FACECUT - A ROBUST APPROACH FOR FACIAL FEATURE SEGMENTATION

Khoa Luu, T. Hoang Ngan Le, Keshav Seshadri and Marios Savvides

{kluu, thihoanl, kseshadr}@andrew.cmu.edu, msavvid@ri.cmu.edu
Department of Electrical and Computer Engineering and CyLab Biometrics Center
Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

Segmentation of facial features is a key pre-processing step in enabling facial recognition, building of 3D facial models, expression analysis, and pose estimation. Recently, graph cuts based algorithms have been adapted to carry out this task but many of these methods require manual initialization of points in the foreground and background. In this paper, we propose a novel and fully automatic approach, named FaceCut, to perform accurate facial feature segmentation. FaceCut combines the positive features of the Modified Active Shape Model (MASM) and GrowCut algorithms to ensure highly accurate and completely automatic segmentation of facial features. We demonstrate the effectiveness of FaceCut on images from two challenging databases.

Index Terms— Face segmentation, facial landmarks, Active Shape Models (ASMs), graph cuts, FaceCut

1. INTRODUCTION

Segmentation of facial features plays a crucial role in numerous face related applications including face recognition, age estimation, expression analysis and construction of 3D facial models. Recent algorithms such as graph cuts [1], GrabCut [2], GrowCut [3], and their improvements have been able to obtain accurate segmentation results in natural images. However, when dealing with facial images, it is hard to determine the boundary between the face and the neck regions by using conventional graph cuts based algorithms. This can be seen in Fig. 1(C), which shows an incorrectly segmented face obtained using the classical GrowCut algorithm. It is also to be noted that GrowCut requires the manual marking of points in the foreground and background and that it can't separately segment facial components such as eyes, nose and mouth.

Active Shape Models (ASMs) [4] have been traditionally used for the task of automatic facial landmark localization. However, ASMs are PCA based approaches and do not always accurately localize landmarks in faces with shapes radically different from those exhibited by faces in their training sets, as shown in Fig. 2(B). Thus, the motivation of this paper is to overcome the shortcomings of both graph cuts based approaches and an ASM based approach in order to ensure that

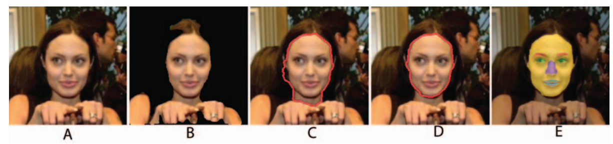

Fig. 1. Comparisons of facial segmentation algorithms: (A) Original image, (B) Face segmentation using color based information, (C) Face segmentation results obtained using GrowCut, (D) Automatic segmentation of the facial boundary using statistical skin information and our modified GrowCut algorithm, (E) Segmentation of facial features segmentation using our FaceCut algorithm.

the facial region and individual facial features are segmented accurately as possible. We propose a novel method named FaceCut to accomplish this, as shown in Fig. 1(E).

The rest of this paper is organized as follows. In section 2, we review prior work on graph cuts based approaches to image segmentation. In section 3, we describe how ASMs are applied to the task of facial landmarking and also mention the key features of the Modified Active Shape Model (MASM) that we use in our segmentation approach. In section 4, we explain our modification of the GrowCut algorithm. The combination of MASM and our modified GrowCut approach is called FaceCut and section 5 presents experimental results obtained by FaceCut on two challenging databases, namely the NIST Multiple Biometric Grand Challenge (MBGC) still face database [5] and the Labeled Faces in the Wild (LFW) [6] database. Finally, we present our conclusions on this work in section 6.

2. PRIOR WORK ON IMAGE SEGMENTATION

In the graph cuts algorithm [1], input images are treated as graphs and their pixels as nodes of the graphs. In order to segment an object, the max-flow and the min-cut algorithms are employed. For the rest of this paper we refer to this original approach to image segmentation as the GraphCuts method.

The GrabCut [2] algorithm was later introduced to improve the GraphCuts method by using an iterative segmentation scheme and applying graph cuts at intermediate steps.

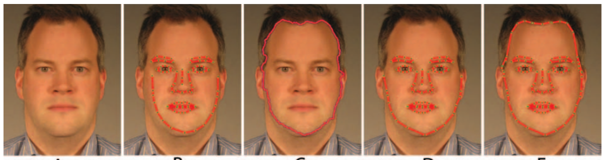

Fig. 2. Comparison of MASM and FaceCut: (A) Original image, (B) 79 landmarks detected by MASM, (C) Complete facial boundary segmentation using our modified GrowCut algorithm, (D) FaceCut segmentation results with overlaid landmarks to benchmark against MASM, (E) FaceCut segmentation results and overlaid facial landmarks with segmentation extended to the upper part of the face.

The user provides a rectangular box around the object to be segmented. The segmentation process is employed using the statistical color information inside and outside the box. The image graph is re-weighted and the new segmentation in each step is refined by using graph cuts.

3. THE MODIFIED ACTIVE SHAPE MODEL (MASM)

An Active Shape Model (ASM) is a deformable template based approach to model the shape and texture of an object using points that lie along its contours and referred to as landmarks. A Point Distribution Model (PDM) of shape variation is built by aligning the landmarks in each of the manually annotated training shapes, calculating the mean shape $\overline{\mathbf{x}}$ of these aligned shapes and then applying Principal Component Analysis (PCA) to obtain an orthonormal basis of vectors $\mathbf{P}$ to approximate any shape $\mathbf{x}$ using the shape model equation in (1), in which $\mathbf{b}$ is a vector of PCA coefficients for the shape in question.

$$\mathbf{x} \approx \overline{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{1}$$

An ASM also builds models of the local texture around each landmark across all training images and at multiple resolutions of an image pyramid.

An unseen image is automatically landmarked by first detecting the face in it and initializing the ASM so that the mean shape is roughly aligned over the face and then iteratively moving the landmarks into locations that best match the training data while still preserving the legality of a facial shape. The process is carried out at multiple image resolutions until convergence is attained at the finest resolution.

The ASM implementation used by us is the Modified Active Shape Model proposed in [7] whose key features include the modeling of local texture around landmarks using a subspace based technique and the use of a new cost function to determine optimal landmark locations during the fitting process. The 79 point landmarking scheme used by MASM to annotate a face is shown in Fig. 2(B).

Algorithm 1 Our Modified GrowCut Algorithm for Facial Boundary Segmentation

$t \leftarrow 0$
while (not converged) $\wedge$ $(t < t_{\max})$ **do**
  for all $\boldsymbol{\omega} \in \Omega$ **do**
    Copy state at t to state at (t+1)
    $\theta_T \leftarrow \theta^t_{\boldsymbol{\omega}}$
    $l_T \leftarrow l^t_{\boldsymbol{\omega}}$
    for all $\rho \in B(\boldsymbol{\omega})$ **do**
      if $(l^t_\rho \neq 0) \wedge (g(\|\mathbf{I}_{\boldsymbol{\omega}} - \mathbf{I}_\rho\|_2)\theta^t_\rho > \theta_T)$ **then**
        if $((l^t_\rho = -1) \vee ((l^t_\rho = 1) \wedge (\text{checkskin}(\mathbf{I}_{\boldsymbol{\omega}}) = true)))$ **then**
          $\theta_T \leftarrow g(\|\mathbf{I}_{\boldsymbol{\omega}} - \mathbf{I}_\rho\|_2)\theta^t_\rho$
          $l_T \leftarrow l^t_\rho$
        end if
      end if
    end for
    $\theta^{t+1}_{\boldsymbol{\omega}} \leftarrow \theta_T$
    $l^{t+1}_{\boldsymbol{\omega}} \leftarrow l_T$
  end for
  $t \leftarrow (t+1)$
end while

4. FACIAL BOUNDARY SEGMENTATION USING OUR MODIFIED GROWCUT ALGORITHM

This section first reviews the GrowCut algorithm and then presents our extensions to it.

4.1. The GrowCut Algorithm

The GrowCut algorithm for image segmentation was first introduced by Vezhnevets and Konouchine in [3]. Unlike graph cuts based approaches, GrowCut employs cellular automata that are appropriate for dealing with digital images due to their discrete properties in both space and time domains. It can segment color images efficiently and produces results comparable to those obtained by the GraphCuts method.

In the algorithm, given an image $\mathbf{I}$, the atomic unit for a pixel $\boldsymbol{\omega}$ is formed by $\mathbf{A}_\omega$ which is composed of three components as shown in (2), in which N is the size of image $\mathbf{I}$, $\boldsymbol{\omega} \in R^N$, $\mathbf{S}_\omega$ is the state of the current automaton, $B(\boldsymbol{\omega})$ is the neighborhood function that usually chooses 4 or 8 nearest neighbor pixels about the center pixel $\boldsymbol{\omega}$, and δ is the mapping function of a transition rule from the cell's state at t to its state at $t+1$.

$$\mathbf{A}_\omega = (\mathbf{S}_\omega, B(\boldsymbol{\omega}), \delta) \tag{2}$$

The state $\mathbf{S}_\omega$ of the automation cell $\boldsymbol{\omega}$ is defined by a triplet as shown in (3), in which $l_{\boldsymbol{\omega}}$ is the label of the current cell $\boldsymbol{\omega}$, $\theta_{\boldsymbol{\omega}} \in [0, 1]$ is the "strength" of the cell $\boldsymbol{\omega}$ used to attack its neighbors, and $\mathbf{I}_{\boldsymbol{\omega}}$ is the cell feature vector defined by the color values of that cell.

$$\mathbf{S}_\omega = (l_{\boldsymbol{\omega}}, \theta_{\boldsymbol{\omega}}, \mathbf{I}_{\boldsymbol{\omega}}) \tag{3}$$

The algorithm places the initially labeled cells, called the bacteria cells, in the image $\mathbf{I}$. Each bacteria cell $\boldsymbol{\omega}$ attacks its neighbors with a strength $\theta_{\boldsymbol{\omega}}$. If its total force is strong enough, it will conquer the neighbors by labeling them with the same label $l_{\boldsymbol{\omega}}$. The total attack force is defined as the product of the current strength $\theta_{\boldsymbol{\omega}}$ and the attack force, computed by a monotonically decreasing function [3]. The strengths of the initial bacteria cells are set to the value one so that their power is strongest at the beginning. The strength of the other cells are set to zero. The image $\mathbf{I}$ is finally segmented with each pixel either assigned foreground or background labels.

4.2. Skin Color Detection

[8] and [9] review several prior approaches to skin detection. In this paper, we use a bag of words model for skin detection [10]. In our approach, the training set includes the textures of two classes, i.e. skin and non-skin. Each sample in this set is divided into a collection of patches or codewords. A skin model is trained on this set using the Naive Bayes classification technique. Given a sample y of K patches, the skin/non-skin conclusion C is decided by the function $\mathrm{checkskin}(y)$ in (4), in which $p(C)$ is the prior probability of the skin/non-skin classes and $p(y|C)$ is the likelihood given the class. Both of these functions are built using the training set.

$$\begin{aligned} \mathrm{checkskin}(y) &= \underset{C}{\mathrm{argmax}} \quad p(C|y) \propto p(C)p(y|C) \\ &= \underset{C}{\mathrm{argmax}} \quad p(C)\prod_{k=1}^{K} p(y_k|C) \end{aligned} \tag{4}$$

We also use an ad-hoc skin color detection model in this work. Numerous color spaces have been used in prior work on human-skin detection, such as: RGB, HSV, CIE-Lab, etc. However, as the experimental comparisons carried out by Phung et al. [8] have shown, compared to other color spaces, the RGB color space can give acceptable skin detection results by imposing a set of rules on its color components (R,G and B) as shown in (5).

$$\begin{gathered} (R > 95) \wedge (G > 40) \wedge (B > 20) \\ \max(R,G,B) - \min(R,G,B) > 15 \\ (|R-G| > 15) \wedge (R > G) \wedge (R > B) \end{gathered} \tag{5}$$

The rules in (5) ensure the correct selection using the skin-tone of the face, that the three color components are sufficiently separated from each other to ensure greyness elimination, and the separation of the R and G components respectively.

4.3. Our Modification of the GrowCut Algorithm

Our extension of the Growcut algorithm for facial boundary segmentation is presented in Algorithm 1. In Algorithm 1, $t_{\max}$ is the maximum number of interactions and g is a monotonically decreasing function bounded to lie in the interval $[0,1]$. Instead of using the function introduced in [3], we use the sigmoid function $g(t)$ shown in (6).

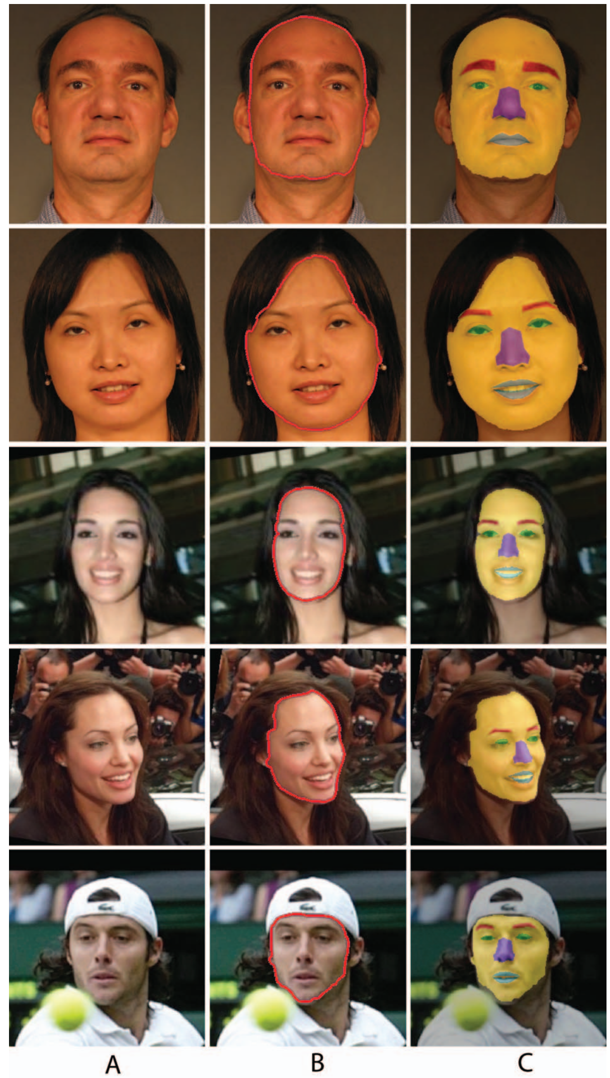

Fig. 3. FaceCut results on MBGC (first two rows) and LFW (last three rows) database images: (A) Original image, (B) Facial boundary segmented, (C) Facial features segmented.

$$g(t) = \frac{1}{1+e^{-t}} \tag{6}$$

The function $\mathrm{checkskin}$ in Algorithm 1 is used to verify if the current pixel belongs to the skin region or not and is described in subsection 4.2. As can be seen from Fig. 2, compared to MASM, FaceCut produces better facial boundary segmentation results. Additionally, FaceCut can extend the segmentation to the upper part of the face without including facial hair in the segmented region. This is something that MASM is not capable of as it can only isolate discrete landmarks along the lower part of the facial boundary.

Algorithms	Mean (all landmarks)	Std. Dev. (all landmarks)	Mean (facial boundary landmarks)	Std. Dev. (facial boundary landmarks)
MASM	9.65	11.66	14.85	12.21
FaceCut	9.56	11.51	14.46	11.59

Table 1. Comparison of the mean (in pixels) and standard deviation (in pixels) of the fitting error across landmarks produced by MASM and our FaceCut approach on images from the MBGC database.

5. FACECUT: ALGORITHM AND EXPERIMENTAL RESULTS

In the first step of our FaceCut algorithm, we use MASM to localize a set of 79 facial landmarks. Based on these detected internal points, we infer a set of interior and exterior points on the face. Following this, our modified GrowCut algorithm is used to segment the facial boundary. FaceCut can extend segmentation to the upper part of the face, but to quantitatively compare it to MASM, only landmarks along the eyes, nose, lips and lower part of the facial boundary are considered.

MASM was trained on images drawn from the frontal view set of the CMU Multi-PIE (MPIE) database [11]. We tested our FaceCut approach on 500 images of 95 subjects from the MBGC still face database to evaluate its robustness to varying illumination and in-plane rotation of faces. As can be seen from Table 1, which shows the mean and standard deviation of the point to point landmark fitting errors (for all 79 landmarks and for just the 15 facial boundary landmarks) obtained by MASM and FaceCut across all MBGC test images when compared to manually annotated images (ground truths), FaceCut results in improved fitting accuracy of the facial boundary landmarks. FaceCut was also tested on images from the LFW database and was again able to accurately segment facial features in these challenging everyday images. Fig. 3 shows sample FaceCut segmentation results on images from the MBGC and LFW databases respectively and demonstrates the effectiveness of the algorithm.

6. CONCLUSIONS

We have proposed a novel method, named FaceCut, to automatically segment human facial features. FaceCut combines positive features of the Modified Active Shape Model (MASM) and GrowCut algorithms to ensure highly accurate and completely automatic segmentation of facial features and the facial boundary. The effectiveness of segmentation using FaceCut is demonstrated on unseen test images from the challenging MBGC and LFW databases.

7. ACKNOWLEDGMENTS

This research was supported in part by CyLab at Carnegie Mellon under grants DAAD19-02-1-0389 and W911NF-09-1-0273 from the Army Research Office and by cooperative agreement W911NF-10-2-0028 from the Army Research Laboratory.

8. REFERENCES

[1] Y. Y. Boykov and M. P. Jolly, "Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images," in *Proc. of IEEE ICCV*, 2001, pp. 105–112.

[2] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": Interactive Foreground Extraction using Iterated Graph Cuts," in *Proc. of ACM SIGGRAPH*, 2004, pp. 309–314.

[3] V. Vezhnevets and V. Konouchine, ""GrowCut" - Interactive Multi-Label N-D Image Segmentation by Cellular Automata," in *Proc. of the Intl. Conf. on Computer Graphics and Vision (GRAPHICON)*, 2005.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application," *CVIU*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[5] NIST, "National Institute of Standards and Technology: Multiple Biometric Grand Challenge (MBGC) - 2008," `http://face.nist.gov/mbgc`, 2008.

[6] G. B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Tech. Rep. 07-49, Univ. of Massachusetts, Amherst, Oct. 2007.

[7] Keshav Seshadri and Marios Savvides, "Robust Modified Active Shape Model for Automatic Facial Landmark Annotation of Frontal Faces," in *Proc. of the* 3^{rd} *IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2009, pp. 319–326.

[8] S. L. Phung, A. Bouzerdoum Sr., and D. Chai, "Skin Segmentation Using Color Pixel Classification: Analysis and Comparison," *IEEE TPAMI*, vol. 27, no. 1, pp. 148–154, Jan. 2005.

[9] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A Survey of Skin-Color Modeling and Detection Methods," *Journal of Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, Mar. 2007.

[10] L. Fei-Fei and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," in *Proc. of IEEE CVPR*, 2005, pp. 524–531.

[11] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Proc. of the* 8^{th} *IEEE Intl. Conf. on Automatic Face and Gesture Recognition (FG)*, 2008, pp. 1–8.