AUTOMATIC PEOPLE SEGMENTATION WITH A TEMPLATE-DRIVEN GRAPH CUT

Cyrille Migniot, Pascal Bertolino, Jean-Marc Chassery

CNRS Gipsa-lab DIS, Grenoble, France

ABSTRACT

This paper presents a new fully automatic method for segmenting upright people in the images. Is is based on the efficient graph cut segmentation. Since colour and texture prevent from discriminating this particular class, silhouette shape is used instead. The graph cut is guided by a non-binary template of silhouette that represents the probability of each pixel to be a part of the person to segment. Subsequently, part-based template is used to better take into account the different postures of a person. Our method is close to real time and is tested on a large person dataset.

Index Terms— people segmentation, graph cut, template of silhouette, part-based template.

1. INTRODUCTION

To detect and segment people, numerous methods use binary templates. A template is a model used to characterize the elements of a class. In the class of upright people, the one we are interested in, the possible postures are fairly varied. In the form of a binary mask, a template represents the silhouette for one of these typical postures. A catalogue of templates contains all the postures that a person may take. These templates are then compared one by one to the attributes of the image (often the edges [1]). If the comparison is positive, a person is detected and the silhouette corresponding to his/her posture is evaluated. The final segmentation is obtained by slightly adapting this silhouette to the edges in the image [2]. To reduce the computing time, [3] makes a hierarchical repartition of the templates. Thus, the comparisons are not performed with all the templates (using a template depends on the result of the previous comparisons). Finally, [4] divides the body in three parts (head-torso, upper legs and lower legs) and searches the best template for each part.

Concerning shape-guided graph cut segmentation, a number of proposals has been done. The original graph cut method was modified by adding a third term to its energy function to take into account the shape with level sets [5]. The region or the boundary term (or both) of the energy function were modified [6, 7, 8]. [9] proposed an interesting but rather time-costly solution where a pre-image obtained from a training set with a Kernel PCA was iteratively updated.

The work presented in this paper only deals with segmentation. The preliminary people detection is carried out with the method by [10] that provides normalized windows centered on the person (see examples figure 5).

The graph cut as defined by [11] is an efficient segmentation method. One of its main advantage is its easy interaction with the user. As a consequence, it is widely used in image segmentation [12].

The graph is made up of nodes corresponding to the image pixels. **Neighborhood links** (n-links) connect the nodes to model the pixel adjacency. Two special nodes called source and sink are added to the graph to materialize the foreground and the background. **Terminal links** (t-links) connect each pixel to the source and to the sink. A weight is associated to each of the links. Then a graph cut that minimizes the sum of the weights of the cut edges is calculated. Pixels that are on the same side of the cut as the source belong to the foreground while the other pixels belong to the background.

In order to take into account the image content, the n-link weights are related to the image boundaries and the t-link weights are related to the probability of a pixel to belong to the foreground and to the background. These probabilities are generally based on the color repartitions since the user may teach the graph how foreground and background look like by drawing scribbles in the image. This method is quite efficient but quite general. In this paper, we present a new technique for adapting graph cut to the characteristics of people segmentation without any user interaction. Our contribution is first to weight the graph cut with a learned mean silhouette called template and secondly to adapt this technique by computing several graphcuts on each body part and keeping the ones that give the best response to build a part-based template.

In section 2, we introduce a non binary template learned from a training dataset, that represents the probability of a pixel to belong to the silhouette of a person. This probability is used to initialize the weights of the t-links.

Then in section 3, the segmentation is refined in order to adapt to the various postures: the image and thus the body is divided in several parts. For each part, several sub-templates are tested and the final template (called part-based template) is obtained by concataining the best sub-templates.

Finally, in section 4 our two approaches (single template and part-based template) are evaluated.

2. GRAPH CREATION

The graph is achieved from a window encompassing the person provided by the detection process. A source F and a sink B represent the foreground and the background. Each pixel of the window is a node of the graph and is connected to its neighbors by n-links and to F and B by t-links. The links are weighed in order to introduce image information in the graph representation. The relative strength between n-links and t-links weights is tuned using two coefficients (α and β).

2.1. Neighborhood links

N-links weights correspond to a penalty for discontinuity between the pixels. Thus it is logical to represent them with the image boundaries. The intensity difference is used like in the work of Boykov et al [11]. Let I_p be the intensity of pixel p and I_q be the intensity of a neighbor pixel q. The weight associated to the n-link between p and q is defined by:

$$\omega_{pq} = \beta e^{-\frac{|I_p - I_q|^2}{2\sigma^2}} \qquad (1)$$

where σ is the filtering realized by the exponential. The small values of this function represent the strong contour probabilities.

2.2. Terminal links

T-links connect each pixel to F and B. For a given pixel, the cut passes through only one of these two links which indicates to which region the pixel is assigned. The way these links are weighted must reflect the probability of the pixel to belong to foreground or background. [11] expresses this probability using color distribution. But the wide variety of colors among people and in the background make this attribute too few discriminative when dealing with people. So we make a template of the size of the detection window that gives for each pixel its probability to belong to a human silhouette. This template is made from a training set of 200 silhouettes [13] well representative of upright postures (figure 1). Let t_p be the probability

Fig. 1. Template (i.e. mean silhouette) made from a dataset of 200 binary silhouettes.

of the template for pixel p. The t-link weights for F and B are defined by:

$$\omega_p^F = -\alpha ln(1 - t_p) \qquad (2)$$

$$\omega_p^B = -\alpha ln(t_p) \qquad (3)$$

3. A PART-BASED TEMPLATE

The template presented in the previous section takes into account all the postures but penalizes the ones with a low occurrence, in particular when legs are spread. The idea is then to use a template adapted to the posture itself. It would be possible to perform a graph cut with as many templates as there exist different postures. But the number of postures is high and the processing cost would be high too. We propose to process independently the different parts of the body: the silhouette is divided into five parts: head, right and left part of the torso including the arm, right and left leg. For each of these parts, some sub-templates are built using the corresponding postures in the people dataset (figure 2). The number of sub-template is proportional to the variability of the body part and can be tuned according to the needs if the dataset contains the corresponding cases. We chose one sub-template for the head, five for each side of the torso including the arm and nine for each leg. For each part, a graph cut is carried out using each of the n sub-templates. The best of the n max-flows (i.e. the lowest sum of the weights among the n graph cuts) indicates which sub-template fits the best any part of the silhouette.
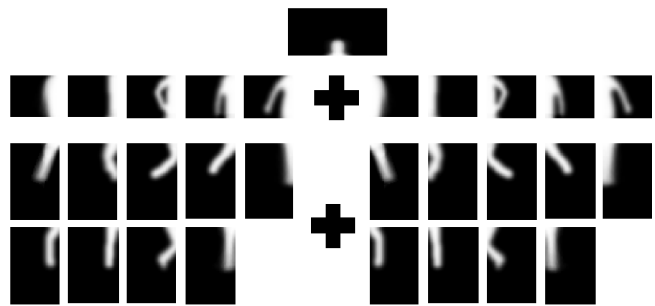

Fig. 2. 1 (head), 2×5 (torso and arms) and 2×9 (legs) sub-templates to adapt to specificities of the five silhouette parts

The part-based template is simply the concatenation of the five best sub-templates (figure 3). Finally, in order to preserve continuity of the silhouette between its different parts, a graph cut is performed on the whole window using the part-based template obtained above. A small number of sub-templates is sufficient. Indeed, since the n-links will adapt the cut to the edges, the templates just have to promote a state (unstuck arms, bended legs, ...) and don't have to perfectly fit the silhouette.
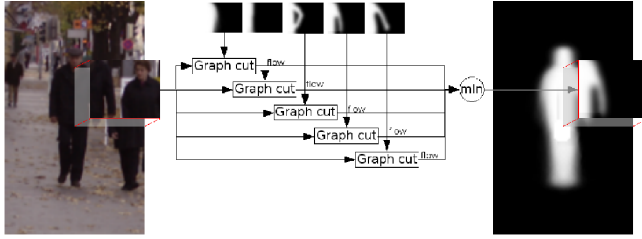
Fig. 3. For each part of the detection window (here the left part of the torso), several graph cuts are carried out using several sub-templates. The one that provides the minimal flow cut is added to the part-based template

4. PERFORMANCES

Our method was evaluated with tests on a set of 400 images of people from the INRIA Person Dataset. The $F_{measure}$ and the Yasnoff measure [14] evaluate our method by comparing a ground truth segmentation performed manually with the results obtained. The two measures are used for each of the 400 segmented images and a mean is calculated for both. The processing times are obtained with a non-optimized C++ implementation running on a 3GHz Pentium D.

4.1. Optimizing the parameters

In order to optimize the process, we must determine the values of parameters α, β and σ that give the best performances. Tests have been made with different values of these parameters. The results given in figure 4 are obtained with a part-based template but the ones obtained with a single template are very similar.

In order to obtain the best segmentation, i.e. to minimize[1] the Yasnoff measure and to maximize the $F_{measure}$, in the sequel the values $\sigma = 9$, $\alpha = 12$ and $\beta = 60$ are chosen.

4.2. Single or part-based template?

Since we have presented two versions of the template, their respective performances must be compared. First, the process with a single template is logically faster: a 96×160 pixel window is processed in a mean of 12 ms as against a mean of 70 ms with the part-based template. The part-based template provides both better $F_{measure}$ and Yasnoff measure as shown in table 1. In the great majority of the cases, the part-based template gives a better adapted and more accurate segmentation (figure 5).

---

[1]in order to facilitate the coming reading, the sign ↓ indicates a measure to minimize and the sign ↑ a measure to maximize

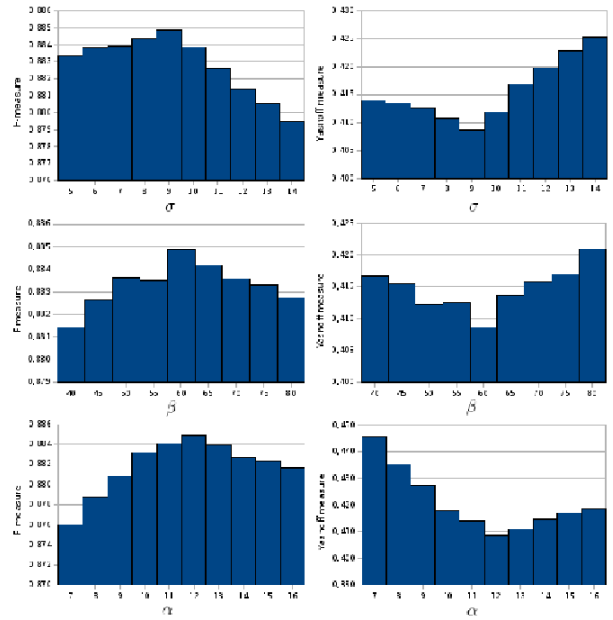**Fig. 4**. Results of the tests carried out to optimize the values of σ (first row), β (second row) and α (third row). The $F_{measure}$ (↑) left column and the Yasnoff measure (↓) right column give an evaluation of the segmentation quality.

Measure	Single template	Part-based template
$F_{measure}$ (↑)	0,8813	0,8849
Yasnoff (↓)	0,4178	0,4087

Table 1. Segmentation performances over a set of 400 images

5. CONCLUSION

In this paper we have proposed a new segmentation method that adapts the well-known graph-cut method to the particular case of persons. To do this, we have introduced templates to evaluate the shape of the silhouette. The template is either generic for the class or adapted to the person posture. The process is efficient and close to real time.

The segmentation of video sequences and the possibility of an easy interaction with the user are among the main advantages of the graph cut approach. Future relevant works would be to adapt our method to videos and to permit an efficient user interaction to segment people in difficult cases.

6. REFERENCES

[1] L. Zhao and L.S. Davis, "Closely coupled object detection and segmentation," in *International Conference on Computer Vision*, October 2005.

[2] M.D. Rodriguez and M. Shah, "Detecting and segmenting humans in crowded scenes," in *International Conference on Multimedia*, 2007, pp. 353–356.

Fig. 5. From left to right: initial image, segmentation obtained with a single template and with a part-based template. In most of the cases, the segmentation is more accurate with the part-based template

[3] D.M. Gavrila and J. Giebel, "Shape-based pedestrian detection and tracking," *Intelligent Vehicle Symposium*, vol. 1, pp. 8–14, June 2002.

[4] Z. Lin, L.S. Davis, D. Doermann, and D. DeMenthon, "Hierarchical part-template matching for human detection and segmentation," *International Conference on Computer Vision*, pp. 1–8, October 2007.

[5] D. Freedman and T. Zhang, "Interactive graph cut based segmentation with shape priors," *International Conference on Computer Vision & Pattern Recognition*, vol. 1, pp. 755–762, 2005.

[6] H. Wang and H. Zhang, "Adaptive shape prior in graph cut segmentation," *International Conference on Image Processing*, pp. 3029–3032, September 2010.

[7] G.G. Slabaugh and G. Unal, "Graph cuts segmentation using an elliptical shape prior," *International Conference on Image Processing*, vol. 2, pp. 1222–1225, September 2005.

[8] P. Das, O. Veksler, V. Zavadsky, and Y. Boykov, "Semi-automatic segmentation with compact shape prior," *Image and Vision Computing*, vol. 27, pp. 206–219, January 2009.

[9] J. Malcolm, Y. Rathi, and A. Tannenbaum, "Graph cut segmentation with nonlinear shape priors," *International Conference on Image Processing*, vol. 4, pp. 365–368, October 2007.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *International Conference on Computer Vision & Pattern Recognition*, vol. 2, pp. 886–893, June 2005.

[11] Y.Y. Boykov and M.P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," *International Conference on Computer Vision*, vol. 1, pp. 105–112, July 2001.

[12] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, pp. 309–314, August 2004.

[13] C. Migniot, P. Bertolino, and J.M. Chassery, "Contour segment analysis for human silhouette presegmentation," *International Conference on Computer Vision Theory and Applications*, May 2010.

[14] S. Philipp-Foliguet and L. Guigues, "Mutlti-scale criteria for the evaluation of image segmentation algorithms," *Journal of Multimedia*, vol. 3, no. 5, pp. 42–56, 2008.