2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)
November 5-6, 2015
(Iran University of Science and Technology) – Tehran, Iran

SDD: A Skin Detection Dataset for Training and Assessment of Human Skin Classifiers

Mohammad Reza Mahmoodi
Department of Electrical and Computer Engineering
University of California Santa Barbara,
Santa Barbara, CA 93106, USA
mr.mahmoodi@umail.ucsb.edu

Sayed Masoud Sayedi, Fariba Karimi, Zahra Fahimi,Vahid Rezai, Zahra Mannani
Department of Electrical and Computer Engineering,
Isfahan University of Technology, Isfahan, Iran
m_sayedi@cc.iut.ac.ir, fariba.karimi@ec.iut.ac.ir,
z.fahimi@ec.iut.ac.ir, mrvahidrezaie@gmail.com,
zahramannani@gmail.com

***Abstract*— Recently, skin segmentation has been utilized in wide variety of biometric applications including face detection, recognition, tracking, image filtering, archival and retrieval, etc. Along with its applications, different methods have been designed in order to segment pixels of an arbitrary image into skin and non-skin classes. However, there is no reliable, accurate, appropriate and applicatory dataset either to train or evaluate these algorithms. To this end, a comprehensive dataset, SDD, is introduced in this paper which addresses the limitations of former image libraries. SDD contains more than 20,000 color images accompanying with their manually annotated ground truth. It is suitable for assessment of skin classifiers since it is a very extensive database in which images are divided distinctively (very important from evaluation and training point of view) and it covers multifarious photos captured in all around the world in different conditions. In addition, unlike many other datasets with semi-automatic ground truth labeling, GTs in SDD are very precise thanks to the use of a professional graphical tool and more importantly, the idea of ternary division. The proposed database has been compared to SFA through which both qualitatively and quantitatively, the appealing features of the SDD are confirmed.**

Keywords—Human Skin Detection; Skin Database; Skin Classification; Image Processing; Biometrics

I. INTRODUCTION

A glance at actuarial of sleepy people driving cars reported by National Sleep Foundation's 2005 Sleep in America poll call up for serious reactions. 60% of adult drivers (about 168 million people) have admitted that they have driven a vehicle while feeling drowsy in the past year, and more than one-third (37% or 103 million people) have actually fallen asleep at the wheel [1]! As a matter of fact, more than 1 tenth of those who have nodded off, say they have done so at least once a month and four percentages of them avouched that they have had an accident or near accident since they were too tired to drive. Fatigue directly affects the driver's reaction response time and drivers are often too tired to realize the level of inattention. A driving monitoring system shall be a promising solution to this dreadful problem. In such systems, several features including different signs and body gestures in the eyes or head such as repetitious yawning, heavy eyes and slow reaction will be exploited as a drowsiness cue. However, an essential step before facial feature extraction in variety of such systems is detection of human skin to limit the search area for facial detection. Here, as the head pose or rotation may vary frequently, using skin color cue to detect facial features could be a promising choice. In such applications, both the accuracy and speed of the algorithm is determinative. Therefore, it is very important to use methods which are verified from reliability, robustness and speed points of view in any imaging condition or testing situation. To this end, two factors are vital and important for appropriate assessment of skin classifiers.

The first one is defining and use of appropriate criteria to evaluate and compare different methods. A number of standard touchstones have been already developed which are generally, applicable for any classification problem [2-5]. The second factor, that might be more important, is the benchmarks and databases which are employed in both evaluation process and training processes. The aim of present paper is to address limitations of available skin datasets and to introduce a new database which is superior to previous ones. As it will be shown, when the proposed dataset is used for training classifiers, yields lower false rejection and false detection rates in compare with previous ones.

The rest of the paper is organized as follows. In section 2, skin classification and its features and challenges are illustrated accompanying with several major problems with existing databases. In the section 3, current databases employed in skin detection systems in the literature are elucidated in detail. In section 4, the SDD database is presented with graphical explanations. In section 5, the performance of the SDD dataset in compare with previous dataset is investigated both quantitatively and qualitatively. Also, the result of simulating several skin detection methods based on proposed database is presented. Finally, conclusion is provided in section 6.

978-1-4673-6506-2/15/$31.00©2015 IEEE

II. SKIN CLASSIFICATION

Skin classification is the act of discriminating skin and non-skin pixels in an arbitrary image. In fact, skin detection is a two-class grouping problem; pixels are all grouped into two categories. The first one includes pixels which belong to human's body such as a face, hand or any part of human body. The other group contains pixels which are not associated with human peel. Skin detection is a very prevalent algorithm as it has potential of high speed processing [6]. Skin complexion is one of the silent characteristics of human being which is not sensitive to the changes of posture and facial expression. It is invariant against rotation [6,7], geometrics [8], is stable against partial occlusion [6,7], scaling and shape and it is somewhat person independent. Skin detectors are easily implementable and they have low computation cost [9].

Though skin segmentation task has several appealing features as mentioned above, there are some challenges involved in degrading their performance including uneven, inconsistent and nonlinear illumination [7], complex and pseudo skin background [6,10], imaging equipment and camera dependency [11], individual and intra-personal characteristics and other less important but effective factors such as quality of images [12], traditional scarf or wear [13], computation inaccuracies and continuous nature of color transformations, movement of an object and blurring and reflection from water and glass [14]. These difficulties are related to those algorithms which are based on visible spectrum imaging. For infrared systems, expensive systems, and tedious setup procedures are two most common problems.

In order to overcome so called problems as much as possible, wide variety of methods has been developed in last two decades. Looking at different surveys on skin classification, one of the main concluded points is that not only it is impossible to evaluate the performance of detectors themselves precisely, but it is also infeasible to compare them fairly due to the unavailability of a standard and unified dataset [7,11,15,16]. Different methods exploit either different existing datasets or use some self-made small databases. Using self-made databases is not reliable for comparison purposes. Globally used databases also have several problems. One problem is related to the exact definition of a skin pixel which is not determined identically among different datasets and authors. For example, some databases consider human lip as a part of skin while some have ignored it. Thus, if in one algorithm, lip pixels are trained to be skin, then they will probably segment lip pixels as skin ones. However, if in evaluation phase, the same logic is not applied (by those who developed the ground truths of the dataset), the result will be affected.

Moreover, ground truth images in current databases are often too inaccurate (as it will be shown in this paper). In fact, datasets are sometimes annotated with hand but, since the task is both boring and time-consuming, it could be done carelessly. In other databases, GT images are generated using a semi-automatic procedure for reducing the annotation time. However, ground truth images are not accurate enough yet and they could be even worse than those manually marked. The effect is misinterpretation of results. In some databases, skin pixels are marked non-skin, while they are not even near skin regions and in some cases, non-skin pixels are marked skin pixels which could lessen the performance to some extent.

Additionally, the impact of both the number and diversity of training and test sets on the performance of classifiers is clear. However, this factor is often ignored in some works. There are many works report their statistical results without considering the effect of both size and diversity of photo sets. Some datasets are compiled in particular imaging and illumination conditions. This can be problematic in both evaluation and training steps. In evaluation, if dataset is not general enough, the algorithm result either too good or too bad outputs which both are unrealistic. In training, using a non-general database optimizes the system just based on a specific kind of skin pixels causing the performance of the algorithm be seriously affected. Some databases are too small to be used for training of various methods. For instance, considering the Bayesian method developed by Jones et al. [17], in one case the size of LUT (look-up-table) achieves 2×256^3 cells for both skin and non-skin histogram. Appropriate training of this algorithm needs a dataset consisting of several times bigger than entire size of possible cells. In some cases even though databases are relatively huge, most of its images are repeated several times. Also, several datasets are neither available for public usage nor they are free for non-commercial academic usage. Based on above problem, the aim of this paper is to compile a semi-ideal applicatory dataset which can be utilized both in training and evaluation of skin detection methods.

III. PREVIOUS DATABASES

A complete skin detection database is required in both training and evaluation phases. Large number of skin and non-skin pixels need to be used in order to develop an optimum classifier in training phase. In evaluation also, an exhaustive dataset of images are required in conjunction with their manually annotated ground truth images in order to effectively enhance the performance of the system. Thus, it is essential to investigate characteristics of current available datasets to either to utilize them or construct a new reliable, publically available, accurate and applicatory image library. Although, a great number of skin detectors have been proposed, many of them have not been trained or evaluated based on standard datasets. Some of the skin detection datasets are those originally developed for face detection, hand tracking and face recognition problems. The most important group of skin databases are those designed especially for training and assessment of skin classifiers rather than other biometric applications. Compaq [17] is the first large skin dataset and perhaps the most widely used database consisting of 9,731 images containing skin pixels and 8,965 images without any skin pixels. The entire database includes approximately 2 billion skin and non-skin pixels collected by crawling Internet Web pages. Skin regions of 4,675 skin images have been segmented which in conjunction with non-skin images leads to 1 billion pixels. Many skin classifiers have been trained and evaluated based on this database [8,18,19,20]. An automatic software tool is also developed in order to generate ground truths. GT images are annotated using a tunable threshold. With this capability, the annotation process is much faster and easier, but still, the accuracy of GTs is not high. Fig. 1-2 shows a set of images from Compaq dataset. This database is no longer available for public use [21,22]. Compaq is relatively an old database and the quality of its images is too low to be trustable anymore. New generation of imaging devices have been developed and commercially utilized which are far different from previous ones. Another point is that Compaq is not distinctly divided i.e. different authors may utilized different images for their test and evaluation processes and this will affect the performance.

Figure 1. Several image used to be available in Compaq dataset with their corresponding GTs

Figure 2. The low quality of images in Compaq dataset

In contrast with poor quality images in Compaq and its semi-supervised ground truth, ECU skin and face datasets [23] is compiled based on near 4,000 high quality color images and relatively accurate ground truth data for direct benchmarking of skin segmentation algorithms. ECU images ensure the diversity in terms of the background scenes, lighting conditions, and skin types. The lighting conditions include indoor lighting and outdoor lighting; the skin types include whitish, brownish, yellowish, and darkish skins. Several works have been evaluated using this database. Though images are diverse enough to be utilized in assessment of a general skin detection system, the size of the dataset is not yet big enough. Also, handing annotation suffers from a particular problem which is related to the fact that, there are certain regions in all images specifically those with lower quality or crowded images in which annotation is not a straightforward task. For example, the points which are indicated with red arrows in the image in Fig. 3 are pixels around hairs, or other objects often impose additional difficulties. This is in fact the source of error in GTs with hand marking schemes.

Skin dataset provided by Schmugge et al. [24] consists of 845 images. The dataset composed of 4.9 million pixels of skin pixels and 13.7 million pixels without skin pixels. This dataset is very general as it contains images with different facial expressions, illumination levels and camera calibrations. The images with skin pixels were collected from the AR face dataset [25], the UOPB dataset [26], and University of Chile database [27]. Non-skin pixels were selected only from the Chile dataset and University of Washington content-based image retrieval database [28]. In Schmugge database, the accuracy of the ground truths are higher, but, yet this database is too small to be used for training and evaluation steps of most methods.

Figure 3. A set of sample images

MCG-skin database [29] contains 1,000 images randomly sampled from social network websites captured in variable ambient lights, confusing backgrounds, diversity of human races and also various resolutions and visual quality. This dataset contains 38,868,720 skin pixels and 139,091,233 non-skin pixels. Ground truth images in this dataset are not accurately labeled, as eyes, eyebrows, and even bracelets are also considered as skin and pixels around edges are not also marked charily. Ling et al. [30] used MCG-skin dataset and additional web collected images to construct a dataset of 37.5 million skin pixels and 135.58 million non-skin pixels. They used half of this dataset to train their SOM-based (Self Organizing Map) classifier and used the other half for evaluation.

Kawulok et al. compiled HGR [31], the database of images developed mainly for hand gesture recognition. The dataset is organized into three series acquired in different conditions and totally including 1558 images. This database only contains hand images and the backgrounds are not complex at all. In addition, all images are captured from limited number of individuals. UCI Machine Learning Repository [32] consists of skin pixels collected by randomly sampling B,G,R values from images of various age groups, races, and genders derived from FERET [33] and PAL [34] databases. Total learning sample size is 245,057; out of which 50,859 are skin samples and 194,198 are non-skin samples. This database is only applicable for training purposes and only for particular methods.

Castai et al. [35] have also created SFA dataset based on FERET (876 images) [33] and AR (242 images) [25] databases. SFA consists of 3,354 samples of skin pixels and

5,590 non-skin samples ranging differently from 1 to 35×35 dimensions. A comparison between UCI and SFA, based on the best topology and threshold of a neural network based skin classifier, yielded to the fact that SFA is slightly more accurate than UCI in evaluation of skin detectors [35]. However, SFA mainly includes semi passport images which are not suitable for evaluation purposes. Also GT annotation in SFA is inconsistent, i.e. there are many images which are not accurately labeled. In addition, faces are encompassed with 2-3 pixel thick white boundaries in some images. This will lead to an inaccurate construction of skin cluster in different color spaces. In Fig. 4, SFA images with corresponding GTs are depicted; they are representative of all images. This dataset is publically available.

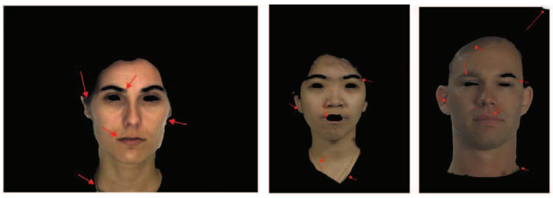

Figure 4. Several GTs in SFA database

Severino et al. [36] used the face databases (FERET and AR) and prepared ground truths manually to corroborate the validity of their skin detection. Db-skin dataset [37] contains 103 skin images annotated by human with relative charily. In some images, eyes and skin and non-skin boundaries are not marked precisely, but they are taken in different lighting conditions and complex backgrounds. In a recent study, Montenegro et al. [38] compiled a dataset in order to compare the performance of 5 common color spaces. This dataset contains 705 RGB images from 47 Mexican subjects with different ages and distinct skin tones. Apart from former databases, images have been all acquired using a single Kinect sensor in completely controlled conditions. This dataset is not available to probe its features.

R. Khan et al. [6] also developed Feeval dataset based on 8991 frames in 25 online videos plus per pixel manually generated ground truths. Three snapshots were selected and shown in Fig. 5. GTs are not annotated precisely, the quality of some video is too low and the diversity of only 25 videos is questionable. TDSD [40] dataset have been also developed for skin segmentation where ground truth labeling for both databases has not been performed with hand as it is observed in Fig. 6. It contains 554 images including 24 million skin pixels and 75 million non-skin pixels. Table. I summarize the characteristics of aforementioned skin datasets.

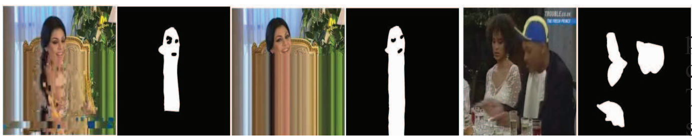

Figure 5. Three snapshots from Feeval dataset with imprecise ground truth and low quality

Figure 6. An image from TDSD and its imprecise GT

TABLE I. COMMON DATASETS IN SKIN SEGMENTATION

Dataset	No. of images	Quality of GT	Quality of Images	Availability	Evaluation	Training
Compaq [17]	13,640	L[a]	L	N	Y	Y
ECU [23]	6,000	M	H	N	Y	Y
MCG [29]	1,000	L	H	N	-	-
Schmugge [24]	845	H	M	N	Y	Y
Ling et al. [30]	1,000	-	-	N	Y	Y
HGR [31]	1,558	M	H	Y	Y	Y
UCI [32]	-	-	-	Y	N	Y
SFA [35]	1,118	M	H	Y	Y	Y
Db-skin [37]	103	M	M	Y	Y	Y
Montenegro [38]	705	-	H	N	Y	Y
Feeval [6]	8,991	M	L	Y	Y	Y
TDSD [40]	554	L	M	Y	Y	Y

a. L, M, H, Y and N stands for low, medium, high, yes and no.

IV. SKIN DETECTION DATASET (SDD)

In previous section, each of the introduced datasets has one or more limitations which make it very hard to train an efficient algorithm and unfair to evaluate a method using them. In order to address so called problems, a new applicatory dataset has been compiled. In developing Skin detection dataset (SDD), limitations of former databases were obviated from different points of view. SDD contain 21,000 images; a large enough database to train all of current methods. To the best knowledge of authors, it is the biggest database introduced in literature to train and apprize the performance of skin classifiers. Images are captured in different illumination conditions, using variety of imaging devices, from diversity of skin tones of people all around the world. Some images are snapshots from online videos and movies, while some are static images acquired from popular face recognition/tracking/detection datasets [40-43]. The high diversity of images in SDD allows its usage on training and evaluation of a general system i.e. when a method is assessed by SDD, the result is reliable, and when a method is trained with SDD, it achieves its maximum potential performance.

Images are divided into 4 sets. First part of the database is particularly considered for training purposes which mainly comprises of single face images in different lighting conditions. This is very important though has been neglected in former studies. The effect of both the number and the content of the training image are influential on the performance of any method. And, due to the fact that the main goal is to compare methods, rather than the datasets, in the proposed dataset, training images are distinguished so that all methods use the same number and the same images. Other sets are to be exclusively used for testing.

GTs have been marked with careful attention to the fact that no pixel is misclassified i.e. all skin pixels annotated as skin are actually skin pixels and non-skin pixels are not. As already mentioned, there are certain pixels in all images in which either there is a question on the skinness of them (such as those regions around eyebrows and eyes, lips and nose holes, etc) or they are located at the boundary of skin and non-skin, and in both situations, it is difficult to discriminate their skinness and in some images, it takes a lot of time to annotated them. Thus, GT images are divided into 3 non-overlapping regions; skin pixels, non-skin pixels and pixels which are considered both in evaluation and training steps (don't care points). It is reasonable that to remove certain number of points in evaluation phase, it does not affect the performance at all since these points will be discarded in all methods. When annotating GTs, the professional CS5 tool has been utilized. Using this graphic tool, it is possible to estimate boundary of the face before coloring so that the result would be much more accurate. All GTs have been marked with hand. Fig. 7 shows the process of ground truth labeling. In Fig. 8, a set of training images are depicted. Black points indicate non-skin regions, red ones are associated with skin pixels and blue pixels show ignored pixels. Using this database allows precise evaluation of performance of skin detection methods. The dataset is available for public use.

V. EXPERIMENTAL RESULTS

In order to evaluate the effectiveness and characteristics of the database, a test has been conducted that measures the performance of two datasets in a common skin detection method. The aim is to measure the quality of training images. In [35], Casati et al. used ANN (Artificial Neural Network) to compare their proposed dataset (SFA) with UCI [32]. However, in training ANN classifier, several other factors including initial weights and the structure to the network are affective and they could lead to misinterpretation. In the contrary, LUT method [17] is free of these limitations and also, it is more powerful to be used for this purpose since it doesn't generalize well and the performance only depends on the training dataset. It is concluded that SFA is more efficient than UCI. Considering the single histogram LUT method, SFA [35] and SDD are utilized for training purposes. The result is investigated both qualitatively and quantitatively.

Figure 7. Ground truth labeling process

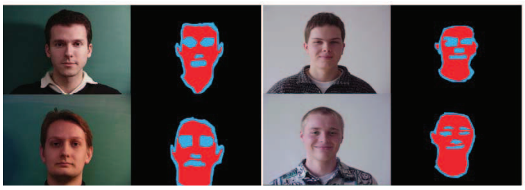

Figure 8. A set of training images in SDD

In Fig. 9, the result of performing single histogram technique is depicted. The first column represents the original images, the second one is the result of applying single histogram trained by SDD and the last column represents the result of using SFA. The threshold which is used in the method is the same for both and the only difference is training dataset. Also, no image including the individuals shown in Fig. 9 is utilized in SDD or SFA for training. As it is observed in the figure, the detection rate of the classifier is much higher when SDD is employed in training and the false detection is also much lower which are both encouraging.

The ROC of histogram method for both SDD and SFA is presented in Fig. 10. As it is observed, utilizing SDD in order to train this method has been yielded to a noticeable improvement in the performance. This is mainly due to the appealing aforementioned features of the image collection which make it favorable over SFA and also other databases.

Figure 9. Skin classification using single histogram based on SDD and SFA

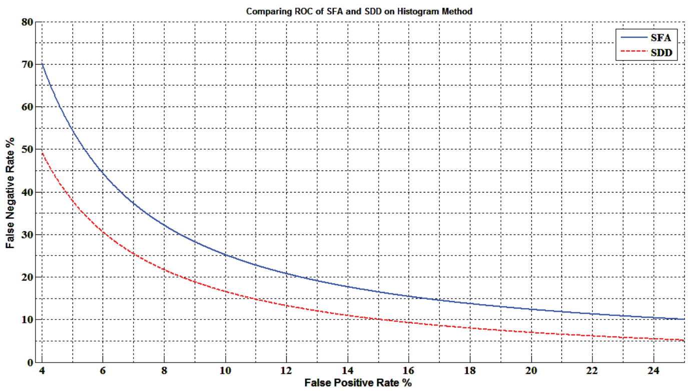

Figure 10. The ROC of single hisotgram method

VI. Conclusion

In this paper, a comprehensive skin dataset is introduced in which addresses the limitations of former databases. It contains more than 20,000 color images which gives a reliable result when it is used for evaluation purposes. The database is well-organized so that when comparing the performance of different methods, the effective factors would be only related to the methods rather than dataset. Furthermore, SDD is suitable for assessment of skin classifiers since it includes multifarious photos captured in different lighting and imaging conditions. Unlike many other datasets with semi-automatic ground truth labeling, in SDD, ground truths are annotated very precise thanks to the usage of a professional graphical tool and more the idea of ternary division. Using the latter, dealing with tedious points in the boundary of skin and non-skin regions would be more convenient and easier. For evaluation of this dataset, single histogram method has been chosen due to its silence characteristics and both qualitatively and quantitatively, it was shown that using SDD is more efficient than SFA.

References

[1] M. R. Mahmoodi, and S. M. Sayedi, “Leveraging Spatial Analysis on Homogeneous regions of Color Images for Skin Classification,” (ICCKE), 4th International Conference on, IEEE, 29-30 Oct, 2014.

[2] M. Di Martino, G. Hernández, M. Fiori, and A. Fernández, "A new framework for optimal classifier design," Pattern Recognition 46, no. 8 (2013): 2249-2255.

[3] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," (2011).

[4] V. Garcıa, J. S. Sánchez, and R. A. Mollineda, "On the suitability of numerical performance measures for class imbalance problems," In International Conference in Pattern Recognition Applications and Methods, pp. 310-313. 2012.

[5] S. Ketenci, and B. Gencturk, "Performance analysis in common color spaces of 2D Gaussian Color Model for skin segmentation," In EUROCON, 2013 IEEE, pp. 1653-1657. IEEE, 2013.

[6] R. Khan, A. Hanbury, J. Stöttinger, and A. Bais. "Color based skin classification," Pattern Recognition Letters 33, no. 2 (2012): 157-163.

[7] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," Pattern recognition 40, no. 3 (2007): 1106-1122.

[8] C. E. Erdem, S. Ulukaya, A. Karaali, and A. T. Erdem. "Combining Haar feature and skin color based classifiers for face detection," In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 1497-1500. IEEE, 2011.

[9] A. Kumar, "An empirical study of selection of the appropriate color space for skin detection: A case of face detection in color images," In Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on, pp. 725-730. IEEE, 2014.

[10] M. Kawulok, J. Kawulok, and J. Nalepa, "Spatial-based skin detection using discriminative skin-presence features," Pattern Recognition Letters (2013).

[11] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," Pattern Analysis and Machine Intelligence, IEEE Transactions on 24, no. 1 (2002): 34-58.

[12] S. Thakur, S. Paul, A. Mondal, S. Das, and A. Abraham, "Face detection using skin tone segmentation," In Information and Communication Technologies (WICT), 2011 World Congress on, pp. 53-60. IEEE, 2011.

[13] A. Sagheer, and S. Aly, "An Effective Face Detection Algorithm Based on Skin Color Information," In Signal Image Technology and Internet Based Systems (SITIS), 2012, pp. 90-96. IEEE, 2012.

[14] A. A. Zaidan, N. N. Ahmad, H. A. Karim, M. Larbani, B. B. Zaidan, and A. Sali, "On the multi-agent learning neural and Bayesian methods in skin detector and pornography classifier: An automated anti-pornography system," Neurocomputing (2013).

[15] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," In Proc. Graphicon, vol. 3, 2003.

[16] E. Hjelmås, and B. K. Low, "Face detection: A survey," Computer vision and image understanding 83, no. 3 (2001): 236-274.

[17] M. J. Jones, and M. R. James, "Statistical color models with application to skin detection," Computer Vision 46, no. 1 (2002): 81-96.

[18] J. Brand, and J. S. Mason, "A comparative assessment of three approaches to pixel-level human skin-detection," In Pattern Recognition, 2000. Proceedings, vol. 1, pp. 1056-1059. IEEE, 2000.

[19] C. Platzer, M. Stuetz, and M. Lindorfer, "Skin sheriff: a machine learning solution for detecting explicit images." In 2nd international workshop on Security and forensics in communication systems, pp. 45-56. ACM, 2014.

[20] S. Mitra, "A probabilistic method of skin detection," Image Information Processing, Second International Conference on , vol., no., pp.73,77, 2013

[21] G. Osman, and M. S. Hitam, "Skin colour classification using linear discriminant analysis and colour mapping co-occurrence matrix," In Computer Applications Technology, 2013.

[22] J. Ruiz-del-Solar and R. Verschae, "SKINDIFF - Robust and fast skin segmentation," Department of Electrical Engineering, Universidad de Chile, 2006.

[23] S. L. Phung, , A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison."Pattern Analysis and Machine Intelligence, IEEE Transactions on 27, no. 1 (2005): 148-154.

[24] S. J. Schmugge, S. Jayaram, M. C. Shin, and L. V. Tsap, "Objective evaluation of approaches of skin detection using ROC analysis," Computer Vision and Image Understanding 108, no. 1 (2007): 41-51.

[25] A. M. Martinez and R. Benavente, “The AR Face Database,” CVC Technical Report#24, June 1998.

[26] E. Marszalec, B. Martinkauppi, M. Soriano, M. Pietikinen, “A physics-based face database for color research,” Journal of Electronic Imaging 9 (1) (2000) 32–38.

[27] Online: http://agami.die.uchile.cl/skindiff/

[28] A. Berman, L.G. Shapiro, “A flexible image database system for content-based retrieval,” Computer Vision and Image Understanding. 75 (1-2) (1999) 175–195.

[29] L. Huang, T. Xia, Y. Zhang, and S. Lin, "Human skin detection in images by MSER analysis," In Image Processing (ICIP), 2011 18th IEEE International Conference on, pp. 1257-1260. IEEE, 2011.

[30] L. Chang, L. Jun-min, and Y. Chong-xiu, "Skin Detection Using a Modified Self-Organizing Mixture Network," In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, pp. 1-6. IEEE, 2013.

[31] Online: http://sun.aei.polsl.pl/~mkawulok/gestures/

[32] B. Rajen, A. Dhall, “Skin Segmentation Dataset,” UCI Machine Learning Repository.

[33] P. J. Phillips et. al. "The FERET evaluation methodology for face-recognition algorithms." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22, no. 10 (2000): 1090-1104.

[34] http://agingmind.utdallas.edu/facedb/

[35] Casati, Joao Paulo Brognoni, Diego Rafael Moraes, and Evandro Luis Linhari Rodrigues. "SFA: A Human Skin Image Database based on FERET and AR Facial Images."

[36] J. Severino, and A. Gonzaga, "A new approach for color image segmentation based on color mixture," Machine vision and applications 24, no. 3 (2013): 607-618.

[37] J. Ruiz-del-Solar, and R. Verschae, "SKINDIFF-Robust and fast skin segmentation," Department of Electrical Engineering, Universidad de Chile(2006).

[38] J. Montenegro, W. Gomez, and P. Sanchez-Orellana, "A comparative study of color spaces in skin-based face segmentation," In Electrical Engineering, Computing Science and Automatic Control (CCE), 2013 10th International Conference on, pp. 313-317. IEEE, 2013.

[39] O. Zhu, C. Wu, K. Cheng, and Y. Wu, "An adaptive skin model and its application to objectionable image filtering," In Proceedings of the 12th annual ACM international conference on Multimedia, pp. 56-63. ACM, 2004.