Identification of Space-Time Flows of Moving Objects

Núbia Rosa da Silva
Faculty of Computer Science – FACOM
Federal University of Uberlandia
Uberlandia, MG, Brazil
nubiasrosa@gmail.com

Celia Aparecida Zorzo Barcelos
Faculty of Mathematics – FAMAT
Federal University of Uberlandia
Uberlandia, MG, Brazil
celiazb@ufu.br

Eraldo Ribeiro
Department of Computer Sciences
Florida Institute of Technology
Melbourne, FL 32901, USA
eribeiro@fit.edu

Marcos Aurélio Batista
Department of Computer Science
Federal University of Goias
Catalao, GO, Brazil
marcos@catalao.ufg.br

Abstract— The goal of our work is to identify flow lines of moving objects in videos. These flows can then be used to represent the underlying movement. Unlike studies using shape-based representations, this proposal uses all the image information to build denser flows for motion representation. We begin by segmenting each video frame into a large set of non-overlapping uniform-color regions called superpixels. These superpixels are then tracked over a number of frames to generate flow lines. The ideas reported in this work can be applied to the task of creating motion patterns, which in turn are important for identifying complex movements of objects and predict behavior. Experiments are performed on a set of synthetic and real-world motion sequences.

Keywords-flow lines; flow; superpixel, motion pattern.

I. INTRODUCTION

In general, movements in videos are complex and articulated. Obtaining a mathematical representation of movement, although ideal for motion recognition, is a hard and challenging task. Despite a number of developments by the computer vision community, action recognition is still an open problem. This problem is an important component of a variety of applications, such as, automated video surveillance, human-computer interface, video indexing, gesture recognition, and the analysis of sport events.

The knowledge of motion patterns has been used to detect complexity in moving objects and to predict behavior, but also to recognize actions events of particular interest in a video. Indeed, an object's movements can be described in a number of ways based on their shapes, silhouettes, primitive geometric shapes, optical flow, and others.

In this work, we investigate the idea of representing object motion using dense trajectories of flow lines. Flow lines describe the space-time variation of motion in videos. The main idea underlying our work is based on the shape-flow descriptors proposed by Jiang and Martin [2] for human motion recognition. Jiang and Martin [2] use the time-varying deformation of the person's silhouette contour to characterize an action.

Nevertheless, contour-based shape flows can be hard to obtain when object's motion happens within contour boundaries. For example, human motions containing self-occluding body will generate complex shape flows. These flows may not represent well the underlying motion. Another example is when the object's contour remains unchanged during motion, e.g., a rotating texture ball. Shape flows would not be able to describe the ball's rotation. In this paper, we investigate a way to improve shape-flow motion representation by using the object's surface texture, i.e., we will produce flow trajectories using features from both inside and outside the object, and not limited to the outline of it. This paper reports preliminary results on this idea.

The remainder of this paper is organized as follows. The method's key elements are detailed in Section III. Section II discusses the related work. Finally, in Section IV, we show the experimental results.

II. RELATED WORK

Jiang and Martin [2] represent human movement in videos by means of flow lines, which are obtained using the Iterative Conditional Modes (ICM) algorithm [1]. Here, ICM is used to characterize motion points in adjacent frames by minimizing the matching cost while maximizing the point-neighborhood consistency at the object's outer contour. To detect contour points, the authors used the Canny edge detector, and neighborhood information was obtained by Delaunay triangulation performed on selected edge points. The final result is a dense motion field. Flow lines are then obtained by connecting motion fields frame by frame. Measuring the similarity of flow lines between the video analysis and template that represents the action is not a simple problem, and the alignment of shape flows is an optimization problem.

Ke et al. [5] presents a motion representation where using volumetric features. They propose a shape-based representation based to align videos. The method does not require background subtraction. The shape-based alignment consists of the extraction of spatial-temporal regions and the subsequent alignment of these regions. Unsupervised clustering is used to segment the video into three-dimensional volumes that maintain the similar appearance within the video volume. These volumes are called supervoxels, based on the same concept of superpixels defined by Ren and Malik (2003) [6]. Usually, the object's boundary in space-time volumes corresponds to the borders of supervoxels. Here, object's boundary mostly matches the superpixels partitioning [3].

III. PROPOSED METHOD

In the study by Jiang and Martin [2], flows are created at points located only at the edge. As a result, we obtain the representation of the motion shape. Tracking points of the object's border can be problematic to represent motion due to the location uncertainty of points along the edge. We propose to modify this representation by including points detected inside the moving object's contour to represent the object's movement. Our final goal is to describe an object's motion through dense set of flow lines obtained by tracking the centroid of superpixels as in [Ke et al. 2007]. Our method can be divided into three parts. The first one is to segment each frame of video in superpixels. After this segmentation, the superpíxel is characterized as a model of appearance, and finally, superpixel regions are traked frame by frame using a metric.

The superpixels segmentation used in our method is the same as the one proposed by Ren and Malik [6], in which a superpixel is a local group of pixels with common characteristics between the grouped elements. In [6], superpixels are fairly homogeneous in shape and size, to simplify computing costs.

Given a set of the superpixels obtained for each video frame in a motion sequence, it is possible to track superpixel locations and speed across video frames. The tracked location is represented by the centroid of each superpixel. Therefore, the set points of each frame T is represented:

$$S^t = \{X_1^t, \dots, X_N^t\}, \quad X_i^t = (x_i^t, y_i^t) \qquad (1)$$

where X_i^t is the centroid of the i-th superpixel at the t-th frame. For each superpixel i in frame t, the local appearance model used is given by:

$$P(X_i^t) = \mu_{Xi} \qquad (2)$$

where μ_{Xi} is the mean of all pixels that form superpixel i. The extension of the flow line that passes through X_i^t is X_k^{t+1}, where

$$k = \arg(\min_j |P(X_j^{t+1}) - P(X_i^t)|) \qquad (3)$$

Flow lines obtained are composed by trajectories of each tracked point. Movements do not need to be defined across the whole duration of the video. Fig. 1 illustrates the concept of motion flow used in this work.

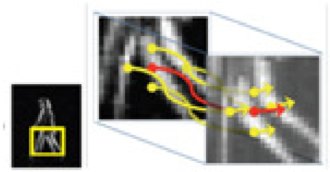

Figure 1. Consistent flow of space-time method used in this work.

Flow lines will be grouped according to their similarity to define motion patterns. These patterns will be used to establish a new representation of movements of objects in video. Unlike studies that use a representation based on the form, this work uses all the image information to build denser flows to better represent the motion.

IV. RESULTS

The aim of our experiments is to demonstrate the good performance of our method for identification of flow lines of moving objects in video. The tracking of areas has been implemented in Matlab. We use the superpixels algorithm implementation available in http://www.cs.sfu.ca/~mori/research/superpixels/ [3]. The images used in our experiments were of size 400 × 300 pixels.

To define the similarity among the superpixels were used, in principle, the concepts of intra-region similarity and inter-region dissimilarity. According to the definition of superpixels proposed by Ren [6], a superpixel should contain elements that preserve texture and brightness and low contour energy inside the region. Additionally, different texture and brightness should only be found in different superpixels. There are two key problems here. First, sometimes superpixels of different colors are used to represent the same texture. Secondly, some superpixels include information across the original object's contour. Fig. 2 shows an example of the object contour within the same superpixel. These may be regions with the same features that are separated into distinct superpixels, where we use a ball set in three colors on a white background to illustrate the problem. Fig. 4 (a) also shows an example of inconsistency in the superpixels segmentation. Fig. 2 shows an example from [4], where the right hand of the man is at the same superpixel with lawn.

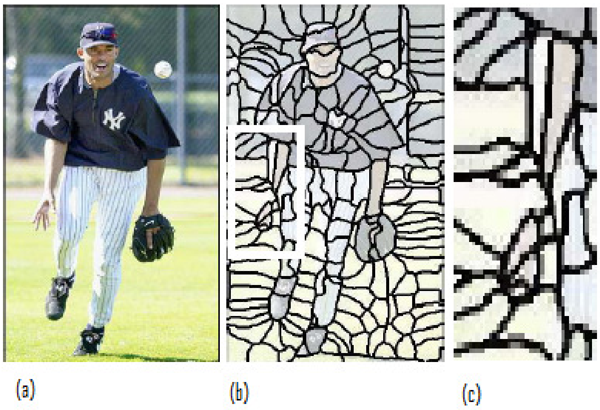

Figure 2. Superpixels segmentation. (a) Original image from [4], (b) Segmentation of (a) and (c) zooms the white box in (b) showing the hand with lawn.

In Fig. 3, it is possible to analyze the result of tracking superpixels, defined by segmenting superpixels. The estimated trajectory is relatively noisy. Moreover, two flow lines join together due to color similarity in different superpixels.

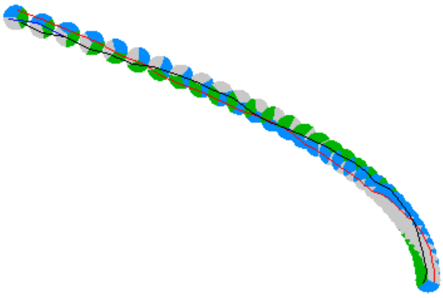

Figure 3. Flow lines using segmentation in superpixels.

This paper defines the concept of ideal superpixel. A superpixel is considered ideal when only it includes pixels having characteristics very similar, as color, texture, and it is not acceptable to have two regions with different textures within the same superpixel. The internal region in an ideal superpixel must maintain its nuance. Fig. 4 (b) shows an example of an ideal superpixel, where we also show the segmentation into superpixels (Fig. 4 (a)), showing the divergence between the two segmentations.

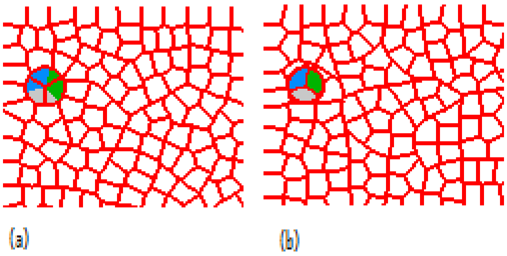

Figure 4. Comparison between the two segmentations. (a) Contours within the same superpixel. (b) Ideal segmentation.

Using ideal superpixel to trace the flow line of a sequence of images, one can obtain more consistent flow lines, as shown in Fig. 5.

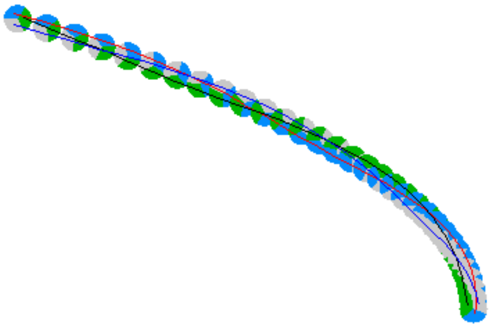

Figure 5. Flow lines using segmentation with ideal superpixel.

Another experiment was conducted with a video with more complex background. The results used in superpixels segmentation were not as satisfactory as those obtained using simple background and using the ideal segmentation. Fig. 6 (top) shows the first frame of the video segmentation that used superpixels segmentation. Fig. 6 (lower left) shows only the zoom of the ball segmentation. And Fig. 6 (bottom right) shows the ball using the ideal segmentation.

The first segmentation cannot define exactly different colors inside the ball and even its contour. Thus the ball three colors were divided into five superpixels, which were tracked by video frames, generating inconsistent flow lines that do not represent the motion. This result is shown in Fig. 7. Moreover, independent of the image background, when the ideal segmentation is used, it is possible to obtain concise flow lines. The flow lines that use ideal superpixel segmentation can be seen in Fig. 8.

The ideal superpixel segmentation was empirically tested only in simple videos. The challenge is to find a segmentation model to identify each superpixel.

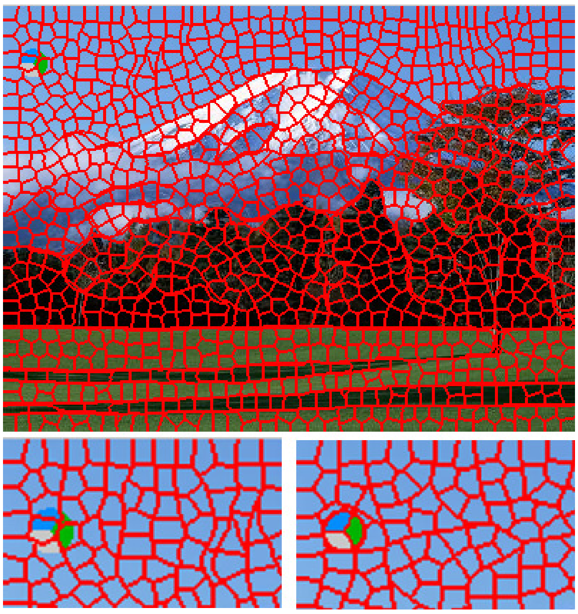

Figure 6. Segmentation in superpixels of [3]. Top: first frame of the video segmentation that used superpixels segmentation. Lower left: zoom of the ball segmentation. Bottom right: ideal segmentation.

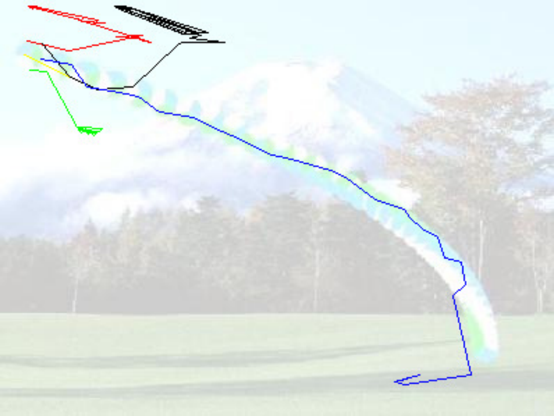

Figure 7. Flow lines using segmentation in superpixels of [3].

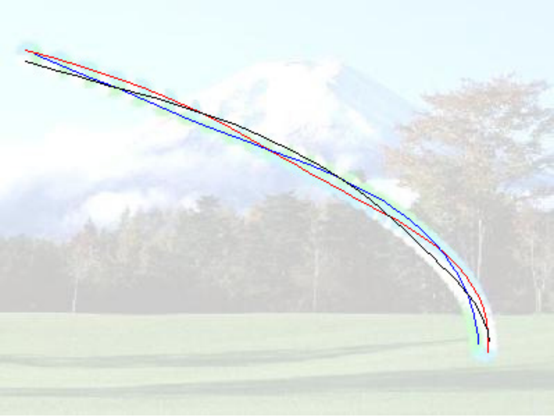

Figure 8. Flow lines using ideal superpixel segmentation.

Fig. 9 shows the comparison of flow lines using segmentation in superpixels of [3] and segmentation using ideal superpixel. We realize that flow lines using ideal superpixels are more consistent.

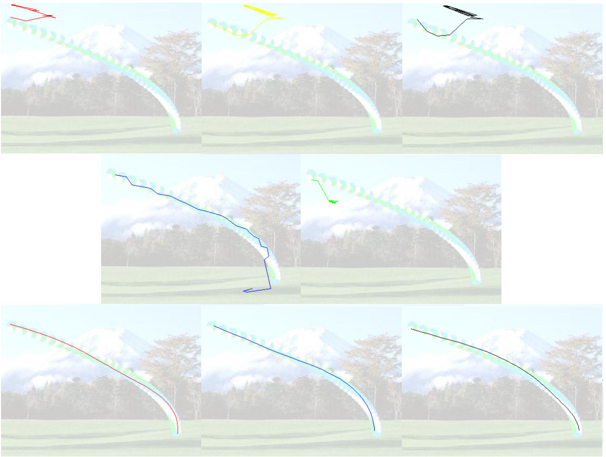

Figure 9. Comparison of flow lines (first and second lines) using segmentation in superpixels of [3] and (third row) segmentation using ideal superpixel.

V. CONCLUSIONS

Video-based motion identification can be computed by analyzing motion patterns obtained from flow lines representing these movements. The work described in this paper focused mainly on flow-line detection, using the concept of superpixels to identify areas of interest on the moving object's surface, and tracking these regions.

It is worth point out that we are presenting our preliminary results on trying to generate denser flows to represent the movement. It is also important to highlight that this work is useful to define motion patterns, and the patterns can be used to detect complex movements of articulated objects and to predict behavior.

However, we encountered a number of problems with the segmentation method. New segmentation methods are currently being considered. Results obtained from the ideal segmentation experiment are promising.

ACKNOWLEDGMENT

This work is supported by CNPq.

REFERENCES

[1] Besag, J. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society B-48, 1986, 259–302.

[2] Jiang, H. and Martin, D. R. Finding actions using shape flows. In ECCV (2), pp. 278–292, 2008.

[3] Mori, G. Guiding model search using segmentation. In Proc. 10th Int. Conf. Computer Vision (2005), vol. 2, pp. 1417–1423.

[4] Mori, G., Ren, X., Efros, A. A., and Malik, J. Recovering human body configurations: Combining segmentation and recognition. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 2 (2004), 326–333.

[5] Ke, Y., Sukthankar, R., and Hebert, M. Spatio-temporal shape and flow correlation for action recognition. In Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on (June 2007), pp. 1–8.

[6] Ren, X., and Malik, J. Learning a classification model for segmentation. In Proc. 9th Int'l. Conf. Computer Vision , 2003, vol. 1, pp. 10–17.