STUDIA INFORMATICA 2012
Volume 33 Number 2B (106)

Michał KAWULOK
Politechnika Śląska, Instytut Informatyki

SKIN REGION DETECTION IN DIGITAL IMAGES USING DISCRIMINATIVE TEXTURAL FEATURES

Summary. This paper presents a novel approach towards human skin regions detection and segmentation. The main contribution is concerned with proposing the discriminative texture analysis performed over skin probability maps obtained using conventional color-based methods. Results of the experimental validation reported in the paper confirm that the texture is an important source of information, neglected by many skin detection techniques.

Keywords: skin detection and segmentation, image processing, texture analysis

DETEKCJA OBSZARÓW SKÓRY W OBRAZACH CYFROWYCH ZA POMOCĄ ROZRÓŻNIAJĄCYCH CECH TEKSTURALNYCH

Streszczenie. Niniejszy artykuł przedstawia nową metodę poprawy dokładności detekcji i segmentacji obszarów ludzkiej skóry w obrazach cyfrowych. Oryginalnym elementem przedstawionych badań jest zastosowanie dyskryminacyjnej analizy teksturalnej, przeprowadzanej w obrazie mapy prawdopodobieństwa występowania skóry. Mapy takie są otrzymywane za pomocą klasycznych metod funkcjonujących na bazie analizy barwy. Przedstawione wyniki walidacji eksperymentalnej potwierdzają, że tekstura stanowi istotne źródło informacji w detekcji skóry, co nie jest wykorzystywane przez większość istniejących metod.

Słowa kluczowe: detekcja i segmentacja obszarów skóry, przetwarzanie obrazów, analiza teksturalna

1. Introduction

Skin region detection and segmentation is an interesting topic of computer vision, having a wide range of potential applications. Its general aim is to classify every individual pixel in a given digital image as belonging to skin or not. Skin detection is an important source of information for locating faces and hands for human-computer interaction purposes, being the first step of the image understanding chain. Segmenting skin regions can also be used for indexing images in multimedia databases. This, in turn, allows search criteria be defined, related to the number of skin blobs in an image or the total area covered by skin. If such information is combined with highly-effective face detectors [8, 20], it is just a matter of defining simple decision rules to retrieve images of specific contents like portraits, crowd or street scenes. Furthermore, this can also be used for content-based filtering and parental control applications.

There have been many skin detection methods proposed which are based on modeling skin color in various color spaces. Such models are later used to determine the probability that a given pixel presents human skin. In this way, the input image is converted into a skin presence probability map. Depending on a particular approach, such a map may have continuous or binary values. Skin color models are usually quite sensitive to illumination variations, and alternative sources of information can be helpful to stabilize the result. This can be achieved by analyzing neighborhood and topology of pixels classified as skin to reduce false positive errors [7, 18]. Also, the features helpful for skin segmentation can be extracted using texture analysis performed in grayscale [21] or color domain [1].

In the work reported here we propose and investigate analysis of discriminative textural features (DTF) performed in the probability map obtained using a statistical approach [5]. We analyze the texture based on simple statistics computed in kernels of various size. This forms basic feature vectors that are later subject to *linear discriminant analysis* (LDA). This generates the LDA feature space, in which we measure the similarity between the pixels with regards to their skin-likelihood. In this way we managed to reduce the segmentation errors substantially, which was confirmed during experimental validation.

The paper is organized as follows. In Section 2 an overview of existing approaches to skin segmentation is given, and statistical skin segmentation that is the basis for the research reported here is described in Section 3. Our contribution is explained in Section 4 and experimental validation results are shown and discussed in Section 5. The paper is concluded and directions of future works are presented in Section 6.

2. Related work

Existing skin segmentation techniques take advantage from observation that skin-tone color has common properties which can be defined in various color spaces. In general, skin color detectors are based on parametric or statistical skin modeling. An interesting, thorough survey which compares various color-based skin detection routines was recently presented by Khan et al. [10].

Parametric skin models are based on fixed decision rules defined empirically in various color spaces after skin-tone distribution analysis. These rules are applied after color normalization to determine if a pixel color value belongs to the skin. Kovac et al. [11] proposed a model defined in *RGB* color space. Skin-tone color was also modeled in *HSV* by Tsekeridou et al. [19]. An approach proposed by Hsu et al. [4] takes advantage of common skin color properties in nonlinearly transformed YC_bC_r, in which elliptical skin color model is defined. Some techniques operate in multiple color spaces to increase the stability, for example a composed skin detector [12] defined in *RGB* and YC_bC_r color spaces.

Statistical modeling is based on analysis of skin pixel values distribution for a training set of images, in which skin and non-skin areas are already identified and marked. This creates a global model of skin color, which allows determining the probability that a given pixel value belongs to the skin class. Skin color can be modeled using a number of techniques, including Bayes classifier [5], Gaussian mixture model [3] or random forests [9]. The first method [5] was used in the research reported here to generate the probability maps and it is given more attention later in the paper.

There are a number of adaptive methods that improve the segmentation accuracy. Lee et al. proposed to extract lighting features from every analyzed image to adjust the skin detector [13]. Phung et al. proposed a method for adapting the segmentation threshold in the probability map [15], and this approach was later extended by Zhang et al. [22]. We have also investigated adaptive approaches during our earlier works. First, we proposed to adjust the global skin model to every face detected in the image [6], and then we managed to decrease false positives using skin blobs analysis [7].

In case of skin segmentation in video sequences rather than in still images, the system may take advantage of dynamic information. Sigal et al. used Markov models to predict illumination changes in subsequent video frames to adjust the skin color model [17]. Furthermore, background extraction techniques and motion detectors may be used to find potential locations of skin pixels.

3. Statistical skin modeling

In the reported research, statistical skin modeling based on Bayes classifier [5] was used to obtain the skin probability maps that were later analyzed. At first, based on a training set, histograms for skin (C_S) and non-skin (C_{NS}) classes are built. For every color value (v) probability of observing it in each class is obtained:

$$P(v \mid C_x) = C_x(v) / N_x, \tag{1}$$

where $C_x(v)$ is a number of v-color pixels in x-th class and N_x is the total number of pixels in that class. Maximal number of histogram bins depends on pixel bit-depth and usually it is 256 × 256 × 256. However, based on our earlier works [7] we determined that detection is more effective if the number of bins is reduced to about 64 per channel. To detect skin with low false positive error, not only is it important that a pixel value appears frequently as skin, but also its density in the non-skin class should be low. Hence, probability that a given pixel value belongs to the skin class is computed using Bayes rule:

$$P(C_S \mid v) = \frac{P(v \mid C_S)P(C_S)}{P(v \mid C_S)P(C_S) + P(v \mid C_{NS})P(C_{NS})}, \tag{2}$$

where $P(C_S)$ and $P(C_{NS})$ are estimated based on number of pixels in both classes. If the training set is big enough, probabilities for all possible color values can be computed to create skin color probability lookup table. However, if a value is not observed in the training set, it is assumed that the probabilities $P(v|C_S)$ and $P(v|C_{NS})$ equal 0.5. After obtaining the probability lookup table, every pixel value can be transformed into the probability that this pixel belongs to the skin class. Skin regions can be later extracted based on a fixed acceptance threshold or more advanced thresholding and blob detection techniques.

4. Texture-based probability map processing

Distribution of skin color has been effectively modeled in various color spaces and chrominance is considered as the most distinctive skin feature. However, its discriminating capability is limited, because skin color is not exclusive for skin regions and other objects can also have similar color to human skin. Hence, color-based skin models are quite effective for detecting skin pixels, however they fail when the background objects are of skin-like color. In order to decrease false positives, the model should be extended to use other distinctive features extracted in pixels' neighborhood, instead of classifying every individual pixel independently.

We explored how to enhance the conventional color model using textural features. Although human skin is characterized by distinctive pattern, it can be observed only in high-quality images of high resolution. Otherwise, in images of medium or low quality, skin regions are rather plain, without any remarkable textural properties. On the other hand, false positives often present relatively small color variations which are well manifested in the skin color probability map. This is illustrated in Fig. 1, in which an original image (a), and corresponding skin probability map (b) are presented. The selected region, characterized by smooth texture with small chrominance variations, is magnified below. It may be noticed that although the values in the probability map are high within the region, they form a distinctive pattern, which cannot be observed for real skin. Such phenomenon can be observed for many images, which leads to a conclusion that textural features extracted from the probability map may be helpful for skin segmentation. During our experiments we found it much more beneficial to extract these features from the probability maps rather than from the original images.

In order to determine, which textural features discriminate well between skin and non-skin regions, we first compute simple image statistics in several kernels of different dimension. This forms *basic image features* which are subsequently subject to linear discriminant analysis (LDA) to obtain *discriminative textural features* (DTF). The details of the proposed feature extraction technique are presented later in this Section.

a) b)

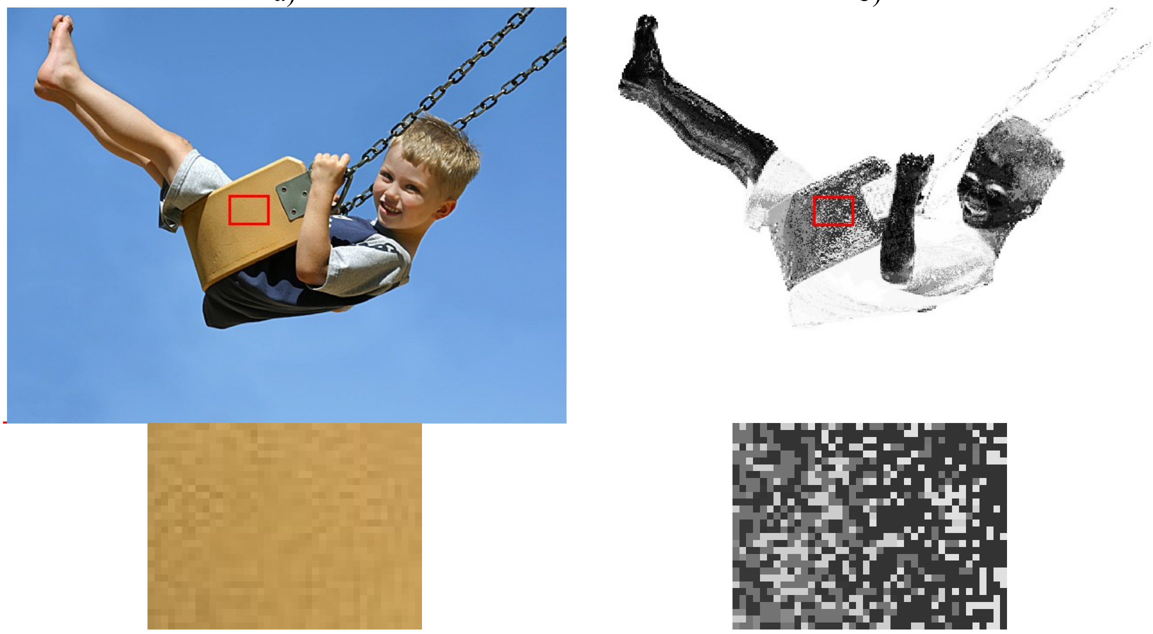

Fig. 1. Skin detection: a) input image, b) obtained skin probability map
Rys. 1. Detekcja skóry: a) obraz wejściowy, b) otrzymana mapa prawdopodobieństwa występowania skóry

4.1. Linear discriminant analysis

Linear discriminant analysis [16] is a supervised statistical feature extraction method frequently used in machine learning. It finds a subspace defined by the most discriminative directions within a given training set of M-dimensional vectors classified into K classes. The analysis is performed first by computing two covariance matrices termed within-class scatter matrix:

$$S_W = \sum_{i=1}^{K} \sum_{u_k \in K_i} (u_k - \mu_i)(u_k - \mu_i)^T , \tag{3}$$

and between-class scatter matrix:

$$S_B = \sum_{i=1}^{K} (\mu_i - \mu)(\mu_i - \mu)^T , \tag{4}$$

where μ is a mean vector of the training set and μ_i is a mean vector of the i-th class (termed K_i). Subsequently, the matrix $S = S_W^{-1} S_B$ is subjected to the eigen decomposition $S = \Phi \Lambda \Phi^T$, where $\Lambda = diag(\lambda_1, \ldots, \lambda_M)$ is the matrix with the ordered eigenvalues along the matrix diagonal and $\Phi = [\upsilon_1 | \ldots | \upsilon_M]$ is the matrix with the correspondingly ordered eigenvectors as columns. The eigenvectors form the orthogonal basis of the feature space. Originally, the feature space has M dimensions, but only those associated with the highest eigenvalues have strong discriminative power, while the remaining can be rejected. In this way the dimensionality is reduced from M to m, where $m < M$. After having built the m-dimensional feature space, the feature vectors are obtained by projecting the original vectors u onto the feature space: $v = \Phi^T u$. The similarity between the feature vectors is computed based on their Euclidean distance in the feature space.

4.2. Discriminative textural features

In order to determine the discriminative textural features, first we calculate basic image features from every pixel in the probability map. They are composed of: a) mean and b) median value, c) standard deviation, d) minimal and e) maximal value, computed in multiple kernels of different size, and f) the difference between maximum and minimum values in each kernel. These features are obtained for four kernel sizes ranging from 5×5 to 11×11. Hence, every pixel x is described with an M-dimensional basic feature vector u_x (M = 24 in the presented case).

Skin and non-skin pixels are transformed into two classes of basic feature vectors which are subsequently subject to LDA in order to increase their discriminative capability. This training process, illustrated in Fig. 2, creates a discriminative feature space that is later used to generate skin probability maps. First, for a given training set of images and associated skin masks, the probability lookup table is obtained using statistical skin modeling [5]. After that, all the images from the training set are converted into probability maps, from which basic image features are extracted. Finally, based on the basic feature vectors, labeled using the ground-truth skin masks, LDA projection matrix Φ is computed.

In most cases, the first three eigenvalues were definitely larger than the rest of them. This means that the discriminating power was cumulated in the first three components, hence the dimensionality of the DTF space was reduced to m=3. This was also confirmed on the experimental way. Using less than three components would affect the detection score, while using more would only increase the computation time without any gain.

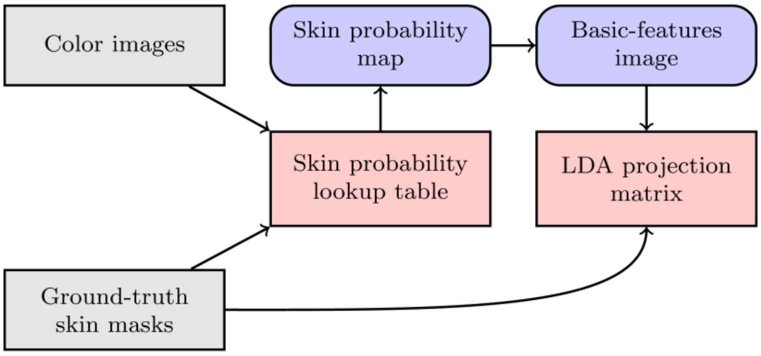

Fig. 2. Flowchart of the training process
Rys. 2. Schemat przebiegu procesu treningu

4.3. Skin detection

Skin detection requires two stages. First of all, an image is converted into the skin probability map using the statistical model (i.e. the lookup table obtained during training). Subsequently, every pixel from the probability map is projected onto the DTF space based on the LDA projection matrix. We investigated a possibility of using support vector machines (SVM) [2] for classifying the vectors in the DTF space to determine whether they belong to skin. This required an additional step during training (i.e. the SVM was trained using labeled pixels from the train set, represented in the DTF space), but it also introduced substantial burden on the skin detector, as every individual pixel had to be classified with the SVM.

We found it equally effective and much more efficient to use the DTF space only to determine similarity between the pixels with regards to their "skinness". We benefit from the observation that for a very low acceptance threshold, false positives are virtually eliminated. Hence, in every image we determine a pixel of maximal skin probability, which is treated as a reference. The final skin probability map is obtained based on the similarity of every individual pixel to the reference pixel in the DTF space. Probability of the reference pixel remains unchanged, and it is maximal in the map after the transformation. The values in the probability map are scaled linearly using the DTF distance of every pixel from the reference pixel. Hence, if there are no skin pixels in the image (i.e. the probability of the reference pixel is low), the values in the final skin probability map will be low as well.

The reference pixel should not be an isolated skin spot, because this would disturb the textural features, so the maximal value is found after filtering the probability map with a minimum filter of a large 15×15 kernel (i.e. larger than the largest kernel used for extracting the basic features). The whole process for detecting skin regions is presented in Fig. 3.

5. Experimental validation

The experiments were carried out for 50 images from ECU database [14], which contains color images associated with ground-truth skin binary masks indicating skin regions. These images were acquired in uncontrolled lighting conditions, and skin-color objects often appear in the background, which altogether makes these images difficult for the analysis. These images were randomly split into two equinumerous sets used for training and testing, and the results presented in the paper were obtained using 10-fold cross validation.

We measured the skin segmentation performance based on two errors, namely: a) *false positive* rate (δ_{FP}), obtained as a percentage of background pixels classified as skin, and b) *false negative* rate (δ_{FN}), i.e. a percentage of skin pixels misclassified as background. Classification was based on a fixed acceptance threshold, which is a simple method, but it demonstrates performance of the core skin detector. False positives and false negatives both depend from the acceptance threshold, and their mutual relation is commonly presented using *Receiver Operating Characteristics* (ROC).

ROC curve, averaged over all ten tests, is presented in Fig. 4. In Tab. 1 we present false positive error rate and its standard deviation obtained for several fixed false negative rates. It may be observed from the figure and table that the proposed method decreases the error at every threshold value. For false negative rates over 10%, the false positives were decreased by around a half, which is a substantial improvement.

An example of error reduction using the proposed approach is demonstrated in Fig. 5. Here, the probability map obtained using the color-based model (a) is thresholded to determine skin and non-skin regions without (b) and with (c) DTF-based analysis. Correctly detected regions are marked with a green border, false positives are presented in red, and false negatives – in blue. It can be noticed that our method eliminated the majority of false positives, and also some small false negatives were removed from the face region. The proposed

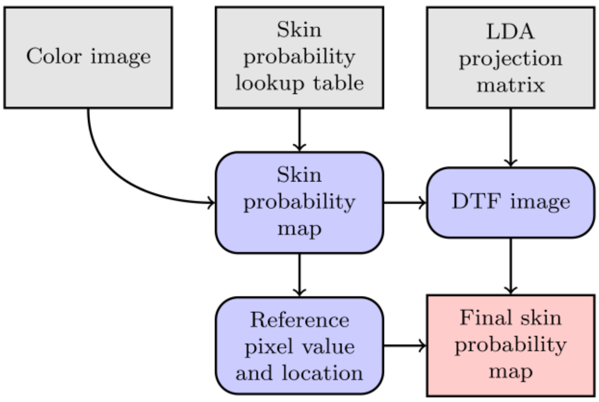

Fig. 3. Flowchart of the skin detection process
Rys. 3. Schemat przebiegu procesu detekcji obszarów skóry

method does not analyze image edges due to the large kernel size used for extracting the basic features, which explains the false positives observed at the left edge in Fig. 5c.

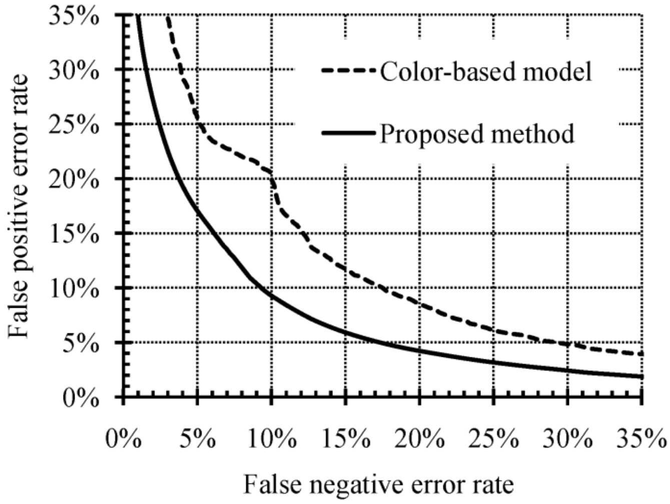

Fig. 4. ROC curves obtained for our method and conventional detector [5]
Rys. 4. Krzywe ROC uzyskane dla zaproponowanej metody oraz detektora statystycznego [5]

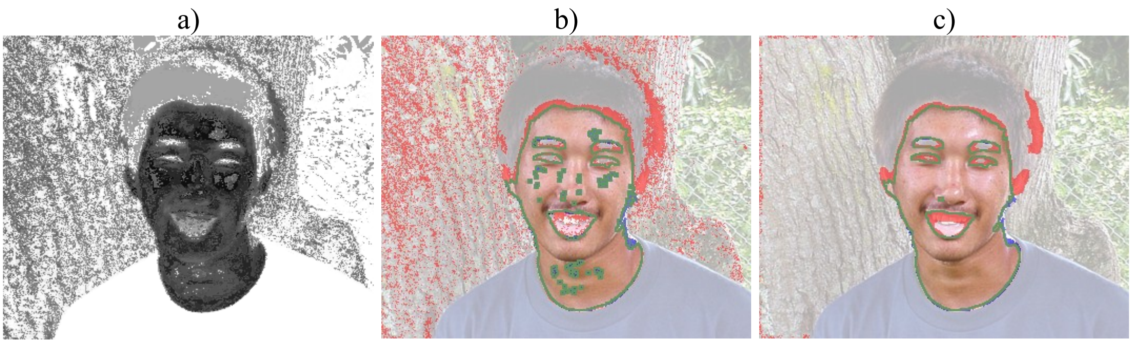

Fig. 5. Skin segmentation results: a) skin probability map, b) skin segmentation performed directly in the map, c) after DTF-based processing
Rys. 5. Segmentacja obszarów skóry: a) mapa prawdopodobieństwa, b) segmentacja na podstawie mapy, c) segmentacja po zastosowaniu analizy tekstury

Table 1
False positive error for various false negative rates

False negative rate (δ_{FN}) ↓	False positive rate (δ_{FP})	
	Color-based detector	Our method
5%	26.19±2.56%	**16.31±3.53%**
10%	19.1±3.71%	**9.52±2.77%**
15%	11.99±2.66%	**6.04±1.84%**
20%	8.43±2.19%	**4.29±1.24%**

6. Conclusions and future work

In this paper it was demonstrated that skin region segmentation can be improved using textural features extracted from the skin probability maps. The proposed method for skin region segmentation using discriminative textural features processes the skin probability maps, which reduces the detection errors. This was confirmed by the experimental results, presented and discussed in the paper.

The obtained detection errors can be further reduced using global-scale methods applied to the probability maps. The main directions of future work include implementation of the blob detection and adaptive thresholding in the DTF space, which should be more effective than performed in the probability maps. Furthermore, simple image statistics were used as the basic features here, and we plan to investigate application of other features, especially those that represent local frequency. This should further increase the discriminating capability of the proposed detector.

BIBLIOGRAPHY

1. Conci A., Nunes E., Pantrigo J. J., Sánchez A.: Comparing color and texture-based algorithms for human skin detection. Proc. 10th International Conference on Enterprise Information Systems HCI (ICEIS), 2008, p. 166÷173.
2. Cortes C., Vapnik V.: Support-Vector Networks. Machine Learning, Vol. 20, No. 3,1995, p. 273÷297.
3. Greenspan H., Goldberger J., Eshet I.: Mixture model for face-color modeling and segmentation. Pattern Recognition Letters, Vol. 22, No. 14, 2001, p. 1525÷1536.
4. Hsu R.-L., Abdel-Mottaleb M., Jain A. K.: Face detection in color images. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, May 2002, p. 696÷706.
5. Jones M. J., Rehg J. M.: Statistical color models with application to skin detection. International Journal of Comp. Vision, Vol. 46, No. 1, 2002, p. 81÷96.
6. Kawulok M.: Dynamic skin detection in color images for sign language recognition. Lecture Notes in Computer Science, Vol. 5099, Springer-Verlag, Heidelberg 2008, p. 112÷119.
7. Kawulok M.: Energy-based blob analysis for improving precision of skin segmentation. Multimedia Tools and Applications, Vol. 49, No. 3, Springer, 2010, p. 463÷481.
8. Kawulok M., Szymanek J.: Precise multi-level face detector for advanced analysis of facial images. IET Image Processing, Vol. 6, No. 2, 2012, p. 95÷103.

9. Khan R., Hanbury A., Stoettinger J.: Skin detection: A random forest approach. Proc. IEEE International Conference on Image Processing, ICIP 2010, Hong Kong 2010, p. 4613÷4616.
10. Khan R., Hanbury A., Stoettinger J., Bais A.: Color based skin classification. Pattern Recognition Letters, Vol. 33, No. 2, 2012, p. 157÷163.
11. Kovac J., Peer P., Solina F.: Human Skin Colour Clustering for Face Detection. Eurocon 2003 Ljubliana, Slovenia 2003, p. 144÷148.
12. Kukharev G., Nowosielski A.: Fast and Efficient Algorithm For Face Detection in Colour Images. Machine Graphics & Vision, Vol. 13, No. 4, 2004, p. 377÷399.
13. Lee J.-S., Kuo Y.-M., Chung P.-C., Chen E.-L.: Naked image detection based on adaptive and extensible skin color model. Pattern Recognition, Vol. 40, No. 8, 2007, p. 2261÷2270.
14. Phung S. L.: ECU Face Detection Database. University of Wollongong, 2002.
15. Phung S. L., Chai D., Bouzerdoum A.: Adaptive skin segmentation in color images. Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2003, p. 353÷356.
16. Seber G.: Multivariate Observations. Wiley, New York 1984.
17. Sigal L., Sclaroff S., Athitsos V.: Skin color-based video segmentation under time-varying illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No.7, 2004, p. 862÷877.
18. Ruiz-Del-Solar J., Verschae R.: Skin detection using neighborhood information. Proc. IEEE International Conference on Automatic Face and Gesture Recognition, FG 2004, p. 463÷468.
19. Tsekeridou S., Pitas I.: Facial feature extraction in frontal views using biometric analogies. Proc. of EUSIPCO ’98, Vol. 1, Rhodes, Greece September 1998, p. 315÷318.
20. Viola P., Jones M.: Robust real-time face detection. Int. J. of Computer Vision, Vol. 57, No. 2, 2004, p. 137÷154.
21. Wang X., Zhang X., Yao J.: Skin color detection under complex background. Proc. International Conference on Mechatronic Science, Electric Engineering and Computer, MEC 2011, p. 1985÷1988.
22. Zhang M.-J., Wang W.-Q., Zheng Q.-F., Gao W.: Skin-Color Detection Based on Adaptive Thresholds. Proc. IEEE International Conference on Image and Graphics, 2004, p. 250÷253.

Wpłynęło do Redakcji 26 stycznia 2012 r.

Omówienie

Detekcja oraz segmentacja obszarów ludzkiej skóry w obrazach cyfrowych stanowią istotne elementy wielu systemów wizyjnych, będąc w nich podstawowym źródłem informacji o lokalizacji dłoni, a także sylwetki człowieka. Ponadto, detektor skóry pozwala na ekstrakcję zaawansowanych informacji dotyczących treści obrazu oraz na automatyczne indeksowanie multimedialnych baz danych, bez konieczności manualnego definiowania metadanych semantycznych. To z kolei umożliwia tworzenie kryteriów wyszukiwania związanych z zawartością obrazów, na przykład powierzchnią lub liczbą obszarów skóry.

Zdecydowana większość istniejących metod funkcjonuje na bazie analizy barwy, która jest uważana za najbardziej charakterystyczną cechę pikseli reprezentujących obszary ludzkiej skóry. Warto przy tym zwrócić uwagę na fakt, że o ile rozkład barwy skóry jest skumulowany w niewielkim obszarze przestrzeni barwnej, to istnieje jednak wiele obiektów posiadających barwę zbliżoną do skóry. To oznacza, że modele skóry oparte na barwie umożliwiają poprawne wskazanie obszarów jej występowania, jednak generują wysokie błędy przyjęcia (tj. wyniki fałszywie dodatnie) dla obiektów tła o kolorze ludzkiej skóry. Błędy te mogą być zmniejszane przez rozszerzenie modelu o dodatkowe źródła informacji, takie jak tekstura bądź geometria obszarów skóry, co jest wykorzystywane przez niektóre z opublikowanych metod.

Niniejszy artykuł przedstawia nowe podejście do ekstrakcji rozróżniających cech teksturalnych, rozszerzających modele barwne ludzkiej skóry. Jednak, w przeciwieństwie do istniejących algorytmów, analizie nie jest poddawany oryginalny obraz barwny lub monochromatyczny, tylko mapa prawdopodobieństwa, otrzymana za pomocą klasycznych modeli statystycznych. W ujęciu standardowym mapy te są poddawane różnym technikom progowania w celu wskazania lokalizacji obszarów występowania skóry. Badania zaprezentowane w artykule wskazują na fakt, że nie tylko wartość prawdopodobieństwa poszczególnych pikseli w tychże mapach, ale również ich tekstura, posiada właściwości rozróżniające pomiędzy skórą a nie-skórą. To z kolei pozwala, przy zastosowaniu zaproponowanej metody, na zmniejszenie błędów segmentacji, co zostało potwierdzone przez przedstawione wyniki badań eksperymentalnych.

Address

Michał KAWULOK: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16,
44-100 Gliwice, Polska, michal.kawulok@polsl.pl.