A Multiple Classifier Approach for Detecting Naked Human Bodies in Images

Luca Giangiuseppe Esposito and Carlo Sansone

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione (DIETI)
University of Naples Federico II
lucag.esposito@gmail.com, carlosan@unina.it

Abstract. With the arise of Web 2.0 and social networks, millions of people upload multimedia contents on the web every day. Images are typically published without control and only if users report a problem, unappropriate or offensive content are removed. So, there is the need of a system for automatically processing such kind of data.

In this paper we propose a system, based on a multiple classifier approach, for detecting naked human bodies in images. It analyzes both body geometric properties and global visual features. A comparison with other state-of-the-art proposals demonstrated the effectiveness of the proposed approach.

1 Introduction

Content-based image retrieval methods [1] are capable of analysing image visual features like colours and shapes. Analyzing digital image content led to understand, completely or partially, the meaning of a scene. Within this framework, because of the need of avoiding the diffusion of offensive contents, many researchers had proposed various methods to detect images depicting naked or scantily dressed people/ human bodies in digital images.

Since there is a great correlation between images with wide skin colored areas and pornographic/erotic images, most of the proposed approach are first based on skin detection techniques that can isolate human skin pixels from the background. In [2] the authors, by collecting and analyzing several human images and their color histograms, empirically generated a cluster region in HSV colour space whose values approximate the skin colour distribution. To reduce noise influence on human skin colour, because of the influence of artificial lights and image forgery techniques, Wang et al. [3] perform a median filter processing to smooth the images before distinguish skin from non-skin pixels, by making a comparison with a skin color model in HIS color space. In [4] and [5] skin/background image segmentation is obtained by generating a skin color model in a combination of color spaces. Human skin has a specific luminance texture and young people skin is usually smooth. Then, to better detect human skin in digital images, in [3] a Gaussian Mixture Model method are employed to model the texture of human skin and to reject all human skin-colored blocks whose texture features do not fit the skin model. Cao et al. [5] adopt instead co-occurence matrix standard

A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 389–398, 2013.
© Springer-Verlag Berlin Heidelberg 2013

feature to measure texture. Once skin pixels have been correctly detected, the percentage of skin pixels can be used as the criteria for detecting if the image under test contains nude bodies. Agbinya et al. [2] assert that if skin colored pixels are more than 20% of entire image, probably it depicts a human body. According to Marcial-Basilio et al. [6], if skin coloured pixels dominate other pixels, the photo is ranked as porn. However, by using this approch most of mug shots can be wrongly classified as porn content [5].

More complex approaches adopt neural networks to discriminate adult images from innocuous one. Total skin area percentage and largest connected skin area percentage are used as useful local features [3,5]. Ruiz et al. [7] propose a content-based method employing a large skin bulk layout. Local measures of the skin blobs (area, orientation, eccentricity, position and average colour of the three largest skin regions) are extracted and then fed into a Support Vector Machine. Other classification methods take into account image-based features like colour and shapes. According with these proposals, nude images can be detected by analyzing colour and spatial information of the entire image by using color histogram and color coherence vector [3]. To effectively describe image contents, the author of [8] extract edge and texture global features, by using, respectively, Edge Directional and Local Binary Pattern histograms.

Differently from the above cited methods, in this paper we propose to detect nude images by analyzing both body geometric properties and global visual features and by exploiting a multiple classifer approach. Features coming from the layout of connected skin regions (provided by a novel segmentation algorithm), as well as global features about colour, texture and shapes of the entire image are fed into different neural network classifiers. Decisions are then combined by using a weighted majority voting approach, in order to establish whether the image under test depicts naked human bodies or not.

The rest of the paper is as follows. In Section 2 the proposed system architecture is presented. In particular, in Subsection 2.1 a novel skin/background segmentation method based on pixel classification is first described, then the set of local features of the skin colored regions is reported. In Subsection 2.2, instead, the proposed global features extracted from the entire image are described. The used dataset and experimental results on both high and low resolution images are reported in Section 3, together with a comparison with some state-of-the art proposals. Some conclusions are eventually drawn is Section 4.

2 System Architecture

Image contents can be seen as a set of objects disposed over the background. We can obtain a characterization of the image contents by detecting objects and their mutual position. In our context, relevant elements in the image are human bodies. An erotic/pornographic photography portrays one or more naked people whose bodies occupy a considerable part of the image. By analyzing human limbs layout it is then possible to establish whether image depicts naked or scantily dressed people. On the other hand, image content analysis can be also performed by gathering image visual information like color, texture and shape.

Therefore, the proposed system analyzes both human body geometric properties (from now on, *local* features) and global visual features: since the considered features are quite complementary each other, we proposed to combine them by using a multiple classifier approach [9]. More in details, local geometrical features are extracted from the skin/background binary map of the image, while global features are obtained from color or grayscale-converted digital images. To perform skin/background segmentation each pixel is classified based on its color by a neural network classifier. To reduce misclassified pixels, the obtained binary map is further processed by using suitable image processing techniques. Starting from this map, a set of nineteen features, carefully selected among those proposed so far in the literature, have been extracted. As regards global features, three different sets of features have been considered which exploit respectively color, texture and shape information of the image under test. In this way, as depicted in Figure 1, we trained four binary classifiers whose decisions were combined in a multiple classifier system, by using a weighted majority voting approach. Classifiers' weights are calculated as suggested in [9].

In the following the above cited feature sets are descibed in details.

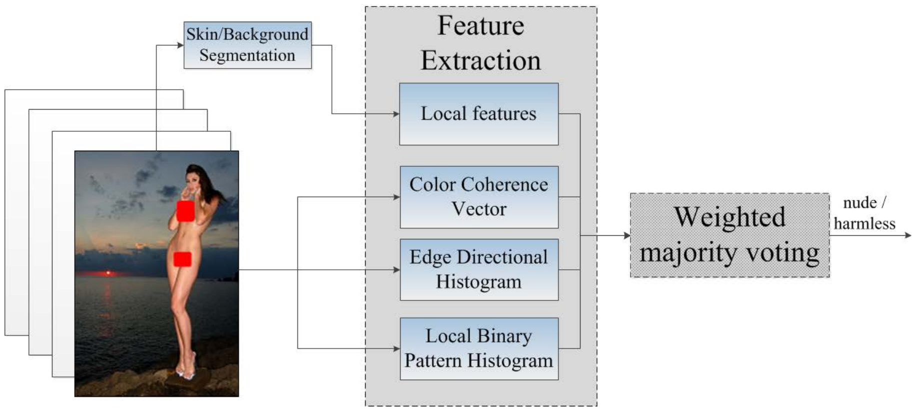

Fig. 1. System Architecture

2.1 Local Features Extracted from the Body Layout

Skin Detection: we proposed a novel approach for detecting skin pixels. Each pixel in the image is first classified based on its color in order to determine whether is part of a human body. In particular, each pixel is described by a feature vector made up of 6 components which are the values of the pixel itself in RGB e HSV colour spaces. A Multi-Layer Perceptron (MLP) is used as classifer. Binary classification of all the pixels produces a skin/background segmentation mask. In order to improve the results coming from the MLP classifier (see Fig. 2a), several image processing techniques are applied on the skin/background mask. Light incidence on human skin can strongly alter colours perceived by image sensors and little shadowed or overexposed pixels could be incorrectly classified as background. Therefore, we first applied a region-growing algorithm to reduce misclassified skin-pixels (false negatives) on the edges of body-corresponding area (see Fig. 2b). A skin region is grown with background-classified pixels that satisfy the chromatic condition in (1). A background pixel x, whose $CbCr$ components are x_{Cb} and x_{Cr} , is reclassified if body-region pixels $y_1, y_2, ..., y_n$ show a similar color:

$$\left| x_{Cb} - \frac{1}{n}\sum_{i=1}^{n} y_{iCb} \right| \leq 1 \wedge \left| x_{Cr} - \frac{1}{n}\sum_{i=1}^{n} y_{iCr} \right| \leq 1 \tag{1}$$

Moreover, some natural areas like rocks, desert, sand or wood often show typical skin-like color but are characterized by a different texture. Skin is typically smooth, especially for young people. Therefore, skin areas show homogeneous color, whereas rocky surfaces, tree trunk, desert sands are composed by a rapid alternation of brighter and darker pixels due to inhomogeneous light incidence on a rough surface. These skin-colored zones correspond to sparse aggregate of skin-classified pixels in skin/background mask. In this binary image, typically, human limbs are made up of large connected aggregate of skin-classified pixel so we can reduce color misclassification by analyzing the skin region size. A connected area of skin-colored pixel is then assigned to background class if its area is less than 5% of skin-area in the image. Skin-area is the number of skin-classified pixel in the entire image. Denoting with $R_1, R_2, ..., R_n$ skin regions and with $A_1, A_2, ..., A_n$ their areas, pixels belonging to region R_i are considered as background (see Fig. 2c) if :

$$A_i \leq 0.05 \sum_{j=1}^{n} A_j \tag{2}$$

Human limbs in skin/background mask correspond to compact skin regions. An hole in a connected skin-classified region corresponds to suits or to overexposed or shadowed skin pixels. We want to reduce the number of false negatives internal to a skin region. False negatives completely surrounded by skin-classified pixels are named *skin-holes*. We can decide to fill all holes internal to a large skin region: in this case, however, we could run the risk of misclassifying suits as skin region. Usually, overexposed or shadowed skin zones are not much wide. Therefore, we just reclassify internal aggregate of background-classified pixels whose area is less then 3% of the skin-area (Fig. 2d). Finally, the remaining skin-holes are selectively filled. In order to do that, we adopt another segmentation method to detect body pixels. According to [10] human skin colour model can be defined in RGB colour space such as following:

$$\begin{cases} R > 95 \wedge G > 40 \wedge B > 20 \\ max(R, G, B) - min(R, G, B) > 15 \\ R - G > 15 \wedge R > G \wedge R > B \end{cases}$$

By using this approach, we obtain another skin/background binary map (Fig. 2e - note that it gives rise to several false positives if used alone) to be used for closing *skin-holes* provided by our approach. In fact, only the remaining *skin-holes* of our mask that correspond to skin pixels of that RGB-based map are filled and removed from the background pixels. The final result is shown in Fig. 2f.

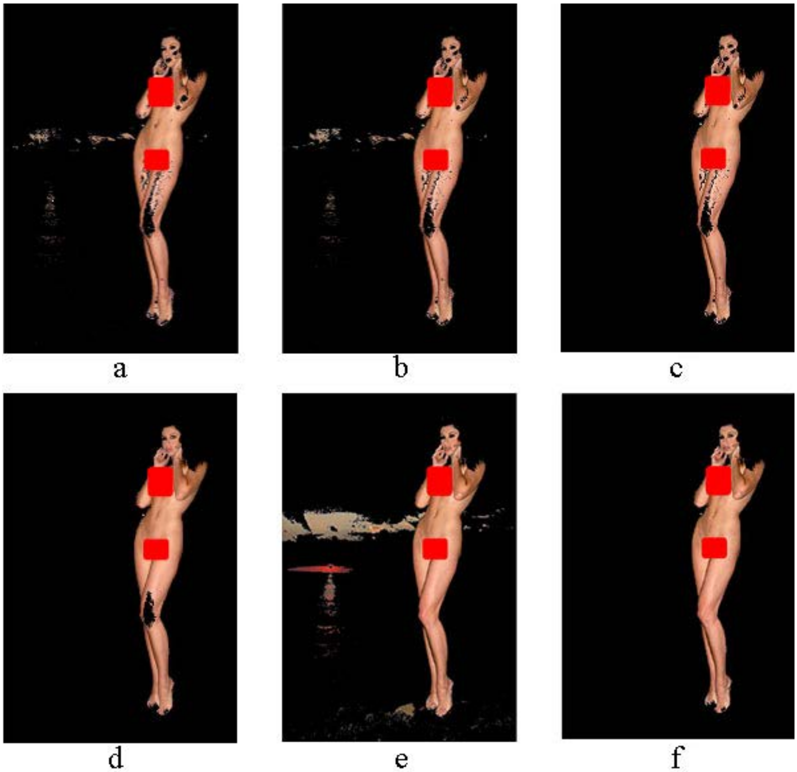

Fig. 2. Skin/background map (a) coming from the MLP classifier for the image of Fig. 1; map after the region-growing step (b); map after the smallest regions clearing (c); map after the small holes filling (d); the RGB-based skin/background map (e) given by [10]; the final skin/background map (f) provided by our approach

Local Feature Extraction: by comparing the skin/background masks coming from nude and harmless images we can assert that: *i*) clothes divide a skin region in several compact parts - therefore a naked body is made up of very few compact and wide skin regions; *ii*) in a nude image human body is often near to the image centre; and *iii*) nude images contain uncovered human parts whose orientation matches the main orientation of the image. Starting from these considerations and after a careful analysis of the literature, we have chosen 19 geometrical features that can be seen as clues of nudity and can discriminate between naked and suited people. Two of them are proposed in this paper for the first time. Their description and the corresponding motivations are reported in Tab. 1.

2.2 Global Features

In order to describe the image content at a global level we considered colour, edge and textural features.

Color Features: a colour histogram H represents the distribution of colours in a digital image. Colour Histograms are widely used in image-content retrieval but they lack of spatial information. Colour Coherence Vector [3] (CCV) is a colour histogram that provides spatial information of colour similar pixels. Spatial information is taken into account by using the coherence: a pixel is coherent if it belongs to a connected as well as uniformly coloured region, whose area exceeds $M \times N \times 0.01$, where M and N are the image dimensions. In order to discriminate similar coloured pixels into coherent and incoherent, we considered a CCV vector of 1024 dimensions.

Shape Features: a global representation of shapes in an image can be obtained from a directional histogram. We consider 4 main directions (horizontal, vertical and two oblique). An image is converted into grayscale and then subdivided in 16 macro blocks. By chaining local directional histogram of 16 macro blocks, we obtain an 80-dimensional histogram. Each Edge Directional Histogram (EDH) [11] is computed by counting the number of sub-blocks in a macro block that match one of the directions. Each of these sub-blocks is convolved with five filters (one for each direction and another one isotropic). We assign to a block the direction of the maximum response filter.

Textural Features: an object can be characterized by its material. Each material shows a peculiar texture, e.g., visible fibers for wood. Textures can be discriminated by using the Local Binary Pattern (LBP) operator that derives from a general definition of texture in local neighborhood [8]. LBP histogram is then computed and normalized by dividing for $M \times N$, so obtaining a feature vector of 256 dimensions.

3 Experimental Results

Our dataset is made up of 1023 color images (min resolution: 0.75 MPixels - max resolution: 9.5 MPixels) downloaded from Internet. It comprises 698 harmless images and 325 nude images of Caucasian, African and Asian models. Nude photos have been captured both in outdoor and indoor context. Within the harmless partition, 231 images contain completely-covered human beings, while other 218 depict human bodies wearing swimsuit or underwear. Finally, 249 images contain subjects like animals, cars, artificial and natural landscapes. Details of the used dataset are summarized in Tab. 2. We also considered a low-resolution version (min resolution: 31K - max resolution 65K) of the entire dataset, obtained by scaling to 256 pixels the maximum dimension of high-resolution images. In this way, we can test the robustness of the proposed approach with respect to different image resolutions. In both cases, 25% of the whole dataset was used as training set, while remaining images are used as test set.

Table 1. The description of the local features used in this work

Feature description	Motivation	Referred paper
Number of connected skin regions	In a nude image there are few compact skin regions	[12]
Wider region area / entire image area		[7]
Second widest region area / entire image area		[7]
Third widest region area / entire image area		[7]
Fourth widest region area / entire image area		This paper
Fifth widest region area / entire image area		This paper
Number of skin pixels in the widest region bounding box / bounding box area	A naked body part corresponds to a compact skin region	[13]
Number of skin pixels in the widest region / filled region area		[7]
Number of skin pixels in the second widest region / filled region area		[7]
Number of skin pixel in the third widest region / filled region area		[7]
The proportion of the number of skin colour pixels in the external rectangular region and the total number of pixels of the external rectangular region		[11]
Normalized horizontal and vertical distances from the widest region gravity centre	A naked body is often near to image centre	[14]
Widest region equivalent aspect ratio	Limbs are elongated	[14]
Eccentricity of the best ellipse that contains the widest region		[7]
Eccentricity of the best ellipse that contains the second widest region		[7]
Eccentricity of the best ellipse that contains the third widest region		[7]
Result of subtracting the orientation of the best ellipse that contains the widest region and the image orientation	Nude images contain uncovered human parts whose orientation matches the main orientation of the image	[7]
Result of subtracting the orientation of the best ellipse that contains the second widest region and the image orientation		[7]
Result of subtracting the orientation of the best ellipse that contains the third widest region and the image orientation		[7]

To extract the skin/background mask from color images, the first step of our approach (as stated in Section 2.1) considers the use of an MLP classifier. In order to train it, a training set made up of 500,000 manually selected pixels (11% of them were skin pixels) has been built. We have chosen skin pixels from Caucasian, African and Asian models' bodies. Background pixels have been extracted instead in random position from harmless images with no human bodies.

Table 2. The used dataset

Category	Subcategory	Ethnicity	# of images
Nude	Female	Asian	90
		Caucasian	116
		African	106
	Male		13
Total			**325**
Harmless	Semi-undressed Female	All	**216**
		Asian	82
		Caucasian	54
		African	80
	Covered Female	All	**223**
		Asian	76
		Caucasian	61
		African	86
	Semi-undressed Male		2
	Covered Male		8
	No bodies	–	**249**
Total			**698**

As described in Section 2, four binary MLP classifiers have been used to discriminate images that depict naked people from others, by employing the feature sets described in the same Section. The MLP parameters have been optimized by performing a 10-fold cross validation on the training set.

Table 3 report the results of the MLP classifiers on the test set, as well as those obtained by the proposed Multiple Classifier System (MCS), in terms of sensitivity, specificity and F-Measure. Both high- and low-resolution images have been considered. As it can be seen from Tab. 3, the proposed MCS outperforms all the single classifiers in terms of F-measure, for both high- and low-resolution images. In particular, it filters out 75% of nude images (77% in case of low-resolution images). Moreover, 91% of the images without human beings (92.5% in case of low-resolution images) are correctly classified. Finally, as it could be expected, our system classifies harmless images of completely-covered bodies better than semi-undressed bodies because of the minor percentage of skin pixels.

In order to verify the goodness of the chosen combining rule, we have compared the performance of the weighted voting with other rules, such as the simple majority voting, BKS and Sum Rule [9]. Results reported in Tab. 4 confirmed that the proposed approach is the most suited one for the problem at hand.

Table 3. Results obtained by the single classifiers and the proposed system

	High resolution images			Low resolution images		
	Sensitivity	Specificity	F-measure	Sensitivity	Specificity	F-measure
Local	0.62	**0.78**	0.69	0.67	**0.73**	0.70
CCV	**0.76**	0.63	0.69	0.75	0.63	0.68
LBP	0.70	0.63	0.66	0.70	0.58	0.63
EDH	0.62	0.52	0.56	0.54	0.51	0.52
Weighted voting	0.76	0.73	**0.74**	**0.77**	0.69	**0.73**

Table 4. Comparing different combining approaches

	High resolution images			Low resolution images		
Combining rule	Sensitivity	Specificity	F-measure	Sensitivity	Specificity	F-measure
Weighted voting	0.76	0.73	**0.74**	0.77	0.69	**0.73**
Majority voting	0.60	**0.80**	0.68	0.57	**0.78**	0.66
BKS	0.65	0.77	0.71	0.80	0.56	0.66
Sum	**0.80**	0.64	0.71	**0.79**	0.61	0.69

Finally, Tab. 5 shows a comparison among our method and other state-of-the-art algorithms. Once again, our approach gave rise to the best results in terms of F-measure. It demonstrated also to be more independent of the image resolution and guaranteed a better balancing between specificity and sensitivity.

Table 5. Comparison of the proposed approach with other methods

	High resolution images			Low resolution images		
Method	Sensitivity	Specificity	F-measure	Sensitivity	Specificity	F-measure
This paper	0.76	0.73	**0.74**	0.77	0.69	**0.73**
Wang [3]	0.84	0.64	0.72	**0.87**	0.47	0.61
Marcial-Basilio [6]	**0.86**	0.50	0.63	0.86	0.50	0.64
Ruiz del Solar [7]	0.65	**0.78**	0.71	0.69	**0.71**	0.70

4 Conclusions and Future Works

In this paper we proposed a system, based on a multiple classifier approach, for detecting naked human bodies in images. It uses both features extracted from skin pixels obtained by a novel skin segmentation algorithm, and global visual features. A comparison with other state-of-the-art proposals demonstrated the effectiveness of the proposed approach.

As future works, we have planned to test the proposed approach on a wider set of images and to extend it in order to be able to process videos, too.

References

1. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys 40(2), 1–60 (2008)

2. Agbinya, J., Lok, B., Wong, Y.S., Da Silva, S.: Automatic Online Porn Detection and Tracking. In: Faculty of Engineering, University of Technology, pp. 1–5 (2007)
3. Wang, S.-l., Hui, H., Li, S.-H., Zhang, H., Shi, Y.-Y., Qu, W.-T.: Exploring content-based and image-based feature for nude image detection. In: Wang, L., Jin, Y. (eds.) FSKD 2005. LNCS (LNAI), vol. 3614, pp. 324–328. Springer, Heidelberg (2005)
4. Duan, L., Cui, G., Gao, W., Zhang, H.: Adult Image Detection Method Base-on Skin Color Model and Support Vector Machine. In: 5th Asian Conference on Computer Vision, Australia, pp. 1–4 (2002)
5. Cao, L.L., Li, X.-L., Yu, N.H., Liu, Z.K.: Naked people retrieval based on adaboost learning. In: 1st Int. Conf. on Machine Learning and Cybernetics, pp. 1133–1138 (2002)
6. Marcial-Basilio, J.A., Aguilar-Torres, G., Sanchez-Perez, G., Karina Toscano-Medina, L., Perez-Meana, H.M.: Detection of Pornographic Digital Images. International Journal of Computers 5(2), 298–305 (2011)
7. Ruiz del Solar, J., Castaneda, V., Verschae, R., Baeza-Yates, R., Ortiz, F.: Characterizing objectionable image content (pornography and nude images) of specific web segments: Chile as a case study. In: 3rd Latin American Web Congress, pp. 269–278 (2005)
8. Zhao, Z., Cai, A.: Combining multiple SVM classifiers for adult image recognition. In: 2nd IEEE International Conference on Network Infrastructure and Digital Content, pp. 149–153 (2010)
9. Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. Wiley-Interscience (2004)
10. Boirouga, H., El Fkihi, S., Jilbab, A., Aboutajdine, D.: Skin detection in pornographic videos using threshold technique. Journal of Theoretical and Applied Information Technology 35(1), 7–19 (2012)
11. Ap-apid, R.: An algorithm for nudity detection. In: 5th Philippine Computing Science Congress, pp. 201–205 (2005)
12. Wang, X., Li, X., Liu, X.: Nude image detection based on SVM. In: Int. Conference on Computational Intelligence and Natural Computing, pp. 178–181 (2009)
13. Zheng, Q.-F., Zeng, W., Gao, W., Wang, W.-Q.: Shape-based adult images detection. International Journal of Image Graphics 6(1), 115–124 (2006)
14. 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 40(8), 2261–2270 (2007)