FCV2016, the 22nd Korea-Japan Joint Workshop on Frontiers of Computer Vision

Human Body Segmentation Using Texture Aware Grab-cut and Statistical Shape Models

Esmaeil Pourjam*, Yasutomo Kawanishi*, Ichiro Ide*, Daisuke Deguchi†, Hiroshi Murase*
*Graduate School of Information Science, Nagoya University
Email: esmaeilp@murase.m.is.nagoya-u.ac.jp, {kawanishi,ide,murase}@is.nagoya-u.ac.jp
†Information Strategy Office, Nagoya University
Email: ddeguchi@nagoya-u.jp

Abstract—Segmentation is one of active areas in computer vision eld with application in many areas from entertainment to intelligent vehicles (IVs). Among the objects, humans themselves have always been among the most interested subjects because of their special features.

Since human body has an articulated structure, modeling and recognizing di erent variations in the body has proved to be very di cult. Wearing various kinds of clothes in di erent situations which can have a completely dissimilar appearance based on the clothing type, makes the modeling much more di cult. Add to this, the common problems of vision like illumination changes, blurring due to camera movements, etc. make the problem even more di cult. Thus having a system that can segment human subjects accurately can be useful in many applications.

In this paper, we propose a system for segmenting human subjects using Statistical Shapes Models (SSM) feedback and a texture aware version of Grab-cut which incorporates texture feature for improving the segmentation accuracy. Our experiments show that the proposed system has an acceptable accuracy compared to the state-of-the art interactive methods and much better than the conventional ones.

I. Introduction

For a long time, human subject segmentation has been one of active areas in the field of vision. Making systems and machines capable of interacting with humans or understanding the presence of human can be very useful in various types of applications. Face recognition for security systems, pose estimation, intention recognition for Human Machine Interfaces or Intelligent Vehicles (IVs) are some samples among the list of imaginable applications for human segmentation. Especially, recent trends in IVs have focused on driver assistance systems, object avoidance systems, and automatic driving systems in which understanding pedestrians presence and their intention can have a great impact on their performance.

Object segmentation methods can be categorized into two major categories: automatic, and interactive (semi-automatic) algorithms. The difference is that in automatic methods, usually human interference is unnecessary while it is opposite in interactive ones. Still, in contrast to automatic segmentation, in recent years, interactive image segmentation has shown some potential in the field of segmentation. Different methods have already been introduced in the literature such as graph-cut [1], obj-cut [2], lazy snapping [3], intelligent scissors [4], Grab-cut [5], TVSeg [6] and Geodesic matting [7]. Among these, those utilizing the Markov random field framework (graph-cut, Grab-cut, etc.) have shown more potential in comparison with the others. Based on that, some works like [8], [9], [10] have studied the application of this framework for automatic segmentation.

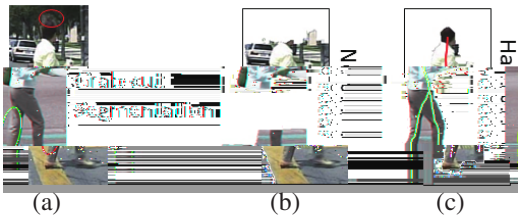
One of major problems when targeting human body is the large variations in its shape. This makes the task of modeling difficult and has led to different types of models to be introduced in the literature from simple skeletal to sophisticated probabilistic shape or part models. In addition, the variations of human clothings make the task much more complex because aside from the numerous combination to the texture and color of the cloths changes in the illumination can affect the perceived color to a great degree. As a result, many of the methods that use color as their primary parameter for differentiating between foreground and background like Graph-cut [1], Grab-cut [5], Gulshan et al.'s method [8], and some others will have serious problems in such cases. The problem also manifests itself when the texture of different regions on the body changes along side the color as depicted in Fig. 1.

In this paper, we propose an automatic system for segmenting human subjects using Statistical Shapes Models (SSM) feedback and an upgraded version of Grab-cut which incorporates texture feature. By incorporating color, shape, and texture factors, we show that the segmentation accuracy of the system can be improved and the above mentioned problems can also be avoided. Our experiments shows that the proposed system has an acceptable accuracy compared to the state-of-the art interactive methods and much better than the conventional ones.

The paper is organized as follows: In section 2, we will consider some background researches done in this field. Section 3 will explain the proposed method alongside explanations on statistical shape models, texture feature and modified Grab-cut. Section 4 will show the experimental results of the proposed method.

II. Related Works

As mentioned earlier, various segmentation methods exist in the literature. In this section, first some related methods will be reviewed. After that, a brief introduction to SSM and Grab-cut will be presented.



(a) (b) (c)

Fig. 1. An example of how using just color factor for segmentation may fail in Grab-cut [5], (a) input image in which some parts have similar color distribution to the background. (b) segmentation result using just a rectangle around the subject of interest, (c) adding some foreground seeds to the selected region.

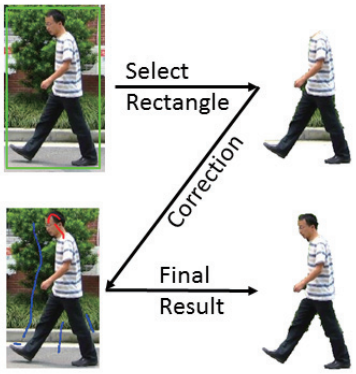


Fig. 2. An example of the segmentation process by the Grab-cut method.

A. Interactive Segmentation

Rother et al. [5] in 2004, introduced the now famous Grab-cut segmentation algorithm which tries to segment the foreground object based on a simple polygon (usually a rectangle) drawn by the user. For this, their system first learns two Gaussian Mixture Models (GMMs); one for foreground and one for background based on user selection. By turning the image into a graph and using max-flow/min-cut method iteratively, these models are refined until the object is segmented. The user can make some corrections in case of need. An example of the segmentation process by the Grab-cut method is presented in Fig. 2. M. Tang et al. [11] have tried to replace the energy model being used in Graph-cut based methods like Grab-cut [5] with a new, much simpler energy formula and also proposed an L_1 distance measure for minimizing the appearance overlap between a foreground object and its background. Utilizing this, their method segments a foreground object in one iteration and iterative energy minimization like the one proposed in Grab-cut is unnecessary.

Kuang et al. [12] use user input polygons (one for foreground, one for background) to learn color feature, texture feature and a smoothing parameter from the input image. Their method maximizes a weighted energy function margin for estimating the parameters iteratively, and at the same time segments the image. The notable point of this method is that, it learns aforementioned parameters (color, texture, smoothness) specific to the input image. This is mentioned in the work to be much better than pre-setting parameters by training the system beforehand.

(a) (b)

Fig. 3. Difference between using just color and using color & texture information. (a) Grab-cut (just color). (b) Proposed method (color & texture).

B. Automatic Segmentation

Zhang et al. [13] have proposed a video object segmentation method using a Directed Acyclic Graph (DAG) for object detection and segmentation. They assume as a basic principal that generally objects can be described with locally smooth motion trajectories, and try to find the primary object in the image series from available proposals. The proposals are created using optical flow, and by using dynamic programming, they try to find the best candidate in series of images.

Gulshan et al. [8] have utilized Microsoft Kinect for generating a training dataset, using depth information. They then used the generated dataset to train a classifier and proposed an automatic segmentation algorithm. For that they first extract HOG features from images in the dataset and then train a classifier. In the segmentation stage, the classifier generates a rough segmentation of the input image which is then refined using a local Grab-cut segmentation.

Prakash et al. [14] proposed an automatic object segmentation system by combining Active contour (snakes) [15] with the Grab-cut algorithm. The segmentation process is divided into two parallel procedures in their work; The active contour will try to find the boundary of the object from outside while the grab-cut tries to do it from inside. By combining the results of the two procedures, they then show that the segmentation accuracy increases in comparison with both original methods.

III. Texture Aware Grab-cut Segmentation

A. Main Idea

Trying to segment human subjects (other objects are the same) just based on the color distribution (like the color distance used by grab-cut) would not always result in desired results. Especially, since the human body is composed of different parts and humans wear various clothes, using just color for segmentation will lead to miss-segmentations like the one presented in Fig. 3. On the other hand, even if the color changes under different illuminations, the texture is known to be invariant. Also as mentioned in the work of Zhou et al. [16], texture is of a semi-local nature, so using it in combination with color information could be a very helpful asset in segmentation.

TABLE I
Accuracy [%] comparison between comparative methods.

	Dataset		
Method	Private dataset	PennFudan dataset	Average
Grab-cut	72.22	79.00	75.61
Watershed	79.33	80.74	80.04
One-cut	83.89	85.76	84.83
SSFSeg	85.43	80.54	82.99
Proposed	87.07	83.05	85.06

TABLE II
Comparison of F1 measure [%] between comparative methods.

	Dataset		
Method	Private dataset	Pennfudan dataset	Average
Grab-cut	67.21	75.57	71.89
Watershed	76.01	78.92	77.46
One-cut	80.41	82.81	81.61
SSFSeg	75.44	68.74	72.09
Proposed	80.99	76.77	78.88

samples for training the SSM model have been used. At each stage of sample generation, we created $N = 50$ samples from which, one was selected randomly for segmentation (the mean shape was used for the first segmentation). As for the number of times for local and global refinements, $N_l = 4$ and $N_g = 3$ were selected in this experiment, respectively.

Two datasets have been used for our experiments. The first dataset was a private set of 180 images from different human subjects (full body) in different situations which we created based on data available in our laboratory. All images were taken with an in-vehicle camera and were color images with different sizes. The images were all taken during day time.

The second data set was derived from "Penn-Fudan database for pedestrian detection and segmentation" introduced by Wang et al. [21]. It contains 230 image with different sizes and all taken during day time.

The comparison was done between the original Grab-cut segmentation [5], Watershed [22], One-cut [11], SSFSeg [17], and the proposed method. The average segmentation accuracy and F1 measure was calculated for each method as depicted in Tables I and II.

For Grab-cut and Watershed segmentation methods, the code provided by the OpenCV open source library [23], and for One-cut Segmentation, the code provided by Tang el al. [24] were used.

As for the accuracy we used:

$$\text{Accuracy } (\%) = \frac{\text{TN} + \text{TP}}{\text{TN+TP+FP+FN}} \times 100 \tag{10}$$

And for F_1 measure, we used:

$$F_1 \ (\%) = \frac{2\text{TP}}{2\text{TP} + \text{FN+FP}} \times 100 \tag{11}$$

in which TP, TN, FP and FN show respectively the number of foreground pixels segmented correctly, the number of background pixels segmented correctly, the number of foreground pixels segmented as background, and the number of background pixels segmented as foreground. Figures 8 and 9 show the segmentation results by the proposed system and its comparison to other methods. As it can be seen, the system performs with higher accuracy compared to conventional methods. Tables I and II also present the average accuracy and F1 measure which also proves that the proposed method

(a) (b) (c) (d) (e)

Fig. 8. Comparison between segmentation methods (Private Dataset). (a) Proposed method, (b) SSFSeg, (c) One-cut, (d) Grab-cut, (e) Watershed.

have better performance on accuracy while improving the F1 measure for SSFSeg [17] by 6%.

V. Conclusions

In this paper, we presented a method that can automatically segment pedestrians in images. The system incorporates texture alongside the color feature to improve the segmentation results. Our experiments show that using texture can help to improve the segmentation accuracy especially in cases where the color information is not enough to segment the desired object correctly.

It is also good to note that the proposed system uses and generates full body silhouettes at SSM stage so it does not consider the case of occlusions which is one of the cases we want to include in our future work. Also, the method explained in this paper expects the output of a human detector algorithm