Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 3164 – 3174

doi:10.3906/elk-1607-332

Multiclass semantic segmentation of faces using CRFs

Khalil KHAN[1,*], **Nasir AHMAD**[2], **Khalil ULLAH**[3], **Irfanud DIN**[4]
[1]Department of Electrical Engineering, University of Poonch, Rawlakot, Pakistan
[2]Department of Computer Engineering, University of Engineering & Technology, Peshawar, Pakistan
[3]Department of Electrical Engineering, National University of Computer & Emerging Sciences, Peshawar, Pakistan
[4]Department of Information Engineering, Inha University, Tashkent, Uzbekistan

Received: 30.07.2016 • **Accepted/Published Online:** 23.12.2016 • **Final Version:** 30.07.2017

Abstract: Multiclass semantic image segmentation is widely used in a variety of computer vision tasks, such as object segmentation and complex scene understanding. As it decomposes an image into semantically relevant regions, it can be applied in segmentation of face images. In this paper, an algorithm based on multiclass semantic segmentation of faces is proposed using conditional random fields. In the proposed model, each node corresponds to a superpixel, while the neighboring superpixels are connected to nodes through edges. Unlike previous approaches, which rely on three or four classes, the label set is extended here to six classes, i.e. hair, eyes, nose, mouth, skin, and background. The proposed framework is evaluated on standard face databases FASSEG, FIGARO, and LFW. Experimental results reveal that the performance of the proposed model is comparable with state-of-the-art techniques on these standard databases.

Key words: Multiclass face segmentation, conditional random fields, feature extraction, classification

1. Introduction

Face segmentation is useful in many facial applications of computer vision, such as estimation of gender, expression, age, and ethnicity. Multiclass face segmentation is used as a front-end for the estimation of all midlevel vision features for these applications. In the recent years, face segmentation techniques have attracted much attention with the development of many new algorithms [1–3]. Notable factors influencing face segmentation are variations in lighting conditions, facial expressions, face orientation, occlusion, and image resolution. These and many more factors make the development of an efficient segmentation algorithm a challenging task.

Many researchers around the world have solved many complicated problems of segmentation using the idea of semantic segmentation. Extensive research work has been carried out to investigate the problem with a major contribution from the PASCAL VOC challenge [4].

Huang et al. [5] tackled the joint study of face segmentation and pose estimation. The authors have suggested that high-level features such as pose, gender, and expression can be predicted easily based on labeling face image into hair, skin, and background. They proved that such segmentation provided useful information for the estimation of pose. Experiments were performed on a small database of 100 images. They worked on three simple poses, i.e. left profile, right profile, and portrait.

The relationship between face parts and pose is well established from psychology literature as well [6].

*Correspondence: e.khalilkhan@gmail.com

Moreover, there is compelling evidence that facial features provide useful information for the human visual system to recognize identity [7,8]. Hair modeling, synthesis, and animation are already active research topics in computer graphics [9,10]. Research on face-processing applications, such as virtual make-up

[11], skin color beautification [12], and skin smoothing [13] have also been reported. All these applications require precise knowledge for each face segment at the pixel level. We argue that the proposed framework is a better solution compared to the state-of-the-art options for all these applications.

In this paper, an algorithm for face segmentation using the idea of semantic segmentation and CRFs has been developed. This work is based on previous research wherein a new method of face segmentation was introduced, called multiclass face segmentation (MFS) [14]. In previous work, the problem of face segmentation was thoroughly investigated using a small database of high-resolution frontal images. A built model returns a class label and probability value for each pixel. The present work is an extension of the MFS work and tries to cover the main weaknesses of MFS. Unlike the previous work, here experiments are performed on a large database of low-resolution images. Manual labeling of the face segments is performed with an excellent manual labeling tool. One of the main problems of MFS is the processing time. To solve the speed problem, the pipe-line was integrated with the superpixel segmentation algorithm. Similarly, a conditional hierarchy for various face segments is added to the proposed new framework.

2. Related work

A number of models for face parts segmentation and face labeling have been proposed in the literature. Yacoob and Davis [15] addressed problem of hair labeling. The authors adapted a region-growing algorithm by building a Gaussian mixture model (GMM). They compared the appearance of different people's hair using their model. However, the performance of their proposed method was affected badly when faced with significant changes in hair color. The GMM model was further extended by Lee et al. [11]. Their algorithm segmented an image of a face into background and hair regions. They also contributed a database of 150 manually labeled images (hair, face, and skin). A superpixel-based CRF [15] was introduced by Huang et al. [5]. They trained standard CRFs on images taken outside of the laboratory to provide facial image labels for hair, skin, and background. Kae et al. [16] combined the strength of CRFs and shape Boltzmann machine [17], introducing a new model named GLOC (GLObal and LOCal). They claimed that this hybrid model produces better results than CRFs alone.

Yali et al. [18] focused mainly on the hair style representation and its segmentation from facial regions. Scheffler and Jean [19] studied the segmentation of hair, skin, background, and clothing. Local label consistency was enhanced by the combination of CRFs and spatial prior of each label. Matteo et al. [1] introduced a multiclassifier approach for face segmentation. They exploited color and texture information to partition a face image into four classes (skin, hair, clothes, and background). Their study focused on the adaptation of the proposed technique in electronic identity documents.

A deep-learning–based face labeling method was proposed by Luo et al. [2]. They combined several trained models separately, in which facial parts are labeled only. The method proposed by the authors does not provide complete face labeling. Liu et al. [3] proposed a deep convolution network that models likelihoods (pixelwise) and label dependencies through an objective learning method called multiobjective through GraphCut. The framework proposed in this method uses a single deep convolutional network. Two nonstructured loss functions were used: the first one encodes the label likelihoods and the second one encodes label dependencies. To the best of our knowledge, this is the latest proposed method providing face labeling to date.

Differently from all the mentioned approaches, MFS is a new method for face segmentation that extends the label set into six semantic classes. A dataset of 70 manually labeled images was built and made publicly available. A new model was trained using the extracted features. The best possible configuration was investigated by changing various parameters and spatial setting in those experiments. MFS faced three major problems during experiments. Firstly, we did not include any kind of conditional hierarchy or global modeling of face regions in the framework. In the proposed MSS-CRFs model, we included a conditional hierarchy for six facial regions, which boosted the performance of the whole framework. Secondly, MFS processing time is very long, due to labeling each pixel individually. MSS-CRFs uses superpixels, which reduces the processing time of a testing image. Lastly, the testing set of the MFS is only 70 images (high-resolution frontal images), out of which 20 were used for training and 50 for testing. Along with MFS comparison, we also performed experiments on three other datasets: FASSEG V-4, FIGARO [20], and LFW [3]. FASSEG V-4 consists of low-resolution front-facing images taken from the Pointing'04 [21] and SliblingDB [22] databases, with an image dataset of 182 images.

3. Proposed face segmentation model

MSF divides a given image into patches with a fixed step size. After patch creation, features are extracted from each patch. Using the extracted features, a random decision classifier is trained and tested. This method does not consider any conditional hierarchies, such as the locations of various face parts and their relationship with each other. For example, it is very unlikely to have a mouth region near an eye region. Unlike MSF, we formulate a CRF model that couples labels of face parts in a scale hierarchy. Another serious problem with MSF is speed, since labeling each pixel within an image takes a long time. Instead of labeling each pixel individually, a given image is first divided into superpixels. All pixels within the superpixel get the same class label and as a result, the processing time of the framework is reduced.

The presentation of the proposed algorithm is divided into two parts: feature extraction is presented in subsection 3.1 and segmentation via CRF and energy optimization is explained in subsection 3.2.

3.1. Feature extraction methods

The superpixel algorithm oversegments an image by grouping pixels into small meaningful patches that belong to the same object. Instead of using just pixels, many image-processing applications benefit from working with superpixels. The number of entities to be labeled in semantic segmentation is reduced immensely by superpixels. Each superpixel has multiple visual features. A single image is represented by multiple visual feature spaces after segmentation. We used the SEEDS [23] algorithm to oversegment an image into superpixels. SEEDS is faster than previously proposed superpixel segmentation methods [23]. Moreover, according to standard error metrics, the quality of superpixel segmentation in SEEDS is also higher than SLIC and other methods [23]. The main problem with the previously proposed MFS method is speed; hence, SEEDS is the best choice in our experiments.

To determine the optimal amount of superpixels, we did a large number of experiments. During these experiments, we noted better results with 700 superpixels. The actual number of superpixels is of course smaller than this due to certain restrictions. The actual number of superpixels depends on the image size and the number of block levels used in the superpixel extraction process. The number of block levels defines the blocks that the algorithm uses in the optimization process. If the number of levels is increased, the superpixel segmentation is more accurate, but this results in more memory and time consumption by the CPU. The SEEDS parameters we set are block levels = 3 and histogram bins = 5; each block level is iterated twice for better accuracy.

For node features, we use three different feature-extraction methods: color, shape, and spatial information. Different parameterization and settings for features are explored to find the best possible configuration. We investigated these parameters in our previous work with MFS.

For spatial information, the relative location of the center pixel of each patch is used as a feature. Relative location of a pixel at position $(x,\ y)$ is defined as $f_{loc} = [x\ /\ \text{W},\ y\ /\ \text{H}] \in R^2$; where W is the width and H is the height of the image.

An HSV color histogram is adapted as color features. All three values in HSV (hue, saturation, and variance) are concatenated to form a single feature vector. Patch dimension of 16×16 ($D_{HSV} = 16 \times 16$) is used with 32 bins ($N_{bins} = 32$). Using these values, each patch generated a feature vector $F_{HSV} \in R^{96}$ for color information.

To account for shape features, the widely used HOG [24] is utilized. The dimension of the patch for extracting HOG is kept at 64×88 ($D_{HOG} = 64 \times 88$). With this dimension, a feature vector $f_{HOG} \in R^{2520}$ was produced.

Spatial, color, and shape feature vectors were concatenated to form a single feature vector $f \in R^{2618}$.

3.2. Multiclass segmentation via CRFs and energy optimization

To estimate face segments, we use CRFs. The proposed CRF model encodes the probability of segmentation S with image features Z. S is represented by $S = \{\ s_1,...,s_m\}$, where m is the total number of super-pixels in the image. s_i can take one of the six values corresponding to "mouth," "eyes," "background," "nose," "hair," and "skin." Z consists of node features Z^m and edge features Z^e. We compute F_m features for the ith superpixel and so Z_i^m is a vector having length F_m. For pairing neighboring superpixels $i,\ j$ we compute F_e features, resulting in a single vector $Z_{i,j}^e$ having length F_e.

Now the log linear CRFs model developed will have node energies $\psi(s_i, Z_i^m)$ and edge energies $\psi(s_i,\ s_j,\ Z_{i,j}^m)$. Both of these quantities can be represented as follows:

$$\psi\left(s_i = l, Z_i^m\right) = \sum\nolimits_{f=1}^{F_m} \left(X_l^m\right)_f \left(Z_i^m\right)_f$$

$$\psi\left(s_i = l_1, s_j = l_2, Z_{i,j}^e\right) = \sum\nolimits_{f\,=\,1}^{F_e} \left(X_{l_1,l_2}^e\right)_f \left(Z_{i,j}^e\right)_f$$

where a set of node weights is represented by X^m and edge weights X^e for each label $l$and pair of labels (l_1, l_2), respectively.

Now the probability of S if Z is given will be

$$p\left(S \setminus Z\right) = \frac{exp?\left(-\sum_{i\,=\,1}^{m} \psi\left(s_i, Z_i^m\right) - \sum_{i,j} \psi\left(s_i, s_j, Z_{i,j}^m\right)\right)}{N\left(Z\right)}$$

The second sum in the above equation is for neighboring superpixels and $N(Z)$ is the partition function used to normalize the distribution.

For the partition function, log likelihood through the Bethe approximation [25] is used. Similarly, for the marginal approximation of each $s_{i,}$ loopy belief propagation is used. Gaussian is added first to regulate weights. For estimating segmentation, loopy belief propagation was used to find the maximum posterior marginals. To evaluate the labeling accuracy of the segmentation estimates, an L1 error was applied to each segmentation estimate. This way, each superpixel was penalized according to the difference between probability of correct label and probability value 1.0. For example, if the estimated superpixel had a probability of 0.7 being skin, and was in fact skin, a penalty of 0.3 would be incurred as a result.

4. Experimental results and discussion

The only dataset available for six classes is FASSEG [14]. FASSEG is available in four different versions. It can be downloaded from http://khalilkhan.net/face-segmentation-dataset/. FASSEG V-2 contains high-resolution front-facing portrait images with low variability. FASSEG V-4 contains low-resolution images with variability factors such as beards, moustaches, and glasses. We performed our experiments with FASSEG V-2 and FASSEG V-4. The promising results show that the proposed model is capable of segmenting facial parts successfully from facial images.

Some of the images segmented with proposed MSS-CRFs model are shown in Figure 1. Images shown in Figure 1 are efficiently segmented into their corresponding face parts. However, in some cases, the segmentation results of the proposed MSS-CRFs algorithm are comparatively poor. Figure 2 show images from the database with poor results. The testing image shown in row 1 is a case where the face passed to the framework is not compatible with the training data images. The nose, eyes, and eyebrows are more concentrated to the upper part of the image. As a result, segmentation results of the eyes and eyebrows are very poor. If a testing image has glasses, there are problems in segmentation with nose, eyes, and eyebrows, specifically (testing image in row 2). Similarly, if a testing image has a beard or moustache, there are also segmentation problems (testing image in row 3). The proposed framework is unable to segment face parts such as moustaches and beards.

In the following paragraphs the results obtained during experiments using FASSEG V-2, FASSEG V- 4, FIGARO, and LFW databases are presented.

4.1. Face segmentation V-2

FASSEGV-2 contains 70 images. This version of the database was used with the MFS. Figure 3 shows a comparison of the MFS and the proposed MSS-CRFs results. From Figure 3, it is clear that there is improved pixel-labeling accuracy (PLA) for all classes with the proposed method.

MFS performance was not poor in the majority classes (hair, background, and skin); however, results for the minority classes (eyes, nose, and mouth) were not satisfactory. The main target in the present work was improving PLA of the rare and difficult classes. The most advantageous classes in MSS-CRFs are the eyes, nose, and mouth regions. PLA of the minor classes increased in the present work for two reasons. First, manual labeling was not performed properly in MFS. Rare classes were not properly labeled due to their complex shapes. Due to their limited area in the whole face image, training data for these classes were not provided properly. Here manual labeling was performed with extreme care, using Photoshop, particularly for the nose. Just the tip of the nose was labeled previously in MFS, but labeling followed a different convention here, i.e. extending the nose label to the midpoint of the two eyes. As a result, PLA of the nose jumped from 29.83% to 68.97% (Figure 3). Secondly, MFS does not consider any conditional hierarchy with respect to various face parts; previously, these minor classes were mostly misclassified as majority classes.

Moreover, the processing time for single-image segmentation is reduced with proposed method. A substantial increase in speed by an order of magnitude is obtained by using superpixels, since the number of patches to be classified by the model is greatly reduced. In MFS, a class label is individually provided for each pixel, while MSS-CRF assigns a class label to superpixels only. All pixels within the superpixel then get

Figure 1. Images from the FASSEG V-4 database. First column shows original RGB images, second column shows ground truth images, and third column shows results obtained with the proposed MSS-CRF (better segmentation results).

the same class label. A single CPU (2.8 GHz Core i7 and 8 GB RAM) was used, without any GPU or dedicated hardware. A single 520 $\times$ 480 pixel image was divided into superpixels in 1.51 s with SEEDS.

The total framework runs in 2 s in the proposed approach, compared to 49 s in the previously proposed MFS method.

4.2. Face segmentation V-4

Along with FASSEG V-2 images, we added 182 more frontal images to the database. These images were taken from Pointing'04 [21] and SliblingDB [22] databases. The size of the images was kept the same as in MFS (height = 512 and various widths to keep the ratio of the original image). Out of the total images, 20 were taken randomly and used for training. The remaining 152 images were used for testing. Figure 4 shows the confusion matrix for the results obtained for every class. From Figure 4, it is clear that the obtained PLA for all classes except the nose is really impressive.

Figure 2. Images from the FASSEG V-4 database. First column shows original RGB images, second column shows ground truth images, and third column shows results obtained with the proposed MSS-CRF (poor segmentation results).

4.3. FIGARO and LFW-PL databases

Along with the FASSEG database, experiments were conducted with two other databases, FIGARO [20] and LFW [3]. For a fair comparison, the same settings and the same set of images in testing phases as in Svanera et al. [20] and Liu et al. [3] were kept. However, for the training phase, the same images were used as in the experiments conducted in the first phase.

Only the hair class was used in experiments with FIGARO [20]. FIGARO is a comparatively small database with 840 images in total. All these images are collected from web pages. Different variations in hair styles were included, for a total of seven hair classes (straight, curly, wavy, kinky, short-men, braids, dreadlocks).

LFW is a large database with a variety of images. All images in the LFW database are captured in an unconstrained environment where a large number of variations are present due to various environmental factors. Experiments with three classes (hair, skin, and background) were conducted using the LFW database, as in Liu et al. [3]. Reported accuracy for this case is at pixel level for all the three classes.

Figure 5 shows a comparison of the proposed method with FIGARO and LFW database results. From Figure 5, it is clear that we have better results with the FIGARO database. However, the reported results from LFW are lower than previously reported results. All the training images in FASSEG are captured in a controlled lab environment, while the testing images in LFW are from unconstrained conditions. If such variations are included in the training data, better results than state-of-the-art can be obtained with the LFW database as well.

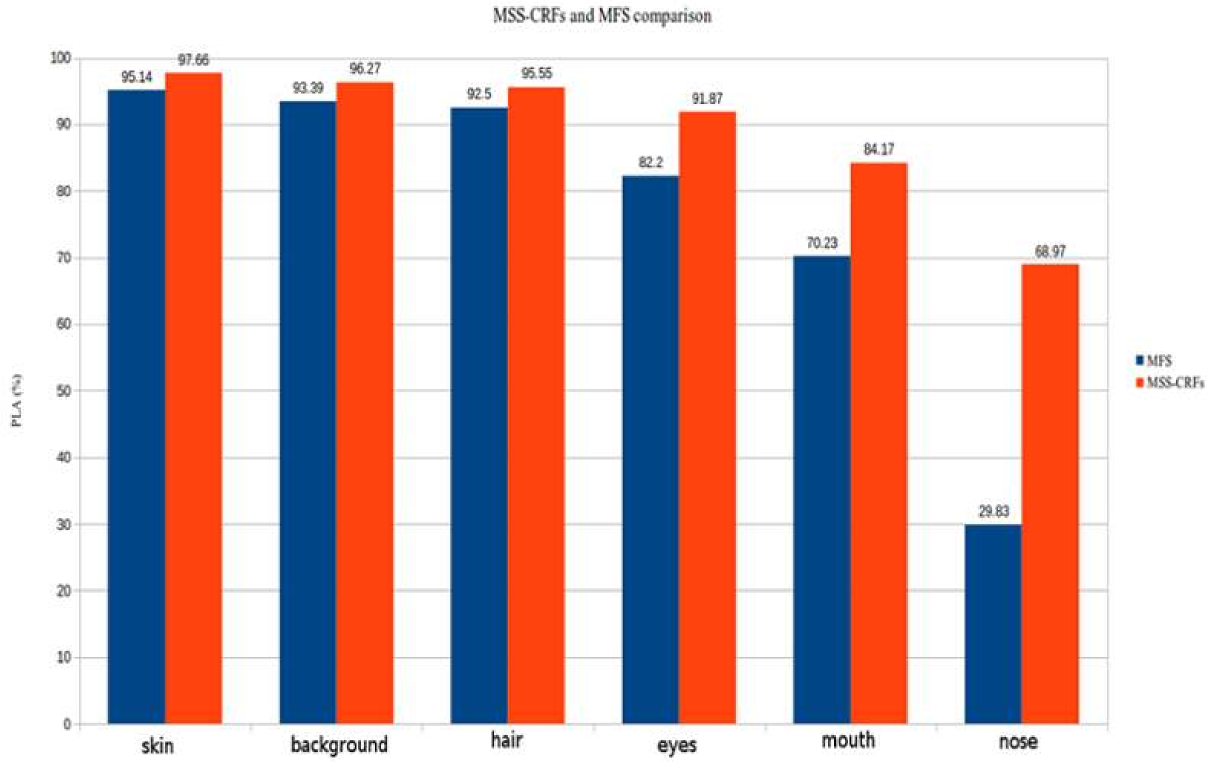

Figure 3. Proposed MSS-CRF and MFS results comparison using the FASSEG V-2 database.

The main advantage of the proposed method is providing class labels for the complete face. Unlike state-of-the-art-methods that only consider a few classes, MSS-CRF provides segmentation of all face parts. Hair segmentation is a comparatively difficult task in previous literature [26,27]. Previously reported methods were not able to segment hair properly due to its complex geometry and larger variability from person to person. However, the reported results for hair are encouraging and confirm the effectiveness of the proposed method. The reported results also show that the proposed method is robust with respect to lighting variations, as some of the images used in the testing phase were captured in uncontrolled lighting conditions. The proposed method provides class labels for all six face parts. In some applications, a class label is needed for a specific part only. In that case, the proposed algorithm can be used according to need and application.

The proposed method has some minor drawbacks as well. While creating the database, the labeling is performed manually by a human. Providing a class label in the transition region between two classes is very uncertain in such conditions. Similarly, patch sampling for the training phase is based on random criteria. However, the number of pixels from the minority classes is insufficient for training. This results in poorer performance than the majority classes, which have sufficient training data. In addition, the proposed framework is unsuitable in cases with beards or moustaches in the images. Providing a separate class label for each of these parts may solve the problem.

5. Conclusion

Semantic segmentation of faces using CRFs is introduced in this paper. Position, HSV color, and shape information are combined to build a CRF model. A great deal of information is provided about the face

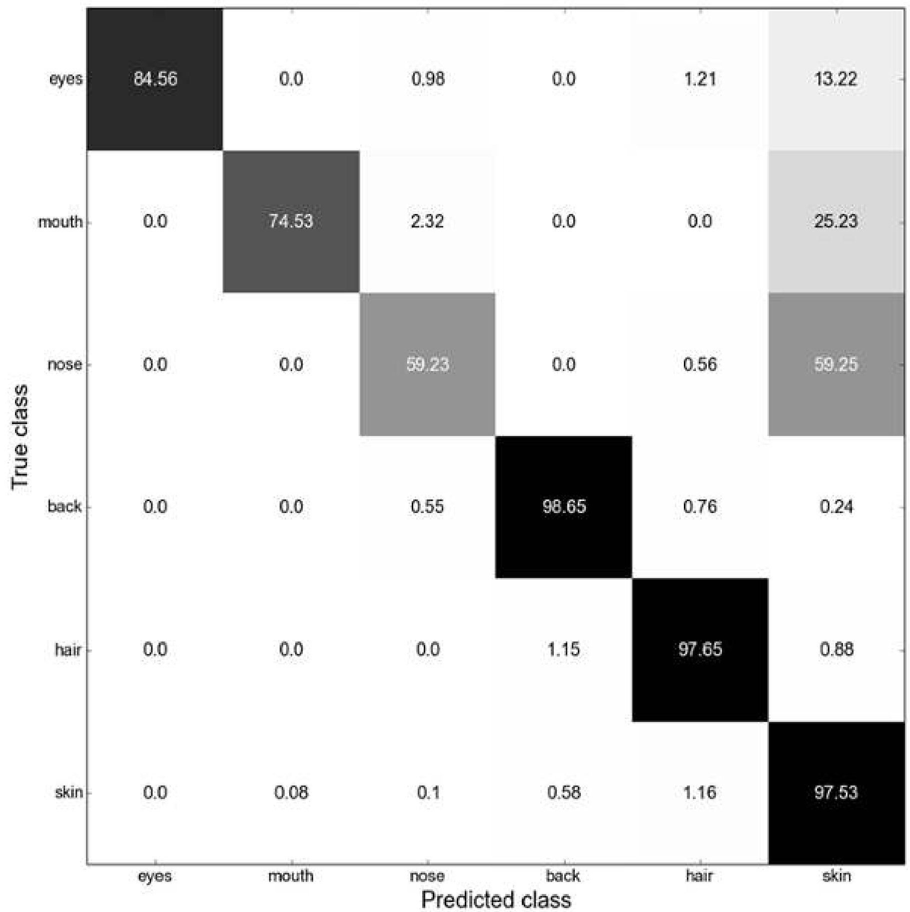

Figure 4. Confusion matrix obtained for all six classes using MSS-CRF and the FASSEG V-4 database.

Database	Method used	Accuracy (Percentage)
FIGARO	Proposed Approach	94.56
FIGARO	FIGARO	86.20
LFW-PL	Proposed Approach	92.47
LFW-PL	MO-GC with prior	95.12
LFW-PL	MO-GC	95.24

Figure 5. MSS-CRF compared with previously reported methods.

parts (skin, hair, nose, eyes, mouth) and background by a CRF estimation model. Experimental results show that the proposed model not only outperforms state-of-the-art results in FASSEG and FIGARO databases, but also improves over the previous results by a large margin.

Future work can be extended in two directions. First, improving the current model to get better pixel-labeling accuracy. A higher level of variability can be added to training and testing data to make the framework suitable for unconstrained conditions. Secondly, applying the current segmentation model to certain midlevel vision feature estimation. Immense sources of information are provided for many hidden variables, such as pose, gender, expression, ethnicity, age, beardedness, and balding.

References

[1] Ferrara M, Franco A, Maio D. A multi-classifier approach to face image segmentation for travel documents. Exper Syst Appl 2012; 9: 8452-8466.

[2] Liu C, Yuen J, Torralba A. Nonparametric scene parsing via label transfer. IEEE T Pattern Anal 2011; 33: 2368-2382.

[3] Liu S, Yang J, Huang C, Yang MH. Multi-objective convolutional learning for face labeling. In: Proc CVPR IEEE; 8–10 June 2015; Boston, MA, USA: IEEE. pp. 3451-3459.

[4] Everingham M, Van GL, Williams CK, Winn J, Zisserman A. The Pascal visual object classes challenge. Int J Comput Vision 2010; 2: 303-338.

[5] Huang GB, Narayana M, Learned-Miller E. Towards unconstrained face recognition. In: Proc CVPR IEEE; 24–26 June 2008; Alaska, USA: IEEE. pp. 1-8.

[6] Zebrowitz LA, Montepare JM. Social psychological face perception: why appearance matters. Soc Personal Psychol Compass 2008; 3: 1497-1517.

[7] Shepherd J, Ellis H, Davies G. Perceiving and Remembering Faces. 1st ed. Chicago, IL, USA: Academic Press, 1981.

[8] Sinha P, Balas B, Ostrovsky Y, Russell R. Face recognition by humans: nineteen results all computer vision researchers should know about. Proc IEEE 2006; 11: 1948-1962.

[9] Moon JT, Marschner SR. Simulating multiple scattering in hair using a photon mapping approach. ACM T Graphic 2006; 25: 1067-1074.

[10] Ward K, Bertails F, Kim TY, Marschner SR, Cani MP, Lin MC. A survey on hair modelling: styling, simulation, and rendering. IEEE T Vis Comput Gr 2007; 2: 213-234.

[11] Lee C, Schramm MT, Boutin M, Allebach JP. An algorithm for automatic skin smoothing in digital portraits. In: IEEE Image Proc; 7–10 September 2009; Cairo, Egypt: IEEE. pp. 3113-3116.

[12] Chen CW, Huang DY, Fuh CS. Automatic skin color beautification. In: Proceedings of the International Conference on Arts and Technology; 10–12 December 2010; Berlin, Germany: IEEE. pp. 157-164.

[13] Xu L, Du Y, Zhang Y. An automatic framework for example-based virtual makeup. In: IEEE Image Proc; 15–18 September 2013; Melbourne, Australia: IEEE. pp. 3206-3210.

[14] Khan K, Mauro M, Leonardi R. Multi-class semantic segmentation of faces. In: IEEE Image Proc; 27–30 September 2015; Québec, Canada: IEEE. pp. 827-831.

[15] Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labelling sequence data. In: Proceedings of the International Conference on Machine Learning; June 28–July 1 2001; Williamstown, MA, USA: pp. 282-289.

[16] Kae A, Sohn K, Lee H, Learned ME. Augmenting CRFs with Boltzmann machine shape priors for image labelling. In: Proc CVPR IEEE; 25–27 June 2013; Portland, OR, USA: IEEE. pp. 2019-2026.

[17] Eslami SA, Heess N, Williams CK, Winn J. The shape Boltzmann machine: a strong model of object shape. Int J Comput Vision 2014; 2: 155-176.

[18] Li Y, Wang S, Ding X. Person-independent head pose estimation based on random forest regression. In: IEEE Image Proc; 26–29 September 2010; Hong Kong, China: IEEE. pp. 1521-1524.

[19] Scheffler C, Odobez JM. Joint adaptive colour modelling and skin, hair and clothing segmentation using coherent probabilistic index maps. In: Proceedings of the British Machine Vision Conference; 29 August–2 September 2011; Dundee, UK: BMVA Press.

[20] Svanera M, Muhammad UR, Leonardi R, Benini S. Figaro, hair detection and segmentation in the wild. In: IEEE Image Proc; 25–28 September 2016; Phoenix, USA: IEEE. pp. 546-550.

[21] Gourier N, Hall D, Crowley JL. Estimating face orientation from robust detection of salient facial features. In: Proceedings of the International Workshop on Visual Observation of Deictic Gestures; 22 August 2004; Cambridge, UK.

[22] Laurentini A, De SM, Bottino AG, Vieira TF. A new problem in face image analysis: finding kinship clues for sibling pairs. In: Proceedings of the International Conference on Pattern Recognition Application and Methods; 6–8 February 2012; Vilamoura, Algarve, Portugal: Springer. pp. 153-162.

[23] Van DB, Michael, Xavier B, Gemma R, Benjamin DC, Luc VG. Seeds: Super-pixels extracted via energy-driven sampling. In: Proceedings of the European Conference on Computer Vision; 7–13 October 2012; Firenze, Italy: Springer. pp. 13-26.

[24] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc CVPR IEEE; 20–26 June 2005; San Diego, CA, USA: IEEE. pp. 886-893.

[25] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: a literature survey. ACM Computing Surveys 2003; 4: 399-458.

[26] Wang N, Ai H, Lao S. A compositional exemplar-based model for hair segmentation. In: Proceedings of the Asian Conference on Computer Vision; 8–12 November 2010; New Zealand: Springer. pp. 171-184.

[27] Lee KC, Anguelov D, Sumengen B, Gokturk SB. Markov random field models for hair and face segmentation. In: Proceedings of the International Conference on Automatic Face & Gesture Recognition; 17–19 September 2008; Amsterdam, the Netherlands: IEEE. pp. 1-6.