Influence of the amplitude of the Receptive Field on Skin Detection and Segmentation

Giovanni Ciampi[1], Antonio Corsuto[1], and Francesco Mogavero[1]

[1] University of Salerno, via Giovanni Paolo II, 132, 84084 Fisciano SA, Italy
g.ciampi5@studenti.unisa.it
[2] University of Salerno, via Giovanni Paolo II, 132, 84084 Fisciano SA, Italy
a.corsuto@studenti.unisa.it
[3] University of Salerno, via Giovanni Paolo II, 132, 84084 Fisciano SA, Italy
f.mogavero5@studenti.unisa.it

Abstract. Skin detection is the starting point for skin segmentation, that can be useful in a variety of contexts, such as face recognition, hand gesture analysis and naked people detection. Many papers on skin detection focus on algorithms that work on a single pixel: in this paper we investigate how the number of contiguous pixels analyzed at once can impact the performances of skin detection and segmentation.

Keywords: Skin Segmentation · Skin Detection · Skin Recognition · Machine Learning · Deep Learning · Object Detection · Color Spaces.

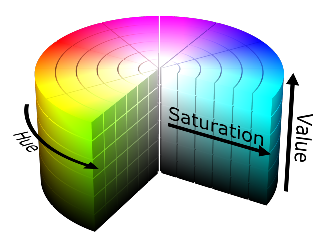
1 Introduction

1.1 The Skin Detection and Segmentation Problems

Skin detection and segmentation are two closely related problem, that concern the individuation of (human) skin on images. Although they are closely related, they are not exactly the same, so, we would like to start this paper by clearly defining these problems, as in [5]. Skin detection is the problem of determining whether a given pixel region presents skin or not, while skin segmentation is the problem of determining the boundaries of skin regions. Such problems are fundamental for a variety of tasks, ranging from face recognition to video surveillance. These problems are a special case of a more general problem, that is the problem of object recognition. Compared to the general case, skin recognition (and by this we mean both skin detection and segmentation) is a much harder task: indeed, skin can assume any shape and a huge variety of colors, making it almost impossible to find a discriminating feature. These intrinsic characteristics of the problem make most of the solutions not very reliable. Even after pointing out the reasons for the difficulty of such problems, it is still reason of great wonder why we are not able to design an effective algorithm to perform what is trivial for almost any kid.

Problem Setting: Features and Color Spaces In order to determine if a certain pixel or group of pixels contains skin or not, one has to identify a characteristic on which to discriminate. As we previously said, this is the key task, and it is indeed the most difficult one. The most natural feature that can be used for this aim is certainly the color, and it also has proved to be the most effective one. In addition to the color, some papers, such as [2], tried to use some statistic properties of neighbor pixels in order to achieve better results: this approach, in general, has not showed significant benefits, although it requires an additional amount of computing power. Although it has been proved in [3] that under certain conditions, all color spaces are equivalent for skin detection, almost every paper published on this matter discusses the benefits of one over another. The reason may be that, since the features of the images are very often used to train machine learning and deep learning algorithms, some color spaces may make it easier for them to converge. Most of the studies that compare the performances of skin detection and segmentation on the various color spaces seem to agree on the choice of YCbCr as the best one. However, even this community is divided in two groups: who claims that the Y (luminance) component has to be kept, and who, like [1] and [7], claims that the best performances are achieved when the Y component has been removed (or at least that Y does not provide useful information). This debate on the Y component may be (partly) due to the varying characteristics of the images: in fact, features such as image quality or lightning condition can make the problem setting completely different, and make one feature more valuable than another. Although the majority claims that YCbCr is the best color space for skin detection and segmentation, it is becoming increasingly popular the use of Hue-based color spaces, such as HSV. Hue-based color spaces are an alternative representation of the RGB color space, and it is said that they are more intuitive to humans since they represent colors in a way similar to how we perceive them.

Fig. 1. The HSV Color Space.



Problem Setting: Human Skin Detection vs Algorithmic Skin Detection We opened this paper by claiming that it is reason of great wonder the fact that we are not able to design an (effective) algorithm to perform what is trivial for kids. It is known (and clear) that the human mind does a huge preprocessing on images before we can consciously realize them, but this may not be the (only) cause of the difference between performances of algorithms and humans on skin detection. There is a profound difference, in general, between how the problem is set for humans and computer programs: the size of the receptive field. Most of the presented algorithms, in fact, work on a very small number of pixels, mostly one at a time. This is completely different to the way humans (that are the best at this task) approach the problem: in fact humans are good only when they can see a conspicuous portion of pixels at once. What if we ask a human being to tell if a single pixel contains skin or not? The performances of the human skin detector will certainly drop. But another question arises: what if we ask a computer program if an increasingly bigger group of pixels contains skin or not?

2 Evaluation

Features chosen The aim of our solution is to be as general as possible. For this reason, we chose to use the (full) YCbCr color space, along with the Hue component of the HSV color space and the Red component of the normalized RGB color space. Although the YCbCr color space has proven to be very effective for skin detection, we chose to add the Hue because intuitively, skin colors are located in very narrow ranges of the Hue circle, and for this reason Hue could be an important discriminating feature. In addition to Hue, we decided to add the normalized red component, because it has been found that there is some correlation between the normalized red component and the heat of a certain spot at the moment the image was taken.

Algorithm In order to answer our question, we trained the same classifier five times, each of which giving a different number of pixels in input. The classifier used was Random Forest, with 50 decision trees and max depth 20. We chose this classifier both because it is immediate, and because many papers, such as [6], show it has amongst the best performances. In order to carry out the training phase, we divided each image in square blocks of pixels, with side length of 1, 2, 5, 10 and 20. When the size of the image was not a multiple of the block size, we added a black border to compensate. This operation is shown in *figure 2*. As regards the training method, once the images were divided in blocks of pixels, we extracted a tridimensional vector of size $block\ side\ length \times block\ side\ length \times 5$. Each pixel was in position $(i \cdot block\ side\ length) + j$, with i representing the row index and j representing the column index, with $i, j \in [0, 1, 2, ..., block\ side\ length - 1]$. Alongside the pixel blocks, we supplied the classifier with another vector, of size $block\ side\ length \times block\ side\ length$, in which every element was 0 for the *non skin* class, and 1 for the *skin* class.

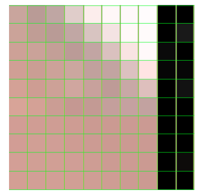


Fig. 2. A block of size 10x10 obtained from a portion of pixels of size 8x10.

Results The results of our tests are summarized in *Figure 3*, and some examples of classified images are presented in *Figure 4*. In order to train the classifiers, we used 900 images from the ECU database; for the testing phase, we used 250 (different) images from the same database. The results show that, as the block size grows, there is a noticeable improvement on the performances of the classifiers: there is an improvement of about 5 points on the F1 score between single-pixels predictions and predictions on blocks of size 5x5.

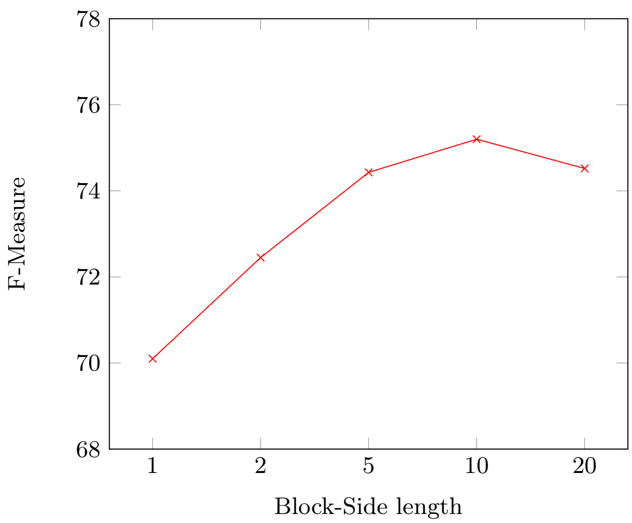


Fig. 3. F-Measure performances of classifiers trained with blocks of different sizes

3 Conclusions

From the presented results, we can conclude that the number of contiguous pixels in input to skin-detection algorithms, has a remarkable impact on the performances. Compared with the results stated in [6], the differences in the performances due to the size of the receptive field, seem to be at least as important as the differences caused by the choice of different color spaces.

References

1. Aibinu, A. M., Shafie, A. A., and Salami, M. J. E. Performance analysis of ann based ycbcr skin detection algorithm. *Procedia Engineering 41* (2012), 1183–1189.
2. Al-Mohair, H. K., Saleh, J. M., and Suandi, S. A. Hybrid human skin detection using neural network and k-means clustering technique. *Applied Soft Computing 33* (2015), 337–347.
3. Albiol, A., Torres, L., and J. Delp, E. Optimum color spaces for skin detection. 122 – 124 vol.1.
4. Géron, A. Hands-on machine learning with scikit-learn and tensorflow: Concepts, tools, and techniques to build intelligent systems. *O'Reilly Media, Inc.* (2017).
5. Kawulok, M., Nalepa, J., and Kawulok, J. Skin detection and segmentation in color images.
6. Khan, R., Hanbury, A., Stöttinger, J., and Bais, A. Color based skin classification. *Pattern Recognition Letters 33*, 2 (2012), 157–163.
7. Khan, R., Hanbury, A., Stöttinger, J., Khan, F. A., Khattak, A. U., and Ali, A. Multiple color space channel fusion for skin detection. *Multimedia Tools and Applications*.
8. Salami, M. J. E., Aibinu, A. M., Mohideen, S. B. O. K., and Mansor, S. A. B. Design of an intelligent robotic donation box a case study.
9. Shaik, K. B., Ganesan, P., Kalist, V., Sathish, B., and Jenitha, J. M. M. Comparative study of skin color detection and segmentation in hsv and ycbcr color space. *Procedia Computer Science* (2015).

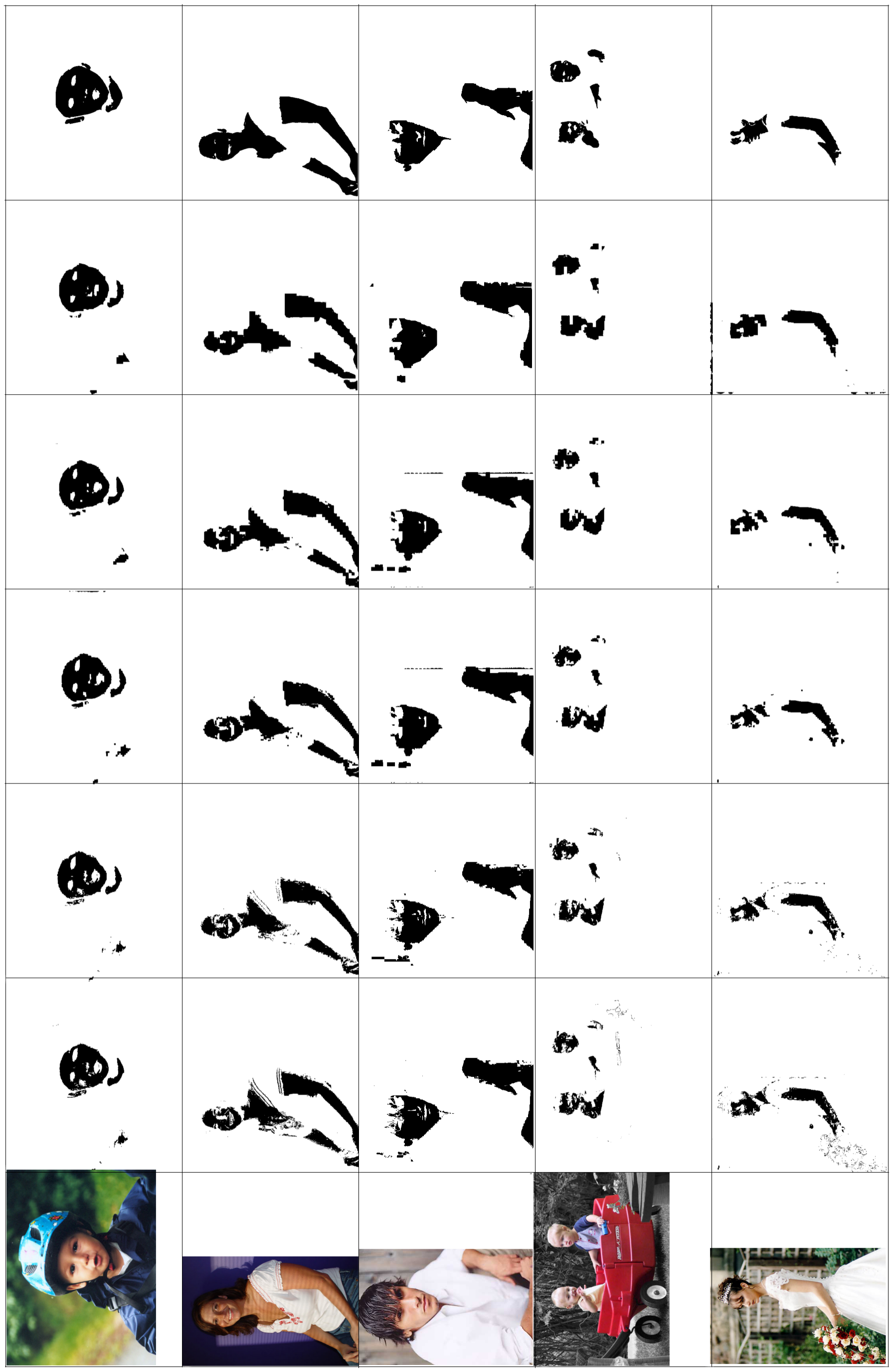


Fig. 4. Some examples on images from the ECU database. From left to right: image, masks obtained with classifiers of block size respectively 1px, 2x2px, 5x5px, 10x10px, 20x20px and ground truth.