Department of Information Engineering and Computer Science DISI

UNIVERSITY OF TRENTO - Italy

DISI - Via Sommarive 14 - 38123 Povo - Trento (Italy)
http://www.disi.unitn.it

COLOR BASED SKIN CLASSIFICATION

Rehanullah Khan, Allan Hanbury,
Julian Stöttinger, Abdul Bais

March 2012

Technical Report # DISI-12-009

Color Based Skin Classification

Rehanullah Khan[a,d], Allan Hanbury[b], Julian Stöttinger[c], Abdul Bais[d]

[a] *CVL, Institute of Computer Aided Automation, TU-Wien, Austria*
[b] *IRF, Vienna, Austria*
[c] *Department of Information Engineering and Computer Science, University of Trento*
[d] *Sarhad University of Science and Information Technology, Peshawar Pakistan*

Abstract

Skin detection is used in applications ranging from face detection, tracking body parts and hand gesture analysis, to retrieval and blocking objectionable content. In this paper, we investigate and evaluate (1) the effect of color space transformation on skin detection performance and finding the appropriate color space for skin detection, (2) the role of the illuminance component of a color space, (3) the appropriate pixel based skin color modeling technique and finally, (4) the effect of color constancy algorithms on color based skin classification. The comprehensive color space and skin color modeling evaluation will help in the selection of the best combinations for skin detection. Nine skin modeling approaches (AdaBoost, Bayesian network, J48, Multilayer Perceptron, Naive Bayesian, Random Forest, RBF network, SVM and the histogram approach of Jones and Rehg [15]) in six color spaces (IHLS, HSI, RGB, normalized RGB, YCbCr and CIELAB) with the presence or absence of the illuminance component are compared and evaluated. Moreover, the impact of five color constancy algorithms on skin detection is reported. Results on a database of 8991 images with manually annotated pixel-level ground truth show that (1) the cylindrical color spaces outperform other color spaces, (2) the absence of the illuminance component decreases performance, (3) the selection of an appropriate skin color modeling approach is important and that the tree based classifiers (Random forest, J48) are well suited to pixel based skin detection. As a best combination, the Random Forest combined with the cylindrical color spaces, while keeping the illumi-

Email addresses: rehanmarwat1@gmail.com (Rehanullah Khan), a.hanbury@ir-facility.org (Allan Hanbury), julian@disi.unitn.it (Julian Stöttinger), bais@ieee.org (Abdul Bais)

nance component outperforms other combinations, and (4) the usage of color constancy algorithms can improve skin detection performance.

Keywords: skin detection, skin classification, color spaces and skin detection, color constancy

---

1. Introduction

Skin detection is a popular and useful technique for detecting and tracking human-body parts. It has received much attention because of its wide range of applications, such as: detecting and tracking faces, naked people detection, hand tracking, people retrieval in databases and the Internet etc. Skin detection can also contribute towards blocking objectionable content from the Internet. The most attractive properties of color based skin detection are the potentially high processing speed and invariance against rotation, partial occlusion and pose change. However, standard skin color detection techniques are negatively affected by changing lighting conditions, complex backgrounds and surfaces having skin-like colors.

For the pixel-wise skin detection or classification considered in this paper, the objective is to build a decision rule that differentiates between skin and non-skin pixels given only a color triplet as input. Kakumanu et al. [16] state that the major difficulties in skin color detection are caused by various effects such as varying illumination, camera characteristics, ethnicity, individual characteristics and other factors like makeup, hairstyle, glasses, sweat, and background colors. An approach for reliable skin detection has therefore to be stable against noise, artifacts and very flexible against varying lighting conditions.

Typically a skin detection framework involves transformation of the RGB color space to another color space, leaving out the illuminance component and using only the color components of a color space, finally classifying skin by an appropriate skin color modeling technique. The objective of these steps is to provide skin detection robustness in varying illumination conditions and different skin tones.

In this paper, our objective is creating a framework for helping in the selection of the best combination of color space and skin modeling approach for skin detection. As such, we examine (1) the effect of color space transformation, (2) the role of the illuminance component of a color space for skin

detection, (3) the skin color distribution, modeling and selection, and finally, (4) the effect of color constancy algorithms on color based skin classification.

We provide a comprehensive evaluation of color based skin detection in six color spaces (IHLS, HSI, RGB, normalized RGB (nRGB), YCbCr and CIELAB) with nine skin color modeling approaches (AdaBoost, Bayesian network, J48, Multilayer Perceptron, Naive Bayesian, Random Forest, RBF network, SVM and Jones and Rehg [15] approach) on a large publicly available dataset with manually generated pixel-level ground truth. Besides the selection of best combination of color spaces and skin color modeling methodology, this study creates a baseline against which more advanced skin modeling (classification) algorithms can be tested.

Results on a database of 8991 images with manually annotated pixel-level ground truth show that (1) color space transformation has profound effect on overall skin detection performance and that the cylindrical color spaces outperform other color spaces, (2) the absence of the illuminance component decreases performance, (3) an appropriate skin color modeling approach selection is important for pixel based skin classification and that the tree based classifiers (Random Forest, J48) due to their simple rules for pixel based features are well suited to pixel based skin classification, (4) the usage of color constancy algorithms can improve skin detection performance.

In Section 2, we summarize some of the related work regarding skin detection. Experimental details are presented in Section 3. We discuss the results in Section 4 and finally the paper is concluded in Section 5.

2. Related Work

In computer vision, skin detection is used as a first step in face detection, e.g. [14], and for localization in the first stages of gesture tracking systems, e.g. [2]. It has also been used in the detection of naked people [8, 18] and for blocking objectionable content [27]. The latter application has been developed for videos.

The approaches to classify skin in images can be grouped into three types of skin modeling: parametric, non-parametric and explicit skin cluster definition methods. The parametric models use a Gaussian color distribution since they assume that skin can be modeled by a Gaussian probability density function [33]. Non-parametric methods estimate the skin-color from the histogram that is generated by the training data used [15].

An efficient and widely method is the definition of classifiers that are built upon the approach of skin color clustering/thresholding. This thresholding of different color space coordinates is used in many approaches, e.g. [21] and explicitly defines the boundaries of the skin clusters in a given color space, generally termed as static skin filters. The static filters used in YCbCr and RGB color spaces for skin detection are reported in [7] and [20]. Its main drawback is a comparably high number of false detections [16]. Khan et al. [17] addressed this problem by opting for a multiple model approach, which makes it possible to filter out skin for multiple people with different skin tones, reducing false positives.

Color is a low level feature that is computationally inexpensive. For many applications in computer vision, it is suitable for real-time object characterization, detection and localization. Color spaces like the HS* family transform the RGB cube into a cylindrical coordinates representation. They have been widely used in skin detection scenarios, such as [4, 9, 10]. Perceptually uniform color spaces like the CIELAB, CIELUV are used for skin detection e.g. in [6]. Orthogonal color spaces like YCbCr, YCgCr, YIQ, YUV, YES try to form as independent components as possible. YCbCr is one of the most successful color spaces for skin detection and used in e.g. [31, 14].

Skin detection under varying illumination in image sequences is addressed in [25, 32, 26]. These approaches try to map the illuminance of the image into a common range. They compensate for the variance of changing lighting to equalize the appearance of skin color throughout different scenes.

In [23], 845 images are used for comparison of nine color spaces with the histogram approach and a normal density approach. It is reported that color space transformation can improve performance in certain instances and that skin color modeling has greater impact than the color space transformation. The best performance is reported on indoor images in HSI color space and modeling the skin color with the histogram approach using a larger size distribution. Similarly in [15], 20,000 images collected from the web are used for measuring skin detection performance. The performance is measured using ROC curve for histogram-based and mixture of Gaussian in the RGB color space with the reported histogram models to be superior in accuracy and computational cost. As a final step, skin detection is used for blocking objectionable images.

Four color spaces in conjunction with histogram based skin detection methods are evaluated in [1], claiming that skin detection performance is independent of a color space transformation that is invertible. It has been

shown that RGB, YCbCr and HSV have similar performance while CbCr is different since being a 'non-invertible' color transformation. The skin probability Map (SPM) with RGB color space is reported as the best combination in [3], compared to Red-Green (rg) ratio and linear color transformation of RGB color space into YIQ color space.

Nine color spaces are compared in [28] for face detection using Mahalanobis metrics. True Positive (TP) and True Negative (TN) rates are used as evaluation measures. It is reported that the normalized color spaces yield best results and are preferable for robust color based skin detection. In [34], five color spaces and two histogram based skin detection methods are considered. It is reported that HSV combined with lookup table method has higher skin detection performance using % skin correct measure and that color space transformation affects the over-all skin detection performance. Neural networks [19], Bayesian networks e.g. [24], Gaussian classifiers e.g. [15], and Self Organizing Maps (SOM) [4] have been used in skin detection applications.

3. Experimental setup

In this section, color spaces, color constancy, skin color modeling approaches, dataset and the evaluation measures used are explained.

3.1. Color spaces

The effect of color transformation on skin detection performance is measured with five color space transformations: RGB to IHLS, HSI, normalized RGB (nRGB), YCbCr and CIELAB. These color models are commonly used in color image precessing. These color models contain both variant and invariant properties with reference to imaging conditions. With the color space transformation, the objective is to decrease the overlap between skin and non-skin pixels thereby maximizing classification performance. The equations for color space transformation do not always yield values in the range 0-255. In our framework the values are not adjusted and are left as the default transformation values. Transformation equations can be found in [12]. The Improved Hue, Luminance and Saturation (IHLS) color space is introduced in [13]. The IHLS model is improved with respect to the similar color spaces (HLS, HSI, HSV, etc.) by removing the normalization of the saturation by the brightness. This property overcomes certain numerical problems on the

limits of the color channels giving a better distribution in our feature space. We for the first time use it for the skin classification problem.

3.2. Illuminance component

It is commonly assumed that variations in skin color occur more in intensity than in chrominance and that robustness in skin detection can be achieved by dropping the illuminance component and using chrominance components only [23]. When we use all the components of a color space, we refer to it as a 3D color space, while a 2D color space is one without the illuminance component. L of IHLS, I of HSI, G of RGB, nG of nRGB, Y of YCbCr, and L of CIELAB are the illuminance components.

3.3. Color constancy

Color constancy is the ability of human vision system to resolve object colors in a scene independent of the illuminant. In other words, the color constancy problem can be defined as the ability to estimate the unknown light of a scene from an image/photograph. Different color constancy algorithms provide different estimation of illuminant. For the role of color constancy for skin detection, we use five color constancy algorithms: Gray-Edge, Gray-World, max-RGB, Shades of Gray and Bayesian color constancy.

Gray-Edge hypothesis assumes that the average of the reflectance differences in a scene is achromatic [30]. The Gray-World hypothesis assumes that the average reflectance in a scene is achromatic [5]. max-RGB is based on the assumption that the reflectance which is achieved for each of the three color channels is equal [30]. The Gray-World and the max-RGB algorithm are two different instantiations of a more general color constancy algorithm based on the Minkowski norm. A Shades of Gray is computed by [30]:

$$\left(\frac{\int (f(x))^p dx}{\int dx}\right)^{\frac{1}{p}} = ke \tag{1}$$

In the Bayesian color constancy approach, the observed image pixels are modeled with a probabilistic generative model, decomposing them as the product of unknown surface reflectances with an unknown illuminant [11]. Using Bayes rule, a posterior for the illuminant is obtained and from this the estimate with minimum risk is extracted.

3.4. Skin color modeling

We use nine skin color modeling techniques for pixel based skin classification: AdaBoost, Bayesian network, J48, Multilayer Perceptron, Naive Bayesian, Random Forest, RBF network, SVM and histogram approach [15]. These approaches are the commonly preferable choices for classification problems. For classifiers, for each pixel in every color space, a feature vector is created by using all the three color channels in case of 3D color spaces and two color channels for 2D color spaces. For the skin color modeling techniques considered, the performance is affected by the parameter settings. In most cases, the parameters affect precision and recall in such a way that if changing a parameter increases precision, the recall is decreased and vice versa, but the overall F-measure still remains the same. We discuss the parameters for each classifier below.

3.4.1. AdaBoost

We use the multi-class case which requires the accuracy of the weak hypothesis greater than 0.5. During the training of 10-fold cross validation, using Decision Stump as the base classifier, the weight threshold of 100 and the number of iterations of 10 reports overall best performance in the case of 3D and 2D color spaces.

3.4.2. Bayesian network (BayesNet)

The Bayesian network is a representation for random variables and conditional independences within these random variables. During the training of 10-fold cross validation, we obtain high F-measure by setting the estimator parameter to 0.5 and using a hill climbing algorithm as the searching algorithm.

3.4.3. J48

Builds decision trees (binary trees) from a set of training data using the concept of information entropy [22]. The confidence factor and the minimum number of instances per leaf has an effect on the performance of skin classification. We achieve highest performance with confidence factor of 0.25 and minimum instances per leaf to be 2.

3.4.4. Multilayer Perceptron (MLP)

In the MLP, learning occurs in the perceptron by changing connection weights after each piece of data processed and is carried out through backpropagation. The optimum performance is obtained by setting the learning

rate (the amount the weights are updated) to 0.3 and the momentum applied to the weights during updating to 0.2. The hidden nodes are set to 5 for 3D color spaces and 4 for 2D color spaces.

3.4.5. Naive Bayesian (NaiveBayes)

A simple probabilistic classifier providing maximum a posteriori probability for each testing instance. We find that using supervised discretization (to convert numeric attributes to nominal ones) and kernel estimation (for numeric attributes) rather than a normal distribution increases performance.

3.4.6. Random Forest

Grows many classification trees and chooses the classification having the most votes (overall trees in the forest). The most important parameter is the number of trees grown for the classification. We find that the highest F-measure is reported for 10 trees grown. Less than 10 trees decreases overall performance and greater than 10 does not increase performance but rather converges to a stable performance. Also, limiting the depth of the tress grown decreases performance.

3.4.7. RBF network (RBF)

A neural network using radial basis functions as activation functions. We find that the performance is independent of ridge value for the linear regression of RBF. The performance is however affected by the number of clusters selected with optimum performance observed for number of clusters being 2.

3.4.8. SVM

Finds a hyperplane for inter-class separation with the objective being the maximal margin. During the training by 10-fold cross validation, the performance is increased by using the polynomial kernel. The complexity parameter C is found to yield maximum performance with $C = 1$ and the tolerance parameter of 0.9. For the complexity parameter, values below $C = 1$ decreased the overall performance, while for any increase above 1, we attained a stable performance close to that of $C = 1$.

3.4.9. Histogram approach (Hist.)

Jones and Rehg [15] constructed a 3D RGB histogram model of skin and non-skin from 18,696 web images. The skin and non-skin histograms can be

used to build fast and accurate skin classifiers. With the class conditional probabilities of skin and non-skin color models, a skin classifier is built using Bayes Maximum Likelihood (ML) approach. An image pixel is classified as skin, if:

$$\frac{P(c|skin)}{P(c|non-skin)} \geq \theta \tag{2}$$

where θ is a threshold value which can be adjusted to trade-off between true positives and false positives. We use $\theta = 0.3$ using number of histogram bins of 64.

3.5. Dataset

We use 8991 images extracted from 25 videos provided by an Internet service provider. The sequences contain scenes with multiple people and/or multiple visible body parts and scene shots both indoors and outdoors, with steady or moving camera. The lighting varies from natural light to directional stage lighting. The data set is available on-line[1]. Ground truth has been generated for all of the 25 videos on a per pixel basis annotating 8991 frames manually.

3.6. Evaluation measures

For both 3D and 2D color spaces, there are 54 combinations of the six color spaces and nine skin color modeling approaches. For a particular combination, we follow a training and testing paradigm using 10-fold cross validation. In the 10-fold cross-validation setup, we make sure that for each experiment, each video (and all frames from that video) is entirely allocated to the test set/training set. The performance measure is based on F-measure calculated by evenly weighting precision and recall.

4. Results

In this section, we report the effect of color space transformation, the role of the illuminance component, role of skin color modeling approaches and the effect of color constancy on skin classification.

[1]http://www.feeval.org

4.1. Effect of color space transformation

We investigate if a color space transformation improves skin and non-skin separability. As such, we examine it by F-measure, comparing RGB with non RGB color spaces. Refer to Table 1 and Figure 1 for the effect of color space transformation on skin performance. In Table 1 and Figure 1, we find that compared to RGB, the IHLS color space has better performance in eight skin modeling approaches except the histogram approach. The HSI color space also shows improvement in eight cases with the exception of histogram approach. The worst performing of all the color spaces, independent of the skin color modeling came out to be nRGB. YCbCr shows significant increase in almost all the skin color modeling approaches with the exception of Random Forest and histogram approach. CIELAB also showed improved performance in eight skin color modeling approaches.

For 2D color modeling (see Table 2), compared to RGB, the IHLS shows improvement in only five skin color modeling approaches, whereas HSI shows increase in the cases of Bayesian network, J48, Random Forest and SVM. nRGB and YCbCr show significant improvement over 2D RGB in four skin color modeling approaches. The most significant improvement in 2D color spaces over the 2D RGB is exhibited by the CIELAB color space.

For transformed color spaces, what we experience is that independent of the skin color modeling, nRGB shows to be unsuitable for robust skin classification. This is contrary to prior experiments in the literature, but can be explained with the very noisy dataset of on-line videos, which shows many dark colors where nRGB becomes unstable. IHLS reports overall highest performance, outperforming other color spaces. Finally, we see that the color spaces improve more with the tree-based skin modeling (Random Forest and J48) than with other approaches.

4.2. Role of illuminance component

We evaluate the effect of removing the illuminance component of a color space on skin detection performance. The F-measure values for 2D color spaces i.e. without the illuminance component are shown in Table 2. The difference of the F-measure for the 2D and 3D colors is shown in Table 3, reported for each color and classifier combination. The difference is computed as the 3D F-measure minus the 2D F-measure multiplied by 100. We find that in almost all cases, the 3D color spaces perform better than 2D color spaces. There are 4 cases (negative values in Table 3) where the performance of 3D color space is slightly less than 2D. There are 50 cases (positive values

Table 1: F-measure for 3D color spaces with different skin color modeling approaches. Bold indicates increased performance compared to RGB color space.

	IHLS	HSI	RGB	nRGB	YCbCr	CIELAB
AdaBoost	**0.320**	**0.300**	0.260	0.250	**0.270**	**0.276**
BayesianNet	**0.590**	**0.570**	0.321	**0.370**	**0.490**	**0.560**
J48	**0.684**	**0.680**	0.662	0.626	**0.680**	**0.660**
MLP	**0.650**	**0.591**	0.590	0.569	**0.627**	**0.600**
NaiveBayes	**0.466**	**0.450**	0.255	**0.408**	**0.427**	**0.454**
Random Forest	**0.745**	**0.741**	0.710	0.700	0.705	**0.740**
RBF	**0.467**	**0.430**	0.389	**0.420**	**0.490**	**0.510**
SVM	**0.503**	**0.471**	0.360	**0.370**	**0.385**	**0.400**
Hist.	0.409	0.408	0.418	0.399	0.390	0.400

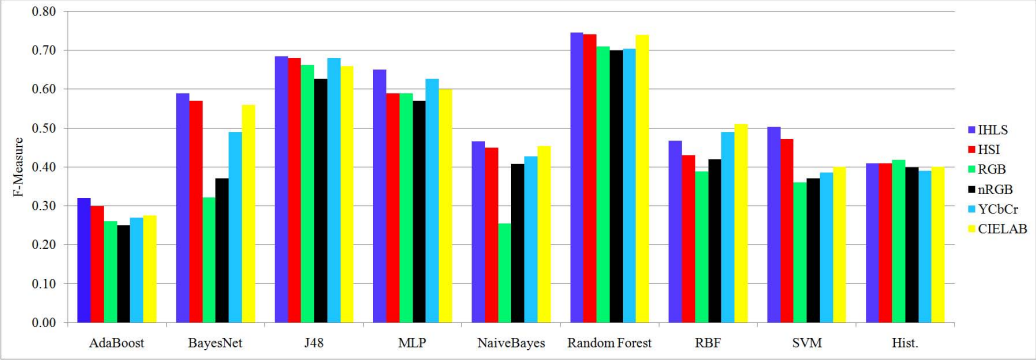


Figure 1: F-measure for skin color modeling approaches with 3D color spaces.

in Table 3), where the performance of 3D color space is higher than 2D. We represent significant improvement as a value greater than 10. There are total of 19 cases (values with asterisk in Table 3) where the performance of 3D color space is significantly higher than their 2D counterpart. The difference of F-measure is also related to the color space transformation. We find (Table 3) that the effect of the illuminance component is more dominant in the cases of IHLS and HSI color spaces. We argue that the illuminance component adds more information for skin and non-skin separability in the cases of IHLS and HSI color spaces.

Table 2: F-measure for 2D color spaces (without the illuminance component) for nine skin color modeling approaches. Bold indicates increased performance compared to 2D RGB.

	IHLS	**HSI**	**RGB**	**nRGB**	**YCbCr**	**CIELAB**
AdaBoost	**0.250**	0.257	0.259	0.213	**0.280**	0.160
BayesNet	**0.413**	**0.400**	0.292	**0.332**	**0.391**	**0.413**
J48	0.540	**0.588**	0.550	**0.593**	**0.560**	**0.567**
MLP	0.549	0.481	0.602	0.527	0.571	0.534
NaiveBayes	0.284	0.220	0.292	**0.389**	0.227	**0.423**
Random Forest	**0.531**	**0.657**	0.415	**0.670**	**0.539**	**0.670**
RBF	**0.418**	0.186	0.351	0.327	0.333	**0.501**
SVM	**0.462**	**0.438**	0.312	0.204	0.227	**0.418**
Hist.	0.370	0.352	0.401	0.333	0.347	0.354

4.3. Role of skin color modeling

How does skin-color modeling affect the skin detection performance? For the role investigation, we consider 3D color spaces. For the visual interpretation of results, see Figure 1. We see that the Random Forest dominates other classifiers and has stable performance for all the color spaces. Of most importance is the combination of Random Forest and IHLS color space outperforming all possible combinations. The second best combination with the Random Forest is the HSI and CIELAB color spaces. The other dominant cluster is exhibited by the J48, which belongs to the same group of tree based classifiers as the Random Forest. In the case of J48, the highest F-measure is reported by the IHLS and YCbCr color spaces. Regarding AdaBoost and MLP, the combination with cylindrical color spaces like HSI and IHLS achieves state of the art results. The histogram approach reports good stable performance independent of the color space used and on average, the performance is close to that of Bayesian approach. Opposed to our expectations, even after an extensive parameter tuning, we were not able to achieve comparable results with SVM, although there is a significant boost in performance using cylindrical color spaces. Figure 2 shows sample skin detection using nine skin-color modeling methods in the IHLS (selected based on higher F-measure) color space.

4.4. Effect of color constancy

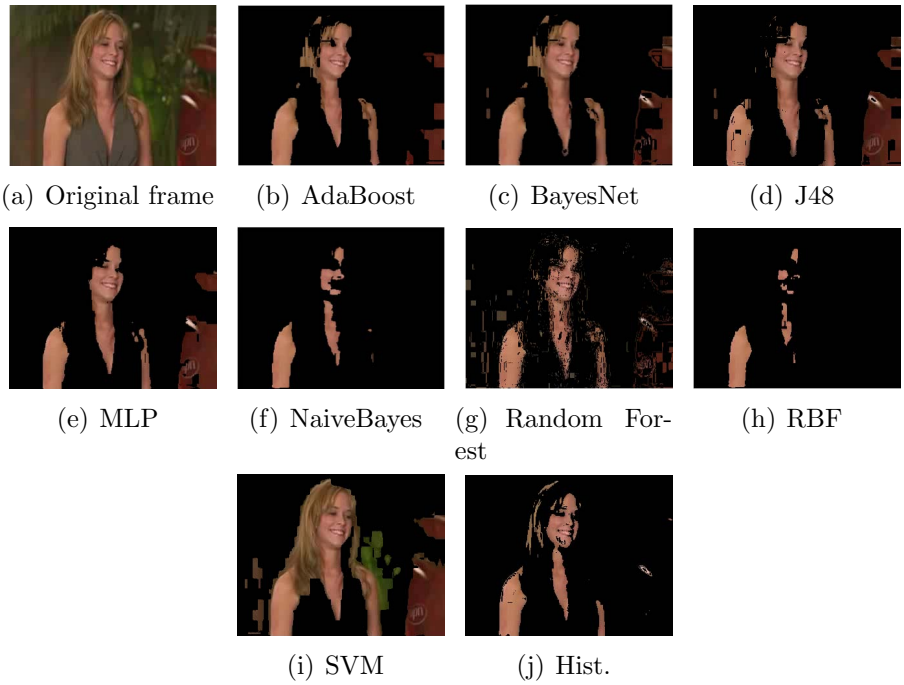
In this section, we show the effect of color constancy algorithms, using Gray-Edge [30], Gray-World [5], max-RGB [30], Shades of Gray [30] and

Table 3: Difference of F-Measure (multiplied by 100) in the 3D color compared to the 2D color space. An asterisk represents a significant performance difference (greater than 10) between the 3D and 2D color spaces.

	IHLS	**HSI**	**RGB**	**nRGB**	**YCbCr**	**CIELAB**
AdaBoost	7.00	4.27	0.14	3.70	-1.00	11.58*
BayesNet	17.70*	17.00*	2.99	3.76	9.88	14.73*
J48	14.42*	9.29	11.17*	3.35	12.04*	9.26
MLP	10.07*	10.92*	-1.19	4.24	5.61	6.57
NaiveBayes	18.17*	22.98*	-3.70	1.89	20.0*	3.09
Random Forest	21.40*	8.32	29.46*	2.97	16.59*	6.96
RBF	4.92	24.35*	3.79	9.32	15.69*	0.90
SVM	4.08	3.36	4.78	16.61*	15.77*	-1.84
Hist.	3.89	5.58	1.75	6.60	4.21	4.51

Bayesian color constancy [11] for color based skin classification. For this purpose, we fix YCbCr as the color space because of its wide usage for skin detection [29] and Random forest as the classifier because of its overall increased classification performance.

Figure 3 displays the skin locus in the YCbCr color space. Figure 3(a) shows skin locus without applying lighting correction. Figure 3(b-f) reports skin locus after applying Gray-Edge, Gray-World, max-RGB, Shades-of-Gray and Bayesian color constancy respectively. In comparison with the original skin locus, the skin locus in Figure 3(b-f) is compact compared with Figure 3(a). Figure 4 shows Accuracy, Precision, Recall and F-measure for the dataset with and without lighting correction using the Random forest classifier. We find that the skin locus in Figure 4(b) for Gray-Edge is much different than the original and therefore, we get decreased classification with F-measure of 0.69 for the Gray-Edge algorithm. Figure 3(c) shows the Gray-World skin locus which is compact compared to the uncorrected case, reporting an increase in classification with F-measure of 0.74. Regarding max-RGB, we get almost identical results to the Gray-Edge for Accuracy, Precision, Recall and F-measure. The Shades-of-Gray gives a compact locus with an increased performance, having F-measure of 0.78. Bayesian color constancy also reports an increase in performance with an F-measure of 0.76. From the results, it can be concluded that the lighting correction for skin classification can improve performance.



(a) Original frame (b) AdaBoost (c) BayesNet (d) J48

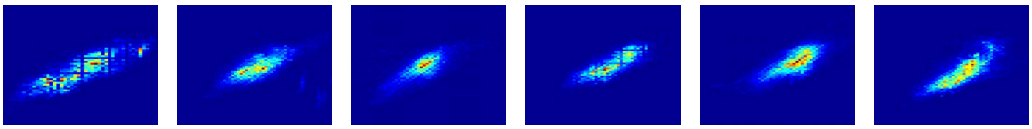
(e) MLP (f) NaiveBayes (g) Random Forest (h) RBF

(i) SVM (j) Hist.

Figure 2: Skin detection using different classifiers in IHLS color space. Non-skin is black.

5. Conclusion

We evaluated color based skin classification using different skin-color modeling techniques. Six color spaces (IHLS, HSI, RGB, normalized RGB, YCbCr and CIELAB) and nine skin color modeling approaches (AdaBoost, Bayesian network, J48, Multilayer Perceptron, Naive Bayesian, Random Forest, RBF network, SVM and Histogram approach) are evaluated on 8991 manually per-pixel annotated images on the basis of F-measure. We observe that (1) color space transformation does affect the overall skin performance,



(a) Original (b) GE (c) GW (d) mRGB (e) SG (f) Bayes

Figure 3: Skin spread and color constancy. (GE: Gray-Edge, GW: Gray-World, mRGB: max-RGB, SG: Shades-of-Gray, Bayes: Bayesian color constancy)

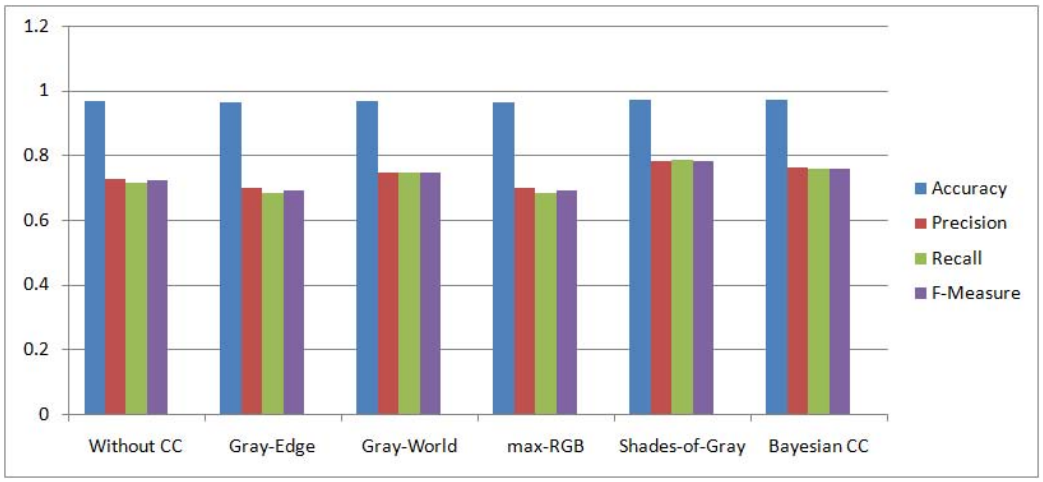


Figure 4: Results of the Random Forest classifier for color constancy (CC: color constancy).

(2) performance degrades with the removal of the illuminance component, (3) the proper selection of the skin color modeling approach is important for skin detection, and (4) skin classification performance can be increased with the usage of lighting correction algorithms. Lighting correction can also have negative effect on the results. This is due to the fact that color constancy algorithms produce a compact representation of skin locus (skin color ranges in a color space) but the skin locus can also be shifted and deviated in the chromaticity space, resulting in varying performance.

We find that the decision tree based classifiers, especially the Random Forest, are preferable for pixel-based skin classification, independent of the color spaces. For noisy and dark visual material, as we are using, nRGB is not the color space of choice. Cylindrical color spaces (IHLS, HSI) outperform other color spaces providing robust data for classification. The best performing combination is IHLS with Random Forest. The best performing skin detection system can be used for example, as a pre-processing step for graph cuts approach which takes into account the neighborhood relationship of the pixels. With such a thorough evaluation, we have provided a framework for selecting the best approach for skin detection. An interesting question that remains to be investigated is whether these conclusions hold for other types of material, such as photo collections and collections where there is potentially greater changes in illumination.

References

[1] Albiol, A., Torres, L., & Delp, E. J. (2001). Optimum color spaces for skin detection. In *Proceedings of the ICIP* (pp. 122–124).

[2] Argyros, A. A., & Lourakis, M. I. (2004). Real-time tracking of multiple skin-colored objects with a possibly moving camera. In *ECCV* (pp. 368–379).

[3] Brand, J., & Mason, J. (2000). A comparative assessment of three approaches to pixel-level human skin-detection. In *ICPR* (pp. 1056–1059). volume 1.

[4] Brown, D., Craw, I., & Lewthwaite, J. (2001). A SOM based approach to skin detection with application in real time systems. In *BMVC'01* (pp. 491–500).

[5] Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, *310*, 1 – 26.

[6] Cai, J., & Goshtasby, A. (1999). Detecting human faces in color images. *Image and Vision Computing*, *18*, 63–75.

[7] Chai, D., & Ngan, K. (1998). Locating facial region of a head-and-shoulders color image. In *Int. Conf. Automatic Face and Gesture Recognition* (pp. 124–129).

[8] Fleck, M. M., Forsyth, D. A., & Bregler, C. (1996). Finding naked people. In *ECCV* (pp. 593–602).

[9] Fu, Z., Yang, J., Hu, W., & Tan, T. (2004). Mixture clustering using multidimensional histograms for skin detection. In *ICPR* (pp. 549–552). Washington, DC, USA.

[10] Garcia, C., & Tziritas, G. (Sep 1999). Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia*, *1*, 264–277.

[11] Gehler, P. V., Rother, C., Blake, A., Minka, T., & Sharp, T. (2008). Bayesian color constancy revisited. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1–8).

[12] Gonzalez, R. C., & Woods, R. E. (2001). *Digital Image Processing*. Addison-Wesley.

[13] Hanbury, A. (2003). A 3d-polar coordinate colour representation well adapted to image analysis. In *SCIA* (pp. 804–811).

[14] Hsu, R., Abdel-Mottaleb, M., & Jain, A. (2002). Face detection in color images. *PAMI*, *24*, 696–706.

[15] Jones, M. J., & Rehg, J. M. (2002). Statistical color models with application to skin detection. *IJCV*, *46*, 81–96.

[16] Kakumanu, P., Makrogiannis, S., & Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *PR*, *40*, 1106–1122.

[17] Khan, R., Stöttinger, J., & Kampel, M. (2008). An adaptive multiple model approach for fast content-based skin detection in on-line videos. In *ACM MM, AREA workshop* (pp. 89–96).

[18] Lee, J.-S., Kuo, Y.-M., Chung, P.-C., & Chen, E.-L. (2007). Naked image detection based on adaptive and extensible skin color model. *PR*, *40*, 2261–2270.

[19] Lee, J. Y., & Yoo, S. (2002). An elliptical boundary model for skin color detection. In *ISST* (pp. 579–584).

[20] Peer, P., Kovac, J., & Solina, F. (2003). Human skin colour clustering for face detection. In *EUROCON* (pp. 144–148, vol.2).

[21] Phung, S. L., Bouzerdoum, A., & Chai, D. (2005). Skin segmentation using color pixel classification: Analysis and comparison. *PAMI*, *27*, 148–154.

[22] Quinlan, R. J. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc.

[23] Schmugge, S. J., Jayaram, S., Shin, M. C., & Tsap, L. V. (2007). Objective evaluation of approaches of skin detection using ROC analysis. *Computer Vision and Image Understanding*, *108*, 41 – 51.

[24] Sebe, N., Cohen, I., Huang, T. S., & Gevers, T. (2004). Skin detection: A Bayesian network approach. In *ICPR* (pp. 903–906).

[25] Sigal, L., Sclaroff, S., & Athitsos, V. (July 2004). Skin color-based video segmentation under time-varying illumination. *PAMI*, *26*, 862–877.

[26] Störring, M., Andersen, H., & Granum, E. (2000). Estimation of the illuminant colour from human skin colour. In *IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 64–69).

[27] Stöttinger, J., Hanbury, A., Liensberger, C., & Khan, R. (2009). Skin paths for contextual flagging adult videos. In *International Symposium on Visual Computing* (pp. 303–314).

[28] Terrillon, J.-C., & Akamatsu, S. (2000). Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images. In *Proceedings of the 12th Conference on Vision Interface* (pp. 180–187).

[29] Vezhnevets, V., Sazonov, V., & Andreev, A. (2003). A survey on pixel-based skin color detection techniques. In *GraphiCon* (pp. 85–92).

[30] van de Weijer, J., Gevers, T., & Gijsenij, A. (2007). Edge-based color constancy. *IEEE Transactions on Image Processing*, *16*, 2207–2214.

[31] Wong, K., Lam, K., & Siu, W. (2003). A robust scheme for live detection of human faces in color images. *Signal Processing: Image Communication*, *18*, 103–114.

[32] Yang, J., Lu, W., & Waibel, A. (1997). Skin-color modeling and adaptation. In *ACCV* (pp. 687–694).

[33] Yang, M., & Ahuja, N. (1999). Gaussian mixture model for human skin color and its application in image and video databases. In *SPIE* (pp. 458–466).

[34] Zarit, B. D., Super, B. J., & Quek, F. K. H. (1999). Comparison of five color models in skin pixel classification. In *International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems* (pp. 58–63). USA.