An Adaptive Multiple Model Approach for Fast Content-Based Skin Detection in On-Line Videos

Rehanullah Khan
Vienna University of Technology
Institute of Computer Aided Automation
Pattern Recognition and Image Processing Group
Favoritenstr. 9/1832, A-1040
Vienna, Austria
rehan@prip.tuwien.ac.at

Julian Stöttinger
Vienna University of Technology
Institute of Computer Aided Automation
Pattern Recognition and Image Processing Group
Favoritenstr. 9/1832, A-1040
Vienna, Austria
stoett@prip.tuwien.ac.at

Martin Kampel
Vienna University of Technology
Institute of Computer Aided Automation
Pattern Recognition and Image Processing Group
Favoritenstr. 9/1832, A-1040
Vienna, Austria
kampel@prip.tuwien.ac.at

ABSTRACT

We propose a straightforward skin detection method for online videos. To overcome varying illumination circumstances and a variety of skin colors, we introduce a multiple model approach which can be carried out independently per model. The color models are initiated by skin detection based on face detection and adapted in real time. Our approach outperforms static approaches both in precision and runtime. If we detect a face in a scene, the number of false positives can be diminished significantly. Evaluation is carried out on publicly available on-line videos showing that adaptive multiple model outperforms static methods in classification precision and suppression of false positives.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms

Keywords

Real Time, Face Detection, Skin-Color, Adaptive Modeling.

1. INTRODUCTION

With the international success of "web 2.0" sites, the amount of publicly available content from private sources is growing steadily [13]. The number of videos that are being uploaded every day is way beyond the means of the operating companies that are in charge of classifying the content properly.

The approach most often used for blocking contents on the internet is based on contextual keyword pattern matching technology that categorizes URLs by means of checking contexts of web pages or video names and then traps the websites [13]. This does not hold for websites which allow uploading videos like Google Videos and YouTube, because the videos uploaded have different names from the contents they contain. Due to the fact that there is no automated process, Google and YouTube rely on the user community. Therefore, an automated method to detect and categorize videos based on skin color would help the service providers and can provide control over the videos contents.

We propose an adaptive skin detection method for videos. As a classification problem, we face a two class classification problem. We decide on every pixel if it is a skin-pixel or a non-skin-pixel. We adapt our decision rules upon a prior face detection using the approach from Viola-Jones [21]. Our contribution is twofold: We propose a method which takes advantage of the temporal relationship between frames in an image sequence while taking into account possible time dependent illumination changes. Based on detected faces in the video, we can adapt the classification model and precisely specify possible skin color in the scene. It can be carried out in real-time and is useful for an automated preselection and classification for large video databases. Further, we deal with multiple models at one time. Per detected face, one new decision model is applied and broadens classification. The drawback of using thresholds is its high number of false positive detections [18]. The dynamic multiple model approach reduces false positives detection which is shown in our evaluation. For the task of filtering adult content from videos, we suggest a method for extracting meaningful keyframes from potentially adult content: Based on the results of the skin coverage graph, we are able to extract the key frames for further manual classification.

In a first step we selected 20 videos from YouTube[1], see Figure 1 for samples. The sequences span a wide range of environmental conditions. People of different ethnicity and various skin tones are represented. Sequences also contain scenes with multiple people and/or multiple visible body parts and scenes shot both indoors and outdoors. The lighting varies from natural light to directional stage lighting. Sequences contain shadows and minor occlusions.

In Section 2 we give a survey of the state of the art in skin detection and its adaption towards time-varying color

AREA'08, October 31, 2008, Vancouver, British Columbia, Canada.

[1] `http://www.youtube.com`, last visit: Aug, 1st 2008.

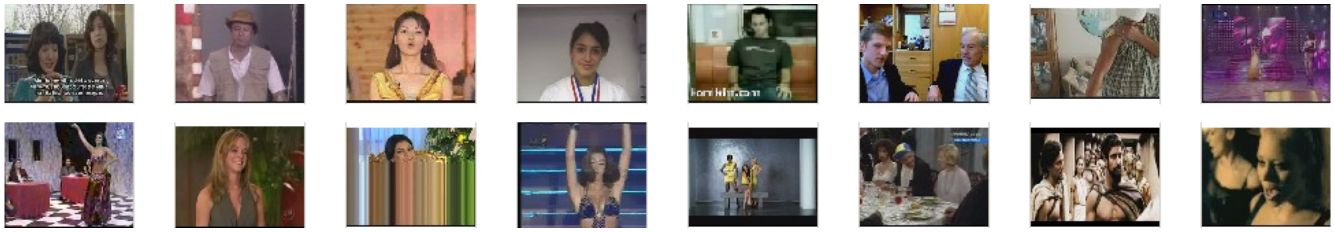


Figure 1: Samples of video data used

circumstances and video segmentation. Section 3 describes the multiple model approach for fast skin detection. The experiments and results are outlined in Section 4. A conclusion is given in Section 5.

2. RELATED WORK

Color is a low level feature, which makes it computationally inexpensive and therefore suitable for real-time object characterization, detection and localization [13]. The goal of skin color detection or classification is to build a decision rule that will discriminate between skin and non-skin pixels. Kakumanu et al. [11] describe the major difficulties in skin-color detection are caused by effects like *Illumination circumstances*, *camera characteristics*, *ethnicity*, and *individual characteristics* like age and sex.

The choice of an appropriate color space for modeling and classifying skin-color is crucial [24]. Yang et al. [24] observed that for color spaces, intensity is more likely to change than chrominance. Additionally they showed that clusters in normalized rgb are an appropriate model for skin-color and approaches that rely on this color space are [16, 2, 3]. Still, the normalized rgb color space suffers from instability in dark colors.

Perceptual color spaces like the $HS*$ color spaces model the RGB cube onto a transformed color space following perceptual features [7]. The Hue component gives the perceptual idea of a color as humans are able to define the hue value as a color tone. The Saturation gives a perceptual measure of the colorfulness. As it is invariant to illumination change, it is regarded as good measurement for skin detection. It is broadly used, e.g. in [2, 6, 7].

To simulate the primate's visual attention, perceptually uniform color spaces like the $CIE-Lab$ or $CIE-Luv$ are used for skin detection e.g. by [4]. Orthogonal color spaces like $YCbCr$, $YCgCr$, YIQ, YUV, YES form as independent components as possible. $YCbCr$ is used by e.g. [23, 17].

According to e.g. [1, 8, 9] a single color space may limit the performance of the skin color filter. Better performance can be achieved using two or more color spaces. Using the most distinct invariant color coordinates of different color spaces increases the detection rate under certain conditions.

Opposed to the application addressed in this paper, [4, 5, 7, 23] and others use the detection of static skin-color for the estimation of faces and further face detection, face recognition and gesture tracking. Viola and Jones [21] introduced a stable face detection algorithm based on illumination which can be applied in real-time. This inspired authors for using this approach as an initial step for further skin color estimation [14, 22]. Contrary to our approach, they use skin-trained classifiers which rely on assumptions and one model at a time. Further, they aim for a precise and reliable skin and face segmentation in high resolution and high quality visual data. Our approach is used for low quality videos and applied in real-time for video categorization and skin detection.

Skin detection under varying illumination in image sequences is addressed in [18, 24, 19]. Static approaches e.g. [15] involve the clustering of skin color with predefined boundaries. The drawback of this method is the false positive detection rate [11]. Neural networks, Bayesian Networks, Gaussian classifiers or Self organizing Maps (SoM) are classifiers to overcome this drawback and achieve better classification accuracy [12, 16, 10, 2]. Inherently, they rely on preceding off-line training phase and/or higher computational complexity.

3. METHODOLOGY

We address the problem of changing lighting conditions in videos by adapting the skin-color model according to reliably detected faces. Prior to any face detected, the static $YCbCr$ skin model [20] is applied for skin detection.

After a successful face detection, a new model is created for the skin being detected based on this updated model. Due to its real-time performance, we use the Viola-Jones [21] face detector for providing parameters for updating: any detected face introduces a new skin-color model, which allows to detect skin of different color and under different lighting conditions even within one single frame (see Figure 2)

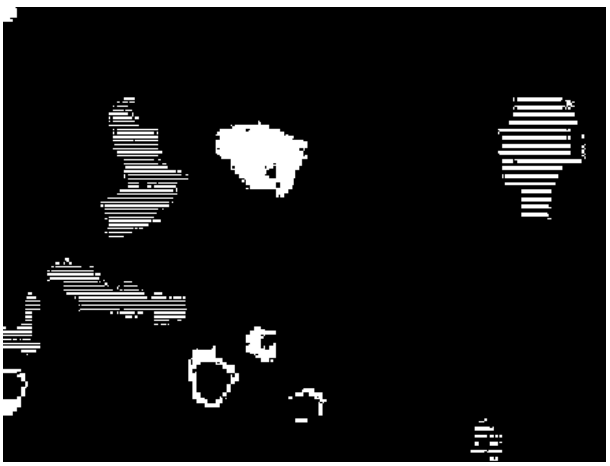
3.1 Viola-Jones Face Detection

The Viola-Jones object detection framework can be used for object detection rates in real-time. The frame work can be trained for a variety of objects, primarily motivated for the problem of face detection. The features used by the detection framework involve the sums of image pixels within rectangular areas [21]. They also bear resemblance to Haar basis functions. Features used by Viola-Jones rely on multiple different rectangular areas per detection.

With the use of the *integral image* as image representation, rectangular features can be evaluated in constant time, which gives them a considerable speed advantage. The integral image can be computed efficiently in a single pass over the image.

The object detection framework employs a variant of the learning algorithm AdaBoost[25] to both select the best features and to train classifiers that use them. AdaBoost is sensitive to noisy data and outliers as it is using fast and simple classifiers and subsequent classifiers are adapted based on the success of previous classifiers. This results in processing gain, and can be embedded in realtime applications.

(a) An example frame with 3 detected faces. (Source: image extracted from a YouTube video)



(b) Shadings of white indicate the applied model. (Source: image extracted from a YouTube video.)

Figure 2: Multiple model skin detection in a frame of an on-line video. (Source: www.youtube.com)

3.2 Color Space for Skin-Color Tracking

The simplicity of its transformation and the explicit separation of luminance and chrominance components makes $YCbCr$ attractive for skin color modeling [20]. The favorable property of this color space for skin-color detection is the stable separation of luminance, chrominance, and its fast conversion from RGB. These points make it suitable for our realtime skin detection. Evaluated extensively [23, 17], the static values used when initially no face is detected are

$$Cr_{max} = 173 \tag{1}$$

$$Cr_{min} = 133$$

$$Cb_{max} = 127$$

$$Cb_{min} = 77$$

where Cr and Cb are the red and blue chroma components in the $YCbCr$ color space. These values cover a broad variety of lighting conditions and skin colors. This results in a large number of false detection, predominantly being background detections. We reduce this number by narrowing down the values based on adaptive skin color modeling.

3.3 Adaptive Skin Color Modeling

Before any successful face detection, the static skin color model is applied to the frames for skin detection. If a face is detected, the face region is extracted for further processing. We discard pixels in the detected regions for the pixel which are most likely non-skin pixel based on the static model. Then we calculate the average over the remaining pixels. The reason for using simple mean calculation is for being fast enough for real time usage. The average color is used to model the new boundary values.

A comprehensive testing was carried out [17] to calculate the dynamic Cr and dynamic Cb values. We use dCr for dynamic Cr and dCb for dynamic Cb. It was found out that dCr has more sensitivity than dCb value, for representing the skin areas. The following boundary gives true positives while minimizing the false positives, and is based on extensive experimentation on the test set described in Section 4. They coincide with the values presented in [23].

$$dCr_{max} = dCr + 9 \tag{2}$$

$$dCr_{min} = dCr - 3$$

$$dCb_{max} = dCb + 10$$

$$dCb_{min} = dCb - 10$$

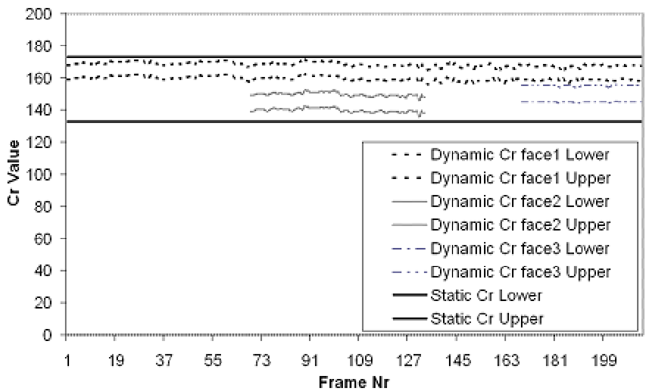
Equation 2 constitute the boundary values for the adaptive skin filter.

3.4 Multiple Models

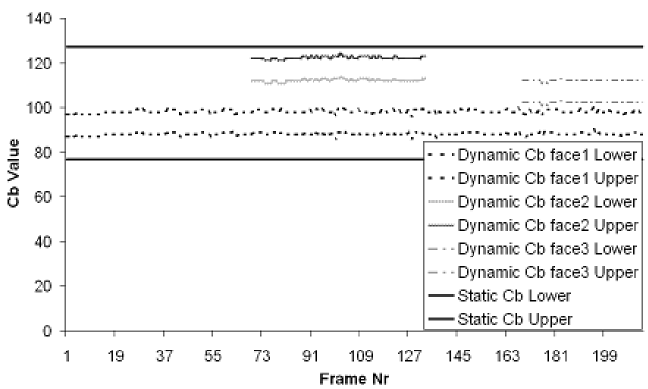
We capture skin information from a reliably detected face for using it to adapt our model parameters for the skin-color detection. At a time when more than one detected face exists in a frame and face information is properly extracted, we use multiple adapted models. Our multiple model approach makes it possible to filter out skin for multiple people with different skin tones and reducing its false positives. In Figure 2(a), three faces are detected which are indicated by white rectangles. Based on this data, three models are estimated and applied (see Figure 2(b)). Black indicates non-skin pixels. For every detected region, the shading shows for which model the most pixels apply to.

Scenes having multiple faces at a particular time contribute multiple models to the detection mechanism. A separate model is used for each separate face as shown in Figure 2. The development of different models can be seen in Figure 3(a) and Figure 3(b): the upper and lower borders for dCr and dCb values of each applied model per frame are shown. Figure 3(a) shows the model parameters for the dCr values and Figure 3(b) for dCb values.

The thresholds/boundaries are determined by contracting and averaging the skin area of face and provide us dynamic paramters i.e. dCr and dCb. These dynamic parameters are scaled up and down using equations 2 for final dynamic boundary values. Note that the adaptive model just covers a small part of the static one. This multiple adaptive model approach is therefore more precise than the static one in-terms of false negatives detection. The models are processed independently, partially resulting in a per skin class segmentation of the image. At any particular time there could be n number of skin models being present in the scene. This can be carried out in parallel per frame. When all faces are lost the last model which existed will be used as a default filter for subsequent frames. This approach overcomes problems of skin-color variance of multiple people with different skin tones.



(a) Model parameters for the dCr values per frame



(b) Model parameters for the dCb values per frame

Figure 3: The temporal development of models over one whole video clip. The thresholds/boundaries are obtained by scaling up and down the dCr and dCb values using equation 2.

3.5 Model Update

Initially when there is no face detected, the static skin filter (equation 1) is applied to the frames for skin detection. Same is true for non-face skin objects. When a face is detected the static skin filter is dropped and dynamic filter based on dCr and dCb boundaries is used for skin detection. For more than one face, a separate dynamic filter/model is created and used. When all the faces are lost the last model is stored and used until new reliable faces found.

4. EXPERIMENTS

In the following, we evaluate our approach and compare it with the static model approach which we are enhancing by adapting it dynamically. All experiments are carried out on a standard desktop computer on an average frame rate of 26,6 fps. 20 videos are manually annotated and the skin detection is evaluated per pixel. Based on these results we evaluate the use of significant changes in the model's appearance and changes in the skin detection for scene change detection. In the following we subjectively categorize online video in 3 classes of skin appearance to evaluate the approach's performance for adult video detection.

4.1 Video categorization

The following experiment gives a proof of concept for the usability of our approach for video portals. The results are solely based on color based skin detection. We divided another 27 testvideos into three categories depending on the amount of skin pixels in the video. Videos which contain no skin, totaling 7 are labeled as NSV (no skin video). 9 videos are labeled as PSV (partial skin video) contain partial nude people and therefore more skin pixels The remaining 11 videos are labeled as LSV(lot of skin video) containing a huge amount of promiscuous skin appearance.

Table 1 shows the results. The algorithm incorrectly catorized two out of 27 videos. Video 12 and Video 24 was reported as LSV. The reason being the skin colored background present in these videos. When there is a sufficient match between the color of skin and the color of non-skin objects, the algorithm incorrectly reports it as skin. In such situation color based skin categorization is misleading and approaches like texture analysis, use of semantics, and object recognition could improve the robustness.

Video#	Video Category	Skin %	Result
1	LSV	18.20	LSV
2	LSV	16.05	LSV
3	LSV	15.90	LSV
4	LSV	30.05	LSV
5	LSV	20.50	LSV
6	LSV	19.76	LSV
7	LSV	21.61	LSV
8	LSV	26.10	LSV
9	LSV	17.03	LSV
10	LSV	20.30	LSV
11	LSV	25.80	LSV
12	PSV	16.59	**LSV**
13	PSV	08.14	PSV
14	PSV	07.65	PSV
15	PSV	06.90	PSV
16	PSV	11.15	PSV
17	PSV	10.87	PSV
18	PSV	09.91	PSV
19	PSV	08.79	PSV
20	PSV	07.69	PSV
21	NSV	01.15	NSV
22	NSV	00.91	NSV
23	NSV	02.10	NSV
24	NSV	19.12	**LSV**
25	NSV	01.17	NSV
26	NSV	02.01	NSV
27	NSV	01.08	NSV

Table 1: Result of Videos Categorization. NSV contain no skin, PSV non promiscuous parts of the body, LSV lot of skin.

4.2 Pixel classification

In order to evaluate our approach we build an annotated ground truth data-set manually marking 4315 frames in 20 videos.

Video	Dyn. Mult. Model		Static Approach	
	Recall	rFP	Recall	rFP
1	0,76	0,93	0,98	5,95
2	0,15	0,49	0,34	1,23
3	0,81	0,58	0,93	1,85
4	0,88	1,75	0,97	3,98
5	0,75	1,19	0,76	1,71
6	0,57	0,22	0,62	0,68
7	0,88	0,42	0,99	1,01
8	0,51	1,60	0,73	3,31
9	0,57	0,11	0,76	0,28
10	*0,93*	*0,86*	*0,93*	*0,86*
11	0,97	4,90	1,00	6,16
12	0,91	4,14	0,57	4,31
13	*0,62*	*0,76*	*0,62*	*0,76*
14	0,96	2,24	1,00	3,46
15	0,67	2,38	0,78	8,00
16	0,96	2,61	0,98	3,94
17	*0,91*	*3,50*	*0,91*	*3,50*
18	0,26	1,93	0,34	1,36
19	0,96	2,22	0,98	5,45
20	0,64	1,66	0,41	2,28
Avg	**0,73**	**1,66**	**0,78**	**3,00**

Table 2: Precision and false positives relative to the amount of marked (skin) pixels. *Italics* lines are videos with no face detection thus providing identical results.

We see that the drawback of static model clustering is false positive detection. Our approach reduces false positives. Table 2 shows the classification results of our experiments per video and are obtained by dividing the corresponding category values by the total number of marked (skin) pixels.

We use the following metrics for pixel-wise evaluation between ground truth and observations in a frame.

$$Recall = \frac{N_dp}{N_{tp}} \tag{3}$$

$$rFP = \frac{N_{dfp}}{N_{tp}}$$

$$Precision = \frac{N_{dp}}{N_{dp} + N_{tn}}$$

$$Fall - out = \frac{N_{dfp}}{N_{tn}}$$

where N_{dp} and N_{dfp} denote the number of detected true positives and false positive pixels respectively. N_{tp} and N_{tn} are the number of total true positives and number of total true negatives. The measure rFP gives a value for the precision of the retrieved result relative to the amount of skin visible in the frame.

In Fig. 4 the weighted harmonic mean of precision and recall (also known as the F-score) is given per annotated video. On average, the dynamic multiple model approach outperforms the static approach. For two videos, the static approach leads to a better F-score. Recall is weighted twice as much as precision in this measurement.

Video	Dyn. Mult. Model		Static Approach	
	Precision	Fall-out	Precision	Fall-out
1	0,45	0,12	0,14	0,80
2	0,23	0,12	0,22	0,29
3	0,58	0,07	0,33	0,24
4	0,33	0,10	0,20	0,23
5	0,39	0,16	0,31	0,23
6	0,72	0,07	0,47	0,20
7	0,68	0,03	0,49	0,08
8	0,24	0,09	0,18	0,18
9	0,83	0,07	0,73	0,18
10	*0,52*	*0,37*	*0,52*	*0,37*
11	0,16	0,34	0,14	0,42
12	0,18	0,18	0,12	0,18
13	*0,45*	*0,55*	*0,45*	*0,55*
14	0,30	0,45	0,22	0,69
15	0,22	0,13	0,09	0,43
16	0,27	0,17	0,20	0,26
17	*0,21*	*0,69*	*0,21*	*0,69*
18	0,37	0,10	0,20	0,32
19	0,33	0,15	0,15	0,42
20	0,22	0,51	0,15	0,53
Avg	**0,38**	**0,22**	**0,28**	**0,36**

Table 3: Precision and Fall-out rate for videos used for testing.

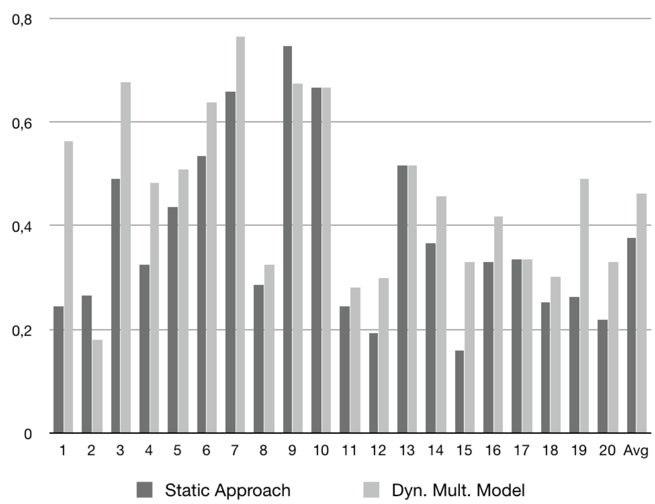


Figure 4: The Weighted harmonic mean of precision and recall is given per video.

4.3 Scene classification

We present the idea and usage of skin graphs. The skin graph shows the peaks related to the skin pixels detected per frame. Figure 5 shows the skin graph for an evaluated video. Per frame the number of detected skin is shown. For peaks in this graph, we extract according frames which are shown in Figure 5 above the graph. This leads to a scene classification which could be used for the final videos classification and categorization done by humans. Presence of strong skin peaks in a video could be marked for manual classification. Probably merged with other keyframe extraction algorithms, these keyframes give a hint towards the nature of a specific video for a final human classification for conspicuous video clips.

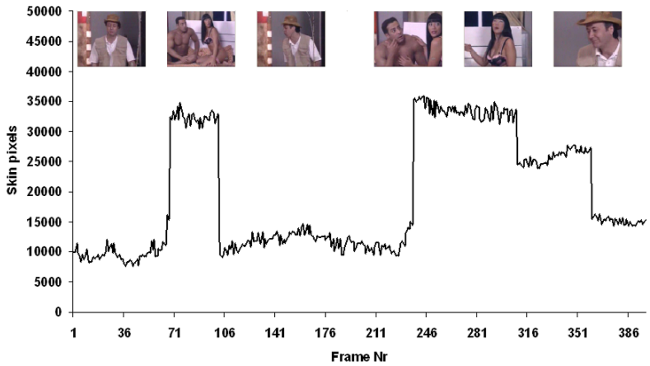


Figure 5: Number of detected skin-color pixels per frame. For large changes in the graph, we extract a keyframe giving a compact representation of the video clip. (Source: www.youtube.com

5. CONCLUSION

The changing environmental conditions and person specific property of skin-color make the precise classification a challenge for static skin color models. We used face detector to dynamically adapt skin color model and use it to significantly reduce the number of false positive detection. Multiple independent models are used. The runtime is realtime and can be carried out in parallel. The extraction of keyframes based on peaks in the skin-color detection give a compact video representation which can be used for final human video classification.

6. ACKNOWLEDGMENT

This work was partly supported by the Austrian Research Promotion Agency (FFG), the national funding institution for applied industrial research in Austria and the CogVis[2] Ltd. However, this paper reflects only the authors' views; the FFG or CogVis Ltd. are not liable for any use that may be made of the information contained herein.

7. REFERENCES

[1] J. Brand and J.S. Mason. A comparative assessment of three approaches to pixel-level human skin-detection. *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 1:1056–1059, 2000.

[2] D. Brown, I. Craw, and J. Lewthwaite. A som based approach to skin detection with application in real time systems. In *BMVC'01: Proceedings of the British Machine Vision Conference*, pages 491–500, 2001.

[2] `http://www.cogvis.at/`

[3] T.S. Caetano and D.A.C. Barone. A probabilistic model for the human skin color. *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*, pages 279–283, 26-28 Sep 2001.

[4] J. Cai and A. Goshtasby. Detecting human faces in color images. *Image and Vision Computing*, 18:63–75, December 1999.

[5] D. Chai and K.N. Ngan. Face segmentation using skin-color map in videophone applications. *Circuits and Systems for Video Technology, IEEE Transactions on*, 9(4):551–564, Jun 1999.

[6] Zhouyu Fu, Jinfeng Yang, Weiming Hu, and Tieniu Tan. Mixture clustering using multidimensional histograms for skin detection. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4*, pages 549–552, Washington, DC, USA, 2004. IEEE Computer Society.

[7] C. Garcia and G. Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *Multimedia, IEEE Transactions on*, 1(3):264–277, Sep 1999.

[8] Moheb R. Girgis, Tarek M. Mahmoud, and Tarek Abd-El-Hafeez. An approach to image extraction and accurate skin detection from web pages. In *Proceedings of World Academy of Science, Engineering and Technology*, pages 367–375, 2007.

[9] G. Gomez, M. Sanchez, and Luis Enrique Sucar. On selecting an appropriate colour space for skin detection. In *MICAI '02: Proceedings of the Second Mexican International Conference on Artificial Intelligence*, pages 69–78, London, UK, 2002. Springer-Verlag.

[10] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.

[11] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recogn.*, 40(3):1106–1122, 2007.

[12] Jae-Young Lee and Yoo Suk-in. An elliptical boundary model for skin color detection. In *International Conference on Imaging Science, Systems and Technology*, pages 579–584, 2002.

[13] Jiann-Shu Lee, Yung-Ming Kuo, Pau-Choo Chung, and E-Liang Chen. Naked image detection based on adaptive and extensible skin color model. *The journal of pattern recognition society, Pattern Recognition*, 40(8):2261–2270, 2007.

[14] Shi lin Wang, Hong Hui1, Sheng hong Li, Hao Zhang, Yong yu Shi, and Wen tao Qu. Exploring content-based and image-based features for nude image detection. In *Fuzzy Systems and Knowledge Discovery*, pages 324–328, 2005.

[15] Member-Son Lam Phung, Sr. Member-Abdesselam Bouzerdoum, and Sr. Member-Douglas Chai. Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(1):148–154, 2005.

[16] Nicu Sebe, Ira Cohen, Thomas S. Huang, and Theo Gevers. Skin detection: A bayesian network approach. In *ICPR '04: Proceedings of the Pattern Recognition,*

17th International Conference on (ICPR'04) Volume 2, pages 903–906, Washington, DC, USA, 2004. IEEE Computer Society.

[17] Andrew Senior, Rein-Lien Hsu, Mohamed Abdel Mottaleb, and Anil K. Jain. Face detection in color images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):696–706, 2002.

[18] L. Sigal, S. Sclaroff, and V. Athitsos. Skin color-based video segmentation under time-varying illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(7):862–877, July 2004.

[19] M. Storring, H.J. Andersen, and E. Granum. Estimation of the illuminant colour from human skin colour. *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 64–69, 2000.

[20] Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreev. A survey on pixel-based skin color detection techniques. In *Graphicon-2003, 13th International Conference on the Computer Graphics and Vision*, pages 85–92, 2003.

[21] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 1:I–511–I–518 vol.1, 2001.

[22] Matthias Wimmer, Bernd Radig, and Michael Beetz. A person and context specific approach for skin color classification. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 39–42, Washington, DC, USA, 2006. IEEE Computer Society.

[23] K.W. Wong, K.M. Lam, and W.C. Siu. A robust scheme for live detection of human faces in color images. *SPIC*, 18(2):103–114, 2003.

[24] Jie Yang, Weier Lu, and Alex Waibel. Skin-color modeling and adaptation. In *ACCV '98: Proceedings of the Third Asian Conference on Computer Vision-Volume II*, pages 687–694, London, UK, 1997. Springer-Verlag.

[25] Yoav Freund, and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *EuroCOLT '95: Proceedings of the Second European Conference on Computational Learning Theory*, pages 23–37, London, UK, 1995. Springer-Verlag.