Automatic Skin Tone Extraction for Visagism Applications

Diana Borza[1], Adrian Darabant[2] and Radu Danescu[1]
[1]*Computer Science Department, Technical University of Cluj-Napoca, 28 Memorandumului Street , 400114, Cluj Napoca, Romania*
[2]*Computer Science Department, Babes Bolyai University, 58-60 Teodor Mihali Street, C333, Cluj Napoca 400591, Romania*

Keywords: Skin Tone, Color Classification, Support Vector Machine, Convolutional Neural Networks.

Abstract: In this paper we propose a skin tone classification system on three skin colors: dark, medium and light. We work on two methods which don't require any camera or color calibration. The first computes color histograms in various color spaces on representative facial sliding patches that are further combined in a large feature vector. The dimensionality of this vector is reduced using Principal Component Analysis a Support Vector Machine determines the skin color of each region. The skin tone is extrapolated using a voting schema. The second method uses Convolutional Neural Networks to automatically extract chromatic features from augmented sets of facial images. Both algorithms were trained and tested on publicly available datasets. The SVM method achieves an accuracy of 86.67%, while the CNN approach obtains an accuracy of 91.29%. The proposed system is developed as an automatic analysis module in an optical *visagism* system where the skin tone is used in an eyewear *virtual try-on* software that allows users to virtually try glasses on their face using a mobile device with a camera. The system proposes only esthetically and functionally fit frames to the user, based on some facial features –skin tone included.

1 INTRODUCTION

In modern society, physical look is an essential aspect and people often resort to several fashion tips to enhance their appearance. Recently, a new concept based on the search of beauty has emerged – *visagism* (Juillard, 2016). Its main goal is to ensure the perfect harmony between one`s personality and appearance, by using some tricks (shape and color of the eyeglasses, hairstyle, makeup etc.) to attenuate or, on the contrary, to highlight some features of the face.

Spectacles sales represent more than 50% of the overall market share. In 2015 the global eyewear market was valued 102.66 billion USD and is continuously expanding (Grand View Research, 2016). An important step in the proposal and selling of eyeglasses is the choice of the frame; this decision must take into account several factors (such as the shape of the face, skin tone, and the eye color etc.), and opticians do not have knowledge in handling these aspects. An automatic framework that accurately classifies these features can assist customers in making the appropriate choice at a much smaller cost than training multiple employees in the field of visagism or using fully qualified estheticians.

Soft biometrics complement the identity information provided by traditional biometric systems using physical and behavioral traits of the individuals (iris and skin color, gender, gait etc.); they are non-obtrusive, don't require human cooperation and can still provide valuable information. Skin tone represents a valuable soft biometric trait.

In this paper, we propose an automatic skin tone classification system mainly intended for the specific use case of eyeglasses selection. Recently, several virtual eyeglasses try-on applications have been developed and their databases typically contain thousands of digitized 3D frames. Physically trying a large number of frames in reality is time challenging and the prospective eyewear buyer often loses interest early in the process. A virtual try-on system assisted by an intelligent module that is able to select the frames esthetically and physically adapted to the user transforms this choice in a playful recreation. Each pair of virtual eyeglasses is given a score for each facial feature trait (gender, skin tone, hair color, shape of the face) and its total score is computed by a

Borza, D., Darabant, A. and Danescu, R.
Automatic Skin Tone Extraction for Visagism Applications.
DOI: 10.5220/0006711104660473
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 4: VISAPP*, pages 466-473
ISBN: 978-989-758-290-5

weighted average of these features and the glasses are displayed to the used in decreasing order of their score.

The proposed system automatically detects the skin tone and it is integrated in this frame selection method. For this particular use-case, the differentiation of skin tones into three classes (dark, medium and light) is sufficient.

Color classification is highly sensitive to capturing devices and illumination conditions; in the case of skin tone classification, the problem is even harder, due to the fact the skin tones are very close and similar, and these perturbing factors have an even higher impact. Moreover, skin color labelling is often found subjective even by trained practitioners (Fitzpatrick, 1998).

We propose and compare two methods for classifying the skin tone. The first method uses the conventional stages of machine learning: region of interest selection, feature extraction and classification. Its main contributions consist in combining and organizing multiple color-spaces into histograms of skin patches and reducing the resulting feature space so that only the features with high discriminative power are kept for voting. A Support Vector Machine (SVM) classifier is trained and used to assign the skin color label to each skin patch.

The second method uses deep learning: the classical stages of machine learning replaced by the convolutional neural network (CNN) which also learns the features which are relevant in the classification problem.

For the training and testing steps, we have gathered and annotated facial images from the Internet and from different publicly available databases.

The remainder of this paper is organized as follows: in Section 2 we describe the state of the art methods used for skin detection and skin color classification and in Section 3 we detail the proposed solution. The experimental results are discussed in Section 4. Section 5 provides the conclusions and directions for future work.

2 STATE OF THE ART

Most of the research conducted on skin color analysis is focused on skin detection (Kakumanu et al., 2007) because of its usefulness in many computer vision tasks such as face detection and tracking (Pujol et al., 2017).

The first attempt to create a taxonomy for skin color was made in 1897 by Felix von Luschan who defined a chromatic scale with 36 categories (von Luschan, 1879). The classification was performed by comparing the subject`s skin with painted glass tiles. This color scale is rather problematic as it is often inconsistent: trained practitioners give different results to the same skin tone. Although it was largely used in early anthropometric studies, nowadays the von Luschan chromatic scale is replaced by novel spectro-photometric methods (Thibodeau et al., 1997). The Fitzpatrick scale (Fitzpatrick, 1988) is a recognized dermatological tool for skin type classification. This classification schema was developed in 1975 and it uses six skin color classes to describe sun-tanning behavior. However, this scale needs training and is subjective.

Color is a prominent feature for image representation and has the important advantage of being invariant to geometrical transformations. However, color classification proves to be a difficult task mainly due to the influence of the illumination conditions: a simple change in the light source, its nature or illumination level can strongly affect the color appearance of the object. Moreover, the classification performance is also limited by the quality of the image capturing devices.

Recently, with the new developments in computer vision several works attempted to classify skin color from images. In (Jmal et al., 2014) the skin tone is roughly classified into two classes: light and dark. The face region is first extracted with a general face detector (Viola and Jones, 2001) and the skin pixels are determined by applying some thresholds on the R, G, B channels. To classify the skin tone, several distances between the test frame and two reference frames are analyzed. This method achieves 87% accuracy on a subset of the Color FERET image database.

In (Boaventura et al., 2006) the skin color is differentiated into three classes: dark, brown and light, using 27 inference rules and fuzzy sets generated from the R, G, B values of each pixel. The method was trained and evaluated on images from the AR dataset and images from the Internet and it achieves a hit rate above 70%. Finer classifications (16 skin tones) are proposed in (Harville et al., 2005) and (Yoon et al., 2006), but these methods involve the use of a color calibration target that contains several predefined colors arranged in a distinctive pattern. The calibration pattern is used for color normalization and skin tone classification.

In the context of racial or ethnical classification from facial images (Fu et al., 2014), some methods use the skin tone as a cue for the race (Xie et al., 2012), especially in the case of degraded facial images, where other features cannot be exploited.

Skin color tones are close to each other and illumination changes make color even harder to distinguish. Although, in the field of dermatology (Fitzpatrick, 1988) and anthropometry (von Luschan, 1897) the skin color is classified at a higher granularity level (using 6 and 36 skin tones, respectively), for the particular applications of visagism and soft biometrics and a simple taxonomy with three classes (light, medium and dark) is sufficient. More complex classification schemes are highly subjective and pose problems even for trained human practitioners. A model with six colors would probably be ideal as the six tones would closely match what we can distinguish visually amongst different regions and human races as predominant skin colors. However, practical studies (Boaventura et al., 2006) show that natural, non-influenced classification of skin colors as performed by humans would contain only three classes: white/light, brown and black. However, even with three classes, the classification is subjective.

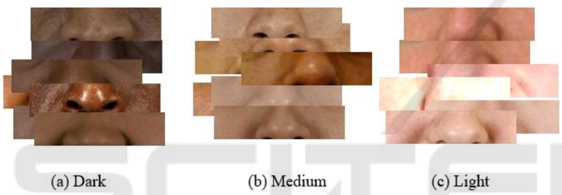

Figure 1: Example of skin patches belonging to the three skin tone classes.

Given these aspects, in the remaining of this paper we only consider the three above classes, based on the idea that when a finer classification is needed it can be derived from these three. Figure 1 shows some skin tone examples belonging to each skin tone class.

3 PROPOSED SOLUTION

This paper presents a fully automatic skin tone classification framework that does not require any prior camera calibration or additional calibration patterns. We propose and compare two methods for classifying the skin tone into three classes: light, medium and dark.

The first method uses support vector machine and histograms of local image patches, while the later one is based on convolutional neural networks.

3.1 Skin Tone Classification using SVM

A moving window is used to compute color histograms of local image patches in several color spaces on a skin region from the face. The histograms are concatenated into a single feature vector and Principal Component Analysis (PCA) is used to reduce its dimensionality of the feature vector.

The reduced histogram from each facial skin patch is fed to a Support Vector Machine (SVM) to determine the skin tone of that region, and, finally the skin color is determined using voting. The outline of the proposed solution is depicted in Figure 2.

The choice of the color space is critical in color classification. Each color-space represents color features in different ways, so that colors are more intuitively distinguished or certain computations are more suitable. However, none of the color-spaces can be considered as a universal solution. In this work, we classify the skin tone by combining the histograms of the most commonly used color-spaces: RGB, HSV, Lab and YCrCb.

As illustrated on Figure 3, using the three simple skin tone classes we proposed, the problem is not trivial as the samples cannot be clearly separated in different color spaces. Figure 3 (a), (b) and (c) illustrates the color distribution of the three selected skin tones in RGB, HSV and YCrCb color spaces. Figure 3 (d) plots the grayscale values of the three classes in order to determine if the intensity feature could bring any additional information to the classification problem. Finally, in Figure 3 (e) the 3D color points from the RGB color-space are projected onto the 2D space using PCA by preserving only the two axes with the most data variation.

The first step of the algorithm is face detection: we use the popular Viola-Jones (Viola and Jones, 2001) algorithm for face localization. However, the face region contains several non-skin pixels, such as eyes, lips, hair etc. that could influence the classification performance. In order to avoid these non-relevant features, we crop the face to a region of interest (ROI) right beneath the eyes and above the center of the face and only the pixels in this area will be analyzed to classify the skin tone.

The crop proportions were heuristically determined as $[0.2 \times w, 0.3 \times h, 0.6 \times w, 0.5 \times h]$, where w and h are the width and the height of the face region.

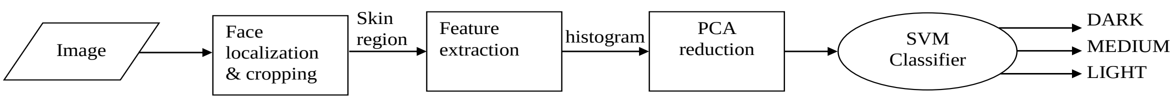

Figure 2: Flowchart of the SVM based-classification method.

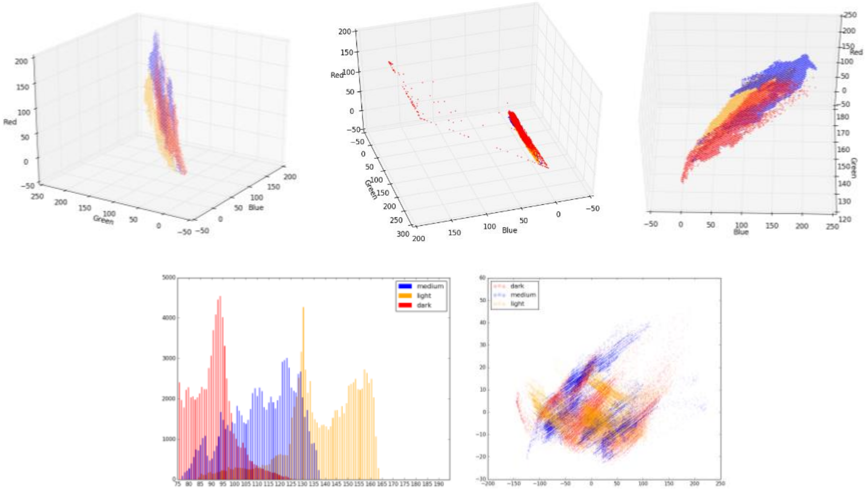

Figure 3: Color distribution for each skin tone. Red points represent skin pixels of dark skin tone, blue points represent skin pixels of medium skin tone and orange points represent pixels of light skin tone.

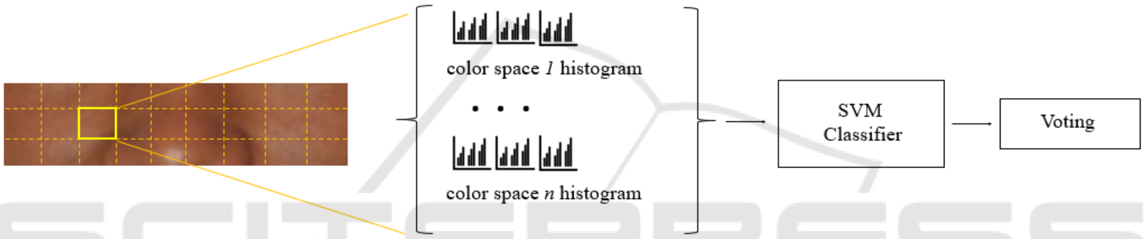

Figure 4: Feature extraction.

Next, a moving window of size *s* x *s* ($s = 21\times$ pixels in our experiments) iterates over the ROI and the histograms for each color component of the 4 selected color spaces are computed and concatenated. The feature vector (if all the 4 color spaces are used) is composed of 13 histograms (3 color-spaces and grayscale) and has a high dimensionality (i.e. $12 \times 256 = 3328$ bins). In such cases the training problem can be difficult and the high dimensional input space can "confuse" the learning algorithm and lead to over-fitting. Therefore, we apply a dimensionality reduction pre-processing step (Principal Component Analysis) in order to increase the robustness towards illumination conditions and to reduce the time complexity of the classifier. Using PCA, the original high dimensional input space is reduced (with some data loss, but retaining as much variance as possible) to a lower dimensional space determined by the highest eigenvectors. We retain the first principal components such that $p = 98\%$ of the variance of the data is preserved; the number of retained principal components is computed on the training set and once computed it is fixed. This value was determined heuristically through trial and error experiments and is fixed once determined.

Figure 4 illustrates the feature extraction process.

The reduced feature vector resulted from each position of the sliding window is fed to a Support Vector Machine classifier in order to obtain the skin tone of the region. A pre-processing step is applied, by scaling the input such that each feature from the training set has zero mean and unit variance.

SVM classifiers are supervised learning algorithms originally developed for binary linear classification problems. To adapt the classifier to our 3 class classification problem, we used "one versus one" approach: $n\times(n - 1)/2$ classifiers are constructed (each SVM must learn to distinguish between two classes) and at prediction time a voting scheme is applied.

We use the Radial basis function (RBF) kernel or Gaussian kernel:

$$K(x,x') = exp\left(\frac{\|x-x\prime\|^2}{2\sigma^2}\right) \quad (1)$$

Finally, the skin tone classification is the dominant skin tone within the selected ROI computed by majority voting. For each sliding window position, the classifier gives a skin label and the final skin tone is selected as the class that occurs most often.

3.2 Skin Tone Classification using CNN

Recently, in the field of computer vision CNN have achieved impressive results in a variety of classification tasks. The CNN structure is based on mechanisms found in the visual cortex of the biological brain. The neurons in a CNN are arranged into 3 dimensions: (width, height and depth) and the neurons within a layer are only connected to a small region of the previous layer, the receptive field of the neuron. Typically, CNNs have three types of basic layers: the *convolutional layer* (*CONV*) – followed by a non-linearity, e.g. Rectified Linear Unit (ReLU) –, the *pooling layer* and the *fully connected* (*FC*) *layer*.

The VGG (Simonyan and Zisserman, 2014) network modified the traditional CNN architecture by adding more convolutional layers (19 layers for the VGG-19 network) and reducing the size of the convolutional layers (to 3x3 convolutional filters with 1 stride in all layers). The structure of the VGG-19 network is the following: *2×conv3_64*, *maxpool*, *2×conv3_128*, *maxpool*, *4×conv3_256*, *maxpool*, *4×conv3_512*, *maxpool*, *4×conv3_512*, *maxpool*, *2× FC_4096*, *FC_1000* and softmax; where conv*s_d* is a convolutional layer of size *s* and depth *d*, *maxpool* is a max-pooling layer and FC_*n* is a fully connected layer with *n* neurons.

We finely tuned a VGG-19 network which was previously trained for gender recognition from facial images (Rothe et al., 2016). The network was trained from scratch using more than 200.000 images and uses RGB, 224×224 images as input.

The only constraint we apply on a new image is that it contains a face which can be detected by a general face detector (Viola and Jones, 2001). The face region is further enlarged with 40% horizontally and vertically and this region is fed to the convolutional neural network. The training is performed using the *rmsprop* optimizer; the batch size was set to 128 and the *rmsprop* momentum to 0.9; the learning rate is initially set to 10^{-2} and then exponentially decreased after each to epochs.

To tune the network for the skin tone classification problem, we removed the last two fully-connected layers of the trained CNN and the remaining part of the convolutional neural network is treated as a fixed feature extractor. Finally, a linear classifier (softmax) is trained for the skin tone classification problem using the features previously learned by the CNN.

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Training and Tests Datasets

To create the training dataset, we have fused together images from multiple face databases ((Caltech, 1999), (Ma et al., 2015), (Minear and Park, 2004), (Thomaz and Giraldi, 2010)) and labelled the images according to the skin color of the subject. The Caltech (Caltech, 1999) face database contains 450 outdoor and indoor images captured in uncontrolled lighting conditions. The Chicago Face Database (Ma et al., 2015) comprises more than 2000 images acquired in controlled environments. The Minear-Park database (Minear and Park, 2004) contains 576 facial images captured in natural lightning conditions and the Brazilian face database (Thomaz and Giraldi, 2010) is composed of 2800 frontal facial images captured against a homogenous background.

In addition, the training dataset was enlarged with images of celebrities captured in unconstrained conditions. For this, we extracted the names of celebrities with different skin tones (www.listal.com) and we crawled Internet face images of the selected celebrities.

To determine the actual skin tone, each image sample was independently annotated by three different persons and the ground truth was determined by merging the results from the independent annotations (majority voting). First, the annotators had a training with a visagist expert who instructed them with the rules two follow in the annotation process. During the training, they also annotated together with the visagist a subset of the images and discussed how to handle boundary cases and uncertainties. After merging the human labeling results, we observed that most of the annotation inconsistencies appeared between the medium and the dark skin tones.

We have used several augmentation techniques: contrast stretching and brightness enhancement. As some of the datasets are captured in controlled scenarios ((Ma et al., 2015), (Minear and Park, 2004)), these augmentation techniques will make learning algorithm more robust to illumination conditions.

We have extracted subsets of each of the four databases such that the distribution of the skin tone classes is approximately even. The final training dataset consists of 8952 images. The classifier was evaluated on 999 images that were not used in the training process; the test dataset is balanced: we use 333 images for each class.

4.2 SVM Classification Results

We trained the SVM classifier using different color features: first we have used the histograms of the four selected color spaces (RGB, HSV, Lab and YCrCb) independently, next we have also added the grayscale histogram (when not redundant) to each color space, and finally we have combined multiple color-spaces. Table 1 shows the classification accuracies for the different color features used.

Table 1: Classification performance using different color spaces.

	Color channels	Accuracy
1	R, G, B	74.65%
2	H, S, V	83.30%
3	L, a, b	72.46%
4	Y, Cr, Cb	83.30%
5	L, a, b, H, S, V	85.18%
6	L, a, b, Grayscale	84.19%
7	L, a, b, H, S, R, G, B, Y, Cr, Cb	**86.67%**

Table 2: Confusion matrix for the VGG-19 convolutional neural network.

		Predicted Value		
		Dark	**Light**	**Medium**
Actual value	**Dark**	271	8	37
	Light	3	302	27
	Medium	19	14	299

The best results are obtained by combining multiple color-spaces (L, a, b, H, S, R, G, B, Y, Cr, Cb): 86.67% accuracy. The confusion matrix for this experimental setup is reported in Table 2.

4.3 CNN classification results

The classification report using the convolutional neural network is reported in Table 3.

Table 3: Classification report for the VGG-19 convolutional neural network.

Class	Precision	Recall	F1-Score
Dark	0.9410	0.9099	0.9252
Light	0.9868	0.8979	0.9403
Medium	0.8289	0.9309	0.8769

The overall accuracy obtained using the VGG-19 network is 91.29%. By using a convolutional neural network to classify the skin tone we obtained a performance boost by 4.62%.

Table 4 shows the confusion matrix for the CNN classification. From the confusion matrix, it can be noticed that the majority of “confusions” occurred between medium-light skin tones and dark-medium skin tones. This behavior is very similar to what we observed in the annotations of the ground truth by the three human labelers. As opposed to the SVM skin classification approach, it can be noticed that there no confusions between the Dark-Light and Light-Dark classes.

Table 4: Confusion matrix for the VGG-19 convolutional neural network.

		Predicted Value		
		Dark	**Light**	**Medium**
Actual value	**Dark**	303	0	30
	Light	0	299	34
	Medium	19	4	310

4.4 Comparison to State of the Art

Next, in this section we compare the proposed solution with other state of the art works that tackle the problem of skin tone classification.

We obtain the best using the VGG-19 convolutional neural network (accuracy 91.29%). Not all the proposed methods classify the skin tones at the same granularity level. In (Jmal et al., 2014) the obtained classification accuracy is 87% but the skin tone is distinguished into only dark and light. In (Boaventura et al., 2006) the skin color is divided into the same 3 classes used in the proposed solution: light, medium and black, but the obtained accuracy is only 70%. However, the method (Boaventura et al., 2006) was tested on images captured in the authors' laboratory and on a subset of the AR dataset, so the testing benchmark is not publically available. Finally, in (Harville et al., 2005) the authors make use of a color calibration pattern that must be held by the subjects in each test image. The color calibration pattern arranges the primary and secondary colors, and 16 patches representative of the range of human skin color into a known pattern. The main disadvantage of that method is that is assumes a controlled image capturing scenario (the user must hold the color chart).

Our method does not impose any restrictions of the image capturing scenario and attains an accuracy rate of 91.29%.

Some examples of correctly classified samples are depicted in Figure 5.

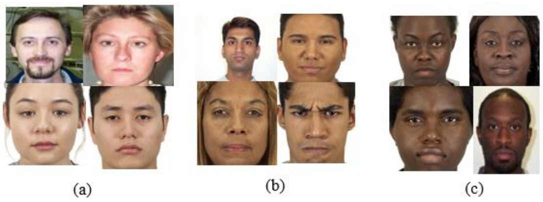

Figure 5: Examples of correctly classified images: (a) dark as dark, (b) medium as medium, (c) dark as dark.

Figure 6 shows some examples of incorrectly classified images.

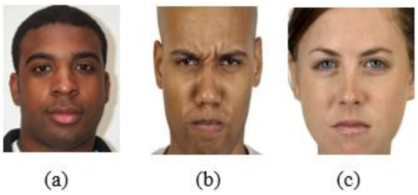

Figure 6: Examples of incorrectly classified images: (a) dark as medium, (b) medium as dark, (c) light as medium.

4.5 Applications

The proposed approach is intended for a facial attribute analysis system used in virtual eyeglasses try-on (Figure 7).

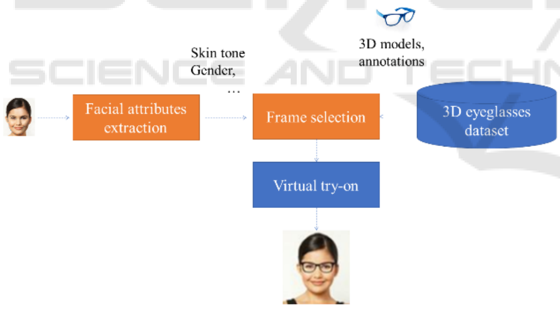

Figure 7: Outline of the eyewear-proposal system.

First, a facial image of the subject is captured and the system (*Facial attributes extraction* module) automatically determines the skin tone (and other demographical attributes: gender, age, eye color etc.). Based on these attributes, the *Frame Selection* module queries the 3D eyeglasses database and selects the accessories that are esthetically and functionally in harmony with the user`s face. Each 3D eyeglasses pair was previously annotated by a specialized visagist/esthetician with a score for each facial attribute and only the eyeglasses with the highest scores are displayed to the user. Typically, the 3D eyeglasses dataset contains more than several thousand eyeglasses models.

More specifically, the system is implemented in Objective-C for an iPad application: first, a picture of the user is analyzed in order to determine the skin tone (and other demographical attributes) and the most appropriate pairs of glasses are selected from the database. Next, the system starts to track the user`s face and uses augmented reality to place the selected eyeglasses on the subject`s face. The system was also tested on images captured in this scenario.

Of course, other applications can be envisioned: the data extracted by the *Facial attributes extraction* module can be used, for instance, to suggest the appropriate make-up or hair color, to estimate skin tone distribution over populations over geographical areas etc.

5 CONCLUSIONS

This paper presented an automatic skin tone classification system that doesn’t require any additional color patterns or prior camera color calibration. We proposed and compared two methods for classifying the skin tone in facial input images.

The first method uses conventional machine learning techniques: histogram of various skin patches from the face and a SVM to determine the skin tone. First, the face is localized in the input image and it is cropped to a region that is most likely to contain only skin pixels. Next, a window slides over this region and color histograms in different color spaces are computed and concatenated for each window position. The feature vector is reduced using PCA and a SVM classifier determines the skin tone of the window. The skin tone is determined using a simple voting procedure on the result of each histogram patch from the region of interest.

The later method uses convolutional neural networks, which also learn the relevant chromatic features for the skin tone classification problem. We finely tuned a neural network which was previously trained on the problem of gender detection from facial images. The system was trained and tested on images from four publicly available datasets and from images crawled from the Internet.

As a future work, a more complex method to determine the skin pixels within the face area is envisioned. In addition, we plan to integrate the current method with a full visagism analysis system that also determines the eyes color, the hair color and the face shape.

ACKNOWLEDGEMENTS

This work was supported by the MULTIFACE grant (Multifocal System for Real Time Tracking of Dynamic Facial and Body Features) of the Romanian National Authority for Scientific Research, CNDI–UEFISCDI, Project code: PN-II-RU-TE-2014-4-1746.

REFERENCES

Boaventura I. A. G., Volpe V. M., da Silva I. N., Gonzaga A., 2006. Fuzzy classification of human skin color in color images. In *IEEE International Conference on Systems, Man and Cybernetics*, 5071– 5075.

Caltech University, 1999. Caltech Face Database - http://www.vision. caltech. edu. *Image Datasets/faces*.

Fitzpatrick, T. B., 1988. The validity and practicality of sun-reactive skin types I through VI. *Archives of dermatology*, *124*(6), pp.869-871.

Fu, S., He, H. and Hou, Z. G., 2014. Learning race from face: A survey. IEEE transactions on pattern analysis and machine intelligence, 36(12), pp. 2483-2509.

Grand View Research, 2016. Eyewear Market Analysis by Product (Contact Lenses, Spectacles, Plano Sunglasses) And Segment Forecasts To 2024, Available at: *http://www.grandviewresearch.com/industry-analysis/ eyewear-industry*, [Accessed: 13.04.2017]

Harville, M., Baker, H., Bhatti, N. and Susstrunk, S., 2005, September. Consistent image-based measurement and classification of skin color. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on* (Vol. 2, pp. II-374). IEEE.

Jmal M., Souidene W., Attia R., Youssef A., 2014. Classification of human skin color and its application to face recognition. *In: The Sixth International Conferences on Advances in Multimedia MMEDIA 2014*.

Juillard C., 2016. Brochure Méthode C. Juillard 2016, Available at*: http://www.visagisme.net/Brochure-Methode-C-JUILLARD-2016.html*, [Accessed: 30.07. 2017]

Kakumanu, P., Makrogiannis, S. and Bourbakis, N., 2007. A survey of skin-color modeling and detection methods. *Pattern recognition*, *40*(3), pp.1106-1122.

von Luschan B., 1897. *Beitrage zur Volkerkunde der deutschen Schutzgebiete*. D. Reimer.

Ma, D. S., Correll, J. and Wittenbrink, B., 2015. The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, *47*(4), pp.1122-1135.

Minear, M. and Park, D. C., 2004. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), pp.630-633.

Pujol, F. A., Pujol, M., Jimeno-Morenilla, A. and Pujol, M. J., 2017. Face detection based on skin color segmentation using fuzzy entropy. *Entropy*, *19*(1), p.26.

Rothe R., Timofte R., van Gool L.. 2016, "Deep expectation of real and apparent age from a single image without facial landmarks." International Journal of Computer Vision (2016): 1-14.

Simonyan, K., Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Thibodeau, E. A. and D'Ambrosio, J. A., 1997. Measurement of lip and skin pigmentation using reflectance spectrophotometry. *European journal of oral sciences*, *105*(4), pp.373-375.

Thomaz, C. E. and Giraldi, G. A., 2010. A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, *28*(6), pp.902-913.

Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.

Yoon, S., Harville, M., Baker, H. and Bhatii, N., 2006, October. Automatic skin pixel selection and skin color classification. In *Image Processing, 2006 IEEE International Conference on* (pp. 941-944). IEEE.

Xie, Y., Luu, K. and Savvides, M., 2012. A robust approach to facial ethnicity classification on large scale face databases. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on* (pp. 143-149).