram adaptation. The blue line indicates the goodness of the tracking result ('A') and the dashed line shows the error count value (anything over 1 means a frozen bounding box).

Fig. 33. Face tracking with fixed histogram. The magenta box shows the search area while the white box and cross display the face localization.
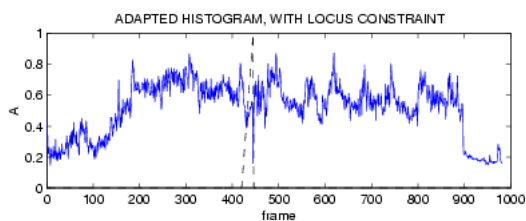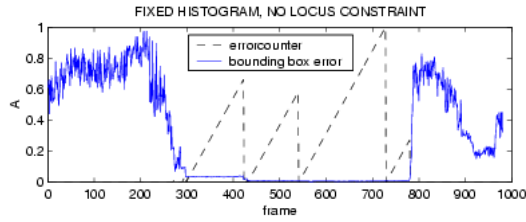
Fig. 34. Numerical evaluation of tracking results. The skin locus was used as a constraint on the histogram adaptation. The blue line indicates the goodness of tracking result ('A') and the dashed line shows the error count value (anything over 1 means a frozen bounding box).

Fig. 35. Adaptive tracking using an elliptical constraint. The magenta box shows the search area while the white box and cross display the face localization.

6.6 Robustness to localization errors

Tracking algorithms can often produce results in which the localization of the face is not perfect. The robustness towards these imperfect localizations improves reliability of the performance. This robustness is investigated with an adaptive face tracking algorithm with a spatial and chromaticity constraint. The spatial constraint used selects pixels for refreshing the skin colour model on the oval region placed within the face localization found.

The tracking results are compared to a ground truth skin colour model which was obtained from the manually defined face regions. The selected face regions include non-skin regions like eyes and mouth. Because they were not excluded from the ground truth calculation, the experiment is biased against a chromaticity constraint.

The new, erroneous localization box is obtained for testing a displacement of the ground truth bounding box. The displacement is done in increments of 10 % from the original cen-

tre in the range of 0-100 % and in four directions: top, left, bottom, and right. This is visualized in Fig. 36 for the oval (or ellipse) spatial constraint.
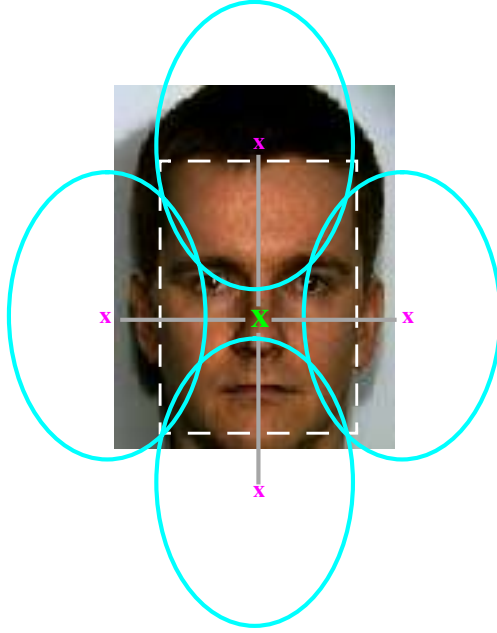
Fig. 36. An example of displacement for elliptical constraint. Ellipses show a displacement in four directions, up, down, right and left, while the x are the middle points and the white dashed line displays the manually selected ground truth bounding box.

To quantify the results, the mean error for a whole movie was calculated as a function of the displacement distance. The mean displacement error D_{c2c} for the video is

$$D_{c2c} = \frac{1}{F \cdot 4 \cdot BINS} \cdot \sum_{f}^{F} \sum_{dir}^{4} \sqrt{\sum_{j}^{BINS} (R_{GT} - R_c)^2} \tag{26}$$

where $c2c$ = the percentage centre-to-centre displacement from ground truth bounding box,

f = frame index,

F = total number of frames,

dir = direction of the displacement, together 4,

j = histogram bin,

$BINS$ = total number of bins (64x64 = 4096),

R = ratio histogram, and

c = constraint (spatial or chromaticity) used in tracking.

The mean errors as a function of displacement for three movies are shown in Fig. 37. The trend of histogram errors for the spatial constraint increases proportionally with the

displacement distance, whereas the results based on chromaticity constraints are stable. The behaviour of spatial constraint is according to intuition: the larger the localization error, the less skin falls within the ellipse. This demonstrates the capability of the skin locus to recover localization mistakes better than the spatial constraint. The bigger error for the skin locus with small dislocations is due to a ground truth histogram containing non-skin objects. Fig. 37 indicates that the elliptical constraint is sensitive to localization errors.
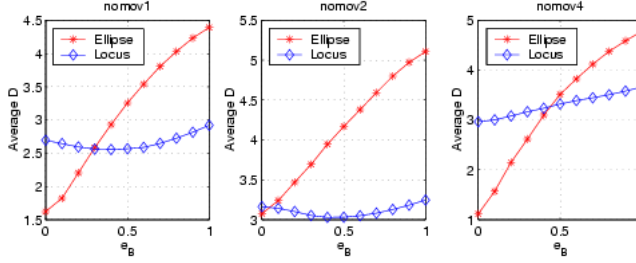
Fig. 37. Error D due to relative displacement e_B for adaptive tracking schemes based on elliptical and skin locus constraints.
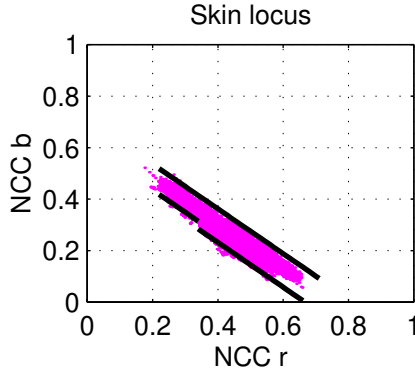
6.7 Mean shift with skin locus

The skin locus constraint can be implemented as a colour distribution adaptation module also for other tracking algorithms (Paper VIII), like mean-shift (Comaniciu *et al.* 2000). Appendix 3 contains a detailed description of this algorithm. Basically, the mean-shift algorithm searches for the nearest place in which the difference between the computed colour distribution and the defined object colour distribution is smaller than some fixed threshold. For maintaining good performance of tracking, the defined distribution should be adapted to environmental changes or environmental changes should be cancelled. The cancellation can be by a colour constancy algorithm or by using only very restricted environments where the illumination does not change. Because the performance of colour constancy algorithms has not been satisfactory so far and robust methods for normal, unrestricted environments are sought, the adaptation of the object colour distribution is a very attractive solution.

The skin locus is now modelled in rb-space with three straight lines (Fig. 38). The limits are modelled by giving the range of chromaticity *b*:

$$b = \begin{cases} 0.71 - \dfrac{r}{1.15}, & 0.22 < r < 0.71 \quad (upper\ \ limit) \\[2mm] 0.61 - \dfrac{r}{1.15}, & 0.22 < r < 0.34 \quad (lower\ \ limit) \\[2mm] 0.58 - \dfrac{r}{1.15}, & 0.34 \le r < 0.71 \quad (lower\ \ limit) \end{cases} \qquad (27)$$

in which the chromaticity *r* can vary between 0.34 and 0.71.

Fig. 38. Skin locus for the Alaris camera.

Initialization for the adaptive mean-shift with skin locus was done by hand and the initialization coordinates were saved. The manually selected area S was then subjected to filtering using the whole skin locus in order to remove non-skin coloured pixels. From the filtered result, the colour distribution for the object was calculated. Unlike with backprojection tracking, only a subsection of the skin locus was used at a time. The range of the subsection was dynamically updated from each segmented image. Instead of using a fixed threshold for excluding those skin chromaticities which appear too often in the background, a video independent dynamic threshold T was used:

$$T = \left(\frac{K_C}{K_{FRAME}}\right)^x \tag{28}$$

where K_C = the size of the bounding box obtained from initialization or algorithm calculations,

K_{FRAME} = the size of the whole frame, and

x = a parameter to be set.

The positive parameter x defines the shape of the threshold function. If x is not near 0 or 1, the thresholding is a nonlinear function of the size ratio. After initialization, the following procedure was repeated until the end of the sequence:

1. A new frame is filtered with a subsection of the skin locus.

2. Threshold T is calculated and applied to the current skin colour model for the object. Those skin chromaticities which have background support more than the threshold value, are excluded.

3. Mean shift is allowed to localize the bounding box and a new skin colour model is calculated on the segmented image.

4. A new, updated skin colour model for the object is obtained as an average of the "old" and "new" colour models. Those chromaticities which appear only in the "old" colour model are removed.

5. Skin locus subsection range is updated using the localized bounding box.

Table 18 shows the localization and segmentation for tracking after applying the previ-

ous algorithm on a video. It is not argued here that this is the optimal implementation; the purpose is to mainly show that even with this implementation, the results are better than for the non-skin locus based methods. The size of the bounding box was frozen after the first frame to make possible comparison with the fixed skin colour model tracking and adaptive tracking with spatial constraint for the whole sequence (Tables 19 and 20). Otherwise, the performance of these two tracking methods became unstable. The colour model for the static distribution tracking was obtained from the first frame of the video. For the spatially restricted adaptive tracking (after Y. Raja *et al.* 1998), the pixels for updating the colour model are obtained from a smaller bounding box centred inside the localization bounding box and 1/3 its size. As it is quite obvious that locus based tracking cannot separate the face area from the neck. This is also true for the other methods. The segmentation results are best with a skin locus and they depend on the background. In some frames the segmentation is so good that the size of the face could be reliably determined from them. The worst result is with a fixed colour model. Both fixed and spatially adaptive methods have a tracking failure. The spatial adaptation method is prone to adapt something which is not skin, as demonstrated in the results for frame 300 of Table 19. It seems to track better when the illumination field is uniform (the pixels selected for model updating represent well the total colour distribution of the face) and the background does not contain similar colours. The fixed method works when the illumination field is quite stable.

For numerical evaluation of tracking results from the methods, three different metrics were used. The first one is the localization quality, for which the following equation can be used to express the goodness G of the tracking:

$$G = \frac{A_{GT} \bigcap A_c}{A_{GT} + A_c - A_{GT} \bigcap A_c} \tag{29}$$

where A_{GT} = area size for the manually selected ground truth bounding box, and
A_C = area size for the computed bounding box.

This metric evaluates the tracking goodness as an intersection of the found and ground truth bounding boxes against the total area which the bounding boxes cover.

From goodness measure *G* it is possible to calculate the error measure *E = 1-G*. Fig. 39 displays examples of a possible ground truth and computed bounding boxes at different goodness and error values. The error in tracking localization for video 1 tracked with three different methods is shown in Fig. 40. The static model has big errors when the object model is no longer valid. Adaptive methods produce similar results in frames but with geometrical constraint there are more spurious error spikes. The locus constraint produces the most stable performance.

Table 18. Adaptive face tracking with a skin locus (video 1).

Frame: 100

Frame: 200

Frame: 300

Frame: 400

Frame: 500

Frame: 600

Frame: 700

Frame: 900

Table 19. Adaptive face tracking with a spatial constraint (video 1).

Frame: 100	Frame: 200
Frame: 300	Frame: 400
Frame: 500	Frame: 600
Frame: 700	Frame: 900

Table 20. Face tracking with a static object model (video 1).

Frame: 100   Frame: 200

Frame: 300   Frame: 400

Frame: 500   Frame: 600

Frame: 700   Frame: 900

.

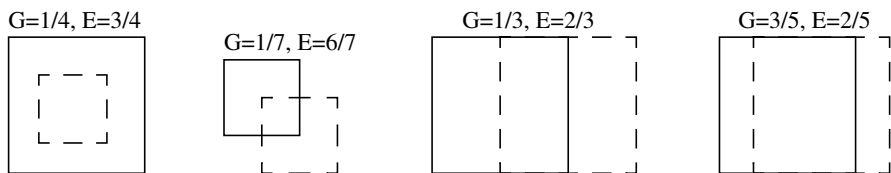G=1/4, E=3/4   G=1/3, E=2/3   G=3/5, E=2/5
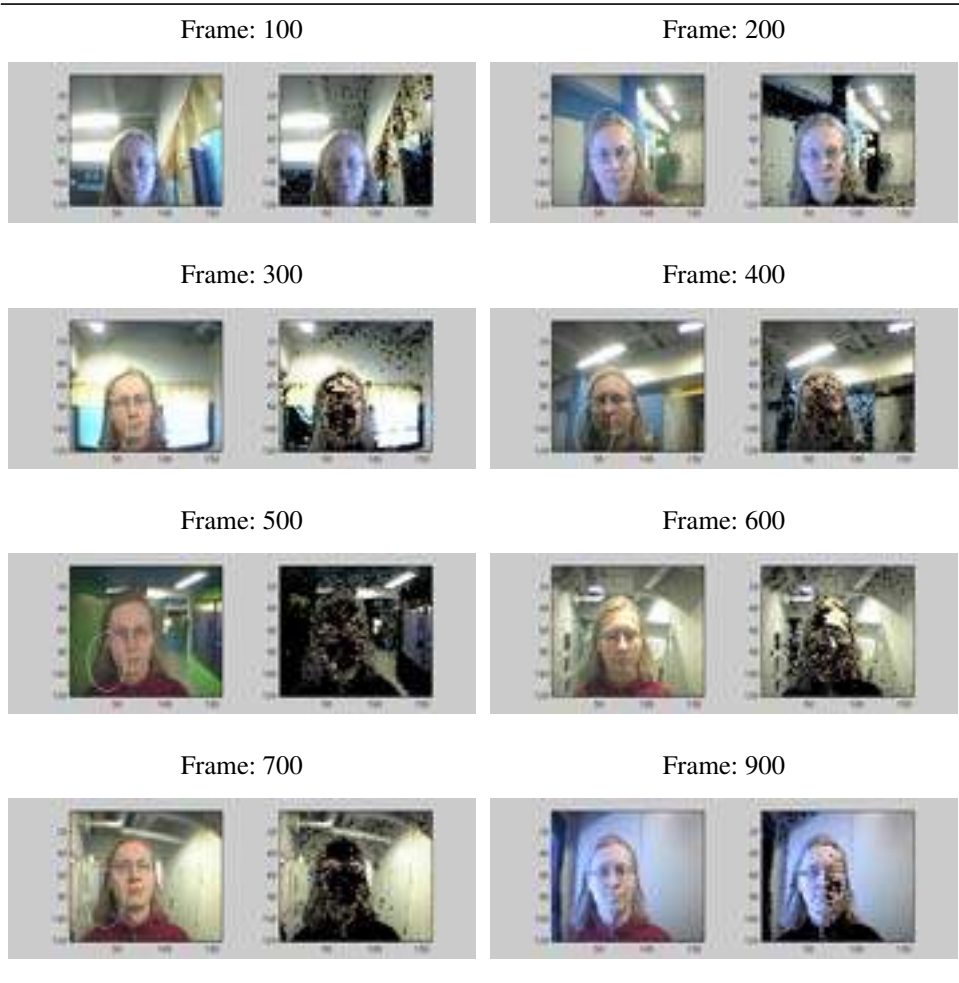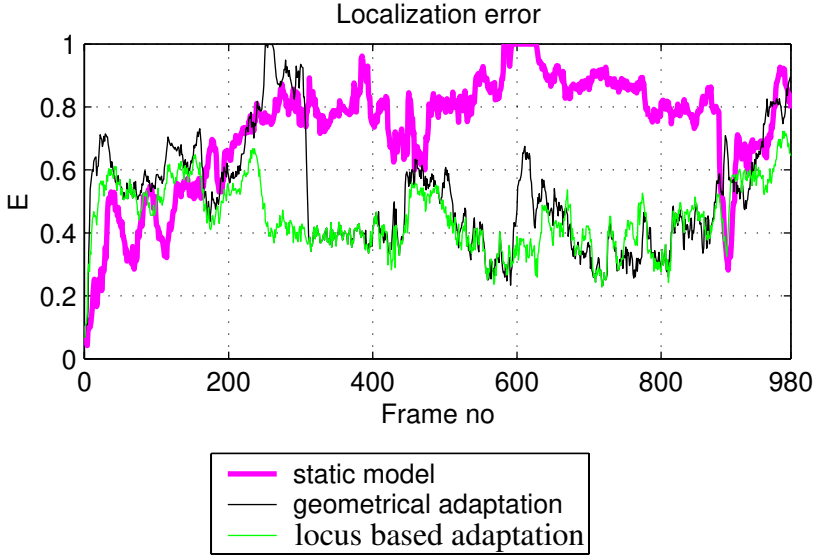   G=1/7, E=6/7

**Fig. 39. A few visualizations of the goodness and error measure.**

Fig. 40. Localization error of the bounding box for three different tracking methods with a static colour model, an adapted model based on geometrical constraint and an adaptive model based on chromaticity constraint.

The segmentation results were evaluated using the true positive and false positive metrics. The formula of true positive *TP* metric is

$$TP = \frac{N_{TP}}{N_{IN}}, \tag{30}$$

where N_{TP} = the number of detected face pixels, and
N_{IN} = the number of pixels inside the face polygon.

TP describes how many true objects pixels were found in relation to the total number of the object pixels. The nearer the value to one, the more object pixels are found. The value one means all object pixels were found. Fig. 41 shows *TP* metric results for video 1. The static model tracking found almost all face pixels when the illumination conditions are near those ones in the first frame. Because with geometrical constraint only pixels inside the face localization are used in colour model adaptation, not all facial chromaticities are present in the model even though they might belong to skin. This might be the reason for the tracking failure around frame 300: when the illumination field is nonuniform and changing, the selected pixels no longer correspond to true colour distribution. This also shows the susceptibility to adapt something else which is not skin. The operation of the locus based adaptation is most stable and best over the whole sequence even though the non-skin regions of the face were not removed.
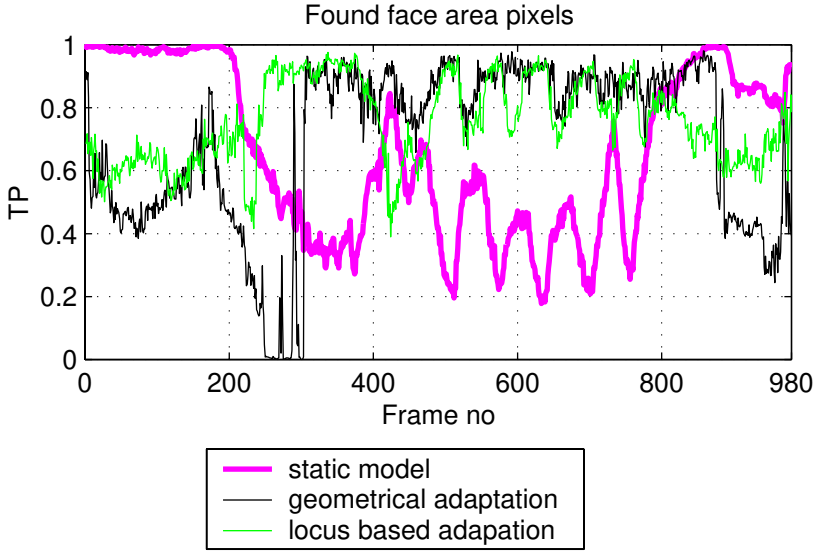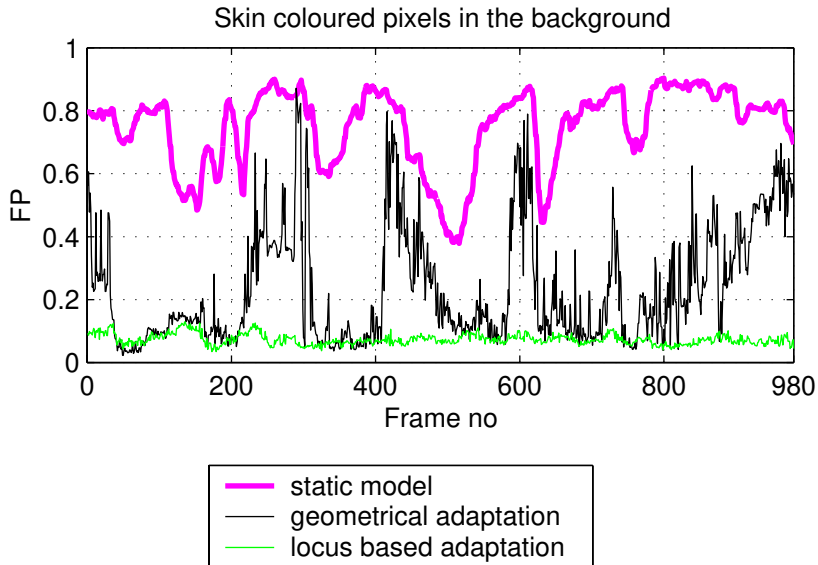
Fig. 41. True positives for video 1.

The false positives *FP* are defined as

$$FP = \frac{N_{FP}}{N_{OUT}},$$

(31)

where N_{FP} = the number of detected facially coloured pixels in the background, and N_{OUT} = the number of pixels outside the area defined by the face polygon.

FP indicates how many background pixels are activated by the colour model. When the *FP* value is zero or a value near zero and the *TP* value is big, then the object is well separable from the background, leading to better and more reliable localization. For video 1, the *FP* values can be seen in Fig. 42. The skin locus based method is superior because from the adaptation, those pixels which occur too often in the background are excluded from the colour model adaptation. Since the underlying assumptions of static model methods are uniqueness of the colour distribution in the scene and unchanging illumination, it gives the poorest results. Also the geometrically adaptive model relies on the assumption of the uniqueness of the chromaticities. As can be observed from Fig. 42, the uniqueness of chromaticities is highly dependent on the background and uniformness of the illumination field

over the face.

Fig. 42. False positives for video 1.

As a conclusion from the results presented in Figs. 40-42, the adaptive tracking method using the skin locus proved to have the most stable and best performance on a video sequence with true, drastic illumination changes, It was necessary to keep the bounding box size fixed because the static and geometrically adaptive tracking methods behaved unstably.

7 Conclusions

Colour is a very useful cue for object characterization, detection and localization. Because it is a low level feature, it is computationally inexpensive and therefore suitable for real-time applications. In addition, it is robust to occlusion and rotation. Unfortunately, colour has also a serious drawback, illumination dependency, and due to this it had limited use in practical applications. Many colour correction and colour constancy schemes have been presented but up to now, they have been unsuccessful for machine vision applications working in real world conditions. It is therefore important to develop techniques which can cope with these situations.

This thesis investigates the use of illumination robust colour techniques on facial skin. Faces were selected as a target because they are very important and common objects in images and video sequences. First, the data on the facial skin needs to be collected under varying illumination conditions. To do this, two databases have been created: the Physics-based Face Database and the Face Video Database. The Physics-based Face Database has images of 125 persons with pale, yellow and dark skin tones. The light sources and their SPDs are known. The images are taken with four white balancing illumination and four prevailing light sources. This makes it possible to compare not only between different canonical images but also between different unbalanced conditions. The skin chromaticities in these unbalanced images differ strongly from those of skin tones. It was also noticed that the direction of illumination colour temperature change effects the results. If the prevailing light source has a higher colour temperature, then the colours in the images shift towards blue. In the opposite case, a shift toward red is observed. In addition to images, the reflectances of skins were measured with a spectrophotometer. Because also the responses of the camera were known, this makes possible a physical modelling of image formation. The Face Video Database contains images as well as videos taken by several cameras. The images are taken under the same conditions as earlier with the Physics-based Face Database. The videos are imaged both under indoor and outdoor lighting conditions, with challenging illumination changes.

Next, the collected data, spectral reflectances and images are analysed and shown to be useful in different applications. The evaluation of skin spectral uniformity revealed that in general, it is quite uniform over the pure facial skin. In some cases, the cheeks' and forehead's reflectances have minor differences but this might be due to difference in the struc-

ture of the positions. The matt approximation of skin seems to be quite valid because the difference between SCI and SCE measurements was only 1-1.8 % per wavelength. This is very close to the result (5 %) given in the literature. The reflectances from facial skin seem to be very smooth, slowly varying and similar in shape for different skin tone groups.

Before studying the possibility to model skin colour change, the different colour spaces were evaluated. The choice of a proper colour space is an important part of any algorithm and it can be evaluated using criteria like discriminability of colours or effect of colour changes caused by illumination. For example, if the colours cannot be separated in RGB, they are not that even after conversion to a human colour space. On the other hand, this non-linear conversion can make two discriminable colours in RGB more difficult to separate. In addition, the transformation matrix to a human colour space is illumination dependent and therefore it cannot be used for colour constancy. Due to these reasons, one should consider very carefully the use of human colour space in those applications which do not require evaluation of human vision point of view. If the evaluation is important, then the method presented in Paper I can be used as a selection criteria for cameras. Because in our research evaluation was not important, the skin chromaticities from different balanced and unbalanced cases were investigated only in 17 device dependent colour spaces. The intensity data was excluded because it is very sensitive to environmental changes. It was noticed that many colour spaces are suitable for skin colour modelling if the criteria were overlapping between skin chromaticity regions of two different skin tone groups, the total area occupied in the space and the uniformity of the chromaticity blob in the space. The overlap of skin chromaticities has some dependence on the camera used and between all cameras it was not very high. But if the comparison was done between cameras with a gain controller and without a gain controller, the overlap increased. From these colour spaces, NCC rgb were selected for further use because of good overall performance. The skin chromaticities in NCC rg were found to be modellable with two quadratic functions and in NCC rb with three straight lines. The purpose of the modelling was to find a limited region for possible skin chromaticities under varying illumination (chromatic constraint). It was also shown that basis functions of skin colour signals can be used for this kind of modelling and evaluation of colour appearance under mixed illumination condition.

The knowledge obtained from data is shown to be useful in three applications. The first application is skin colour correction in colour images with severe overclipping. This technique requires for the overclipped image that there is an unclipped area and at least one channel is free from clipping in the other areas. The unclipped area is used to define ratios between channels and these ratios are used to approximate the pixels' values in overclipped channels. After removing overclipping, the image is subjected to PCA. To correct skin colours so that they have canonical appearances, the first coefficients of the obtained eigen-faces are replaced by those obtained after applying PCA to a canonical image. In the second application, faces of persons are tracked in video sequences under drastic illumination changes using a chromaticity constraint. The chromaticity constraint is used to select pixels for updating the skin colour model of the face to be tracked. It does not assume any probability for a skin colour; it just defines if a camera can perceive a skin with a chromaticity pair. This adaptive tracking is shown to be superior to fixed model based tracking and to adaptive tracking with spatial pixel selection. In the last application, the tracking and segmentation are combined: the chromatic constraint is used to filter out non-skin coloured pixels to achieve the segmented image. The segmented image is again used in face tracking,

and the localization is utilized in updating the colour model for the face and approximate the possible range of chromaticities in the next frame.

These results confirm the validity of the main thesis statement: knowledge about an object's colour, like skin colour changes under different illumination conditions, can be used to develop more robust techniques against illumination changes.

References

Abdel-Mottaleb M & Elgammal A (1999) Face detection in complex environments from color images. Proc. IEEE International Conference on Image Processing, Kobe, Japan, 3: 622-626.

Angelopoulou E (2001) Understanding the color of human skin. SPIE 4299 Human Vision and Electronic Imaging VI, 243-251.

Belhumeur P, Hespanha J & Kriegman D (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7): 711-720.

Berens J & Finlayson GD (2000) Log-opponent chromaticity coding of colour space. Proc. 15th International Conference on Pattern Recognition, Barcelona, Spain, 1: 206-211.

Bergasa LM, Mazo M, Gardel A, Sotelo MA & Boquete L (2000) Unsupervised and adaptive Gaussian skin-color model. Image and Vision Computing 18(12): 987-1003.

Brainard DA & Wandell BA (1986) Analysis of the retinex theory of color vision. Journal of the Optical Society of America A 3(10): 1651-1661.

Bridgeman T & Hudson N (1969) Colour 69. Göttingen.

Broncolor FCC operating instructions. Bron Elektronik AG, Switzerland (written document).

Buchsbaum G (1980) A spatial processor model for object color perception. Journal of the Franklin Institute 310(1): 1-26.

Buck GB II & Froelich HC (1948) Color characteristics of human complexions. Illumination Engineering 43: 27-49.

Cai J & Goshtasby A (1999) Detecting human faces in color images. Image and Vision Computing 18(1): 63-75.

Chai D & Ngan KN (1998) Locating facial region of a head-and-shoulders color image. Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 124-129.

Chang YC & Reid JF (1996) RGB calibration for color image analysis in machine vision. IEEE Transactions on Image Processing 5(10): 1414-1422.

Chen Q, Wu H & Yachida M (1995) Face detection by fuzzy pattern matching. Proc. 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA, 591-596.

Cho KM, Jang JH & Hong KS (2001) Adaptive skin-color filter. Pattern Recognition 34(5): 1067-1073.

Choudhury R & Chatterjee S (1992) Quantifying metamerism. Review of Progress in Coloration 22:

42-27.

Choudhury R & Chatterjee S (1996) Evaluation of the performance of metameric indices. Color Research and Application 21(1): 26-34.

CIE (1974) Method of Measuring and Specifying Colour Rendering Properties of Light Sources (2nd ed.). CIE No. 13.2 (TC-4.2), Bureau Central de la CIE, Paris..

Comaniciu D & Ramesh V (2000) Robust detection and tracking of human faces with an active camera. Proc. 3rd IEEE International Workshop on Visual Surveillance, Dublin, Ireland, 11-18.

Comaniciu D, Ramesh V & Meer P (2000) Real-time tracking of non-rigid objects using mean shift. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA, 2: 142-149

Crowley JL & Bedrune JM (1994) Integration and control of reactive visual processes. Proc. 3rd European Conference on Computer Vision, Stockholm, Sweden, 2: 47-58.

Crowley JL & Berard F (1997) Multi-modal tracking of faces for video communications. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 640-645.

Dai Y & Nakano Y (1995) Extraction for facial images from complex background using color information and SGLD matrices. Proc. 1st International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, 238-242.

Dai Y & Nakano Y (1996) Face-texture model based on SGLD and its application in face detection in a color scene. Pattern Recognition 29(6): 1007-1017.

Debevec P, Hawkins T, Tchou C, Duiker HP, Sarokin W & Sagar M (2000) Acquiring the reflectance field of a human face. Proc. SIGGRAPH 27th International Conference on Computer Graphics and Interactive Techniques, Los Angeles, California, USA, 145-156.

DeCusatis C (1998) Handbook of Applied Photometry. OSA and Springer-Verlag, New York.

Edwards EA & Duntley SQ (1939) The pigments and color of living human skin. The American Journal of Anatomy 65(1): 1-33.

Finlayson GD, Dueck J, Funt BV & Drew MS (1996) Color eigenfaces. Proc. 3rd International Workshop on Image and Signal Processing, Manchester, United Kingdom, 607-610.

Foley JD, van Dam A, Feiner SK & Hughes JF (1996) Computer Graphichs Principles and Practice: Second Edition in C. Addison-Wesley, New York.

Fortner B & Meyer TE (1997) Number by Colors. Springer-Verlag, New York.

Funt B, Barnard K & Martin L (1998) Is machine colour contancy good enough?. Proc. 5th European Conference on Computer Vision, University of Freiburg, Germany, 445-459.

Georghiades AS, Belhumeur PN & Kriegman DJ (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6): 643-660.

Gevers T & Smeulders AWM (1999) Color-based object recognition. Pattern Recognition 32(3): 453-464.

Gong S, McKenna SJ & Psarrou A (2000) Dynamic Vision from Images to Face Recognition. Imperial College Press, London.

Goudail F, Lange E, Iwamoto T, Kyuma K & Otsu N (1996) Face recognition system using local autocorrelations and multiscale integration. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(10): 1024 -1028.

Graf HP, Chen T, Petajan E & Cosatto E (1995) Locating faces and facial parts. Proc. 1st International Workshop Automatic Face and Gesture Recognition, Zurich, Switzerland, 41-46.

Graf HP, Cosatto E, Gibbon D, Kocheisen M & Petajan E (1996) Multimodal system for locating

heads and faces. Proc. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, USA, 88-93.

Graham DB & Allinson NM (1998) Characterizing virtual eigensignatures for general purpose face recognition. In: Wechsler H, Phillips PJ, Bruce V, Fogelman-Soulie F & Huang TS (eds) Face Recognition: From Theory to Applications 163: 446-456. NATO ASI Series F, Computer and Systems Sciences, Springer-Verlag, New York.

Hallinan P (1995) A deformable model for face recognition under arbitrary lighting conditions. PhD. thesis, Harvard University.

Harwood LA (1976) A chrominance demodulator IC with dynamic flesh correction. IEEE Transactions on Consumer Electronics CE-22: 111-117.

Hauta-Kasari M, Parkkinen J & Jääskeläinen T (2000) Unsupervised spectral image segmentation. Proc. 1st International Conference on Color in Graphics and Image Processing, Saint-Etienne, France, 84-88.

Hidai K, Mizoguchi H, Hiraoka K, Tanaka M, Shigehara T & Mishima T (2000) Robust face detection against brightness fluctuation and size variation. Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, Takamatsu, Japan, 2: 1379-1384.

Holst GC (1998) CCD Arrays Cameras and Displays. SPIE Optical Engineering Press, Bellingham.

Hsu RL, Abdel-Mottaleb M & Jain AK (2002) Face detection in color images. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5): 696-706.

Hunke M & Waibel A (1994) Face locating and tracking for human-computer interaction. Proc. Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, USA, 2: 1277-1281.

Hunt RWG (1987) Measuring Colour. John Wiley & Sons, Chichester.

Hyvärinen A, Karhunen J & Oja E (2001) Independent Component Analysis. John Wiley & Sons, New York.

Imai FH, Tsumura N, Haneishi H & Miyake Y (1996) Principal Component Analysis of Skin Color and Its Application to Colorimetric Reproduction on CRT Display and Hardcopy. The Journal of Imaging Science Technology 40(5): 422-430.

Jebara TS & Pentland A (1997) Parametrized structure from motion for 3D adaptive feedback tracking of faces. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 144-150.

Jebara TS, Russell K & Pentland A (1998) Mixture of eigenfeatures for real-time structure from texture. Proc. 6th International Conference on Computer Vision, Bombay, India, 128-135.

Jones S, Martin R & Pilbeam D (1992) The Cambridge Encyclopedia of Human Evolution. Cambridge University Press, Cambridge.

Jones MJ & Rehg JM (1999) Statistical color models with application to skin detection. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA, 1: 274-280.

Jones MJ & Rehg JM (2002) Statistical color models with application to skin detection. International Journal of Computer Vision 46(1): 81-96.

Karlekar J & Desai UB (1999) Finding faces in color images using wavelet transform. Proc. 10th International Conference on Image Analysis and Processing, Venice, Italy, 1085 -1088.

Kawato S & Ohya J (2000a) Automatic skin-color distribution extraction for face detection and tracking. Proc. 5th International Conference on Signal Processing, Beijing, China, 2: 1415-1418.

Kawato S & Ohya J (2000b) Real-time detection of nodding and head-shaking by directly detecting and tracking the between eyes. Proc. 4th IEEE International Conference on Automatic Face and

Gesture Recognition, Grenoble, France, 40-45.

Kim SH, Kim NK, Ahn SC & Kim HG (1998) Object oriented face detection using range and color information Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 76-81.

Kjeldsen R & Kender J (1996) Finding skin in color images. Proc. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, USA, 312-317.

Klette R, Schlüns K & Koschan A (1998) Computer Vision Three-Dimensional Data from Images. Springer-Verlag, Singapore.

Klinker G (1993) A Physical Approach to Color Image Understanding. AK Peters, Wellesley.

Laamanen H, Jääskeläinen T & Parkkinen JPS (2000) Comparison of PCA and ICA in color recognition. SPIE 4197 Intelligent Robots and Computer Vision, 367-377.

Land EH (1977) The retinex theory of color vision. Scientific American 237(6): 108-129.

Land EH (1986) Recent advances in retinex theory. Vision Research 26: 7-21.

Land EH & McCann JJ (1971) Lightness and retinex theory. Journal of the Optical Society of America 61(1): 1-11.

Lee EJ & Ha YH (1997) Automatic flesh tone reappearance for color enhancement in TV. IEEE Transactions on Consumer Electronics 43(4): 1153-1159.

Lenz R, Meer P & Hauta-Kasari M (1999) Spectral-based illumination estimation and color correction. Color Research and Application 24(2): 98-111.

Lomheim TS & Kalman LS (1992) Analytical modeling of digital simulation of scanning charge-coupled device imaging systems. In: Karim MA (ed) Electro-optical Displays. Marcel Dekker, New York.

Loui AC, Judice CN & Liu S (1998) An image database for benchmarking of automatic face detection and recognition algorithms. Proc. IEEE International Conference on Image Processing, Chicago, Illinois, USA, 1: 146-150.

Luo H & Eleftheriadis A (2000) On face detection in the compressed domain. Proc. 8th ACM International Multimedia Conference, Los Angeles, California, USA, 285-294.

Macbeth (1997) Macbeth Spectralight-II. Division of Kollmorgen Instruments Corporation, United Kingdom.

Maloney LT (1986) Evaluation of linear models of surface spectral reflectance with small number of parameters. Journal of the Optical Society of America A 3(10): 1673-1683.

Maloney LT & Wandell BA (1986) Color constancy: a method for recovering surface spectral reflectance. Journal of the Optical Society of America A 3(1): 29-33.

Martinez AM & Benavente R (1998) The AR Face Database. CVC Technical Report #24, AVL, Purdue University.

Matas J, Marik R & Kittler J (1994) Illumination invariant colour recognition. Proc. 5th British Machine Vision Conference, University of York, York, 469-479.

McKenna S, Raja Y & Gong S (1999) Tracking colour objects using adaptive mixture models. Image and Vision Computing 17(3-4): 225-231.

Menser B & Müller F (1999) Face detection in color images using principal component analysis. Proc. 7th International Conference on Image Processing and Its Applications (Conf. Publ. No. 465), Manchester, United Kingdom, 2: 620-624.

Minolta (1991) Spectrophotometer CM-2002 instruction manual. Minolta Camera Co., Japan.

Minolta (1996) Spectroradiometer CS-1000 instruction manual. Minolta Co., Japan.

Miyake Y & Miyata K (1999) Color image processing based on spectral information and its application. Proc. IEEE International Conference on Image Processing, Kobe, Japan, 3: 41-44.

Miyake Y, Saitoh H, Yaguchi H & Tsukada N (1990) Facial pattern detection and color correction from television picture for newspaper printing. Journal of Imaging Technology 16(5): 165-169.

Moon H & Phillips PJ (1998) Analysis of PCA-based face recognition algorithms. Proc. Workshop on Empirical Evaluation Techniques in Computer Vision, Los Alamitos, California, 57-71.

Nakai H, Manabe Y & Inokuchi S (1998) Simulation and analysis of spectral distribution of human skin. Proc. 14th International Conference on Pattern Recognition, Brisbane, Queensland, Australia, 2: 1065-1067.

NCS (1989) NCS Natural Colour System® colour block with 8 fan-blocks. Scandinavian Colour Institute/Skandinaviska Färginstitutet Ab, Stockholm.

Nienstedt W, Hänninen O & Arstila A (1984) Ihmisen Fysiologia ja Anatomia. WSOY, Porvoo.

Nimeroff I & Yurow J (1965) Degree of metamerism. Journal of the Optical Society of America 55(2): 185-190.

Novak CL & Shafer SA (1992) Supervised color constancy for machine vision. In: Healey G, Shafer SA & Wolff L (eds.) Physics-Based Vision Principles and Practices Color. Jones and Barlett Publishers, London, United Kingdom, p. 284-299.

Novak CL, Shafer SA & Wilson R (1992) Obtaining accurate color images for machine vision research. In: Healey G, Shafer SA & Wolff L (eds.) Physics-based Vision Principles and Practice: Color. Jones and Bartlett Publishers, London, United Kingdom, p. 13-27.

Ohta YI, Kanade T & Sakai T (1980) Color information for region segmentation. Computer Graphics and Image Processing 13(3): 222-241.

Ohtsuki T & Healey G (1998) Using color and geometric models for extracting facial features. The Journal of Imaging Science and Technology 42(6): 554-561.

Oliver N, Pentland A & Berard F (1997) Lafter: lips and face real time tracker. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 123-129.

Parkkinen JPS, Hallikainen J & Jääskeläinen T (1989) Characteristic spectra of Munsell colors. Journal of the Optical Society of America A 6(2): 318-322.

Parkkinen J & Jääskeläinen T (1989) Color vision: machine and human. SPIE 1199 Visual Communications and Image Processing IV, 1184-1192.

Philips Lamppuluettelo 1998-1999. Philips, Finland.

Phillips PJ, Moon H, Rizvi S & Rauss PJ (2000) The FERET evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10): 1090 -1104.

Pierce PE & Marcus RT (1994) Color and Appearance. Federation of Societies for Coatings Technology, Blue Bell.

Pigeon S & Vandendrope G (1997) The M2VTS multimodal face database. Proc. 1st International Conference on Audio- and Video-based Biometric Person Authentication, Crans-Montana, Switzerland, 403-409.

Piirainen T, Silvén O & Tuulos V (2000) Layered self-organizing maps based video content classification. Proc. Workshop on Real-time Image Sequence Analysis, Oulu, Finland, 89-98.

Powell MW, Sarkar S & Goldgof D (1999) Color correction using explicit illumination models, color and registered range. Proc. IEEE Workshop on Photometric Modeling for Computer Vision and Graphics, Fort Collins, Colorado, USA, 64-71.

Poynton CA (1996) A Technical Introduction to Digital Video. John Wiley & Sons, New York.

Qian RJ, Sezan MI & Matthews KE (1998) A robust real-time face tracking algorithm. Proc. IEEE International Conference on Image Processing, Chicago, Illinois, USA, 1: 131-135.

Raja Y, McKenna SJ & Gong G (1998) Tracking and segmenting people in varying lighting conditions using colour. Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 228-233.

Rasmussen C & Hager CG (1997) An adaptive model for tracking objects by color. Techical Report, DCS-TR1200, Yale University.

Redfield S & Harris JG (2000) The role of extreme color quantization in object recognition. Proc. International Conference on Color in Graphics and Image Processing, Saint-Etienne, France, 225-230.

Romero J, García-Beltrán A & Hernández-Andrés J (1997) Linear bases for representation of natural and artificial illuminants. Journal of the Optical Society of America A 14(5): 1007-1014.

Ryer A (1998) Light Measurement Handbook. International Light Inc., available from http://www.intl-light.com/handbook/.

Saber E & Tekalp AM (1998) Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. Pattern Recognition Letters 17(8): 669-680.

Sahbi H & Boujemaa N (2000) From coarse to fine skin and face detection. Proc. 8th ACM international conference, Marina del Rey, California, USA, 432-434.

Samaria F & Harter A (1994) Parameterisation of a stochastic model for human face identification. Proc. 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, 138-142.

Satoh S, Nakamura Y & Kanade T (1999) Name-it: naming and detecting faces in news videos. IEEE Multimedia 6(1): 22-35.

Satyanarayana S & Dalal S (1996) Video color enhancement using neural networks. IEEE Transactions on Circuits and Systems for Video Technology 6(3): 297-307.

Saxe D & Foulds R (1996) Toward robust skin identification in video images. Proc. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, USA, 379-384.

Schiele B & Waibel A (1995) Gaze tracking based on face-color. Proc. International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, 344-348.

Schuster R (1994) Color object tracking with adaptive modeling. Proc. IEEE Symposium on Visual Languages, St. Louis, Missouri, USA, 91-96.

Shafer SA (1992) Using color to separate reflection components. In: Healey G, Shafer SA & Wolff L (eds.) Physics-based vision principles and practices color. Jones and Barlett Publishers, London, United Kingdom, p. 43-51.

Shimizu H, Uetsuki K, Tsumura N & Miyake Y (2001) Analysing the effect of cosmetic essence by independent component analysis for skin color images. Proc. 3rd International Conference on Multispectral Color Science, Joensuu, Finland, 65-68.

Sigal L, Sclaroff S & Vassilis A (2000) Estimation and prediction of evolving color distribution for skin segmenation under varying illumination. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA, 2: 152-159.

Sobottka J & Pitas I (1996a) Segmentation and tracking of faces in color images. Proc. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, USA, 236-241.

Sobottka K & Pitas I (1996b) Face localization and feature extraction based on shape and color information. Proc. IEEE International Conference on Image Processing, Lausanne, Switzerland, 483-486.

So-Ling C & Ling L (2001) A multi-layered reflection model of natural human skin. Proc. Computer Graphics International, Hong Kong, China, 249-256.

Sony (1991) Sony 3 CCD Video Camera DXC-755P. Sony Corporation, Japan.

Sony (1989) Sony DXC-750P service manual TGR-750 3CCD Video camera title generator. Sony Corporation, Japan.

Soriano M, Marszalec E & Pietikäinen M (1999) Color correction of face images under different illuminants by RGB eigenfaces. Proc. 2nd Audio- and Video-based Biometric Person Authentication, Washington DC, USA, 148-153.

Stimson A & Fee E (1953) Color and reflectance of human flesh. Journal of the Society of Motion Picture and Television Engineers 60: 553-558.

Starner T & Pentland A (1996) Real-time ASL recognition from video using HMM's. Technical Report 375, MIT Media Lab.

Störring M, Andersen H & Granum E (1999) Skin color detection under changing lighting condition. Proc. 7th Symposium on Intelligent Robotics Systems, Coimbra, Portugal, 187-195.

Störring M, Andersen HJ & Granum E (2001) Physics-based modelling of human skin colour under mixed illuminants. Robotics and Autonomous Systems 35(3-4): 131-142.

Sun QB, Huang WM & Wu JK (1998) Face detection based on color and local symmetry information. Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 130-135.

Swain MJ & Ballard DH (1991) Color indexing. International Journal on Computer Vision 7(1): 11-32.

Terrillon JC, Niwa Y & Yamamoto K (2001) On the selection of an efficient chrominance space for skin color-based image segmentation with an application to face detection. Proc. International Conference on Quality Control by Artificial Vision, Bourgougne, France, 2: 409-414.

Terrillon JC, Shirazi MN, Fukamachi H & Akamatsu S (2000) Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 54-61.

Tominaga S (2000) Illuminant estimation of natural scenes from color images. Proc. International Conference Color in Graphics and Image processing, Saint-Etienne, France, 35-40.

Torres L, Reutter JY & Lorente L (1999) The importance of the color information in face recognition. Proc. IEEE International Conference on Image Processing, Kobe, Japan, 3: 627-631.

Tsapatsoulis N, Avrithis Y & Kollias S (2001) Facial image indexing in multimedia databases. Pattern Analysis and Applications 4(2-3): 93-107.

Tsumura N, Haneishi H & Miyake Y (1999) Independent-component analysis of skin color image. Journal of the Optical Society America A 16(9): 2169-2176.

Tsumura N, Miyake Y & Imai FH (2001) Medical vision: measurement of skin absolute spectral-reflectance-image and the application to component analysis. Proc. 3rd International Conference on Multispectral Color Science, Joensuu, Finland, 25-28.

Turk M & Pentland A (1991) Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1): 71-86.

TVI (1995) T.V.I. MSC-camera. TVI Temet Vision Industry, Helsinki.

Vertan C, Ciuc M & Boujemaa N (2000) On the introduction of a chrominance spectrum and its applications. Proc. International Conference on Color in Graphics and Image Processing, Saint-Etienne, France, 214-218.

Vora PL, Farrell JE, Tietz JD & Brainard DH (1997) Linear models for digital cameras. Proc. IS&T 50th Annual Conference, Cambridge, Massachusetts, USA, 377-382.

Wandell BA (1987) The synthesis and analysis of color images. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(1): 2-13.

Wang H & Chang SF (1997) A highly efficient system for automatic face region detection in MPEG video. IEEE Transactions on Circuits and Systems for Video Technology 7(4): 615-628.

Wu H, Chen Q & Yachida M (1999) Face detection from color images using a fuzzy pattern matching method. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(6): 557-563.

Wu Y & Huang TS (2000) Color tracking by transductive learning. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA, 1: 133-138.

Wyszecki G & Stiles WS (2000) Color Science Concepts and Methods, Quantitative Data and Formulae (2nd edition). John Wiley & Sons, New York.

Yang J, Stiefelhagen R, Meier U & Waibel A (1998) Visual tracking for multimodal human computer interaction. Proc. Conference on Human Factors and Computing Systems, Los Angeles, California, USA, 140-147.

Yang J & Waibel A (1996) A real-time face tracker. Proc. 3rd IEEE Workshop on Applications of Computer Vision, Florida, USA, 142 -147.

Yang MH & Ahuja N (1998) Detecting human faces in color images. Proc. IEEE International Conference on Image Processing, Chicago, Illinois, USA, 1: 127-130.

Yang MH & Ahuja N (2001) Face Detection and Gesture Recognition for Human-computer Interaction. Kluwer Academic Publishers, New York.

Yoo TW & Oh IS (1999) A fast algorithm for tracking human faces based on chromatic histograms. Pattern Recognition Letters 20(10): 967-978.

Zarit BD, Super BJ & Quek FKH (1999) Comparison of five color models in skin pixel classification. Proc. International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece, 58-63.

Zhang X & Wandell BA (1996) A spatial extension of CIELAB for digital color reproduction. Proc. Society for Information Display Symposium, San Diego, USA, 731-734.

Transforms from RGB to other colour spaces

The device dependent color spaces can be divided into two different groups depending on whether they can be obtained from RGB values by a linear transformation or by a nonlinear transformation. Color spaces obtained by a linear transform from RGB are I1I2I3 (Ohta's features), YES, YIQ, YUV, YCrCb1 (Rec 601-5), and YCbCr2 (Rec 709) whereas the following ones use nonlinear transforms: NCC rgb or Normalized Color Coordinates rgb, modified rgb, ln-chromaticity (referred in this thesis as log-chroma color space), P1P2, l1l2l3, ratio between channels (G/R, B/R and B/G), HSV, HSL, ab, TSL and Yuv.

The RGB data was scaled so that values were between 0 and 1 except in the case of logchroma in which case the range was 0-255. The scaling is not essential for most of the colour spaces. Some transformations needed preprocessing for certain RGB values: 1) pixels with 0 intensity - a special handling was needed for transformation to NCC rgb and these pixels were removed before transformation to modified rgb, P1P2, uv and ab; 2) pixels with achromatic color - transformations to HSV and HSL needed special attention before calculating output for achromatic colors and it was necessary to remove these pixels before making transformation to l1l2l3; 3) pixels with a 0 green channel - they were removed before transformation to ab color space, and 4) one or more channels have 0 value - for transformation to logchroma space these pixels were removed.

COLOR SPACES BY A LINEAR TRANSFORM FROM RGB

I1I2I3 or Ohta's features (Ohta *et al.* 1980) were first introduced for segmentation as optimized color features and are shown in Eqs. A1.1-3:

$$I1 \ = \ \frac{(R + G + B)}{3} \tag{A1.1}$$

$$I2' \ = \ (R - B) \tag{A1.2}$$

$$I3' \ = \ \frac{(2G - R - B)}{2} \tag{A1.3}$$

YES space (Eqs. A1.4-6) has been developed by Xerox:

$$Y = 0.253R + 0.684G + 0.063B \tag{A1.4}$$

$$E = 0.500R - 0.500B + 0.000B \tag{A1.5}$$

$$S = 0.250R + 0.250G - 0.500B \tag{A1.6}$$

YIQ space, which is derived from YUV, can be used optionally by the NTSC composite TV standard presented in Eqs. A1.7-9:

$$Y = 0.30R + 0.59G + 0.11B \tag{A1.7}$$

$$I = 0.60R - 0.28G + 0.31B \tag{A1.8}$$

$$Q = 0.21 - 0.52G + 0.31B \tag{A1.9}$$

YUV space in Eqs.A1.10-12 is the basic format for the composite color television standard for NTSC, PAL and SECAM. uv space as shown in Eqs. A1.13-14 and Eq. A1.10 has been presented as luminance normalized UV color information:

$$Y = 0.299R + 0.587G + 0.114B \tag{A1.10}$$

$$U = -0.147R - 0.289G + 0.437B \tag{A1.11}$$

$$V = 0.615R - 0.515G - 0.100B \tag{A1.12}$$

$$u = U/Y \tag{A1.13}$$

$$v = V/Y \tag{A1.14}$$

General formulae for calculating YCbCr are shown in Eqs. A1.15-17 and specific coefficients for different standards are in Table 1. The general YCbCr definition is:

$$Y = c_1 R + c_2 G + c_3 B \tag{A1.15}$$

$$Cb = \frac{B - Y}{2 - 2c_3} \tag{A1.16}$$

$$Cr = \frac{R - Y}{2 - 2c_1} \tag{A1.17}$$

Table 1: Coefficients for two YCbCr standards

Standard	Red coefficient c_1	Green coefficient c_2	Blue coefficient c_3
Rec.601	0.2989	0.5866	0.1145
Rec 709	0.2126	0.7152	0.0722

The chromaticities for the normalized colour coordinates NCC or normalized RGB (Eqs. A1.18-19)are obtained by normalizing the RGB with intensity I (Wyszecki & Stiles 2000):

$$I = R + G + B \tag{A1.18}$$

$$x = \frac{X}{I} \tag{A1.19}$$

where x = r,g,b and X = R, G, B respectively.

As shown by Eqs. A1.19-20 for modified rgb space (Tominaga 2000), the definition of intensity differs from the one of the NCC :

$$I = \sqrt{R^2 + G^2 + B^2} \tag{A1.20}$$

Log-chromaticity space (Berens & Finlayson 2000) has been applied to image indexing and is presented in Eqs. A1.21-22.

$$r\_g = \frac{\ln R}{\ln G} = \ln R - \ln G \tag{A1.21}$$

$$rg\_b = \frac{\ln(R \cdot G)}{\ln B^2} = \ln R + \ln G - 2\ln B \tag{A1.22}$$

P1P2 color space (Vertan *et al*. 2000) has been used for the making of a chromaticity Fourier spectrum:

$$P_1 = \frac{1}{\sqrt{2}} \frac{G - R}{R + G + B} \tag{A1.23}$$

$$P_2 = \frac{1}{\sqrt{6}} \frac{2B - R - G}{R + G + B} \tag{A1.24}$$

l1l2l3 color space (Gevers & Smeulders 1999) as presented in Eq. A1.25-27 has been used for color-based object recognition.

$$l_1 = \frac{(R - G)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2} \tag{A1.25}$$

$$l_2 = \frac{(R - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2} \tag{A1.26}$$

$$l_3 = \frac{(G - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2} \tag{A1.27}$$

Foley *et al*. (1996) give for HSV Eqs. A1.28-30 which show one of the possible transformation methods:

$$V = max(R, G, B) \tag{A1.28}$$

$$S = \begin{cases} \dfrac{max(R, G, B) - min(R, G, B)}{max(R, G, B)}, max(R, G, B) \neq 0 \\ \\ 0, max(R, G, B) = 0 \end{cases} \tag{A1.29}$$

$$H = acos \dfrac{0.5((R - G) + (R - B))}{\sqrt{(R - G)(R - G) + (R - B)(G - B)}} \tag{A1.30}$$

and in the case of B>G, then H = 360-H. (A1.31)

Many people find HS-spaces (HSL, HSV, HSB, HSI) intuitive for colour definition. In Foley *et al*. (1996), HSL (Eqs. A1.32-33, Eq. A1.30) color space differs from HSV on how S and L are defined:

$$L = \dfrac{max(R, G, B) + min(R, G, B)}{2} \tag{A1.32}$$

$$S = \begin{cases} \dfrac{max(R, G, B) - min(R, G, B)}{max(R, G, B) + min(R, G, B)}, L \leq 0.5 \\ \\ \dfrac{max(R, G, B) - min(R, G, B)}{2 - (max(R, G, B) + min(R, G, B))}, L > 0.5 \end{cases} \tag{A1.33}$$

Kawato and Ohya (2000b) have used ab space which is derived from NCC rg-chromaticities and shown in Eq. A1.34. (Note: in their ab space : b=((3 1/2)/2)*g).:

$$a = r + \dfrac{g}{2} \tag{A1.34}$$

$$b = \dfrac{\sqrt{3}}{2g} \tag{A1.35}$$

TSL (Tint - Saturation - Lightness) colorspace was developed by Terrillon *et al*. (2000) and it is also derived from NCC rg-chromaticities as can be seen from Eqs. A1.36-38:.

$$\begin{aligned} r' &= r - 1/3 \\ g' &= g - 1/3 \end{aligned} \tag{A1.36}$$

$$T = \begin{cases} arc\tan\left(\dfrac{r'}{g'}\right)/(2\pi) + \dfrac{1}{4}, \ g' > 0 \\ \\ arc\tan\left(\dfrac{r'}{g'}\right)/(2\pi) + \dfrac{3}{4}, \ g' < 0 \\ \\ 0, \ g' = 0 \end{cases} \tag{A1.37}$$

$$S = [9 \cdot (r'^2 + g'^2)/5] \tag{A1.38}$$

Visualization of skin chromaticities at different colour spaces
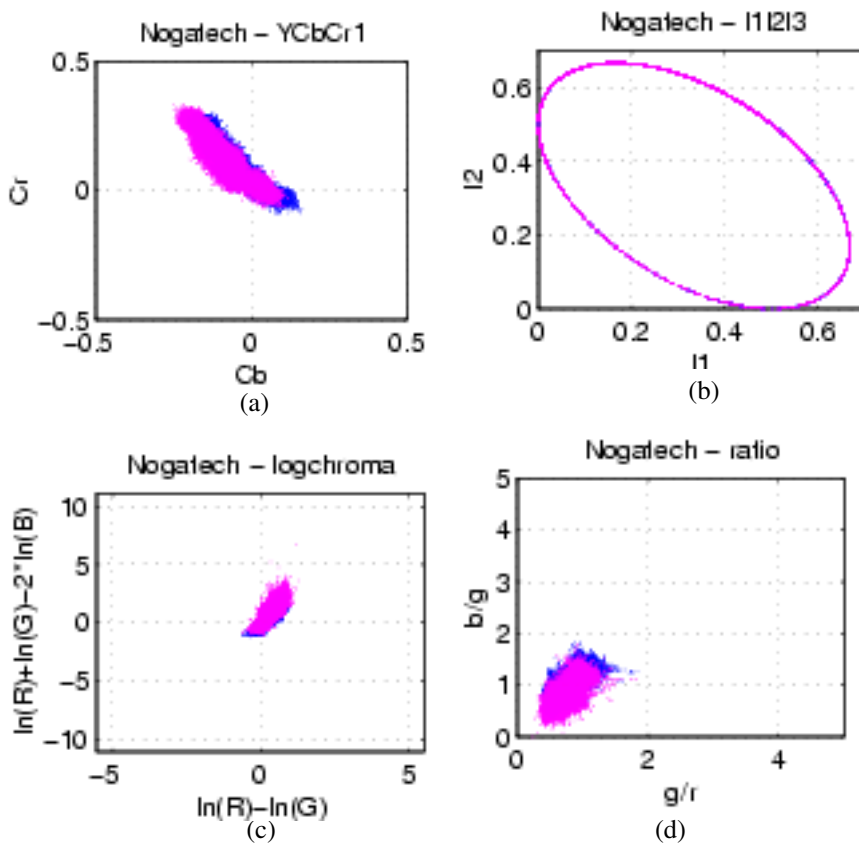

(a)

(b)

(c)

(d)

Fig. A2.1. displays images of skin chromaticities plotted at different colour spaces (continues),
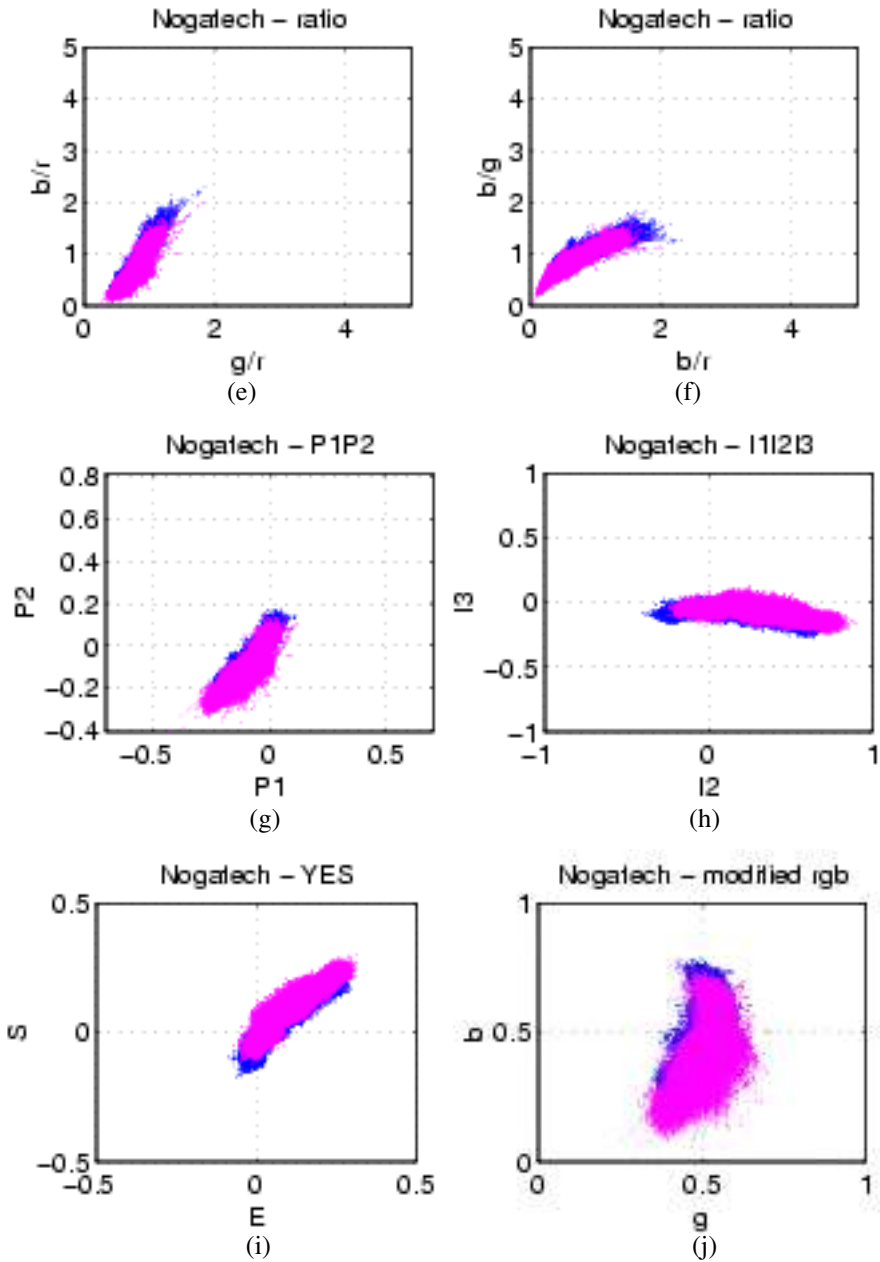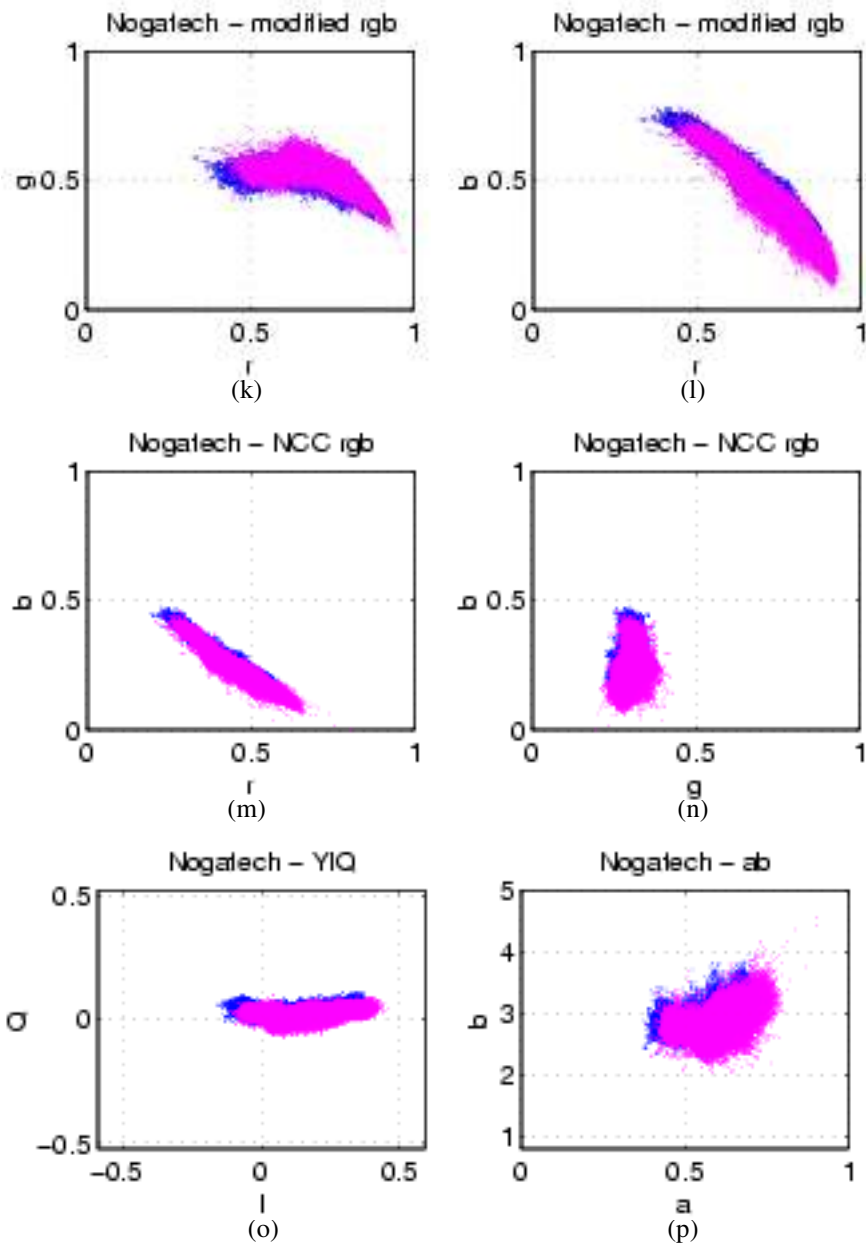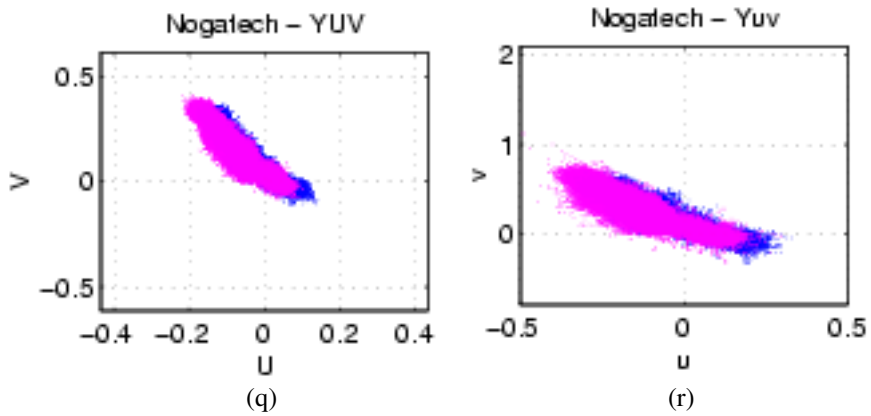
Fig. A2.1. Continues,

Fig. A2.1. Continues,

(q)

(r)

Fig. A2.1. Plotted skin chromaticities at different colour spaces: (a) YCbCr1, (b) l1l2l3, (c) logchroma, (d) ratio b/g-g/r, (e) ratio b/r-g/r, (f) ratio b/g-b/r, (g) P1P2, (h) Ohta I1I2I3, (i) YES, (l) modified rgb, (j) modified rgb, (k) modified rgb, (l) modified rgb, (m) NCC rgb, (n) NCC rgb, (o) YIQ, (p) ab, (q) YUV, and (r) Yuv.

Mean shift algorithm

The mean shift algorithm presented here was used by Comaniciu et al. (2000) and Comaniciu and Ramesh (2000) for face tracking. The multivariate kernel density estimate is

$$\hat{f}(x) \ = \ \frac{1}{nh^d} \sum_{i \, = \, 1}^{n} K\!\left(\frac{x - x_i}{h}\right) \tag{A3.1}$$

where x = a point,
n = number of points,
K = selected kernel,
d = dimension of the space, and
h = window radius or bandwith.
The kernel is Epanechnikov kernel, and its formula is

$$K_E(x) \ = \ \begin{cases} \dfrac{1}{2} C_d^{-1}(d + 2)(1 - \|x\|^2), \ \ if \ \ \|x\| < 1 \\[2mm] \qquad 0, \ \ otherwise, \end{cases} \tag{A3.2}$$

where C_d = the volume of the unit d-dimensional sphere.
The colour distribution of the object is converted to 1D distributions. The multivariate density estimate is used to weight colours based on their appearance inside the kernel. A Bhattacharyya coefficient is used to evaluate the difference δ between two distributions:

$$\delta(y) \ = \ \sqrt{1 - \sum_{u \, = \, 1}^{m} \sqrt{p(y)q}}, \tag{A3.3}$$

where q = target model,
p = calculated distribution, and
y = location.
The tracking starts initialization from which the target histogram is set. In the next frame, the surroundings of former localization is sought to find a position in which the dif-

ference δ is smaller than a certain threshold.

Errata

Paper 1
   Page 349, 4th paragraph: $\lambda_1 = 400$ nm should be $\lambda_1 = 700$ nm.
   Page 353, 3rd paragraph: Birdgeman's should be Bridgeman's.
   Page 353, 3rd paragraph, 16th row: Table 2 should be Table 3.
   Page 356, Table 6: $\Delta E^*_{ab} <= 1.5$ should be $\Delta E^*_{ab} <= 3.0$.

Paper II
   Page 35, 3rd paragraph: Fig. 1 should be Fig. 2.

Paper VIII
   Page 44, 3rd paragraph: [5] should be [4].
   Page 47, 2nd paragraph: vol. 2, pp. 142-149 should be pp.11-18.
   Page 48, 1st paragraph: pp. 11-18 should be pp. 142-149.