International Journal of Engineering & Technology, 7 (2.5) (2018) 1-3

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

FaceParser – A new face segmentation approach and labeled database

Khalil Khan[1]*, Nasir Ahmad[2], Irfan Uddin[3], Muhammad Ehsan Mazhar[4], Rehan Ullah Khan[5]

[1]*Department of Electrical Engineering, The University of Poonch, Azad Kashmir, Pakistan*
[2]*Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar, Pakistan*
[3]*Department of Information Engineering, Inha University, Tashkent, Uzbekistan*
[4]*Bahauddin Zakria University, Multan, Pakistan*
[5]*College of Computers, Qassim University, KSA*
**Corresponding author E-mail:* e.khalilkhan@gmail.com

Abstract

Background and objective: A novel face parsing method is proposed in this paper which partition facial image into six semantic classes. Unlike previous approaches which segmented a facial image into three or four classes, we extended the class labels to six. **Materials and Methods:** A data-set of 464 images taken from FEI, MIT-CBCL, Pointing'04 and SiblingsDB databases was annotated. A discriminative model was trained by extracting features from squared patches. The built model was tested on two different semantic segmentation approaches – pixel-based and super-pixel-based semantic segmentation (PB_SS and SPB_SS). **Results**: A pixel labeling accuracy (PLA) of 94.68% and 90.35% was obtained with PB_SS and SPB_SS methods respectively on frontal images. **Conclusions**: A new method for face parts parsing was proposed which efficiently segmented a facial image into its constitute parts.

Keywords: *Face segmentation, pose estimation, gender classification, expression classification*

1. Introduction

Parsing a facial image into facial components analyze the semantic constitutes such as nose, hair, eyes, mouth, skin and hair. Face segmentation is useful in variety of tasks such as face recognition, synthesis etc. Haung et al. [1] proved that such parsing is extremely helpful in mid-level vision features estimation such as pose estimation. We argue that face segmentation will solve not only problem of pose estimation but also other mid-level vision tasks such as gender classification, facial expression recognition, age estimation and race classification.

The problem of face segmentation is already addressed by the computer vision community thoroughly. But most of these works consider face segmentation as three class or four class problem. We evaluated our work on a database we manually labeled. In three-class labeling algorithms, semantic label is assigned to three prominent classes, skin, hair and background [2, 3, 4, 5, 6]. Four-class labeling algorithms assign semantic labels to four classes -- hair, skin, background, and clothing [7, 8, 9]. Differently from all the mentioned approaches, we introduce a new method for face segmentation which extends the label set into six semantic classes. We performed experiments on low and high resolution frontal and profile images. We considered square patches as processing primitives during this work. Three kinds of features: location, color, and shape are extracted from each patch. A Random Decision Forest (RDF) classifier is trained using the extracted features. The testing phase is accomplished with two approaches PB_SS and SPB_SS. Contribution of this work is twofold -- firstly, contributing a labeled face database to the research community. Secondly proposing a multi-class face segmentation method.

Fig. 1: PB_SS and SPB_SS results comparison. Testing images on the first column, down-sampled images to patches pixel and super-pixels on the second column, labeled ground truth on the third column and algorithm outputs on the fourth column.

2. Materials and method

Most of the segmentation algorithms work at pixel or super-pixel level. We used pixel-based approach in training phase. However testing phase is accomplished with both PB_SS and SPB_SS. We considered square patches as processing primitives. We extracted patches from training and testing images while keeping a fixed step size. Every patch is classified by transferring the class label to the center pixel of each patch.

To capture information from spatial relationship between different classes, relative location of the pixel is used as a feature vector. Coordinates of the central pixel of each patch are extracted. Relative location of the central pixel at position (x, y) of a patch can be represented as;

$$floc = [\, x/W, y/H \,] \in R^2 \tag{1}$$

Where W is the width and H height of the image.

Color features are extracted from patches by using HSV color space. A single feature vector is created for color by concatenating hue, saturation and value histograms. We noted that dimension of the patch and number of bins has immense impact on the performance of segmentation algorithm. We did experimentation to investigate the best possible setting for patch size and number of bins. Based on those experimental results, we concluded that patch size 16×16 and number of bins 32 is the best possible combination for our work.

For extracting shape information we use Histogram of Oriented Gradient (HOG) [10] features. Each patch is transformed into HOG feature space. Like HSV color, we also investigated for proper patch dimension for HOG features. We found that results are best on a patch size 64×88.

All the three features spatial prior, HSV and HOG are concatenated to form a single unique feature vector fI ∈ R2618 -- where fI is the feature vector for patch I(x, y).

An RDF classier is trained using ALGLIB [11] library implementation. The trained RDF classifier predicts probability value for each pixel. The class label is assign to every pixel based on maximum probability;

We investigated two setting for classification: as multiple feature concatenation (MFC) and as spatial prior setting (SPS). In MFC setting all the three features (spatial information, HSV color and HOG) are concatenated with each other to form a single feature vector. This feature vector is input to the RDF classifier. Shape information is provided as a prior and then classification is performed accordingly in SPS.

Most of the semantic segmentation algorithms work at pixel or super-pixel level. In PB_SS approach, square patches are extracted from images with a step size 1. Each patch is classified by a trained model separately. The class label is assigned to the center pixel of each patch. By such a way, every pixel gets a class label from a trained model. SPB_SS algorithm over-segments an image by grouping pixels into small meaningful patches that belong to the same object. A single image is represented by a multiple visual feature space after segmentation. We used SEEDS [12] algorithm to over-segment an image into super-pixels. Center of each super-pixel is found and a patch is extracted around the center point. Each patch is passed to a classifier which returns a class-label. The overall multi-class semantic segmentation (MC-SS) algorithm is summarized in Fig. 2.

Algorithm 1 MC-SS algorithm

Input: $\boldsymbol{T}_{train} = \{(I_n, L_n)\}_{n=1}^{m}$, $\boldsymbol{T}_{test}$

where $\boldsymbol{T}_{train}$ is the training data and $\boldsymbol{T}_{test}$ is the testing image. I is the input training image, L grounds truth corresponding labeling image with $L(i,j) \in \{1,2,3,4,5,6\}$, and n is the number of training pairs each for I and L.

Training Phase:

1. Extract patches from $\boldsymbol{T}_{train}$ using a step size 8.
2. Extract position, HSV color and shape features from training patches.
3. Train an RDF classifier using $\boldsymbol{T}_{train}$.

Testing Phase:

1. ***If:*** *PB_SS* ***Then***
   Extract patches from $\boldsymbol{T}_{test}$ using a step size 1.
   Else-if *SPB_SS* ***Then***
   Extract patch from each super-pixel considering the central pixel.
2. Pass each patch to the classifier for probability prediction.
3. Get the p_{max} (maximum probability) and its corresponding class label for each patch.
5. Assign to each pixel the corresponding label and color.

Output: Semantically segmented face image.

Fig. 2: Flow of Multi Class Semantic Segmentation algorithm.

3. Results

Fig. 1 shows images taken from our database V-4.0. Testing images are shown in column 1. Images divided into patches and super-pixels for PB_SS and SPS_SS approaches are in column 2 respectively. Ground truth images are in column 3 and images segmented by PB_SS and SPB_SS are in column 4. From the last column images, it is very clear that image segmentation results for PB_SS are comparatively better than SPB_SS.

Our manually annotated database is a combination of three databases V-2.0, V-3.0 and latest version V-4.0. These images are taken from four different databases FEI [13], MIT-CBCL [14], Pointing'04 [15] and SiblingsDB database. Detail about downloading information is available on the website http://khalilkhan.net/. From each of the three databases, six images are taken randomly. These images are used to train an RDF classifier. Remaining 446 images are used for testing purposes.

Predicted class

True class	*eyes*	*mouth*	*nose*	*back*	*hair*	*skin*
eyes	**67.64**	0	0	0	8.74	24.33
mouth	0	**86.33**	0	0.00	0.00	13.23
nose	0	0	**69.23**	0.00	0.00	30.23
back	0	0	0.55	**87.88**	9.54	2.22
hair	0.57	0.2	0	1.16	**97.65**	0.57
skin	0.39	0.17	0.00	0.48	3.02	**95.54**

Fig. 3: Segmentation results for frontal images using PB_SS approach.

Predicted class

True class	*eyes*	*mouth*	*nose*	*back*	*hair*	*skin*
eyes	**52.88**	0	0	0	17.33	29.86
mouth	0	**73.23**	0.48	0.00	1.64	24.33
nose	0.53	0	**43.22**	0.00	1.64	24.33
back	0	0	0	**95.77**	2.41	2.2
hair	0.30	0.22	0.28	3.18	**95.02**	1.43
skin	0.59	0.47	0.23	0.66	5.08	**93.77**

Fig. 4: Segmentation results for frontal images using SPB_SS approach.

4. Discussion

Some conclusion remarks that emerge from experimental results are summarized below.

4.1. Frontal and profile images segmentation comparison

Most of the methods proposed previously does not provide results about pixel based segmentation [1, 2, 3, 4]. In our proposed paper results based on pixel based segmentation are provided in the form of confusion matrices. The segmentation results in the form of confusion matrices are shown in Fig. 3, 4, 5 and 6. Confusion matrices for frontal images are in Fig. 3 and 4. Profile images results are in Fig. 5 and 6. The results shown in tables for frontal images have horizontal and vertical orientation angle 0º each. Results for profile images have vertical orientation angle 0º and horizontal orientation angle +60º. We performed experiments on all images from -90º to +90º with a phase difference of 30º. However results shown in this paper are just for two phases. From confusion matrices it is clear that frontal images results are better than profile images. There are two valid reasons for better results for frontal images. Firstly, training phase is performed with frontal images. Secondly more information is extracted by features for frontal images as compare to profile images.

4.2. Features parameters

Three parameters needed to be chosen with extreme care i.e., size of the patch for HSV and HOG and number of bins in the HSV feature. We performed a series of experiments to choose these parameters. We find that the PLA is highest with DHSV = 16×16, DHOG = 64 × 88 and the number of bins 32.

True class		Predicted class					
		eyes	mouth	nose	back	hair	skin
	eyes	**47.33**	0	0	0	20.23	32.22
	mouth	0	**49.22**	3.25	0.00	12.24	36.23
	nose	8.52	0	**32.23**	0.00	1.36	59.22
	back	0	0	0	**92.23**	4.41	3.20
	hair	1.23	0	0.28	7.21	**93.33**	2.43
	skin	1.12	0	0.00	0.95	7.23	**91.22**

Fig. 5: Segmentation results for profile images using PB_SS approach.

True class		Predicted class					
		eyes	mouth	nose	back	hair	skin
	eyes	**55.26**	0	0	0	10.21	35.22
	mouth	0	**66.85**	2.32	0.00	0.00	30.23
	nose	0	2.45	**59.22**	0.00	0.00	30.23
	back	0	0	2.24	**83.25**	11.54	2.8
	hair	0.98	1.22	0	2.15	**94.58**	1.25
	skin	0.39	0.68	0	1.48	3.88	**94.40**

Fig. 6: Segmentation results for profile images using SPB_SS approach.

4.3. MFC and SPS comparison

The results obtained with MFC are better than SPS. Performance of the SPS is better in case of background however for remaining classes SPS has shown very poor performance. Overall PLA of the MFC is much better than SPS.

4.4. Introduction of HOG features

PLA of the framework increases very less with introduction of HOG. However, PLA of the smaller classes increased enormously with HOG.

4.5. PB_SS and SPB_SS Comparison

Overall PLA of the PB_SS is far better than SPB_SS methods. The PB_SS results are better for all classes except back where a drop in PLA is noted. A significant decrease in PLA is noted particularly in smaller and difficult classes (nose, eyes and mouth) with SPB_SS.

4.6. Computational Comparison

Although overall PLA of the PB_SS is greater but a substantial increase in speed is noted with SPB_SS, since the number of square patches to be classified by the trained model are reduced immensely.

5. Conclusion

The problem of multi-class face segmentation is explored in this work. A database consisting of 464 labeled face images is manually annotated taken from three different databases. An RDF classifier is trained by extracting three different kind of features i.e., location information, shape and color. The segmentation model built is tested on two different semantic segmentation methods PB_SS and SPB_SS. Results obtained with PB_SS are better than SPB_SS. However, computational speed noted for PB_SS is comparatively larger than SPB_SS.

We have planes to explore the current research work in two directions: improving the current segmentation model and secondly exploring some mid-level vision features with the proposed model. In first direction, the accuracy of the most problematic classes' mouth nose and eyes can be improved by the addition of rigid part detector with current framework. Similarly the labeling consistency can be enhanced locally by the integration of the current framework with Conditional Random Field. Secondly the multi-class segmentation model developed can be used for the solution of many problems such as age estimation, gender classification and facial expression recognition.

References

[1] G. B. Huang, M. Narayana, E. Learned-Miller, "Towards unconstrained face recognition, in Computer Vision and Pattern Recognition Workshops", 2008. CVPRW'08. IEEE Computer Society Conference on, IEEE, 2008, pp. 1–8.

[2] Y. Yacoob, L. S. Davis, "Detection and analysis of hair, Pattern Analysis and Machine Intelligence", IEEE Transactions on 28 (7) (2006) 1164–1169.

[3] K.-c. Lee, D. Anguelov, B. Sumengen, S. B. Gokturk, "Markov random field models for hair and face segmentation", in: Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, IEEE, 2008, pp. 1–6.

[4] J. Lafferty, A. McCallum, F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data".

[5] A. Kae, K. Sohn, H. Lee, E. Learned-Miller, "Augmenting crfs with boltzmann machine shape priors for image labeling, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 2019–2026.

[6] S. A. Eslami, N. Heess, C. K. Williams, J. Winn, "The shape Boltzmann machine: a strong model of object shape", International Journal of Computer Vision 107 (2) (2014) 155–176.

[7] Y. Li, S. Wang, X. Ding, "Person-independent head pose estimation based on random forest regression", in: Image Processing (ICIP), 2010 17th IEEE International Conference on, IEEE, 2010, pp. 1521–1524.

[8] C. Scheffler, J.-M. Odobez, "Joint adaptive colour modelling and skin, hair and clothing segmentation using coherent probabilistic index maps", in: British Machine Vision Association-British Machine Vision Conference, 2011.

[9] M. Ferrara, A. Franco, D. Maio, "A multi-classifier approach to face image segmentation for travel documents", Expert Systems with Applications 39 (9) (2012) 8452–8466.

[10] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, IEEE, 2005, pp. 886–893.

[11] S. Bochkanov, Alglib, http://www.alglib.net.

[12] M. Van den Bergh, X. Boix, G. Roig, L. Van Gool, Seeds: "Superpixels extracted via energy-driven sampling", International Journal of Computer Vision 111 (3) (2015) 298–314.

[13] C. U. da FEI, Fei facedatabase.html. database, http://www.fei.edu.br/~cet/

[14] M. C. for Biological, C. L. (CBCL), Mit-cbcl database, http://cbcl.mit.edu/software-datasets/FaceData2.html.

[15] R. Stiefelhagen, "Estimating head pose with neural networks-results on the pointing04 icpr workshop evaluation data", in: Pointing04 ICPR Workshop of the Int. Conf. on Pattern Recognition, 2004.