13-11

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Human Body Region Extraction from Photos

Yi Hu
Fujifilm Corporation, Japan
798, Miyanodai, Kaisei-machi, Ashigarakami-gun, Kanagawa, 258-8538 JAPAN
yi_hu@fujifilm.co.jp

Abstract

This paper presents an approach to automatically extract human body region from color photos, which introduces trimap shape updating into iterated GrabCut image segmentation technique. It is based on an observation that human torso is relatively stable in appearance compared with various human poses formed by hands and feet, and on a fact that estimation on a small region is more accurate than on a large region if a few cues for estimation are just available. At first a human face is found by scanning a face detector across the whole target unknown image. Then a body trimap, an image showing potential body area, is initialized according to the found face. With this trimap, body torso is estimated with GrabCut image segmentation method. After that, the trimap is updated by dynamically growing its contour according to local image information, and new body region is estimated by applying GrabCut to the target image. With the iterated processing of trimap shape updating and GrabCut applying, human body region is finally extracted. The approach has been tested with 400 photo images, and the results show its usefulness.

1. Introduction

With the flooding of digital photo images, more and more intelligence is sought by photo processing applications such as photo classification, retrieval, trimming, clipping and album making. There is an increasing demand for the ability to automatically extract human body from photos so that human pose analysis such as standing, sitting and drinking etc is possible and advanced photo applications centering on human can be realized.

In general, human body extraction from still image is an extremely difficult problem due to various poses of human body and complicated background environment. Some researchers modeled human body as an assembly of parts. Candidate parts are produced from low-level part detectors or come from image segmentation results. Then a top-down procedure makes inferences about these parts and finds the best assembly [1,2,3, 4, 10].

Another approach is to utilize some available cues to guide image segmentation to extract object. Rother et al. [5] proposed an interactive foreground/background segmentation called GrabCut. It is an iterative image segmentation technique based upon the Graph Cut algorithm [6]. Since the cues for image segmentation are given manually, it is mainly used as an interactive image tool for foreground object extraction.

Inspired by the work of Rother et al. [5], we present an approach to automatically extract human body region from color photos, which incorporates dynamically updating trimap contour with iterated GrabCut technique. On considering the diversity and variety of human poses, we constrain our researches on those human poses with frontal/side faces in color photo images and focus on the topic of human body region extraction, which aims to separate human body from background and does not classify human body parts. Different from [5], the trimap guiding the image segmentation in our approach is initialized from the results of detected faces, and the contour of the trimap is updated dynamically during body extraction. This is motivated by a fact that estimation on a small region is more accurate than on a large region if a few cues are just available. And we noticed that human torso is relatively stable in appearance compared with various human poses formed by hands and feet. A body torso is firstly extracted. Then the trimap is updated by dynamically growing its contour according to local image information, and new body region is estimated by applying GrabCut to the target image. With the iterated processing of trimap shape updating and GrabCut applying, human body region is finally extracted.

The details of the proposed approach are described in Section 2. The experimental results are given in Section 3. Finally, we draw up the conclusions in Section 4.

2. Human Body Region Extraction Algorithm

The workflow of human body region extraction from a photo is illustrated in Fig.1. At first we extract face region. According to the face region, we set initial body trimap, an image showing potential body area. Then we do color clustering to initialize GMMs for both candidate body region and background, and form link table of graph cut with the information from GMMs and body skin image. After that we apply GrabCut image segmentation technique to extract the initial body (torso). By growing the body trimap gradually and reapplying GrabCut, we extend the extraction result from torso to the whole body. The extraction is an iteration process with body trimap updating.

The details of the algorithm is explained as following:

1) Face/Eye detection

   Given an unknown image U_{img}, locate the human face F_{img} by scanning a face detector across U_{img}. Then locate two eyes with an eye detector on F_{img}. The face/eye detector is a cascade of classifiers learned from face/eye and non-face/eye training data with boosting methods.

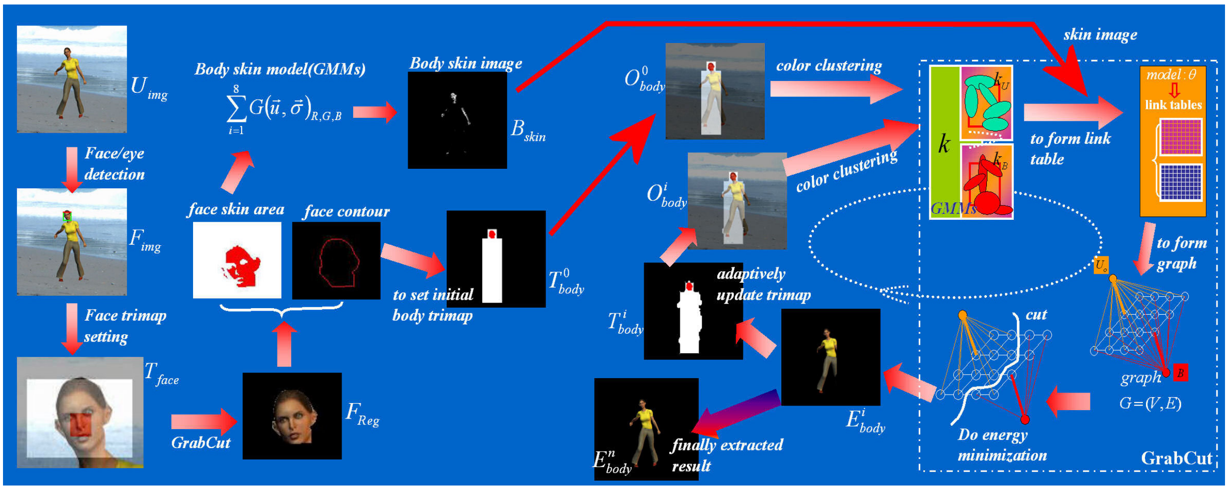

Fig.1 The workflow of our proposed human body region extraction approach.

2) Face region extraction

We apply GrabCut image segmentation technique to extract face region F_{Reg}. The extraction process is shown in Fig.2. According to the distance L_1 between two eyes, we set face trimap T_{face} shown by Fig.2(c), where the region R_A^F (size: $L_1 \times L_1$) corresponds to a sub-area on the face, R_B^F (size: $3.5L_1 \times 3.0L_1$) is the candidate region of F_{Reg} needs to be estimated, and R_C^F is background. The superimposed face trimap image O_{face} is show in Fig.2 (d). Then applying GrabCut to the face work image with T_{face}, we get human face region F_{Reg} shown by Fig.2 (e).

3) Making body skin image

From the extracted face region image, we get face contour and face skin area image. The face skin area is determined by classifying each pixel as skin or not according to the following formula [7]:

$$(R,G,B)\ is\ classified\ as\ skin\ if$$
$$R > 95\ and\ G > 40\ and\ B > 20\ and$$
$$\max(R,G,B) - \min(R,G,B) > 15\ and$$
$$|R - G| > 15\ and\ R > G\ and\ R > B$$

By doing color clustering [8] on all pixels in the face skin area, we represent the body skin with a GMM model:

$$\rho_{bodyskin} = \sum_{i=1}^{8} \eta_i G(u_i, \sigma_i) \qquad (1)$$

where, η_i, u_i and σ_i are the weight, mean and covariance of each GMMs component. $G(\cdot)$ is Gaussian kernel.

By applying Eq.1 to the whole task image U_{img}, we get the body skin image B_{skin}.

4) Initialization of body trimap

Fig.3 show the process of creating initial body trimap T^0_{Body}. By referencing the face contour, we set three regions R_A^B, R_B^B and R_C^B in T^0_{Body}. The region R_A^B is the extracted face area, R_B^B consists of two blocks, one corresponds to a head region (a little bigger than the face area), and another corresponds to the assumed torso region, which size is 7 times high and 3 times wide as the head region. The remained region is R_C^B, assumed to be background. The superimposed initial body trimap image O^0_{Body} is shown in Fig.3 (c). The region R_B^B

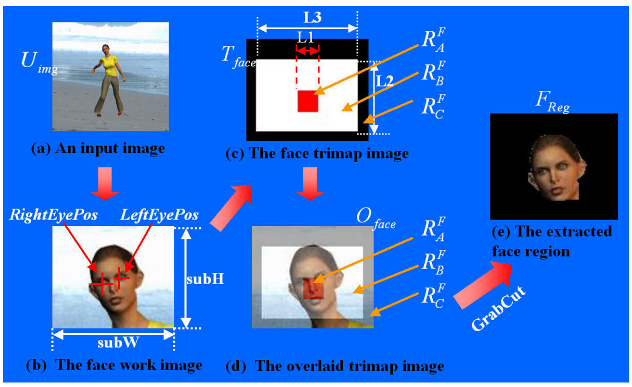

Fig.2 Face trimap setting and face region extraction.

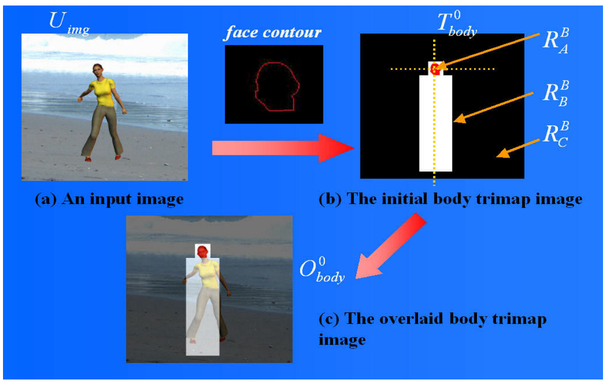

Fig.3 Initialization of the body trimap.

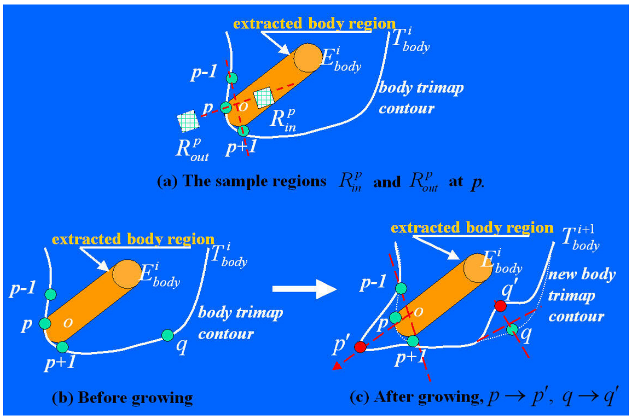

Fig.4 Update the body trimap by growing the trimap contur.

is the region needs to be estimated.

5) Applying GraphCut image segmentation

By regarding regions R^B_A and R^B_B as body region, and R^B_C as background region, we create GMMs by doing color clustering [8] for each region group.

Body region group (R^B_A + R^B_B):

$$g_{obj} = \sum_{i=1}^{8} \lambda_i \frac{1}{(2\pi)^{d/2} \Sigma^{1/2}} \exp\left[-\frac{1}{2}(x-u_i)' \Sigma^{-1}(x-u_i)\right] \quad (2)$$

Background region group (R^B_C):

$$g_{bkg} = \sum_{i=1}^{8} \lambda'_i \frac{1}{(2\pi)^{d/2} \Sigma'^{1/2}} \exp\left[-\frac{1}{2}(x-u'_i)' \Sigma'^{-1}(x-u'_i)\right] \quad (3)$$

where λ_i and λ'_i are the weights of each GMMs component, u_i and u'_i are the mean vectors, Σ and Σ' are the covariance of each GMMs component.

Based on the above GMMs Eq.(2),(3), we form link table for graph [5,6,9]. We also add skin information B_{skin} into the link table.

$if\ (a\ pixel\ p \in R^B_B\ and\ p \in B_{skin})$
$then$
$the\ t-link\ of\ p\ to\ the\ object\ terminal\ is\ set\ as:$

$$t-link_{p-obj} += \alpha\, f_{Kw}$$

$where\ \alpha\ is\ a\ weight\ parameter\ (\alpha = 0.08)$
$and\ f_{Kw}\ is:$

$$f_{Kw} = \max_{p \in U_{img}} \left\{ \sum_{i=1}^{8} n-link_i\ of\ p \right\}$$

Applying GrabCut [5,6,9] on the image U_{img} with the initial body trimap T^0_{Body}, we get the extracted body region E^0_{Body}. (GrabCut itself is an iterated procedure).

6) Growing body trimap contour

We trace the contour of the just used body trimap T^i_{Body} ($i = 0,1,\cdots$) and check each point p on the contour. If p also belongs to the extracted body E^i_{Body}, then we sample two small regions R^p_{in} and R^p_{out} along two sides of p 's normal line shown by Fig.4 (a), where $R^p_{in} \subset E^i_{body}$ and $R^p_{out} \not\subset E^i_{body}$. The mean value of pixels in each region is assumed as MR^p_{in} and MR^p_{out}. With Eq.(2),(3), we determine whether to grow p according to the following conditions:

$if\ \left|g_{bkg}(MR^p_{in}) - g_{bkg}(MR^p_{out})\right| < \theta_1\ and\ g_{bkg}(MR^p_{out}) > \theta_2\ \{$
$to\ grow\ p\ inwards$
$\}$
$else\ if\ \left|g_{obg}(MR^p_{in}) - g_{obg}(MR^p_{out})\right| < \theta_3\ and\ g_{obj}(MR^p_{in}) > \theta_4\ \{$
$to\ grow\ p\ outwards$
$\}$

where θ_i ($i = 1,2,3,4$) are threshold parameters, which are set from experimental results.

If p does not belong to the extracted body

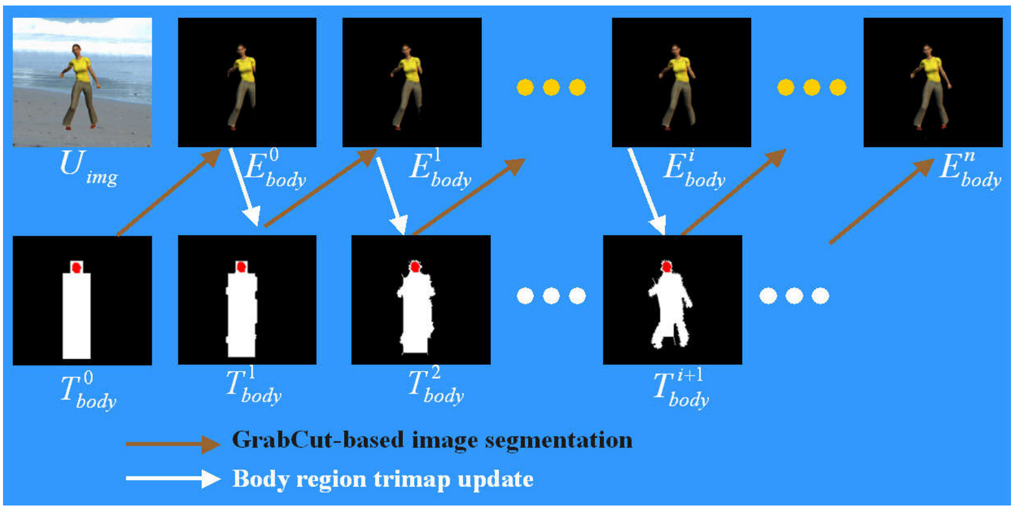

Fig.5 Iterated body region extraction with body trimap updating.

E^{i}_{Body}, then we grow p inwards. Fig.4 (b) and (c) show the growing process of two points along the contour of T^{i}_{Body}, where p is a point on the extracted object and q is not. With this growing, we get a new body trimap T^{i+1}_{Body}:

$$T^{i}_{Body} \xrightarrow{U_{img},\, E^{i}_{Body}} T^{i+1}_{Body}$$

7) Repeat the step 5) to 6) until the extracted body region E^{n}_{Body} converges (no changes). E^{n}_{Body} is the finally extracted body region. Fig.6 shows the iterated body trimap updating and body region extraction process.

3. Experiments and Results

The proposed approach was implemented and evaluated with 400 real-life digital photo images of indoors and outdoors. Nearly 75% human bodies in the photos are extracted. The failed ones are mainly caused by unsuitable body trimap initialized in step 4. The limitation of this approach is that it can not deal with overlapped human bodies. Fig.6 shows some extracted examples. The test results show that the proposed approach is useful.

4. Conclusion

We have presented an approach to automatically extract human body region from color photos. We begin by finding a face in the task photo image. Then we initialize a body trimap according to the found face. By applying GrabCut image segmentation technique, we firstly extract the torso part of a human. After that we dynamically grow the trimap to cover the parts of hands and legs. With the iterated processing of trimap shape updating and GrabCut applying, human body region is finally extracted. The approach has been tested with 400 photo images, and the results show its usefulness.

Future work will focus on how to initialize the body trimap more flexibly, and establishing a model to estimate that R^{p}_{in} in Fig.4 (a) is really a true part of human body or not so that intelligent body trimap updating is possible.

References

[1] Greg Mori, Xiaofeng Ren, Alexei A. Efros and Jitendra Malik, "Recovering human body configurations: combining segmentation and recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, vol. 2, pp. 326-333, 2004.

[2] Qingjie Sun, Enhua Wu, "Probability estimation for people detection," *International Journal of WSCG*, 1(3):448-455, 2003.

[3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient matching of pictorial structures," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2000.

[4] S. Ioffe and D. Forsyth, "Probabilistic methods for finding people," *International Journal of Computer Vision*, 43(1):45-68, Jun. 2001.

[5] Carsten Rother, Vladimir Kolmogorov and Adrew Blake, "GrabCut-Interactive foreground extraction using iterated Graph Cuts," in *ACM Transactions on Graphics (SIGGRAPH'04)*, August 2004.

[6] Yuri Y. Boykov and Marie-Pierre Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in *Proc IEEE Int. Conf. On Computer Vision*, July 2001.

[7] Vladimir Vezhnevets, Vassili Sazonov and Alla Andreeva, "A survey on pixel-based skin detection techniques," in Proceedings of the Graphicon-Moscow, pp. 85-92, September 2003.

[8] Orchard, M.T., and Bounman, C.A. "Color quantization of images," *IEEE Transactions on Signal Processing*, Vol.39, No.12, pp.2677-2690, 1991.

[9] Yuri Boykov and Vladmir Kolmogorov, "An experimental comparison of Min-Cut/Max-Flow algorithms for energy minization in vision," *IEEE Transactions on PAMI*, Vol.26, No.9, pp.1124-1137, Sept. 2004.

[10] Stella X.Yu, Ralph Cross and Jianbo Shi, "Concurrent object recognition and segmentation by graph partitioning", in NIPS'02, November, 2002.

Fig.6 Some test experiment results on human body region extraction from photo images.