

HUMAN IMAGE MATTING BASED ON CONVOLUTIONAL NEURAL NETWORK AND PRINCIPAL CURVATURES

Tamerlan B. Sagindykov, Elena A. Pavelyeva

Laboratory of Mathematical Methods of Image Processing, Faculty of Computational Mathematics and Cybernetics,
Lomonosov Moscow State University, Moscow, Russia - tamerlan.sagindykov@gmail.com, pavelyeva@cs.msu.ru

Commission II

KEY WORDS: Image matting, Image segmentation, Principal Curvatures, Convolutional Neural Network, Deep learning.

ABSTRACT:

Image matting often requires advanced image processing, especially in conditions, when small details such as hair are present in the image. In this article the hybrid method for human image matting based on convolutional neural network and principal curvatures is proposed. The U-Net based neural network is used to predict a rough foreground segmentation mask. Then the obtained foreground mask is refined by principal curvatures method to process the elongated hair-like structures. Test results show that the proposed method can improve the coarse human segmentation.

1. INTRODUCTION

Image matting is an important problem in computer vision and has many applications. It is an ill-posed problem, where only the input image without any other information is available. The goal of image matting is to decompose the input color image I into two components: the foreground F and the background B . At each pixel i :

$$I_i = \lambda_i \cdot F_i + (1 - \lambda_i) \cdot B_i, \quad (1)$$

where $\lambda_i \in [0, 1]$ is the alpha matte for pixel i . The foreground color F_i , background color B_i and matte estimation λ_i are unknown, so at each pixel i we have 3 equations with 7 unknown values (I , F , and B have 3 color channels).

The first approaches to solve the image matting problem were based on classical methods of image processing and various heuristics. In the article (Chuang et al., 2001), the authors apply a probabilistic approach modeling the distribution of the foreground and background colors using Gaussians and then, using the maximum likelihood method, select the foreground and background. Authors of Poisson matting (Sun et al., 2004) formulate the problem of natural image matting as one of solving Poisson equations with the matte gradient field. Poisson matting estimates the gradient of matte from the image, then reconstructs the matte by solving Poisson equations. Their formulation is based on the assumption that intensity change in the foreground and background is smooth. The authors propose a semi-automatic approach to approximate matte from an image gradient given a user-supplied trimap. In the article (Wang and Cohen, 2007), the authors propose a robust matching algorithm to avoid the rapid degradation of performance when foreground and background patterns become complex. The basis of their algorithm contains a robust color sampling method that not only estimates the foreground and background colors for unknown pixels, but also analyze the confidence of the samples. Only high confidence samples are chosen to contribute to the matting energy function which is minimized by a Random Walk. The energy function also contains a neighborhood term to enforce the smoothness of the matting. Combining the optimized color

sampling method with propagation-based approaches, authors propose an iterative optimization process to select truly mixed pixels from all the unmarked ones, and estimate alpha values for them in closed form at each iteration. In the article (He et al., 2010a) the fast algorithm for high quality image matting based on the use of large kernel matting Laplacian matrices is proposed. The algorithm performs an efficient method to solve the linear system with the large kernel matting Laplacian. Using the large kernel helps to speed up the constraint propagation, reduce the time of the linear solver for convergence, and improve the matting quality. To further accelerate the algorithm, the authors use a KD-tree based technique to decompose the trimap so that they can assign an adaptive kernel size to each sub-trimap. Thus, the number of iterations can be fixed beforehand and the running time of the whole algorithm is only linear to the number of the unknown pixels. The algorithm can be also useful in other image processing applications using the matting Laplacian, such as haze removal (He et al., 2010b), spatially variant white balance (Hsu et al., 2008), and intrinsic images (Bousseau et al., 2009).

Recent approaches to deep learning have shown promising results. In the article Deep Image Matting (Xu et al., 2017), authors propose a new deep learning-based algorithm that consists of two parts. The first part is a deep convolutional encoder-decoder network that takes an image and the corresponding cropping as input and predicts the alpha matte color of the image. The second part is a small convolutional network that refines the alpha predictions of the first network to have more accurate alpha values and sharper edges. In the article (Lutz et al., 2018), the authors present the generative adversarial network (GAN) for natural image matting. Generator network is trained to predict visually appealing alphas with the addition of the adversarial loss from the discriminator that is trained to classify well-composited images.

Human image matting enables accurate separation of humans from their backgrounds. The most difficult problem in human image matting is to catch complicated semantic details like human hair. Some of the state-of-the-art matting algorithms require human intervention in the form of trimap or scribbles to generate the alpha matte from the input image. In (Liu et al., 2020), the use of coarse annotated data coupled with fine



Figure 1. Examples of labeled data.

annotated data to boost end-to-end semantic human matting is proposed. No trimaps as extra input are used. The mask prediction network estimates the coarse semantic mask, and then a quality unification network unifies the quality of the previous coarse mask outputs. A matting refinement network takes in the unified mask and the input image to predict the final alpha matte.

In this paper, we propose a new hybrid human image matting method based on deep learning and mathematical methods. Two approaches, convolutional neural network (CNN) and heuristics for refining specific semantic objects like hair are combined. At first, we apply a convolutional U-Net based neural network to get the coarse segmentation of the foreground and background. Then we find the edges of the foreground using the mathematical morphology (erosion and dilation). Next, in the target area obtained from the previous step, we apply the principal curvatures method to process the elongated hair-like structures. The principal curvatures method shows great performance in biometric segmentation (Tikhonova and Pavelyeva, 2020), (Safronova and Pavelyeva, 2020), (Choi et al., 2009). This method is based on calculating the eigenvalues and eigenvectors of Hessian matrix at the target points. The eigenvalues can show the structure is at the current point of image. We find oblong objects using heuristic eigenvalue estimation. Finally, we combine results from CNN and principal curvatures method and show that principal curvatures method improves the results obtained by CNN-only method.

2. HYBRID METHOD FOR IMAGE MATTING

We propose a two-step method for human image matting. The first stage is the selection of the foreground by neural network. We use supervised training with labeled data (Fig. 1). At the second stage the mask is refined using principal curvatures method.

2.1 CNN Segmentation

To determine the preliminary segmentation mask, we use convolutional neural network based on U-Net architecture (Ron-

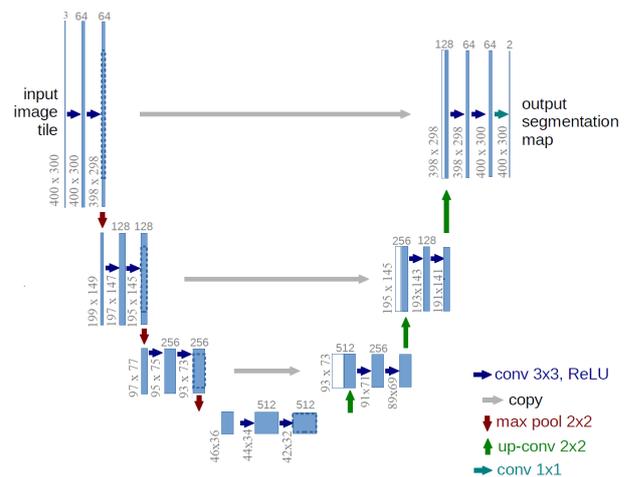


Figure 2. The U-Net based architecture.

neberger et al., 2015) (Fig. 2). Binary cross-entropy is used as a loss function, Adam optimizer (Kingma and Ba, 2014) for optimization. There are many labeled datasets of images in the public domain. We use the Flickr dataset of portrait images collected by (Shen et al., 2016), with 1300 images as train set, 200 images as validation set, and 321 images in the test set. This dataset has the rough ground truth segmentation that can be improved. The training is conducted through 200 epochs. To increase the quality of CNN we use augmentations of images in the training data, such as random crops and horizontal flip augmentation.

However, the neural network trained on the source data is not able to distinguish the thin objects such as hair due to the small size of them and rough labeled segmentation data. Then, to accurately distinguish such details we highlight regions of interest as the border of the foreground obtained by CNN. This border we find using mathematical morphology operations, such as erosion and dilation. Dilation and erosion are the basic operations in mathematical morphology. The dilation operation

usually uses a structuring element for probing and expanding the shapes contained in the input image. An erosion operation typically uses a structuring element to explore and reduce the shapes contained in the input image. The result in Fig. 3 shows the obtained region of interest that is used in principal curvatures method.

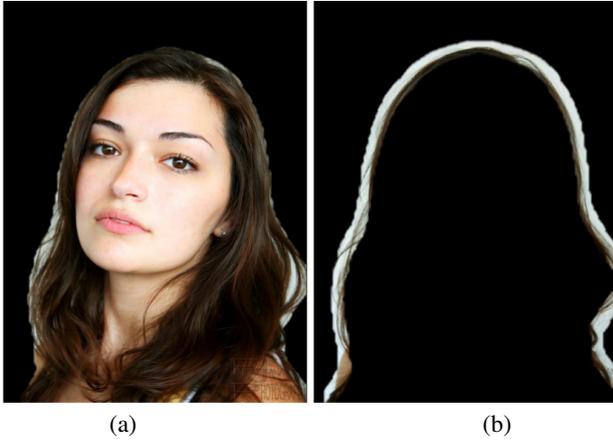


Figure 3. The result of morphology operations: (a) Segmentation result after CNN; (b) Border segmentation with morphology operations.

2.2 Principal curvatures method

The image can be represented as a three-dimensional surface, taking the intensity at each point as the value of the z-coordinate. Let $L(x, y)$ be the Gaussian smoothed image. Local characteristics of the image $L(x, y)$ at the point (x, y) can be determined using the Hessian matrix. The Hessian matrix is a square matrix of second-order partial derivatives of a scalar-valued function, or scalar field. It describes the local curvature of a function:

$$H(x, y) = \begin{pmatrix} L_{xx}(x, y) & L_{xy}(x, y) \\ L_{yx}(x, y) & L_{yy}(x, y) \end{pmatrix}. \quad (2)$$

Let $|\lambda_1| > |\lambda_2|$ be the eigenvalues of $H(x, y)$, and ν_1, ν_2 be the corresponding eigenvectors. Then two principal directions, the directions of the maximum and minimum curvatures, are determined by two eigenvectors ν_1 and ν_2 . Two eigenvalues, λ_1 and λ_2 , represent the principal curvatures (Fig. 4). The tubular-shaped regions have maximum principal curvature higher than other regions, and vector ν_1 is directed across the tubular direction, vector ν_2 - along the tubular direction. At each starting point of the region of interest, found as the border of the CNN-based foreground, the eigenvalues and eigenvectors of the Hessian matrix are calculated. If the point satisfies the rule $\frac{|\lambda_1|}{|\lambda_2|} > \gamma$, the next step of the algorithm is applied at a point located in the direction of the vector ν_2 at a unit distance from the previous point. The algorithm stops if the condition $\frac{|\lambda_1|}{|\lambda_2|} > \gamma$ is not performed.

Thus, in areas that are complicated to segment by the neural network, we obtain segmentation using principal curvatures method by moving along direction of vector ν_2 . Visualization of the vector field of the image corresponding to vector ν_2 is shown in Fig. 5. Fig. 6 shows the curves obtained by the principal curvatures method.

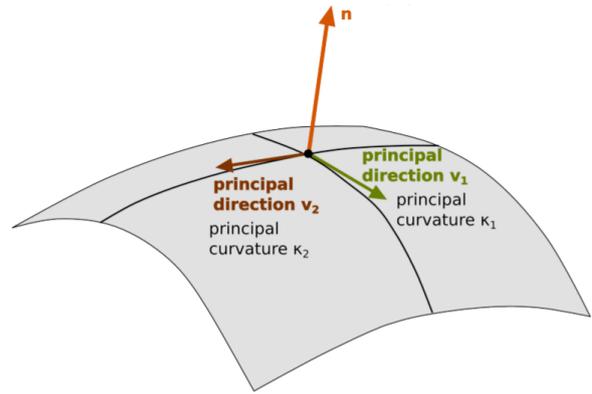


Figure 4. The curvatures along the principal directions.

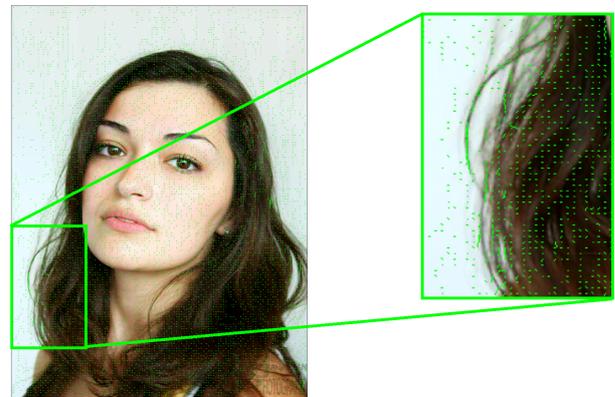


Figure 5. The vector field of the image corresponding to vector ν_2 , the length of vectors is proportional to $|\lambda_1|$.

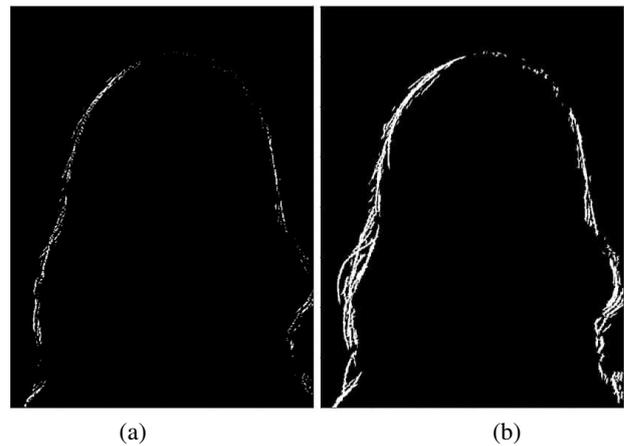


Figure 6. Building the curves: (a) the starting points of curves; (b) the final curves.

3. RESULTS

We combine the results of CNN and principal curvatures method and show that principal curvatures method improves the results obtained by CNN-only method. Test results are shown in Fig. 7. Since the ground truth labeling is rough, the numerical estimates between the ground truth segmentation masks and the obtained results may not be informative.

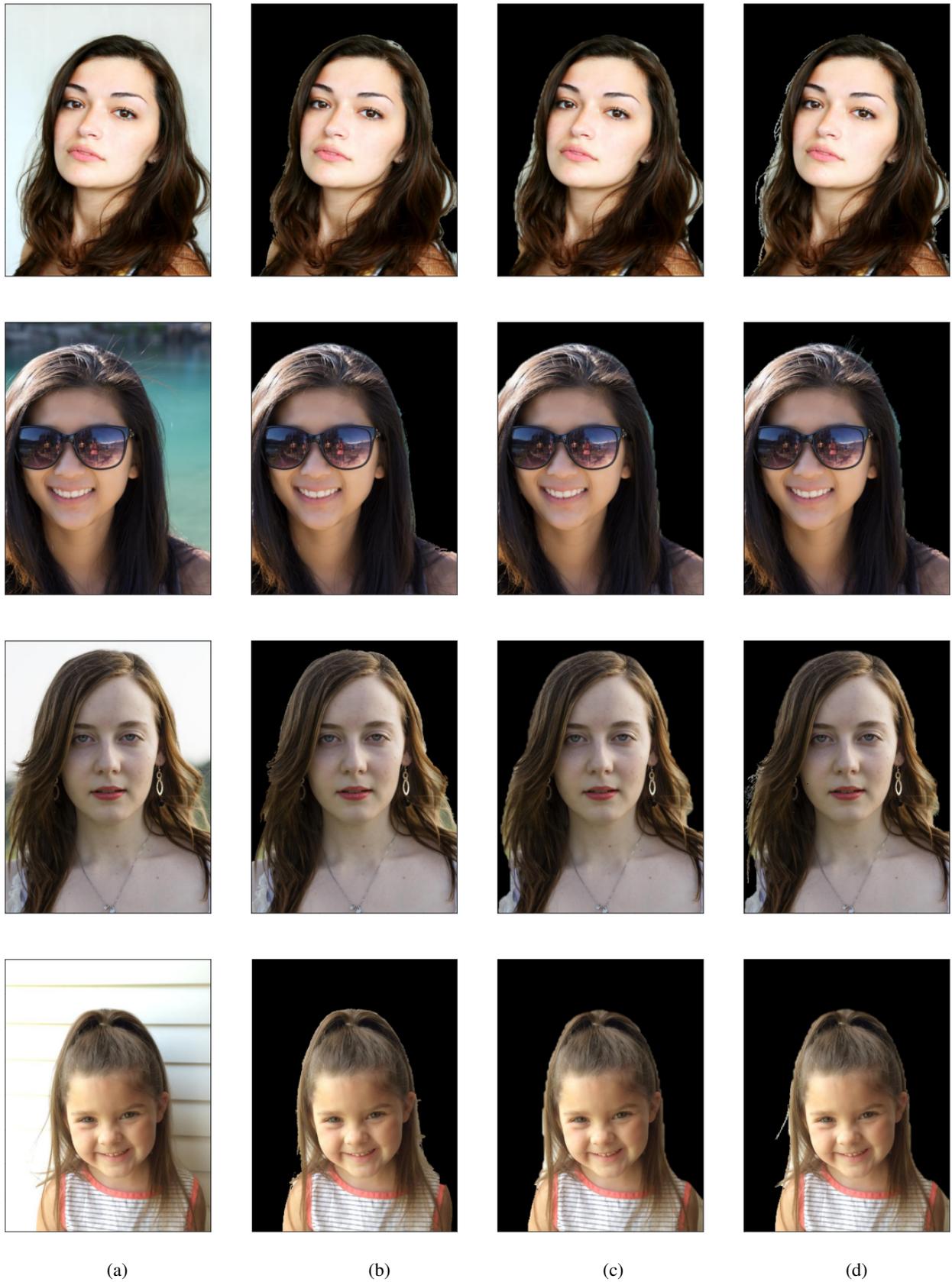


Figure 7. Examples of segmentation:(a) the original image; (b) the ground truth segmentation ; (c) the foreground segmentation by the neural network; (d) the proposed segmentation using the neural network and principal curvatures method.

4. CONCLUSION

In the article the hybrid method for portrait image matting is proposed. It is shown that the principal curvatures method can improve the rough human segmentation. The proposed hybrid approach could be promising for image matting with the more complex convolutional neural network architectures.

REFERENCES

- Bousseau, A., Paris, S., Durand, F., 2009. User-assisted intrinsic images. *ACM SIGGRAPH Asia 2009 papers*, 1–10.
- Choi, J. H., Song, W., Kim, T., Lee, S.-R., Kim, H. C., 2009. Finger vein extraction using gradient normalization and principal curvature. *Image Processing: Machine Vision Applications II*, 7251, International Society for Optics and Photonics, 725111.
- Chuang, Y.-Y., Curless, B., Salesin, D. H., Szeliski, R., 2001. A bayesian approach to digital matting. *Proceedings of IEEE CVPR 2001*, 2, IEEE Computer Society, 264–271.
- He, K., Sun, J., Tang, X., 2010a. Fast matting using large kernel matting laplacian matrices. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2165–2172.
- He, K., Sun, J., Tang, X., 2010b. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12), 2341–2353.
- Hsu, E., Mertens, T., Paris, S., Avidan, S., Durand, F., 2008. Light mixture estimation for spatially varying white balance. *ACM SIGGRAPH 2008 papers*, 1–7.
- Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Liu, J., Yao, Y., Hou, W., Cui, M., Xie, X., Zhang, C., Hua, X.-s., 2020. Boosting semantic human matting with coarse annotations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8563–8572.
- Lutz, S., Amliani, K., Smolic, A., 2018. Alphagan: Generative adversarial networks for natural image matting. *arXiv preprint arXiv:1807.10088*.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.
- Safronova, E., Pavelyeva, E., 2020. Unsupervised Palm Vein Image Segmentation. *CEUR Workshop Proceedings. 2020. Vol. 2744. Paper 40. P. 1-12*.
- Shen, X., Tao, X., Gao, H., Zhou, C., Jia, J., 2016. Deep automatic portrait matting. *European conference on computer vision*, Springer, 92–107.
- Sun, J., Jia, J., Tang, C.-K., Shum, H.-Y., 2004. Poisson Matting. *ACM SIGGRAPH 2004 Papers*. 315-321.
- Tikhonova, V., Pavelyeva, E., 2020. Hybrid Iris Segmentation Method Based on CNN and Principal Curvatures. *CEUR Workshop Proceedings. 2020. Vol. 2744. Paper 31. P. 1–10*.
- Wang, J., Cohen, M. F., 2007. Optimized color sampling for robust matting. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 1–8.
- Xu, N., Price, B., Cohen, S., Huang, T., 2017. Deep image matting. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2970–2979.