Nutrition Assistance based on Skin Color Segmentation and Support Vector Machines

Ermioni Marami, Anastasios Tefas and Ioannis Pitas

Abstract In this paper a new skin color segmentation method that exploits pixels color space information is presented. We evaluate the discrimination strength of features extracted from the RGB and HSV color space and also of a new descriptor generated by combining both spaces. To facilitate our experimental evaluation we have used a linear SVM classifier since it provides certain advantages in terms of computational efficiency compared with its kernel based counterparts. Experiments conducted in video sequences depicting subjects eating and drinking, recorded in complex indoor background and different lightning conditions, where the developed methods achieved satisfactory skin color segmentation.

Key words: skin color segmentation, support vector machines, eating activity recognition, drinking activity recognition

1 Introduction

Skin color segmentation [4] is an important task for various applications such as face detection, localization and tracking [5], skin color enhancement for displays [14] and an essential initial step for many other applications, such as motion analysis and tracking, video surveillance, hand and head gesture recognition [10], video-conference [1], human computer interaction [13], image and video indexing and retrieval [3].

The aim of skin color segmentation is to indicate the presence of human limb or torso within an image or a video frame. Extracted masks, where body parts containing skin are indicated with different color from the background, can be used as input for activity recognition algorithms [7]. The development of a system that automatically recognizes eating and drinking activity, using video processing techniques, would greatly contribute to prolonging independent living of older persons in a non-invasive way aiming at patients in the early stages of dementia. Such a system intends to identify nutrition abnor-

The authors are with the Department of Informatics, Aristotle University of Thessaloniki, Box 451, 54124 Thessaloniki, Greece, e-mail: {emarami,tefas,pitas}@aiia.csd.auth.gr

malities of these persons and assist them primarily in their daily nutrition needs.

In order to preserve the anonymity not only of the final users but also of the participants in training data recordings, privacy preserving human body representations are required. More precisely, binary images where skin parts (silhouettes) appear in white and the rest of the background in black are constructed. It has been noticed that users, especially older persons, resist in having cameras monitoring their daily activities at their home. Although, they were positive in the idea that the monitoring system only analyzes their silhouettes.

In this paper a novel approach for skin color segmentation is proposed. It is based on combining RGB and HSV color spaces and use them as features feeded to a linear Support Vector Machine (SVM). Instead of using non-linear SVMs that increase the computational cost the dimensionality of the RGB features is increased by adding the H and S components which are considered non-linear functions of RGB. That is, the RGB-HS feature space is created where the skin can be more efficiently separated linearly from non-skin regions.

The remainder of the paper is organized as follows. The color spaces used for skin segmentation are reviewed in Section 2 and the SVM classification is described in Section 3. The skin color segmentation approaches examined, as well as the proposed ones, are described in Section 4, where also the post-processing image transformation using mathematical morphology is demonstrated. Experimental results are given in Section 5 and conclusions are drawn in Section 6.

2 Color Spaces

Choosing a color space for skin color segmentation is a controversial issue within the image processing field. Various color spaces with different properties are used distinctly for pixel based skin detection, but sometimes a combination of them can improve performance [8].

RGB (Red, Green, Blue) color space is the most widely used model for processing, representing and storing pixel information of a digital image. In [9], it is stated that the accuracy of the correctly identification of pixels in RGB can be increased if another color space is used. However, RGB color space has been extensively used in skin detection.

HSV (Hue, Saturation, Value) color space is a non-linear transformation of RGB and can be referred to as a perceptual color space due to its similarity to the human perception of color [9]. HSV is widely used for skin detection and it has been found to outperform RGB model in various studies. Hue defines the dominant color (e.g., red, green, purple or yellow) of an area, whereas saturation measures the colorfulness of an area in proportion to its brightness [12]. Value represents brightness along grey axis (e.g., white to

black), but is decoupled from the other two color components. H, S and V components are obtained by applying a non-linear transformation to the RGB color primaries:

$$\begin{aligned} H &= \begin{cases} h, & B \leq G \\ 2\pi - h, & B > G \end{cases} \\ \text{where,} \quad h &= cos^{-1} \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{(R-G)^2+(R-G)(G-B)}} \\ S &= \frac{max(R,G,B) - min(R,G,B)}{max(R,G,B)} \\ V &= max(R,G,B) \end{aligned} \tag{1}$$

3 Support Vector Machines

SVMs are kernel based classifiers, widely applicable in many pattern recognition problems due to their excellent classification performance. Considering the binary separation problem, SVMs aim to determine the *separating hyperplane* with maximum distance (margin) between the closest training points of each class.

Given a set $\mathcal{S}$ of l labeled training points $\mathcal{S} = \{(y_1, \mathbf{x}_1), ..., (y_l, \mathbf{x}_l)\}$, each training point $\mathbf{x}_i \in R^N$ belongs to either of the two classes and is assigned a label $y_i \in \{-1, 1\}$ for $i = 1, ..., l$ [2]. For the linearly separable case, suppose that all the training data can be separated by a hyperplane that is represented by the perpendicular vector $\mathbf{w}$ and the bias b such that:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0 \qquad \forall i \tag{2}$$

Those training points for which the equality in Eq. (2) holds, are the support vectors and their removal would change the solution found.

To form the Lagrangian function of the problem, we introduce positive Lagrange multipliers a_i, $i = 1, ..., l$, one for each inequality constraint in (2) associated with each training sample. Thus, the Lagrangian function L_P is formulated as:

$$L_P \equiv \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} a_i y_i(\mathbf{w}^T\mathbf{x}_i + b) + \sum_{i=1}^{l} a_i \tag{3}$$

Requiring that the gradient of L_P with respect to $\mathbf{w}$ and b vanish we get the following conditions:

$$\frac{\partial L_P}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_i a_i y_i \mathbf{x}_i \tag{4}$$

$$\frac{\partial L_P}{\partial b} = 0 \quad \Rightarrow \quad \sum_i a_i y_i = 0. \tag{5}$$

For non-separable data, we can relax the constraints (2) using positive slack variables $\xi_i, i = 1, \ldots, l$ [6]. The constraints become:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i \geq 0, \qquad \xi_i \geq 0, \qquad \forall i \tag{6}$$

For linearly inseparable data, a non-linear separating hyperplane (non-linear SVM) can be found if we first map those data to a higher dimension feature space $\mathcal{H}$, using a non-linear map function Φ: $\mathbf{R}^d \mapsto \mathcal{H}$, where we assume that data can be linearly separated. According to this, the training algorithm would only depend on the data through dot products in $\mathcal{H}$, i.e., on functions of the form $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. A 'kernel function' K such that $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ can be used. Some examples of kernels used in SVMs and investigated for pattern recognition problems are the polynomial and the Gaussian Radial Basis Function (RBF) kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^p$, $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-||\mathbf{x}_i - \mathbf{x}_j||^2 / 2\sigma^2}$.

It is clear that when linear SVMs are used, the testing procedure requires only a multiplication of the input test vector $\mathbf{x}_j$ with the perpendicular vector $\mathbf{w}$ given in (4) and addition of the bias term b. Thus the predicted label y_j is computed as: $y_j = \mathbf{x}_j^T \mathbf{w} + b$. However, using non-linear kernels the computational cost in the testing procedure depends on the number of support vectors which is prohibitive in many cases.

4 Skin Color Segmentation Techniques

4.1 Thresholds in HSV color space

In order to define skin-like pixels in a digital image we used as a baseline method predefined thresholds for H, S and V components. For every image pixel, each of the three color components is compared with the corresponding thresholds to decide whether a pixel is skin or not. The pixel that fulfills the limitations the thresholds set is considered to be a skin pixel. According to this, a binary image is created where white color (pixel value '1') represents skin regions and black color (pixel value '0') other regions (e.g., clothes, background etc.). The thresholds used for every component are according to [5]: $0 < H < 0.1$, $0.23 < S < 0.68$ and $0.27 < V$. To overcome the limitations of generic thresholds, two approaches are proposed. The first one aims at finding person-specific skin regions by learning the skin distribution inside the facial area whereas the second one uses a combined color space with SVM, for improving the generic thresholds on HSV.

4.2 Adaptive thresholds in HSV color space

In order to calculate person-specific thresholds for skin segmentation a face detector is used in each frame and the area inside the face is used for adjusting the HSV skin thresholds. That is, at each video frame, face detection is applied and the Region Of Interest (ROI) depicting person's face is obtained. The histogram of this ROI for the HSV color space is then calculated. Expecting that most of the pixels in the face's ROI are skin colored pixels the person's skin color can be approximated. This procedure provides a person-specific skin color detection. Furthermore, as face detection techniques used are robust in illumination changes, the problem concerning variable illumination conditions is efficiently tackled by this approach. Afterwards, we estimate the peak of the histogram for the hue channel which is the most important component. We use the width value of the peak to adjust properly the predefined thresholds of hue parameter. The predefined thresholds for hue, saturation and value parameters are used in order to decide whether a pixel is skin or not. If a pixel's value is between the modified thresholds, this is considered to correspond to skin location. We check all the pixels of the image so as to get a binary output.

4.3 SVMs using color space information

The second approach uses SVM combined with color features for skin color classification. A two-class linear SVM classifier was trained to correctly classify pixels in skin and non-skin classes. For the skin class (class label '1') we used pixels from human skin regions (head, hands, neck, arms) and for the non-skin class (class label '-1') we used pixels from other regions except human skin (background, clothes, tables, windows). Firstly, for the training and testing process we used as feature vector the RGB components of each pixel. For every frame of the video, during the testing process, each pixel is classified to one of the two classes (skin, non-skin). A binary image is created as output where skin pixels are represented with white and non-skin pixels with black.

The same procedure was followed using color information from HSV color space. Feature vector was consisted of S and V components. Due to the cylindrical property of the hue component, instead of the H value we use the cosine of this angle as the first feature of this vector. Considering the histogram representation of the three H, S, V components distribution, we also used a non-linear kernel to train and test SVMs. A classical RBF was used as kernel for the non-linear classifier.

Finally, we propose a new descriptor generated by combining both spaces using features extracted from RGB and HSV. To create a feature vector of each pixel for the training and testing of SVMs we used five components: R,

G, B, cosine of H and S. V component is omitted due to its equality with one of the R, G, B components (the maximum of them). The expansion of RGB to RGB-HS can be considered as a non-linear dimensionality increase that allows for linear separation using linear SVM. The resulted separating hyperplanes in the RGB-HS space are indeed non-linear separating surfaces to either RGB or HS alone. That is, a non-linear separation is obtained without explicitly using kernels.

4.4 Post-Processing Morphological Operations

In order to isolate individual elements, join disparate elements and remove noise in the binary output images we apply morphological operations (dilations and erosions) [11]. Dilation generally increases the sizes of objects, filling in holes and broken areas, and connecting areas that are separated by spaces smaller than the size of the structuring element. On the other hand, erosion decreases the sizes of objects and removes small anomalies by subtracting objects with a radius smaller than the structuring element. The application of a dilation followed by an erosion corresponds to a morphological operation named 'closing' and aimed mainly to connect components (interesting parts). The reverse application, where erosion is followed by dilation, is referred to as 'opening' and aims to noise removal.

5 Experimental Results

To compare the performance of the proposed techniques we apply them on data derived for the project MOBISERV (http://www.mobiserv.eu) comprised of 13 video sequences depicting older persons during a meal, performing eating and drinking activity. Thresholds and adaptive thresholds in HSV color space did not manage to provide clear enough binary masks for many of the tested videos. To evaluate SVMs performance, we created a training dataset with 81971 skin pixels and 76311 non-skin pixels selected from the first frame of each video. In order to test the ability of the proposed methods to correctly classify skin and non-skin pixels, a 10-fold cross-validation procedure was applied for a linear and a non-linear SVM and for features from RGB, HSV or both spaces. For each experiment, this procedure excludes one of the ten sets of patterns from the training set and uses this set for validation. This procedure was repeated ten times and an overall accuracy rate was calculated. Tab. 1 presents the classification accuracy rates (%) for each case. We notice that non-linear SVMs achieve higher classification accuracy than linear SVMs. However, since this improvement is minor, we prefer to use linear SVMs which also decrease computational complexity.

Fig. 1 illustrates binary masks extracted from video no. 2 using features from RGB color space (first line), HSV space (second line) and from both

Table 1 Classification accuracy rates (%) for 81971 skin and 76311 non-skin pixels performing a 10-fold cross-validation.

SVM	RGB	HSV	RGB-HS
linear	93.46	89.69	96.11
rbf	94.41	92.11	96.69

color spaces (third line). The features used in the third case were R, G, B components, the cosine of H component and S component. The used classifier was a linear SVM. Column (a) presents the initial classification output for each image pixel, while column (b) shows the final binary masks obtained after morphological operations. Considering an algorithm for eating and drinking activity recognition, the data information revealed from the major ROIs, which include the head and palms regions, is required. The linear SVM with features extracted from RGB color space was unable to define all of these ROIs. Features from HSV color space highlighted irrelevant ROIs, while using features from both color spaces the desired ROIs were accurately maintained.

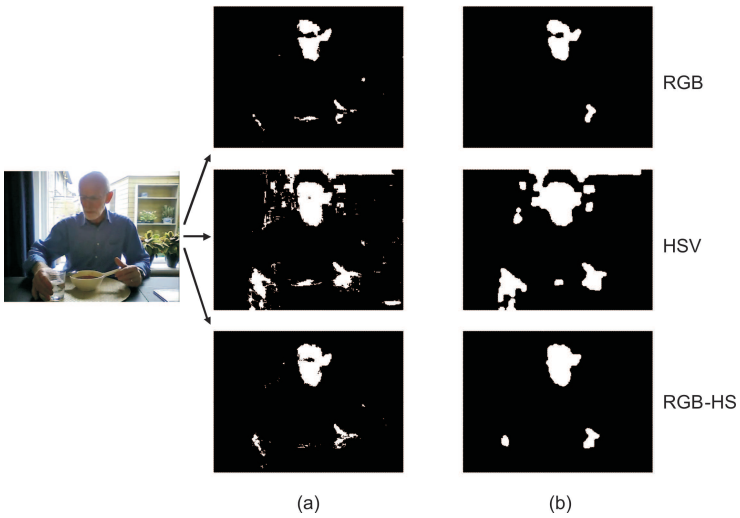

Fig. 1 Binary masks extracted using a linear SVM with features from RGB, HSV or both color spaces from video no. 2 (a) initial classification results, (b) binary masks after morphological operations.

6 Conclusions

Experimental results revealed that thresholds in HSV color space are not effective in all cases. Due to variations in illumination conditions, binary

masks extracted using this technique were not clear enough to be used. Using adaptive thresholds in HSV color space is a technique that remedies the illumination problems but it is person-specific and can not be utilized if the face detector used fails. However, SVMs encounter difficulties arising from variations in lightning conditions due to the diversity in pixels used in the training process and form a more generalized classifier. Final results demonstrate the effectiveness of the new descriptor generated by combining both RGB and HSV color spaces. The conducted experiments verified that the combined descriptor generated most accurate skin color segmentation binary masks compared with the ones attained by the other descriptors.

Acknowledgements This work has been funded by the Collaborative European Project MOBISERV FP7-248434 (http://www.mobiserv.eu), An Integrated Intelligent Home Environment for the Provision of Health, Nutrition and Mobility Services to the Elderly.

References

1. Askar, S., Kondratyuk, Y., Elazouzi, K., Kauff, P., Schreer, O.: Vision-based skin-colour segmentation of moving hands for real-time applications. In: Proc. of 1st European Conf. on Visual Media Production (CVMP), pp. 524–529 (2004)
2. Burges, C.: A tutorial on support vector machines for pattern recognition. Kluwer Academic Publishers, Boston (1998)
3. Cheddad, A., Condell, J., Curran, K., McKevitt, P.: A new colour space for skin tone detection. In: Image Processing (ICIP), 2009 16th IEEE International Conference on, pp. 497–500. IEEE (2010)
4. Cheng, H., Jiang, X., Sun, Y., Wang, J.: Color image segmentation: advances and prospects. Pattern Recognition **34**(12), 2259–2281
5. Cherif, I., Solachidis, V., Pitas, I.: A tracking framework for accurate face localization. Artificial Intelligence in Theory and Practice pp. 385–393 (2006)
6. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297
7. Gkalelis, N., Tefas, A., Pitas, I.: Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition. IEEE Transactions on Circuits and Systems for Video Technology **18**(11), 1511–1521 (2008)
8. Gomez, G., Sanchez, M., Enrique Sucar, L.: On selecting an appropriate colour space for skin detection. MICAI 2002: Advances in Artificial Intelligence pp. 3–18 (2002)
9. Kelly, W., Donnellan, A., Molloy, D.: Screening for objectionable images: A review of skin detection techniques. In: International Machine Vision and Image Processing Conference, pp. 151–158. IEEE (2008)
10. Ong, S., Ranganath, S.: Automatic sign language analysis: A survey and the future beyond lexical meaning. Pattern Analysis and Machine Intelligence, IEEE Transactions on **27**(6), 873–891
11. Soille, P.: Morphological image analysis: principles and applications, 2nd edition edn. Springer-Verlag, New York (2004)
12. Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: Proc. Graphicon, vol. 3, pp. 85–92. Citeseer (2003)
13. Wu, Y., Huang, T.: Hand modeling, analysis and recognition. Signal Processing Magazine, IEEE **18**(3), 51–60
14. Zhang, X., Jiang, J., Liang, Z., Liu, C.: Skin color enhancement based on favorite skin color in HSV color space. IEEE Transactions on Consumer Electronics **56**(3), 1789–1793