Universal Seed Skin Segmentation

Rehanullah Khan[1], Allan Hanbury[3], and Julian Stöttinger[1,2]

[1] Computer Vision Lab, Vienna University of Technology
[2] CogVis Ltd., Vienna, Austria
[3] Information Retrieval Facility, Vienna, Austria

Abstract. We present a principled approach for general skin segmentation using graph cuts. We present the idea of a highly adaptive universal seed thereby exploiting the positive training data only. We model the skin segmentation as a min-cut problem on a graph defined by the image color characteristics. The prior graph cuts based approaches for skin segmentation do not provide general skin detection when the information of foreground or background seeds is not available. We propose a concept for processing arbitrary images; using a universal seed to overcome the potential lack of successful seed detections thereby providing basis for general skin segmentation. The advantage of the proposed approach is that it is based on skin sampled training data only making it robust to unseen backgrounds. It exploits the spatial relationship among the neighboring skin pixels providing more accurate and stable skin blobs. Extensive evaluation on a dataset of 8991 images with annotated pixel-level ground truth show that the universal seed approach outperforms other state of the art approaches.

1 Introduction

Skin detection has a wide range of applications both in human computer interaction and content based analysis. Applications such as: detecting and tracking of human body parts [1], face detection [2], naked people detection, people retrieval in multimedia databases [3] and blocking objectionable content [4], all benefit from skin detection. 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.

We present the idea of skin segmentation based on a global seed which we represent as the universal seed. With the universal seed we successfully remove the need for local foreground seeds from an image. No time consuming training is required and the universal seed can easily be updated with new skin examples under different lighting conditions. The skin segmentation problem is modeled as a min-cut problem on a graph defined by the image color characteristics. We use an efficient algorithm [5] for finding min-cut/max-flow in a graph. Experiments are performed following recent evaluation [6,7,8,9,10]. We select the best performing approaches, namely AdaBoost, BayesNet, NaiveBayes, YCbCr static model, RBF network and J48. The results show that our universal seed approach outperforms other approaches.

Related work regarding skin detection and segmentation is presented in Section 2. Section 3 presents a framework for seed based segmentation, the graph building process,

G. Bebis et al. (Eds.): ISVC 2010, Part II, LNCS 6454, pp. 75–84, 2010.

weights assignment and the universal seed for skin segmentation. Experimental details are given in Section 4 and the results are discussed in Section 5. Section 6 concludes.

2 Related Work

In computer vision, skin detection is used as a first step in face detection, e.g. [11], and for localization in the first stages of gesture tracking systems, e.g. [1]. It has also been used in the detection of naked people [12,13] and for blocking objectionable content [4]. 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 [14]. Non-parametric methods estimate the skin-color from the histogram that is generated by the training data used [15].

An efficient and widely used method is the definition of classifiers that build upon the approach of skin clustering. This thresholding of different color space coordinates is used in many approaches, e.g. [16] 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 [17]. The main drawback of static filters is a comparably high number of false detections [6]. Khan et al [18] 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 and reduce false positives.

The choice of a color space is important for many computer vision algorithms because it induces the equivalence classes to the detection algorithms [19]. 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 [20,21]. Perceptually uniform color spaces like the CIELAB, CIELUV are used for skin detection e.g. in [2]. 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. [22].

Neural networks [23], Bayesian Networks e.g. [8], Gaussian classifiers e.g. [15], and self organizing maps [20] have been used to try to increase the classification accuracy.

In the literature of segmentation, Graph-cuts provide a globally optimal solution for N-dimensional segmentation when the cost function has specific properties as defined in [24]. A semi-automatic method for general image segmentation was created by Boykov et al. [24]. A user puts marks on the image, acting as a cue for being counted as segments and updating the marks without graph reconstruction. The method of Li et al. [25] consists of two steps: an object marking task as in [24] and the pre-segmentation, followed by a simple boundary editing process. The work of Shi & Malik [26] segments the image into many non-overlapping regions. They introduced normalized graph cuts and the method has often been used in combination with computing pixel neighborhood relations using brightness, color and texture cues [27].

Micusik et al. [28] make an assumption that each textured or colored region can be represented by a small template, called the seed and positioning of the seed across the input image gives many possible sub-segmentations of the image. A probability map assigns each pixel to just one most probable region and produces the final pyramid representing various detailed segmentations. Each sub-segmentation is obtained as the min-cut/max-flow in the graph built from the image and the seed. Graph cuts is used for skin segmentation in [29] using both the foreground and background seeds. The foreground seeds are obtained through the face detection algorithm assuming detected faces in the image for skin segmentation, thereby lacking in generic skin detection. We introduce the universal seed concept thereby providing a basis for general skin segmentation when no seeds are available from local images.

3 Universal Seed Skin Segmentation

Using the universal seed based skin segmentation, we exploit the spatial relationship among the skin pixels thereby achieving a performance boost for segmentation. For skin segmentation, a graph is constructed whose nodes represent image pixels and whose edges represent the weights. The min-cut/max-flow algorithm presented in [5] is used for the graph cut.

3.1 Graph Representing the Skin Image

We use a skin segmentation technique based on an interactive graph cut method as used in [28] and [24]. Before segmentation we construct a graph. The graph is shown in Figure 1 for a 9 pixel image and 8-point neighborhood N with representative pixels q and r. For an image, the number of graph nodes equals the pixel count plus two extra nodes labeled as F, B representing foreground and background respectively. There are two types of weights to be set, the foreground/background (skin/non-skin) weights and the neighborhood weights. The foreground weights are computed based on the universal seed. The background weights are calculated from all of the pixels in the image. For min-cut/max-flow a very efficient algorithm [5] is used.

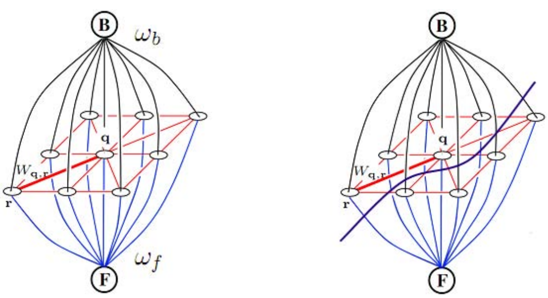


Fig. 1. Left: Graph representation for 9 pixel image. Right: Right: A cut on the graph.

Neighborhood Weights. For a greyscale image the neighborhood weights (weight matrix) $W_{q,r}$, as reported by [24] are

$$W_{q,r} \propto e^{-\frac{||I_q - I_r||^2}{2\sigma^2}} . \frac{1}{||q-r||} \quad (1)$$

where I_q and I_r are the intensities at point q and point r, $||q-r||$ is the distance between these points and σ is a parameter. For color images we modify the above function to take color into account,

$$W_{q,r} = e^{-\frac{||c_q - c_r||^2}{\sigma_1}} . \frac{1}{||q-r||} \tag{2}$$

where c_q and c_r are the YCbCr vectors of points at the position q and r. $||q - r||$ is the distance between these points and σ_1 is a parameter. For skin detection purposes a value of σ_1 = 0.02 is used, which is the optimized value for segmentation in [30] and is obtained experimentally for giving the best performance on a large database of images. We use a neighborhood window of size 21×21. For skin segmentation we use a sampling rate of 0.3. This means that we only select at random a sample of 30% of all the pixels in the window. There are two reasons: Firstly by using only a fraction of pixels we reduce the computational demands and secondly only a fraction of pixels allows the use of larger windows and at the same time preserves the spatial relationship between the neighboring pixels.

Foreground/Background Weights. For foreground/background weights, the regional penalty of a point as being "skin" (foreground) $\mathcal{F}$ or "non-skin" (background) $\mathcal{B}$ [30] is

$$R_{\mathcal{F}|q} = -\ln p(\mathcal{B}|c_q) \quad and \quad R_{\mathcal{B}|q} = -\ln p(\mathcal{F}|c_q) \tag{3}$$

where $c_q = (c_Y, c_{Cb}, c_{Cr})^T$ stands for a vector in $\mathbb{R}^3$ of YCbCr values at the pixel q. The posterior probabilities in Equation 3 are computed as follows,

$$p_{(\mathcal{B}|c_q)} = \frac{p(c_q|\mathcal{B})p(\mathcal{B})}{p(\mathcal{B})p(c_q|\mathcal{B}) + p(\mathcal{F})p(c_q|\mathcal{F})} \tag{4}$$

For the skin segmentation problem we first demonstrate it on $p(\mathcal{B}|c_q)$, for $p(\mathcal{F}|c_q)$ the steps are analogous. Initially we fix $p(\mathcal{F}) = p(\mathcal{B}) = 1/2$ and thus,

$$p_{(\mathcal{B}|c_q)} = \frac{p(c_q|\mathcal{B})}{p(c_q|\mathcal{B}) + p(c_q|\mathcal{F})} \tag{5}$$

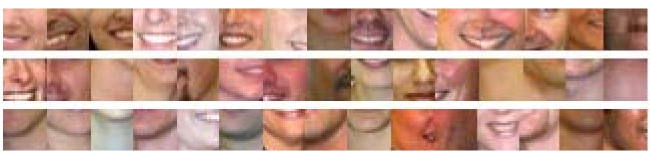
where the "skin" and "non-skin" prior probabilities are

$$p(c_q|\mathcal{F}) = f^{Y}_{c_Y}.f^{Cb}_{c_{Cb}}.f^{Cr}_{c_{Cr}} \quad and \quad p(c_q|\mathcal{B}) = b^{Y}_{c_Y}.b^{Cb}_{c_{Cb}}.b^{Cr}_{c_{Cr}} \tag{6}$$

and $f_i^{\{Y,Cb,Cr\}}$, resp. $b_i^{\{Y,Cb,Cr\}}$, represents the foreground, resp. the background histogram of each color channel separately at the ith bin. $\omega_f = \lambda R_{\mathcal{F}|q}$, where ω_f is the foreground weight, λ is set to 1000 and controls the importance of penalties for foreground and background against the neighborhood weights. Similarly the background weight ω_b is given by $\omega_b = \lambda R_{\mathcal{B}|q}$.

3.2 Universal Seed

A local skin patch from the image can be used as a seed to detect skin in an image. One solution for obtaining local seeds from an image is to use a face detector. A face detector is normally followed by post filtering steps for the removal of non-skin portion from the face for using it as the local skin seed. In case of failure of face detection the local seed based skin segmentation will fail. We propose a concept for processing arbitrary images; using a universal seed to overcome the potential lack of successful seed detections thereby providing basis for using static foreground weights based skin segmentation. With the universal seed, the objective is producing a seed/template that is as general as possible and can be used as skin filter. We base the segmentation process on positive training data samples only, exploiting the spatial relationship between the neighborhood skin pixels. For the universal seed, different skin tones are collected, see Figure 2. These positive skin samples cover different ethnicities in different lighting conditions. For the universal seed we do not use the negative (non-skin) portion of the image. Since there could be infinite background/negative training data, the objective is taking a skin/non-skin decision based on representative skin samples. For the skin scenario this makes sense as the skin covers a well defined region in a color space. We denote these positive representative skin samples as the universal seed. The universal seed is highly adaptive. For adding a new skin patch under different lighting conditions we just have to merge it with skin patches and recalculate the foreground histogram. The foreground histogram in Equation 6 for a new image is calculated based on this seed. The background histogram is calculated form the whole image. Since $\sum_{i=1}^{N}\bar{b}_i$=1, the probability $p(c_q|\mathcal{B})$ gives smaller values than $p(c_q|\mathcal{F})$ for the "skin" colors present in the universal seed therefore we compute the background histogram from all the image pixels.



(a)

Fig. 2. Skin samples used for universal seed

4 Experiments

Universal seed based skin segmentation is compared to skin segmentation using AdaBoost, BayesNet, NaiveBayes, YCbCr static model, RBF network and J48 on the basis of F-measure and specificity. The dataset used is available on-line[1]. A total of 8991 images with annotated pixel-level ground truth are used as test images for evaluation.

Following Kakumanu [6], we select explicit thresholding (YCbCr static filter) because it is a fast and simple rule based filter, successfully used in [7] for skin detection. We select the Bayesian network [8] and neural network [9] based classifiers based on the reported best performance in [6]. Based on the independent feature model we select the Naive Bayesian. J48 is selected based on the superior performance in [31] for tree based classifier. Following [10] boosting is the optimal detection method for skin color and faces.

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

In experiments we aim to combine these evaluations in one experimental setup and evaluate our proposed approach. It is measured precisely per pixel for the dataset.

The evaluation is based on F-measure and specificity for 8991 test images. The F-measure is calculated by evenly weighting precision and recall. The specificity is defined as the true negative rate. For all the experiments, the color space used is YCbCr. The three components of YCbCr are used as feature vectors for AdaBoost, BayesNet, NaiveBayes, RBF network and J48.

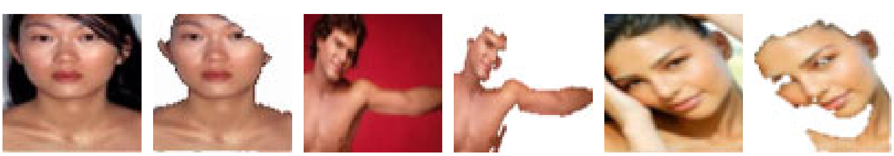
We aim to evaluate the approaches on unseen data. For training data, we choose 118 arbitrary images from the Internet with a great variety of skin colors and lighting conditions. Therefore, the training data is non-overlapping and from different sources with respect to the test dataset. The images are annotated per pixel and provide positive and negative training data. The total number of features for skin pixels is 2113703, the number of negative training samples is 6948697. For the universal seed we do not use the negative (non-skin) pixels. This positive set of skin pixels is denoted as the universal seed. Example positive training data can be found in Figure 2. In this context, the proposed approach uses significantly less training data than other approaches. In the following, the experimental results are presented and discussed.

5 Results

Example results of the proposed approach using the universal seed approach are visualized in Figure 3. Figure 4 reports cases where skin is missed or false detections are considered as skin.

Specificity calculated on a per pixel basis is reported in Figure 5. Figure 6 reports mean and standard deviation for specificity and F-measure calculated on a per image basis. Figure 7 shows the F-measure calculated on a per pixel basis. In Figure 5, it can be seen that the specificity of universal seed approach is higher than YCbCr static approach and less than AdaBoost, BayesNet, NaiveBayes, RBF network and J48. Similarly, mean and standard deviation for specificity in Figure 6 shows that the specificity mean (0.54) of Universal seed approach is lower than that of AdaBoost (0.77), BayesNet (0.63), NaiveBayes (0.68), RBF network (0.85) and J48 (0.83) with standard deviations of 0.35, 0.21, 0.32, 0.30, 0.11, 0.11 for universal seed, AdaBoost, BayesNet, NaiveBayes, RBF network and J48 respectively. This is because the universal seed approach is based on training on positive data only and therefore lower true negative rate. The specificity mean of universal seed is higher than that of YCbCr static approach (0.41).

In terms of increasing/decreasing trend in Figure 6 for specificity, the universal seed approach is similar to BayesNet and NaiveBayes approaches. The mean and standard



(a) (b) (c) (d) (e) (f)

Fig. 3. Universal seed skin segmentation: successful skin detection

Fig. 4. Universal seed skin segmentation: cases where skin is not properly segmented

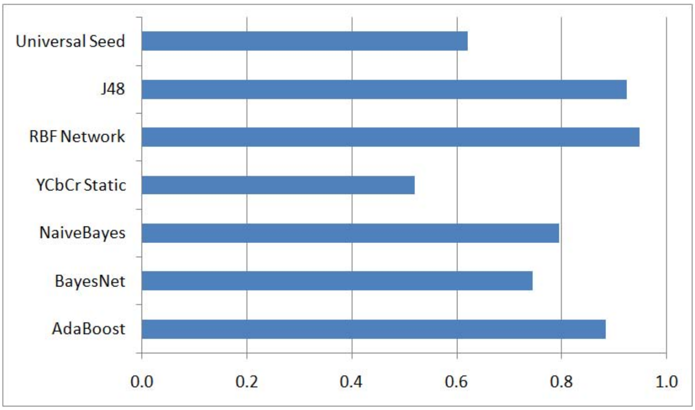


Fig. 5. Specificity for 8991 images. Since the universal seed approach is based on positive data only, the true negative rate is not as high as the F-measure given in Figure 7. The values reported are on a per pixel basis.

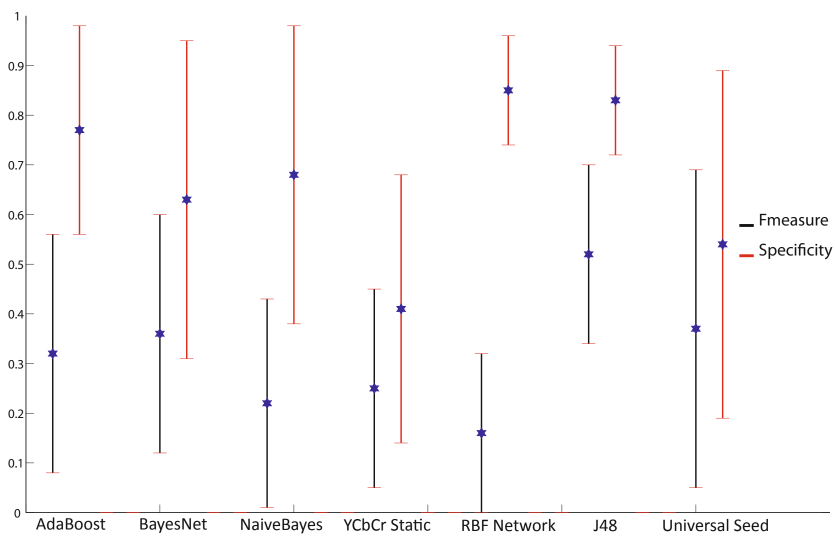


Fig. 6. Mean and standard deviations for F-measure and specificity for 8991 test images. The values reported are on a per image basis.

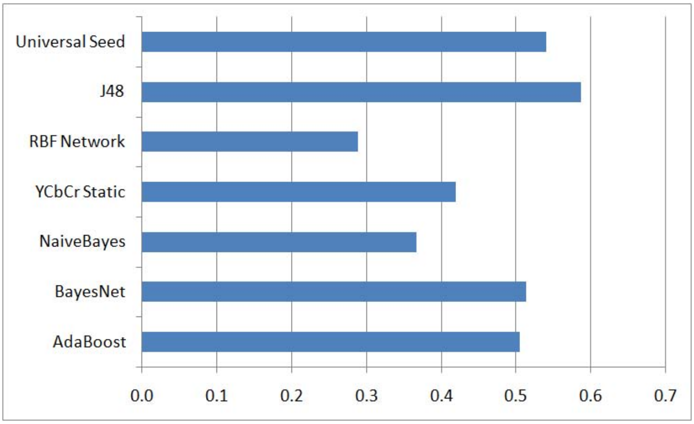


Fig. 7. F-measure for 8991 images. The universal seed approach outperforms other approaches in terms of precision and recall with the exception of tree based classifier (J48). The values are calculated on a per pixel basis.

deviation for F-measure in Figure 6 show that the universal seed mean (0.37) is lower than J48 (0.52) and higher than AdaBoost (0.32), BayesNet (0.36), NaiveBayes (0.22), YCbCr static (0.25) and RBF network (0.16) with standard deviations of 0.32, 0.18, 0.24, 0.24, 0.21, 0.20, and 0.16 for universal seed, J48, AdaBoost, BayesNet, NaiveBayes, YCbCr static and RBF network respectively.

Regarding precision and recall, Figure 7 shows that the universal seed approach has higher F-measure (0.54) compared to other approaches with the exception of the tree based classifier J48 with F-measure of 0.59. The universal seed approach provides increased classification performance of almost 4% to AdaBoost, 3% to Bayesian network, 18% to Naive Bayesian, 12% to YCbCr static, 26% to RBF network and decreased performance of almost 5% compared to J48.

For the test data set the universal seed approach outperforms other approaches in terms of precision and recall with the exception of J48. For the skin detection scenario, the J48 simple rule based decision classification generalizes well for simple feature based classification with high F-score outperforming the universal seed approach.

6 Conclusion

We provide a basis for skin segmentation based on positive training data only using a universal seed to overcome the potential lack of successful seed detections from within an image. The universal seed for skin detection is based on learning a skin model for foreground weights of the graph using skin samples. Experiments on a database of 8991 images with annotated pixel-level ground truth show that using positive training data only, the universal seed approach is well suited to stable and robust skin detection, outperforming other approaches which require negative and positive training data with the exception of J48.

Acknowledgment

This work was partly supported by the Austrian Research Promotion Agency (FFG), project OMOR 815994, the Higher Education Commission of Pakistan under Overseas scholarships (Phase-II-Batch-I) and CogVis[2] Ltd. However, this paper reflects only the authors views; the FFG, HEC or CogVis Ltd. are not liable for any use that may be made of the information contained herein.

References

1. Argyros, A.A., Lourakis, M.I.: Real-time tracking of multiple skin-colored objects with a possibly moving camera. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 368–379. Springer, Heidelberg (2004)
2. Cai, J., Goshtasby, A.: Detecting human faces in color images. Image and Vision Computing 18, 63–75 (1999)
3. Cao, L.L., Li, X.L., Yu, N.H., Liu, Z.K.: Naked people retrieval based on adaboost learning. In: MLC, pp. 1133–1138 (2002)
4. Stöttinger, J., Hanbury, A., Liensberger, C., Khan, R.: Skin paths for contextual flagging adult videos. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Pajarola, R., Lindstrom, P., Hinkenjann, A., Encarnação, M.L., Silva, C.T., Coming, D. (eds.) ISVC 2009. LNCS, vol. 5876, pp. 303–314. Springer, Heidelberg (2009)
5. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. PAMI 26, 1124–1137 (2004)
6. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A survey of skin-color modeling and detection methods. PR 40, 1106–1122 (2007)
7. Chai, D., Ngan, K.: Locating facial region of a head-and-shoulders color image. In: Int. Conf. Automatic Face and Gesture Recognition, pp. 124–129 (1998)
8. Sebe, N., Cohen, I., Huang, T.S., Gevers, T.: Skin detection: A Bayesian network approach. In: ICPR, pp. 903–906 (2004)
9. Phung, S.L., Chai, D., Bouzerdoum, A.: A universal and robust human skin color model using neural networks. In: IJCNN, pp. 2844–2849 (2001)
10. Pavlovic, V.: Boosted detection of objects and attributes. In: CVPR (2001)
11. Hsu, R., Abdel-Mottaleb, M., Jain, A.: Face detection in color images. PAMI 24, 696–706 (2002)
12. Fleck, M.M., Forsyth, D.A., Bregler, C.: Finding naked people. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1065, pp. 593–602. Springer, Heidelberg (1996)
13. Lee, J.S., Kuo, Y.M., Chung, P.C., Chen, E.L.: Naked image detection based on adaptive and extensible skin color model. PR 40, 2261–2270 (2007)
14. Yang, M., Ahuja, N.: Gaussian mixture model for human skin color and its application in image and video databases. In: SPIE, pp. 458–466 (1999)
15. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. IJCV 46, 81–96 (2002)
16. Phung, S.L., Bouzerdoum, A., Chai, D.: Skin segmentation using color pixel classification: Analysis and comparison. PAMI 27, 148–154 (2005)
17. Peer, P., Kovac, J., Solina, F.: Human skin colour clustering for face detection. In: EUROCON, vol. 2, pp. 144–148 (2003)

[2] www.cogvis.at

18. Khan, R., Stöttinger, J., Kampel, M.: An adaptive multiple model approach for fast content-based skin detection in on-line videos. In: ACM MM, AREA workshop (2008)
19. Stokman, H., Gevers, T.: Selection and fusion of color models for image feature detection. IEEE Trans. Pattern Anal. Mach. Intell. 29, 371–381 (2007)
20. Brown, D., Craw, I., Lewthwaite, J.: A SOM based approach to skin detection with application in real time systems. In: BMVC 2001, pp. 491–500 (2001)
21. Fu, Z., Yang, J., Hu, W., Tan, T.: Mixture clustering using multidimensional histograms for skin detection. In: ICPR, Washington, DC, USA, pp. 549–552 (2004)
22. Wong, K., Lam, K., Siu, W.: A robust scheme for live detection of human faces in color images. SPIC 18, 103–114 (2003)
23. Lee, J.Y., Suk-in, Y.: An elliptical boundary model for skin color detection. In: ISST, pp. 579–584 (2002)
24. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In: ICCV-WS 1999, vol. 1, pp. 105–112 (2001)
25. Li, Y., Sun, J., Tang, C.K., Shum, H.Y.: Lazy snapping. In: SIGGRAPH, New York, NY, USA, pp. 303–308 (2004)
26. Shi, J., Malik, J.: Normalized cuts and image segmentation. PAMI 22, 888–905 (2000)
27. Malik, J., Belongie, S., Leung, T.K., Shi, J.: Contour and texture analysis for image segmentation. International Journal of Computer Vision 43, 7–27 (2001)
28. Micusík, B., Hanbury, A.: Automatic image segmentation by positioning a seed. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 468–480. Springer, Heidelberg (2006)
29. Hu, Z., Wang, G., Lin, X., Yan, H.: Skin segmentation based on graph cuts. Science and Technology 14, 478–486 (2009)
30. Micusík, B., Hanbury, A.: Steerable semi-automatic segmentation of textured images. In: Kalviainen, H., Parkkinen, J., Kaarna, A. (eds.) SCIA 2005. LNCS, vol. 3540, pp. 35–44. Springer, Heidelberg (2005)
31. Khan, R., Hanbury, A., Stoettinger, J.: Skin detection: A random forest approach. In: ICIP (to appear, 2010)