

A SELF-CALIBRATING CHROMINANCE MODEL APPLIED TO SKIN COLOR DETECTION

J. F. Lichtenauer, M. J. T. Reinders and E. A. Hendriks

Information & Communication Theory Group, Delft University of Technology, Mekelweg, Delft, The Netherlands

Keywords: Adaptive color modelling, chrominance, chromatic color space, skin detection.

Abstract: In case of the absence of a calibration procedure, or when there exists a color difference between direct and ambient light, standard chrominance models are not completely brightness invariant. Therefore, they cannot provide the best space for robust color modeling. Instead of using a fixed chrominance model, our method estimates the actual dependency between color appearance and brightness. This is done by fitting a linear function to a small set of color samples. In the resulting self-calibrated chromatic space, orthogonal to this line, the color distribution is modeled as a 2D Gaussian distribution. The method is applied to skin detection, where the face provides the initialization samples to detect the skin of hands and arms. A comparison with fixed chrominance models shows an overall improvement and also an increased reliability of detection performance in different environments.

1 INTRODUCTION

Color is an important property of many objects. Therefore, it has been of great interest to researchers in the field of image analysis since the introduction of digital color images. However, the lack of constancy of color between different lightings, camera equipment and settings has challenged researchers ever since. To reduce the problem of color variation between different scenarios, color can be modeled adaptively. But even in the same scenario, color can change due to changing lighting conditions. A significant amount of research has been conducted to find functions of color that are invariant to illumination change. The most prominent functions that are commonly used are the chromatic color models. E.g. normalized rgb, YUV, YCrCb, HSV or CIELAB. In these color spaces, the brightness factor is isolated from the chromatic representation of color. This facilitates adaptive color modeling that is invariant to changes in illumination brightness.

However, brightness invariance is not guaranteed in these models. All these chromatic color space conversions are based on certain assumptions about color appearance in RGB space. Normalized rgb,

HSV and CIELAB assume that black is represented by $[R, G, B] = [0, 0, 0]$ and, as a result, all colors meet each other at this point when their brightness is reduced. Contradictory, YUV and YCrCb assume that a brightness change results in a change of color parallel to the diagonal of RGB space. Furthermore, HSV assumes that gray (unsaturated) colors satisfy $R = G = B$ (correct white balance) and CIELAB even needs a completely calibrated RGB (XYZ) space.

Violation of these assumptions, e.g. due to incorrect white balance, non-ideal camera sensitivity or settings or heterogeneous illumination, can severely degrade performance of color analysis methods based on these color spaces. This was shown by Martinkauppi et al. (Martinkauppi et al., 2003), who have tested robustness of different skin detection methods under large changes of conditions. To our knowledge, the validity in real-world digital imaging situations of the chrominance models that are often used for these methods, and the consequences of violating the model assumptions, have never been studied directly.

If the color spectrum of the illumination is homogeneous, the assumptions can be satisfied by a calibration procedure. However, in many real-world applications, such a procedure is not desirable because

F. Lichtenauer J., J. T. Reinders M. and A. Hendriks E. (2007).

A SELF-CALIBRATING CHROMINANCE MODEL APPLIED TO SKIN COLOR DETECTION.

In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - ICFIA*, pages 115-120

Copyright © SciTePress

it takes extra time and effort. It may not even be possible, e.g. in case of diversity of unknown cameras, non-expert users or processing of video previously recorded under non-ideal circumstances. Furthermore, calibration may not be effective, because of a difference between spectra of direct and ambient light, resulting in correlation between chrominance and brightness.

To avoid the shortcomings of chromatic models in real-world scenarios, we don't want to rely on constraints on white balance, origin offset or correlation between brightness and chrominance. Therefore, we take the principle of chromatic color representation one step further, by presenting our Adaptive Chrominance Space (ACS) method that adaptively finds a linear function that minimizes correlation between brightness and the chrominance representation of the appearance of a specific object color.

Our application for this color model is hand gesture recognition. Skin color is modeled from samples obtained from face detection and applied to detect and track a person's hands. Furthermore, two models are combined, fitted to samples from the left and right part of the face, respectively. This increases robustness to variation in illumination spectra from different horizontal directions.

Section 2 contains a summary of related research on skin color detection, section 3 describes our method, for which experimental results are provided in section 4. Our conclusions are given in section 5.

2 RELATED WORK

Many methods for skin color detection have been proposed in the literature. Surveys can be found in (Yang et al., 2002; Vezhnevets et al., 2003). Most methods learn a general skin color distribution from a database of images (Jones and Rehg, 2002; Phung et al., 2005; Lee and Yoo, 2002), e.g. a selection from the internet. This results in very general models that take into account variation of camera, camera settings and illumination. However, because the variation of skin color appearance between different situations is so large, these general models lack precision to distinguish between real skin and skin-colored objects. This greatly restricts the reliability of skin segmentation, since the false positive rate will be high when other colors in the image are close to skin color.

To overcome this problem, some methods adapt to the specific situation by learning the skin color distribution from samples taken from the same video (Raja et al., 1998; McKenna et al., 1999; Soriano et al., 2000; Fritsch et al., 2002; Argyros and Lourakis,

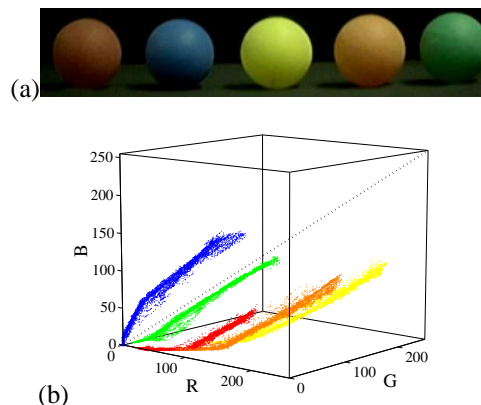


Figure 1: RGB scatter plots of colored balls under office lighting without calibration. The dotted line indicates the diagonal of RGB space.

2004), often combined with a prior skin color distribution learned from a database. The problem with skin model adaptation is that it is difficult to obtain a large and representative sample set of skin color from a video automatically. Adaptive methods do not generalize well to other skin regions in the video if they lack a (realistic) chrominance model.

3 COLOR DETECTION METHOD

Instead of learning a general skin color distribution that generalizes too much, or relying on a chrominance model based on rigid, non-realistic assumptions, we propose to use a general model of skin color variation for not-calibrated camera's with fixed settings and non-changing illumination. We use an automatic procedure that fits this model to a small and noisy sample set.

The general model of skin color variation is explained in paragraph 3.1. Paragraph 3.2 explains how a similarity measure for skin color can be computed using this model, and 3.3 describes how the model is fitted to the sample set.

3.1 Skin Color Appearance Model

We define x as a point at a specific image pixel location, corresponding to a point on the surface of an object with object color ξ and θ as the spectrum of the illumination source. Our appearance model of skin assumes that the appearance in RGB space $\vec{C}_{RGB}(x, \xi, \theta)$ of color ξ is a linear function $\vec{\ell}(x, \xi, \theta)$ of reflected light intensity $I(x)$ plus some independent random zero-mean noise vector $\vec{\eta}(x)$ (assuming Lambertian

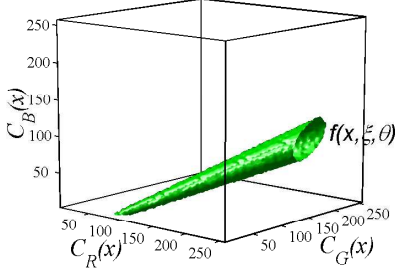


Figure 2: Appearance model of skin color $p(\vec{C}_{RGB}|skin)$ according to the multiplicative noise model, represented by an isosurface.

reflection):

$$\vec{C}_{RGB}(x, \xi, \theta) = \vec{\ell}(x, \xi, \theta) + \vec{\eta}(x), \quad (1)$$

$$\vec{\ell}(x, \xi, \theta) := \vec{C}_{RGB0} + I(x)\vec{c}(\xi, \theta). \quad (2)$$

where $\vec{c}(\xi, \theta)$ is a color vector with unit intensity and \vec{C}_{RGB0} is the (unknown) calibration point for black. $I(x)$ depends on both the light source intensity and the surface tangent, of which the latter depends on x in a non-deterministic way. For brief notation, $\vec{C}_{RGB}(x, \xi, \theta)$ is denoted by \vec{C}_{RGB} in the remainder of this work, silently assuming a specific object color, illumination and image pixel coordinates.

For fixed ξ and θ , $\vec{\ell}(x, \xi, \theta)$ is a straight line in RGB space. If the camera is not calibrated, $\vec{\ell}(x, \xi, \theta)$ does not have to pass through the origin. An example image without calibration is shown in figure 1 (a). This is an image of uniformly colored balls captured by a webcam under office lighting. Although the image quality looks acceptable, the color distributions of the balls are not at all directed towards the origin, as would have been the case if black was calibrated at $\vec{C}_{RGB0} = [0, 0, 0]^T$.

For specific ξ and θ , the dimensionality of the appearance model can be reduced from three to two dimensions by projecting the RGB values \vec{C}_{RGB} onto a plane perpendicular to $\vec{\ell}(x, \xi, \theta)$:

$$\vec{C}_S = [\hat{S}_1, \hat{S}_2]^T \vec{C}_{RGB}. \quad (3)$$

$[\hat{S}_1, \hat{S}_2, \hat{S}_3]$ is the orthonormal basis of the adapted color space, which is a rotated RGB coordinate system with \hat{S}_3 in the direction of $\vec{\ell}(x, \xi, \theta)$, corresponding to the luminance axis, and $[\hat{S}_1, \hat{S}_2]$ spanning the perpendicular plane. The latter can be seen as an intensity-invariant chrominance space for one specific object color in a specific situation (camera, calibration, illumination, etc.), referred to as 'Adaptive Chrominance Space' (ACS). The distribution of an

object color in ACS will be modeled by a Gaussian distribution with mean μ_S and 2×2 covariance matrix Σ_S . The Mahalanobis distance D_S to the closest point on the line $\vec{\ell}(x, \xi, \theta)$ is computed by

$$D_S = \sqrt{(\vec{C}_S - \mu_S)^T \Sigma_S^{-1} (\vec{C}_S - \mu_S)} \quad (4)$$

Because of the Gaussian approximation perpendicular to the luminance axis, the skin color model becomes an infinite elliptic cylinder in RGB space. To save computational load, \hat{S}_1 and \hat{S}_2 can be chosen in the directions of the eigenvectors of Σ_S and divided by the square roots of the respective eigenvalues, leaving the 2×2 identity matrix instead of Σ_S .

3.2 Skin Likelihood

Computation of skin likelihood at image position x is performed according to Bayes' theorem:

$$p(skin|\vec{C}_{RGB}) = \frac{p(\vec{C}_{RGB}|skin)p(skin)}{p(\vec{C}_{RGB})} \quad (5)$$

Where $skin$ is the event that the image really contains skin at the measured location. The prior probability density of $p(\vec{C}_{RGB})$ can be marginalized by

$$\begin{aligned} p(\vec{C}_{RGB}) &= \int_{-\infty}^{\infty} p(\vec{C}_{RGB}|I = \alpha)p(I = \alpha)\delta(\alpha - I)d\alpha \\ &= p(\vec{C}_{RGB}|I = \beta)p(I = \beta), \end{aligned} \quad (6)$$

because $p(\vec{C}_{RGB})$ is zero for $\alpha \neq I$. β is equal to the light intensity of (\vec{C}_{RGB}) , calculated by

$$I = (C_R - C_{R0}) + (C_G - C_{G0}) + (C_B - C_{B0}). \quad (7)$$

Since the real value of black \vec{C}_{RGB0} is unknown, we approximate it by

$$\tilde{C}_{R0} = \tilde{C}_{G0} = \tilde{C}_{B0} = \min\{C_R(X) \cup C_G(X) \cup C_B(X)\} \quad (8)$$

where C_R, C_G, C_B are the red, green and blue values, respectively, of \vec{C}_{RGB} and X is the total set of pixels available, measured in the specific situation. Here we assume that the black origin is on the diagonal of RGB space, which is a reasonable assumption considering the imaging processes of most digital cameras. Unfortunately, in most situations the real value of \vec{C}_{RGB0} is negative and the lower color values are clipped to 0, resulting in a large estimation error in \tilde{C}_{RGB0} .

The probability density of the distance to the RGB diagonal at the brightness plane $C_R + C_G + C_B = I$ is assumed constant and non-zero inside positive RGB space, but zero outside (uniformly distributed saturation). The integral of $p(\vec{C}_{RGB}|I = \beta)$ over all possible values of \vec{C}_{RGB} for which $I = \beta$ must be equal to

1. This set is a plane perpendicular to the diagonal of RGB space, which is an equilateral triangle with its corners at $\vec{C}_{RGB} = \{[\beta, 0, 0]^T, [0, \beta, 0]^T, [0, 0, \beta]^T\}$. Using the area $A(\beta)$ of this triangle and assuming that the prior probability of brightness $p(I = \beta)$ is constant,

$$p(\vec{C}_{RGB}) \propto 1/A(\beta) = 2/(\sqrt{3}\beta^2). \quad (9)$$

We choose to model uncertainty about the actual color of skin ξ , but to neglect the additive noise $\vec{\eta}(x)$. This results in a conical shape with its tip at \vec{C}_{RGB0} , centered around $\vec{\ell}(x, \xi, \theta)$ corresponding to the mean ξ and θ , shown in figure 2. In this case, deviation from the line $\vec{\ell}(x, \xi, \theta)$ is only due to variation of skin or illumination color. The existing appearance model can easily be modified by normalizing the Mahalanobis distance of equation 4 by the brightness I , similar to computing normalized rgb:

$$\hat{D}_S(x) = \frac{D_S}{I}. \quad (10)$$

When $p(\vec{C}_{RGB}|skin)$, in equation 5, is computed with \hat{D}_S instead of D_S , the prior (equation 9) has to be normalized by $1/I^2$, hence becomes a constant. The log likelihood becomes:

$$\ln\{p(skin|\vec{C}_{RGB})\} \propto -\frac{(\vec{C}_S - \mu_S)^T \Sigma_S^{-1} (\vec{C}_S - \mu_S)}{I^2}. \quad (11)$$

Note that, if $\vec{C}_{RGB0} = [0, 0, 0]^T$ (i.e. in case of correct calibration), this approach is very similar to a skin color model in Hue/Saturation space or two dimensions of the normalized rgb space.

3.3 Automatic Model Fitting

The automatic Adaptive Chrominance Space adaptation procedure attempts to find the best fit of the model to a sample set taken from a specific situation. To obtain positive RGB skin samples for model fitting, a face detection method could be used. Pixels from the left and right side of the sample area are modeled separately. Resulting in two separate ACS models. This is because the left and right side of the face are often illuminated with a different θ , causing a different $\vec{\ell}(x, \xi, \theta)$. By modeling both sides of the face separately, the combined model can detect skin illuminated by different light sources.

Skin samples for which at least one of the color channels has an intensity of 0 or 255 are removed. This is because these samples might have been projected from outside the RGB space, leading to a skewed color representation. Note that modeling these projected values is possible, as described in our

previous work (Lichtenauer et al., 2005), however, it is beyond the scope of this work.

The fitting procedure is based on the assumption that the main axis of the positive sample distribution corresponds to $\vec{\ell}(x, \xi, \theta)$. This is usually the case, since the area of the face, used for sampling, contains both dark and light skin values (e.g. due to the nose and its curved shape). Even in case of very uniform illumination, some shading will still be present. Otherwise, one would not see any shape of the face other than the eyes and mouth of a person. These small nuances of skin color are enough to show a main axis in the direction of intensity in RGB space. The main axis is found by line fitting, using RANSAC (Fischler and Bolles, 1981). However, the search of RANSAC is constrained by assumptions about shadowed colors, following (Lichtenauer et al., 2005). Pairs of points for which one of the two can never be a shaded color of the other, are discarded. The regular least square fitting method does not work because of the inhomogeneous distribution of samples along the line. Therefore, the best result of the RANSAC search is refined by computing the means m_L and m_H of two subsets of the inlier samples: First, the samples are projected on the initially fitted line. m_L is the mean of the samples with the lowest 0.1 quantile of projection values, m_H is computed from samples in the highest 0.1 quantile. From the line between m_L and m_H , \hat{S}_3 is derived and two perpendicular vectors \hat{S}_1 and \hat{S}_2 can be determined. After projecting the positive samples onto the AC space spanned by $[\hat{S}_1, \hat{S}_2]$, the mean and covariance of the distribution are estimated by a fixed number of EM iterations, where the 0.9 best fitting samples are used to compute the parameters of the next iteration. The latter is to discard outliers.

To get a combined likelihood of the ACS model for the left and the right side of the face, the two are combined by taking the maximum of the two log-likelihoods, computed with normalized inverse covariance matrices $\hat{\Sigma}_S^{-1}$:

$$\hat{\Sigma}_S^{-1} = \frac{\Sigma_S^{-1}}{\sqrt{|\Sigma_S^{-1}|}}. \quad (12)$$

The measured main axis of the skin color distribution is not always the result from the appearance model in equation 1. For example, when multiple light sources are present, the transition from one light source to the other could contain correlation between $I(x)$ and θ . In these situations, there is no guarantee that the appearance model will continue in the same direction, outside of the intensity range of the training samples. In order to be more robust against these exceptions, also a more conservative Hybrid ACS (HACS) method is considered. This model is split



Figure 3: Some images used for the experiments. The area inside the white rectangle was used for modeling skin color.

in three parts: a middle part for $I(m_L) < I < I(m_H)$, identical to the normal ACS model, and a lower and higher part, for $I < I(m_L)$ and $I > I(m_H)$, respectively, that assume that $\vec{l}(x, \xi, \theta)$ continues from the ends of middle part into the directions corresponding to the estimated \tilde{C}_{RGB0} from equation 8, but using the same covariance as the middle part.

4 EXPERIMENT

We have evaluated our ACS method by applying it to 10 different photos of different persons in different environments (see figure 3). Most photos were taken with a digital photo camera, except for 'Studio', which was captured with a firewire video camera. This camera was calibrated to have $\tilde{C}_{RGB0} = [0, 0, 0]^T$ and homogeneous illumination was used (i.e. ideal circumstances for most color models). The photo cameras used for the other photos automatically adjusted white balance and intensity for each photo. All photos were filtered using a 3x3 median filter, to reduce noise and erroneous colors around sharp edges. A rectangle was manually annotated at each face, to extract the training color samples. These rectangles are shown in the photos of figure 3. The left and right sides of these rectangles were used for training the left and right ACS models, respectively. To remove remaining non-skin pixels from the training samples (e.g. from eyes and mouth), the training is done in two steps. After training the initial ACS models on all respective training pixels, the pixels with the lowest 30% of likelihoods for the respective ACS model are removed from the rectangle and a 3x3 morphological closing is applied on the total remaining pixel mask. The ACS models are re-trained using only the samples selected with this skin mask. For a fair comparison, also the other color models are trained on these filtered samples.

Training and testing of the skin color model was done separately for each image, to evaluate the average performance of ACS over different circumstances. The true positive (TP) rate (number of pixels correctly detected as skin divided by the total num-

Table 1: Area Under the Curve of the ROC for FP rate ≤ 0.1 .

Image	ACS	HACS	HS	rg	CrCb	RGB
Studio	.0992	.0994	.0993	.0994	.0994	.0994
Office	.0935	.0890	.0834	.0814	.0833	.0772
Hallway	.0553	.0547	.0526	.0544	.0907	.0978
Sunset	.0633	.0676	.0523	.0596	.0749	.0701
Restaurant	.0485	.0572	.0601	.0526	.0379	.0543
Window	.0726	.0654	.0486	.0173	.0280	.0257
Living room	.0778	.0557	.0157	.0250	.0599	.0647
Glass wall	.0781	.0832	.0821	.0789	.0711	.0579
Table tennis	.0963	.0961	.0950	.0953	.0953	.0918
Street	.0604	.0726	.0817	.0656	.0845	.0855
average	.0745	.0741	.0671	.0630	.0725	.0724
st.dev.	.0169	.0160	.0245	.0259	.0227	.0217
worst result	.0485	.0547	.0157	.0173	.0280	.0257
average ranking	3.2	2.6	4.0	4.5	3.4	3.3

ber of tested skin pixels) is computed only from the bare skin of the hands and arms in the same image, which was also manually annotated. The false positive (FP) rate (number of pixels wrongly detected as skin divided by the total number of tested non-skin pixels) was computed from all areas in the same photo not containing any bare skin or hair. Note that, although training and test samples were extracted from the same image, they correspond to physically different parts of the scene, subject to possibly different illumination and captured with different parts of the camera sensor. This results in a better separation between training and test set than when they only differ in time.

Four different color spaces were compared to ACS: Hue-Saturation (HS) space, normalized Red-Green (rg) space, CrCb space and RGB space. Histograms of the face samples were used as the skin likelihood models. For the first three color spaces, the histogram sizes are 100×100 bins and for RGB space the histogram was $100 \times 100 \times 100$ bins. The reason for using histograms is because histograms do not assume any type of distribution. Because histograms do not generalize as well as parameterized distributions, the histograms are all smoothed with a Gaussian kernel of 21 bins in all dimensions, with standard deviation 1. HACS and HSV color models and skin detection for the image 'Window' are shown in figure 4.

To get an objective and comparable measurement of performance, the Area Underneath the Curve (AUC) is computed for the Receiver Operation Characteristic (ROC) curves of all methods and photos. Only the part of the curve with small FP rate is relevant, since too many false positives will make it impossible to separate hands from the rest of the image. Therefore, only the AUC for FP rates between 0 and 0.1 is used for comparison. The results are shown in table 1 (the closer to 0.1, the better). The ACS method shows the best mean AUC, however, HACS has the best average ranking over the test images. Further-

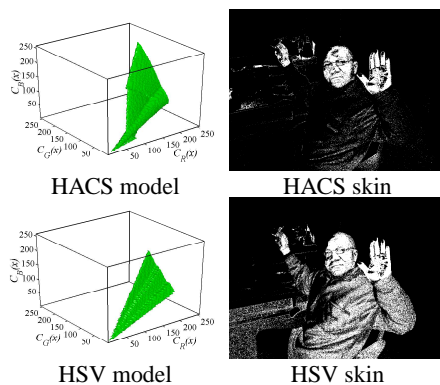


Figure 4: Results for the image 'Window'. The upper left figure shows detection of the main axes of the left (black dots, solid line) and right (blue dots, dashed line) side of the face. The hulls in RGB space correspond to skin color using a threshold that detects 75% of the training samples.

more, these results show that the relative performance between all color models greatly depends on the situation. The low standard deviation of the HACS results indicates the high robustness of the method. Although the other methods have higher results in some specific images, they also have significantly lower results for other images. Surprisingly, the RGB model outperforms the HS, rg and CrCb models, contrary to what could be expected from modeling skin color with prior knowledge about intensity invariance. The violation of the assumptions of these models in everyday situations clearly has a negative effect on their performance.

5 CONCLUSIONS & FUTURE WORK

We have proposed an adaptive chrominance model and automatic fitting procedure that can be used to detect skin color more accurately than the compared methods when no color space calibration is performed and/or heterogeneous illumination is present. This makes our (H)ACS model especially useful for real-world applications. Besides a better overall skin detection performance, HACS also showed a lower standard deviation between different situations, while the other methods showed more unstable results.

Further improvements of the model are possible. First of all, over- or under-saturated colors can be accounted for and assumptions about the model orientation and shape outside of the intensity range of the training samples can be improved. Furthermore, many improvements are possible on the prior prob-

ability model of background color. A histogram of the complete image and/or an off-line image database could be used to exclude colors with a high prior probability.

ACKNOWLEDGEMENTS

We would like to thank the VSB fund and the NSDSK (Dutch Foundation for the Deaf and Hard of hearing Child) for making this research possible.

REFERENCES

- Argyros, A. and Lourakis, M. (2004). Real-time tracking of multiple skin-colored objects with a possibly moving camera. In *ECCV04*, pages Vol III: 368–379.
- Fischler, M. and Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- Fritsch, J., Lang, S., Kleinhagenbrock, M., Fink, G. A., and Sagerer, G. (2002). Improving adaptive skin color segmentation by incorporating results from face detection. In *ROMAN*, pages 337–343, Berlin, Germany.
- Jones, M. J. and Rehg, J. M. (2002). Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96.
- Lee, J. and Yoo, S. (2002). An elliptical boundary model for skin color detection. In *CISST'02*.
- Lichtenauer, J., Hendriks, E., and Reinders, M. (2005). A shadow color subspace model for imperfect non-calibrated cameras. In *ASCI 05*, Heijen, The Netherlands.
- Martinkauppi, B., Soriano, M., and Pietikainen, M. (2003). Detection of skin color under changing illumination: a comparative study. In *CIAP03*, pages 652–657.
- McKenna, S., Raja, Y., and Gong, S. (1999). Tracking colour objects using adaptive mixture models. *IVC*, 17(3/4):225–231.
- Phung, S., Bouzerdoum, A., and Chai, D. (2005). Skin segmentation using color pixel classification: analysis and comparison. *PAMI*, 27(1):148–154.
- Raja, Y., McKenna, S., and Gong, S. (1998). Tracking and segmenting people in varying lighting conditions using colour. In *AFGR98*, pages 228–233.
- Soriano, M., Martinkauppi, B., Huovinen, S., and Laaksonen, M. (2000). Skin detection in video under changing illumination conditions. In *ICPR00*, pages Vol I: 839–842.
- Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Proc. Graphicon-2003*, pages 85–92.
- Yang, M., Kriegman, D., and Ahuja, N. (2002). Detecting faces in images: A survey. *PAMI*, 24(1):34–58.