Skin Color Detection Using Region-Based Approach

Rudra PK Poudel   *rpoudel@bournemouth.ac.uk*
Media School, Bournemouth University
Poole, BH12 5BB, UK

Jian J Zhang   *jzhang@bournemouth.ac.uk*
Media School, Bournemouth University
Poole, BH12 5BB, UK

David Liu   *david-Liu@siemens.com*
Siemens Corporate Research
755 College Road East, Princeton, NJ 08540, USA

Hammadi Nait-Charif   *hncharif@bournemouth.ac.uk*
Media School, Bournemouth University
Poole, BH12 5BB, UK

Abstract

Skin color provides a powerful cue for complex computer vision applications. Although skin color detection has been an active research area for decades, the mainstream technology is based on the individual pixels. This paper, which extended our previous work [1], presented a new region-based technique for skin color detection which outperformed the current state-of-the-art pixel-based skin color detection technique on the popular Compaq dataset [2]. Color and spatial distance based clustering technique is used to extract the regions from the images, also known as superpixels followed by a state-of-the-art non-parametric pixel-based skin color classifier called the basic skin color classifier. The pixel-based skin color evidence is then aggregated to classify the superpixels. Finally, the Conditional Random Field (CRF) is applied to further improve the results. As CRF operates over superpixels, the computational overhead is minimal. Our technique achieved 91.17% true positive rate with 13.12% false negative rate on the Compaq dataset tested over approximately 14,000 web images.

Keywords: Skin Color Detection, Bayes Classifier, Superpixels, MRF.

1. INTRODUCTION

Skin color provides a powerful cue in complex computer vision applications such as hand tracking, face tracking, and pornography detection. Skin color detection is computationally efficient yet invariant of rotation, scaling and occlusion. Which are the major reasons for its popularity. The main challenges of skin color detection are illumination, ethnicity background, make-up, hairstyle, eyeglasses, background color, shadows and motion [3]. Most of the skin color detection problems could be overcome by using infrared [4] and spectral imaging [5]. However, such systems are expensive as well as cumbersome to implement. Moreover, there are many situations such as image retrieval on the internet where such systems cannot be used.

Most of the skin color detection techniques are pixel-based, which treat each skin or non-skin pixel individually without considering its neighbors. However, it is natural to treat skin or non-skin as regions instead of individual pixels. Hence, this research focuses on the region-based skin color detection technique. Surprisingly, there are only few region-based skin detection techniques [6], [7], [8] and [9]. Kruppa [7], Yang and Ahuja [6] searched for elliptical skin color shape to find the face. Sebe [9] used fixed 3x3 pixel patches to train a Bayesian network, and Jedyank [8] smoothed the results using hidden Markov model. This paper proposes a new technique purely

based on the concept of regions, irrespective of the underlying geometrical shape. As such, this technique can be easily integrated into any skin detection based system.

Our technique uses a segmentation technique called superpixel [10] and [11] to group the similar color pixels together. Then each superpixel is classified as skin or non-skin by aggregating pixel-using the evidence obtained from a histogram based (also known as non-parametric) Bayesian classifier similar to [2]. However, any suitable pixel-based or superpixel-based skin color classification technique can be used. The result is further improved with Conditional Random Fields (CRF) which operates over superpixels instead of individual pixels. Even though the segmentation cost is an overhead over the pixel-based approach, it greatly reduces the processing cost further down the line, such as smoothing with CRF. Besides aggregation of pixels into regions helps to reduce local redundancy and the probability of merging unrelated pixels [12]. Since superpixels preserve the boundary of the objects, it helps to achieve accurate object segmentation [13].

The presented technique not only outperforms the current state-of-the-art pixel-based skin color detection techniques but also extracts larger skin regions while still keeping the false-positive rate lower (see Table 1 and Figure 2), providing semantically more meaningful skin regions. This could in turn benefit higher-level vision tasks, such as face, hand or human body detection. Related work is discussed in section 2; section 3 presents the proposed region-based skin color detection technique; experiments and results are discussed in section 4. Finally, we summarize our work in section 5.
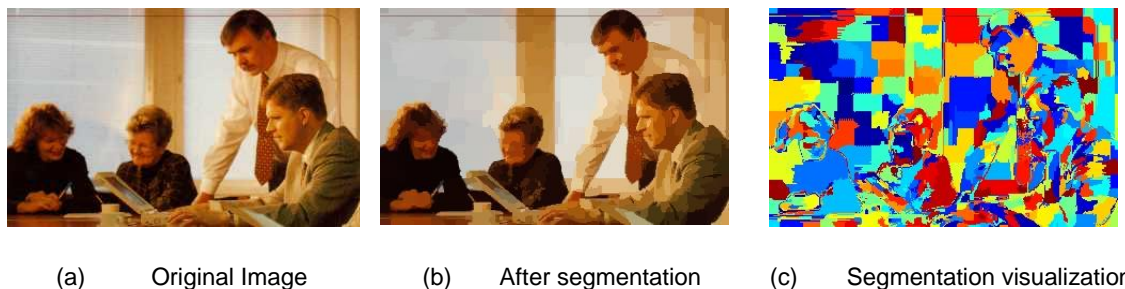

| (a) Original Image | (b) After segmentation | (c) Segmentation visualization |

FIGURE 1: An example of superpixel segmentation. A five dimensional vector is used to extract the superpixels: three RGB color channels and two positional coordinates of the pixel in the image.

2. RELATED WORK

Skin color detection has two important parts: one is color space selection and another is color modeling. RGB: [14], [15], [16], [2], [9], HSV: [17], [18], [19], CIE-Lab: [20], [21], YCbCr: [22], [23], and normalized RGB: [16] are popular color spaces, with RGB and HSV being the most frequently used. CIE-Lab uniformly represents the color based on how two colors differ to the human observer. HSV shows better results under varying illumination [3]. Most systems choose RGB color space because the illumination variation can be eliminated by increasing sample size [2]. Due to this reason the RGB color space is chosen in our experiments.

Skin color modeling techniques fall into three categories: explicitly defined skin region [24], non-parametric and parametric techniques. Histogram based Bayes classifier is a popular non-parametric modeling approach. Jones and Rehg [2] used RGB color space and histograms based Bayes classifier and obtained 90% true positive rate with 14.5% false positive rate on unconstrained web images, a dataset made up of approximately 14,000 images. On parametric skin modeling technique, Gaussian mixture has been found to be producing the best result [25], [26]. However, Jones and Rehg [2] showed that given enough samples, the histogram based Bayes classifier technique is slightly better than Gaussian mixture. Neural Network [27], self-organizing map [16], Bayesian network [9] and a few other techniques have also been used for skin color modeling.

This paper presents a region-based skin color detection technique with no prior knowledge on the geometric shape of the skin regions. The works of Yang and Ahuja [6], Kruppa [7], Jedyank [8] and Sebe [9] are the closest to ours. However, Yang and Ahuja [6] used multi-scale segmentations to find elliptical regions for face detection. Hence, their model is biased toward the skin colored elliptical objects. Kruppa [7] also used a similar concept to find the elliptical regions using color and shape information for the face detection. Sebe [9] used 3x3 fixed size pixel patches. Our presented technique uses patches with varying sizes, which is purely based on the image evidence, i.e. skin color in this case. Also, Jedyank [8] used hidden Markov model at pixel level, while we use conditional random fields and operate on superpixel, as described in the section 3.4.

3. PROPOSED FRAMEWORK

We argue that skin is better presented as regions rather than individual pixels. The proposed region-based approach has four major components: basic skin classifier (section 3.1), extraction of regions called superpixels (section 3.2), superpixels classification (section 3.3), and a smoothing procedure with conditional random fields (CRF) (section 3.4). Each step is discussed in detail below.

3.1 Basic Skin Color Classifier

Any good skin color classification technique can be used as a basic skin color classifier. This paper uses the histogram based Bayesian classifier similar to that of Jones and Rehg [2], a state-of-the-art skin color detection technique.

Learning Skin and Non-Skin Histograms: Densities of skin and non-skin color histograms are learned from the Compaq dataset [2]. The Compaq skin color dataset has approximately 4,700 skin images and 9,000 non-skin images collected from free web crawling. It has images from all ethnic groups with uncontrolled illumination and background conditions. The number of manually labeled pixels is nearly 1 billion. Skin and non-skin histograms are obtained in RGB color space with 32 bins for each color channel, exactly the same to the settings as in Jones and Rehg [2]. Equal numbers of skin images are randomly selected for training and testing. Similarly, equal numbers of non-skin images are randomly selected for training and testing.

Bayesian Skin Classifier: Naive Bayes is used to build the skin and non-skin classifier. The probability of a color being skin s given a color c, P(s|c), is given by

$$P(s|c) = \frac{P(c|s)P(s)}{P(c)} \qquad (1)$$

where, P(c|s) is the likelihood of a given color c being skin, P(s) is skin color prior and P(c) is color prior. Similarly, the probability of a color being non-skin ns given a color c is given by

$$P(ns|c) = \frac{P(c|ns)P(ns)}{P(c)} \qquad (2)$$

where, P(c|ns) is the likelihood of a given color c being non-skin and P(ns) prior for non-skin color. Further P(c) could be calculated as following

$$P(c) = P(c|s)P(s) + P(c|ns)P(ns) \qquad (3)$$

P(c|s) and P(c|ns) are directly calculated from skin and non-skin histograms. Prior probabilities: P(s) and P(ns) can also estimate from the total number of skin and non-skin samples in the training dataset. However, for skin and non-skin classification, we can simply compare P(s|c) to P(ns|c). Using equations (1) and (2), the ratio of P(s|c) to P(ns|c) can be simplified to

$$\frac{P(s|c)}{P(ns|c)} = \frac{P(c|s)P(s)}{P(c|ns)P(ns)} \qquad (4)$$

Equation (4) can be threshold to produce a skin and non-skin classification rule. Further, P(s) and P(ns) are also constant so this can be simplified as follows

$$\frac{P(c|s)}{P(c|ns} > \Theta \qquad (5)$$

where, Θ is a constant threshold value.

In the experiments, equation (5) is used to find the skin and non-skin probability for pixels. The values of P(c|s) and P(c|ns) are directly looked-up from normalized skin and non-skin histograms respectively.

3.2 Superpixels
A region or a collection of pixels is called a superpixel. A five dimensional vector is used to extract the superpixels, three RGB color channels and two positional coordinates of the pixel, using the quick shift [28] image segmentation algorithm. Superpixels generated from this approach vary in size and shape. Hence the number of superpixels in each image is highly dependent upon the complexity of the image. An image with low color variation will have a less number of superpixels than an image with high color variation, as there is no penalty for boundary violation. Generally, the concept of boundary is not used when extracting the superpixels, however different objects have different texture or color which will implicitly act as boundaries. Figure 1 shows an example of superpixels of an image. In our work we have used "the Superpixel extraction library" [29] for superpixel segmentation.

3.3 Superpixel Classification
First, the pixel based skin color classifier defined on section 3.1 is used to classify the pixels of the images. Then the probability of being skin for a given superpixel *sp* with *N* number of color pixels *c* is defined as follows

$$P(s|sp) = \frac{1}{N}\sum_{i=1}^{N} P(s|c_i) \qquad (6)$$

Similarly, the probability of being non-skin for a given superpixel *sp* with *N* number of color pixels *c* is defined as follows

$$P(ns|sp) = \frac{1}{N}\sum_{i=1}^{N} P(ns|c_i) \qquad (7)$$

3.4 Smoothing with CRF
Skin regions have varying size and shape, depending upon the camera angle, distance from the camera and human body factors. Hence, to obtain smooth skin regions but still preserve the skin and non-skin boundaries, it is necessary to introduce some constraints. Conditional Random Field (CRF) provides a natural way of combining pairwise constraints. Color difference and length of boundary between adjacent superpixels are used as pairwise constraints similar to Fulkerson [13]. Optimum skin and non-skin labeling *L* of all superpixels *S* of an image is defined as follows

$$-\log\big(P(L|S;\omega)\big) = -\sum_{s_i \epsilon S}\Psi(l_i|s_i) + \omega \sum_{(s_i,s_j)\epsilon E}\Phi(c_i,c_j\,|s_i,s_j) \qquad (8)$$

where ω is the weight of pairwise constraint, E is the set of edges of superpixel, and *i* and *j* are index nodes in superpixel level graph of an image.

Color potential ($\Psi(l_i|s_i)$): The color potential Ψ captures the skin and non-skin probability of the superpixel s_i. We have used skin and non-skin probability for superpixel directly from superpixel classification defined in the section 3.3 for color potential Ψ as follows

$$\Psi(l_i|s_i) = \log(P(l_i|s_i)) \qquad (9)$$

Edge and boundary potential ($\Phi(c_i, c_j|s_i, s_j)$): Pairwise edge and boundary potential Φ is defined similar to those of [13]

$$\Phi(c_i, c_j|s_i, s_j) = \left(\frac{L(s_i, s_j)}{1+\|s_i - s_j\|}\right), [c_i \neq c_j] \qquad (10)$$

Where, $L(s_i, s_j)$ is the shared boundary length and $\| s_i - s_j \|$ is Euclidean norm of the color difference between s_i and s_j superpixels.

Only one pairwise potential is used to make the system as simple as possible to show that treating skin color with regions is more effective than with pixels. To improve the effectiveness of our skin color detection technique, we could add more pairwise potentials similar to those in Shotton [30]. This implementation has only one weighting factor ω, which is optimized using cross validation. We use the multi-label graph optimization library of [31], [32] and [33] for the inference of skin and non-skin regions. CRF graph is built on the superpixel level hence CRF optimization is fast.

4. EXPERIMENTS AND RESULTS

Method	True Positive	False Positive
Jones and Regh (2002)	90%	14.2%
Our (Superpixel only)	**91.44%**	**13.73%**
Our (superpixel and CRF)	**91.17%**	**13.12%**

TABLE 1: The Results of Pixel-based and Region-based Techniques.

Equal numbers of training and testing sets are randomly chosen from the Compaq dataset [2] and same training and testing sets are used for all experiments. The Compaq dataset has approximately 4,700 skin and 9,000 non-skin images, freely collected from the web. Basic pixel-based skin color classifier mentioned in section 3.1 achieves similar results to those in Jones and Rehg [2]. We have used RGB bin size = 32 for each channels, and threshold constant Θ = 1. It roughly detects 90% skin color with 14.2% false positive rate.

Superpixel extraction using quick shift is controlled by three parameters: (i) λ the tradeoff between spatial and color consistency, (ii) σ the deviation of density estimator, and (iii) τ maximum distance in the quick shift tree. We have used $\sigma = 2$, $\tau = 6$, and $\lambda = 0:9$ for our experiment. Which are chosen using grid search as there is no explicit mechanism to preserve the skin boundaries; with above selected parameters we have noticed that 97.43% skin pixels are correctly grouped into superpixels with 0.35% false positive rate. Average size of the superpixels increases with the larger value of τ and σ and vice versa. Lower value of λ gives importance to the spatial factor while higher value gives importance to the color value. Average size of superpixels are larger when λ is around 0.5. Skin color detection depends upon the values of the color channels, hence higher importance is given to the color consistency in superpixel extraction. Besides experiments show that the skin boundary is not well preserved with higher spatial importance. The average size of superpixel is 65 in our experiments. However, the size of superpixels is not fixed and fully depends on the complexity of the images.

FIGURE 2: Comparison between pixel-based [2] and region-based skin color classification techniques. Left column shows the original images. Middle column shows the result of pixel-based classification technique and right column shows the result of region-based classification technique with CRF.

Table 1 shows the results for both the presented region-based technique and the current state-of-the-art pixel-based skin color detection [2] on unconstrained illumination and background. The region based technique without CRF has 91.44% true positive rate with 13.73% false positive rate and with CRF has 91.17% of true positive rate with 13.12% false positive rate. Simply grouping the pixel-based evidence onto superpixels increased the true positive rate by 1.44% and decreased the false positive rate by 0.48%. This shows treating skin as a region yields better results than using pixels only. Both results from the region-based techniques are better than the pixel-based technique.

The results on figure 2 show the effectiveness of the region-based technique with CRF over pixel-based technique. Region-based technique first groups the skin and non-skin evidence from each pixel into superpixels level using basic skin color classifier, which helps to remove noise. This is the main reason why only grouping the pixel-based evidence into superpixels increases the true positive rate by 1.44% and reduces the false positive rate by 0.5% (see table 1). Moreover, CRF helps further extract larger smooth skin regions by exploiting neighboring color information and boundary sharing between superpixels.

However, there are also some cases where region-based technique performs worse than pixel-based technique when we apply the CRF. Figure 4 and Figure 3 are such examples. Skin-like looking pixels and high boundary sharing between skin and non-skin regions are the main reason of the failure. However, we also experimented using the color difference constraint only on CRF instead of both color difference and boundary sharing constraints and found that it performs better when skin regions are very small and narrow. But overall CRF with both neighbor color difference and length of boundary sharing constraints performed better. Figure 5 shows an example where CRF with both neighbors color difference and length of boundary sharing performs better than only with neighbors color difference.
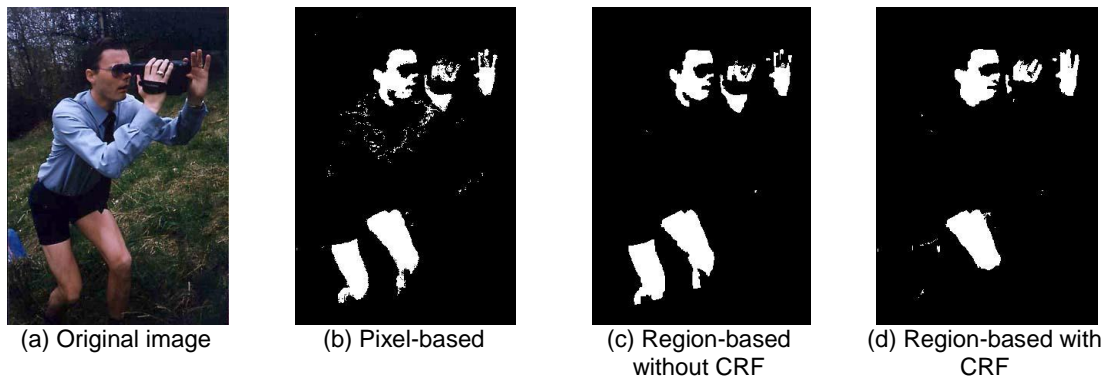
(a) Original image   (b) Pixel-based   (c) Region-based without CRF   (d) Region-based with CRF

FIGURE 3: This example shows the advantages of the region-based approach even without CRF (see sub figures b and c). Sub figures c and d show the failure case when CRF is applied.
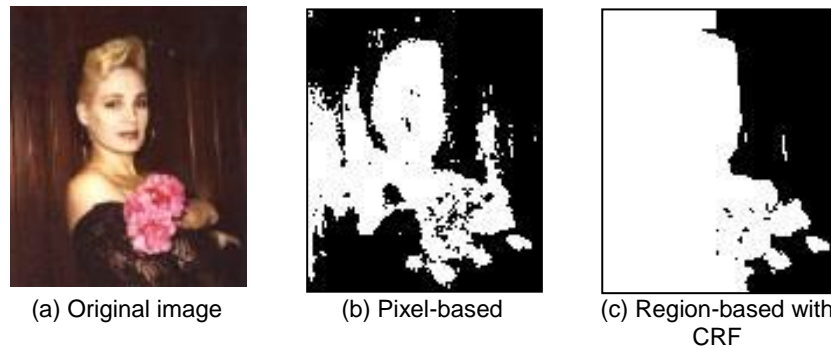

(a) Original image   (b) Pixel-based   (c) Region-based with CRF

FIGURE 4: This example shows the failure of the region-based approach when border information is applied in CRF smoothing.

5. CONCLUSIONS AND FUTURE WORKS

This paper presented a region-based skin color detection technique, which outperforms the current state-of-the-art pixel-based skin color detection technique. Color and spatial distance based clustering technique is used to extract the regions from the images, also known as superpixels. In the first step, our technique uses the state-of-the-art non-parametric pixel-based skin color classifier [2] which we call the basic skin color classifier. The pixel-based skin color evidence is then aggregated to classify the superpixels. Finally, the Conditional Random Field (CRF) is applied to further improve the results. As CRF operates over superpixels, the computational overhead is minimal.

The proposed region-based technique achieved 91.44% true positive rate with 13.73% false positive rate without CRF optimization and 91.17% true positive rate with 13.12% false positive rate with CRF optimization. Grouping the pixel-based evidence into superpixels increased the true positive rate by 1.44% and reduced the false positive rate by 0.48%. Moreover, the region-based approach produced smoother results than the pixel-based techniques. Skin commonly appears as regions of similar pixels, so treating skin as a region is advantageous over treating it as an individual pixel. Due to the illumination, background reflection and other noise factors, pixel values vary greatly and grouping them into a region helps to remove noise by collecting evidence from neighboring pixels.

These results suggest that skin color detection should be region-based rather than pixel-based. Furhter, by adding more constraints on the CRF similar to [30], the detection rate can be

improved. Moreover, any better skin color classification technique can be used as our basic skin color classification module and can be easily combined with our region-based skin color detection framework defined in section 3 to improve the results.

Skin regions do not have the same color values; even the closest skin color pixels within superpixels have different color values. Also, other skin-look-like objects exist. Hence, results can be further improved using texture information. This is left for our future work.
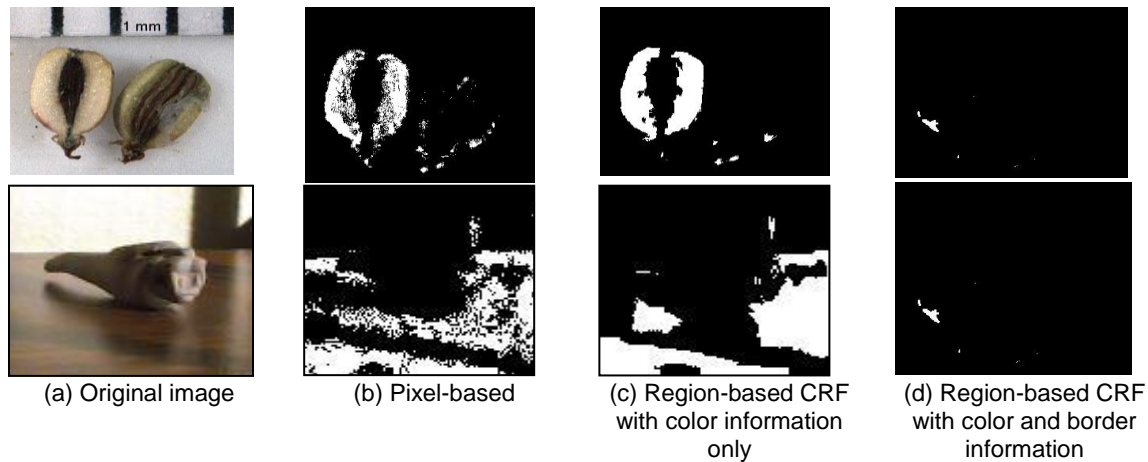

| (a) Original image | (b) Pixel-based | (c) Region-based CRF with color information only | (d) Region-based CRF with color and border information |

FIGURE 5: Example shows the failures of region-based approach when only a color difference constraint is used on CRF optimization.

6. REFERENCES

[1] R. PK Poudel, H. Nait-Charif, J.J. Zhang, D. Liu."Region-Based Skin Color Detection", in VISAPP, 2012.

[2] M.J. Jones,J.M.Rehg."Statistical colo rmodels with application to skin detection". International Journal of Computer Vision 46 (2002) 81–96.

[3] P. Kakumanu, S.Makrogiannis, N.Bourbakis. "A survey of skin-color modelling and detection methods". Pattern Recognition 40 (2007) 1106–1122.

[4] D.A. Socolinsky, A. Selinger,J.D. Neuheisel. "Face recognition with visible and thermal infrared imagery". Computer Vision and Image Understanding 91 (2003) 72–114.

[5] Z. Pan, G. Healey, M.. Prasad, B. Tromberg. "Face recognition in hyperspectral images". IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 1552–1560.

[6] M.H. Yang,N. Ahuja. Detecting human faces in color images". In International Conference on Image Processing,1998. Volume1 (1998) 127–130.

[7] H. Kruppa, M. Bauer, B. Schiele. "Skin patch detection in real-world images". In Van Gool, L.,(Ed.) : Pattern Recognition. Volume 2449 of Lecture Notes in Computer Science. Springer Berlin/Heidelberg (2002)109–116.

[8] B. Jedynak, H. Zheng, M.. Daoudi. "Maximum entropy models for skin detection". In Energy Minimization Methods in Computer Vision and Pattern Recognition (2003) 180–193.

[9] N. Sebe, I. Cohen, T. Huang, T. Gevers. "Skin detection: A Bayesian network approach". In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK (2004) 903–906.

[10] A.P. Moore, S. Prince, J. Warrell, U. Mohammed, G. Jones. "Super pixel lattices". In IEEE Conference on Computer Vision and Pattern Recognition. (2008).

[11] X. Ren, J.Malik. "Learning a classification model for segmentation". In IEEE International Conference on Computer Vision. Volume1 (2003).

[12] S. Soatto."Actionable information invision". In Proceedings of the International Conference on Computer Vision. Volume 25 (2009).

[13] B. Fulkerson, A. Vedaldi, S. Soatto. "Class segmentation and object localization with super pixel neighbourhoods". In Proceedings of International Conference on Computer Vision. Volume 5 (2009).

[14] J.L. Crowley, F. Berard. "Multi-modal tracking of faces for video communications". In Computer Vision and Pattern Recognition, Published by the IEEE Computer Society (1997).

[15] L.M. Bergasa, M. Mazo, A. Gardel, M.A. Sotelo, L. Boquete. "Unsupervised and adaptive gaussian skin-color model". Image and Vision Computing 18 (2000) 987–1003.

[16] D. Brown, I. Craw, J. Lewthwaite. "A som based approach to skin detection with application in real time systems". In Proceedings of the British Machine Vision Conference. Volume 2. (2001) 491–500.

[17] Q. Huynh-Thu, M. Meguro, M. Kaneko. "Skin-color extraction in images with complex background and varying illumination". In Proceedings of IEEE Workshop on Applications of Computer Vision, IEEE (2002) 280–285.

[18] Y. Wang, B. Yuan. "A novel approach for human face detection from color images under complex background". Pattern Recognition 34 (2001) 1983–1992.

[19] Q. Zhu, K.T.Cheng, C.T. Wu, Y.L. Wu. "Adaptiv e-learning of an accurate skin- color model". In Sixth IEEE International Conference on Automatic Face and Gesture Recognition. (2004) 37–42.

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

[21] S. Kawato, J. Ohya." Automatic skin-color distribution extraction for face detection and tracking". In International Conference on Signal Processing. Volume2., IEEE (2002)1415–1418.

[22] R.L. Hsu, M. Abdel-Mottaleb, A.K. Jain. "Face detection in color images". IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 696–706.

[23] K.W. Wong, K.M. Lam, W.C. Siu. "A robust scheme for live detection o fhuman faces in color images". Signal Processing: Image Communication 18 (2003) 103–114.

[24] P. Peer, J. Kovac, F. Solina. "Human skin colour clustering for face detection. In International Conference on Computer Tool. (2003).

[25] M.H. Yang, N. Ahuja. "Gaussian mixture model for human skin color and its application in image and video databases. In Proceedings of SPIE: Storage and Retrieval for Image and Video Databases VII. Volume 3656., Citeseer (1999) 458–466.

[26] J.C. Terrillon, H. Fukamachi, S. Akamatsu, M.N. Shirazi. "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In Fourth IEEE International Conference on Automatic Face and Gesture Recognition (2000) 54.

[27] S.L. Phung, D. Chai, A. Bouzerdoum. "A universal and robust human skin color model using neural networks". In Proceedings of International Joint Conference on Neural Networks Volume 4. (2002) 2844–2849.

[28] A. Vedaldi, S. Soatto. "Quick shift and kernel methods for mode seeking". Proceedings of European Conference on Computer Vision (2008) 705–718.

[29] A. Vedaldi, B. Fulkerson. "VLFeat: An open and portable library of computer vision algorithms". http://www.vlfeat.org (2008).

[30] J. Shotton, J. Winn, C. Rother, A. Criminisi. "Text on boost: Joint appearance, shape and context modelling for multi-class object recognition and segmentation. Proceedings of European Conference on Computer Vision (2006) 1–15.

[31] Y. Boykov, O. Veksler, R. Zabih. "Fast approximate energy minimization via graph cuts". IEEE Transactions on Pattern Analysis and Machine Intelligence (2001) 1222–1239.

[32] Y. Boykov, V. Kolmogorov. "An experimental comparison of min-cut/max-flow algorithms for energy minimization invision". IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 1124–1137.

[33] V. Kolmogorov, R. Zabih, R. What energy functions can be minimized via graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 147–159.