Indonesian Journal of Electrical Engineering and Computer Science
Vol. 15, No. 1, July 2019, pp. 511~516
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v15.i1.pp511-516

Automatic moving foreground extraction using random walks

Idir Boulfrifi, Khalid Housni, Abdelaziz Mouloudi
Ibn Tofail University/Faculty of Science, Morocco

Article Info

Article history:

Received Dec 18, 2018
Revised Jan 21, 2019
Accepted Mar 4, 2019

Keywords:

Automatic motion segmentation
Energy function
Graph
Random walk algorithm

ABSTRACT

In this paper, we propose a method for automatic foreground extraction in video frames by analyzing the spatiotemporal aspect. We divide our contribution to tree steps: Automatic seeds detection, formulating the energy function, and using the random walk algorithm to minimize this function. First, we detect seeds by extracting a sparse of good features to track in the current frame and compute the difference between those pixels and its adjacent in the previous frame, the difference of pixels is treated in HSV color space to make the result more accurate, we thresholds this difference, and we classify moving and stationary pixels. Secondly, we formulate our foreground extraction as a graph based problem, then we define an energy function to evaluate spatiotemporal smoothness. Finally, we applied the random walk algorithm with seeds detected in the first step to minimize the energy function problem, the solution leads to evaluate the potential that every pixel in the video sequences is marked in motion or a stationary pixel. We suggest that our unsupervised method has the potential to be used for many kinds of motion detection and real-time video.

Corresponding Author:

Idir Boulfrifi,
Faculty of Science, Ibn Tofail University,
Kenitra, Morocco.
Email: iboulfrifi@gmail.com

1. INTRODUCTION

Video foreground extraction aims to classify pixels of video frames to pixels belongs to the foreground and background pixels, more generally the most existed approaches are based on optical flow, background subtraction, images difference, graph based approach, clustering algorithm, and deep learning. It's can be used in semantic scene understanding, traffic surveillance, recognition, robotic, video indexing, and many other reel-time application. A lot of research has been focused on motion detection. They can be classified into supervised [1-3] and unsupervised methods [4-11]. In the first category, the segmentation requires some initial seeds to be selected in the first frame to perform segmentation. Therefore Fan and al. [1] use a mask transfer and interpolation method, from foreground mask in source frame he estimates the foreground at an other frame. Wang and al. [3] propose an algorithm to segment video based on a level set framework and an appearance model, this algorithm requires only a single finger touch the object in the first frame. In [17] Rother and al. use iterated graph cut to extract the foreground.

The second category doesn't require any user involvement, over the past decade a lot of works are focused on analyzing information like coherence, motion, and appearance in space-time blob of video [5], [14], [8]. Wu and al. [15] propose a method that uses least squares tracking framework and learned appearance models to segment and track motion. Therefore Khoreva and al. [16] apply a method to learn the graph by exploiting edge topology and weights of the graph. Faktor and al. [4] use re-occurring regions by constructing a graph of the voting scheme of re-occurring regions across the video sequence. Vertens and al. [11] are used the convolutional neural network to predict the object label and motion status of each pixel in an image. This category takes on its importance in real time application requiring an instantaneous understanding of the scene. Motion segmentation still encounters many challenges like occlusion, camera

motion, noise, and many complex situations, specifically the automatic motion segmentation that our contribution will focused.

The first steep in our contribution consist to extract some seeds representing pixels belongs to the foreground and background pixels. the extraction is performed by detecting some good feature to track [12] in the preview frame in RGB space color and make difference between those sparse points and adjacent points in the current frame, to make the result more accurate we use HSV space color to compute the difference. Secondly, we formulate our issues as graph based problem then an energy function is defined to evaluate labeling pixels, by incorporate spatial and temporal information in video sequences. Finally, the random walk algorithm [13] is applied to minimize the energy function and get the final segmentation. The figure (1) explain an overview of our approach.

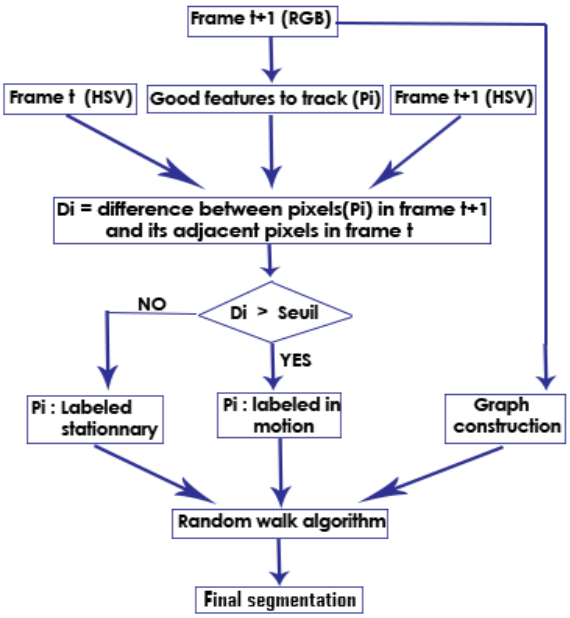

Figure 1. Overview of our approach

2. THE PROPOSED METHOD

Our approach aims to group pixels of frames video into pixels in motion and stationary pixels, The operation of affecting a label to every pixel in the frame we get the motion segmentation. In this paper, we propose a method based on the good features to track and difference to detect initial seeds as illustrated in figure (2), and random walks algorithm to minimize the formulated energy function.

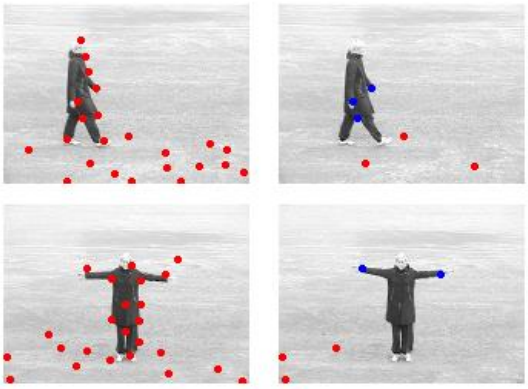

Figure 2. The left frame illustrate the good features to track, and the right frame represent the initial seeds(the blue are in motion and the red are stationary

2.1. Initial Seeds

The extraction of initial seeds still a major challenge, due that the most existing methods are based on the probabilistic model and they are many difficult situations that make the detection of initial seeds inaccurate. In our approach, to detect the initial seeds we are used [12] to extract a sparse of good feature to track in RGB color space in the current frame and performing a difference between those sparse of pixels and his adjacent in preview frame,to increase the result accuracy we have computed the difference in HSV color space. By performing a threshold on this difference we get the classification of those sparse pixels, if it's bigger than a threshold $\alpha 1$, the pixel is labeled as in motion, else if it's smaller than then a threshold $\alpha 2$, the pixel is labeled stationary. Those initial seeds will be incorporated into random walks algorithm to perform our final motion segmentation.

2.2. Energy Function

To achieve our motion segmentation, an energy function is formulated incorporating spatial and temporal information, similar to [19] we get the probability **x** that a pixel is in motion by the minimizing the energy function as follow:

$$Q[x] = Q_S[x] + \lambda Q_T[x] \tag{1}$$

Where λ is a free parameter that controls the weighting between the two energies, and Q_S represent the spatial smoothness, this energy function minimizes the edges weights between neighboring pixels, it's defined as follow:

$$Q_S[x] = x^T L x = \sum_{e_{ij} \in E} w_{ij}(x_i - x_j)^2 \tag{2}$$

The Q_T determine the temporal smoothness, this function minimizes the incoherence with predicted confidence m_i and s_i, it's defined as follow:

$$Q_T[x] = \sum_{v_i} m_i(1 - x_i)^2 + \sum_{v_i} s_i x_i^2 \tag{3}$$

The energy function (9) can be formulated in matricidal form as follow:

$$Q[x] = x^T L x + \lambda((1 - x)^T M(1 - x) + x^T S x) \tag{4}$$

Where m_i represent the probability that the pixel v_i belongs to the foreground, and s_i the probability that the pixel v_i belongs to the background. The optimization of energy leads to resolve the equation as follow:

$$(L_U + \lambda M_U + \lambda S_U)X = -B^T M + M_U \tag{5}$$

The tow matrices M and S are positive and diagonal. To determine the confidence that the pixel v_i belongs to the foreground, we evaluate the gradient to every frame pixels in two directions and we compute the difference between pixels in frame t end his adjacent pixels in frame $t + 1$. Then we can formulate the confidence $m_i \in \{0,1\}$ as follow :

$$m_i = \begin{cases} 1 & \text{if } |\nabla(g_i^t) - \nabla(g_i^{t+1})| > s1, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

Where $\nabla(g_i^t)$ represent the gradient at pixel v_i in frame t. Like (6) the confidence $s_i \in \{0,1\}$ that pixel v_i belongs to background is defined as follow:

$$s_i = \begin{cases} 1 & \text{if } |\nabla(g_i^t) - \nabla(g_i^{t+1})| < s2, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

3. RESEARCH METHOD

Random walks image segmentation have been proposed by Grady [13]. Then image is represented as a graph $G(V, E)$ where each node $v_i \in V$ represent a pixel of image and each edge e_{ij} represent connection between pixel v_i and neighbor pixel v_j. Let $n = |V|$ and $m = |E|$ where $|.|$ denotes the cardinality, edge

weight w_{ij} evaluate the similarity between connected pixels. The basic idea of random walks segmentation consist to starting a random walk from each pixel in the image and compute the probabilities of witch seeds they first arrive at. The edge weight can represent the difference in image intensity, texture information, color or other features. In [13] Grady applies a Gaussian weighting function to construct the graph.

$$W_{ij} = \exp(-\beta(g_i - g_j)^2) \quad (8)$$

$g_i$andg_j: image intensity at pixel i and j. β: free parameter. To deal with other features information we can replace $(g_i - g_j)^2$ with $||g_i - g_j||^2$. The segmentation is achieved by minimizing this energy function:

$$Q[X] = \frac{1}{2}X^T L X = \frac{1}{2}\sum_{e_{ij}\in E} w_{ij}(X_i - X_j)^2 \quad (9)$$

The solution of random walk problem consists to find a harmonic function that satisfies the Laplace equation with respect to the boundary conditions.

$$\nabla^2 x = 0 \quad (10)$$

The combinatorial Laplacian matrix L, defined as follow:

$$L_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -w_{ij} & \text{if } v_i \text{and } v_j \text{are adjacent nodes,} \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

Where L_{ij} is indexed by the vertices v_i and v_j, and $\mathbf{d_i} = \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{w_{ij}}$. The vertices are grouped into seeded nodes V_m and unseeded nodes V_u, without loss of generality nodes in L and X are ordered such that seed nodes are first and unseeded nodes are second. Decomposing equation (9) lead to :

$$Q[X_u] = \frac{1}{2}\begin{bmatrix} X_m^T & X_u^T \end{bmatrix}\begin{bmatrix} L_m & B \\ B^T & L_u \end{bmatrix}\begin{bmatrix} X_m \\ X_u \end{bmatrix} \quad (12)$$

$$Q[X_u] = \frac{1}{2}(X_m^T L_m X_m + 2x_u^T B^T X_m + X_u^T L_u X_u) \quad (13)$$

X_U and X_M correspond to potentials of the seeded and unseeded nodes respectively. Differentiating $Q[X_u]$ with respect to X_u lead to compute X_u by solving this equation :

$$L_u X = -B^T M \quad (14)$$

The final segmentation is obtained by assigning to each node v_i the label s corresponding to $Max(X_i^s)$. Where the probabilities at any node v_i will sum to unity :

$$\sum_s X_i^s = 1 \quad (15)$$

The random walk segmentation has nice properties like robustness to the weak boundary, noise, and avoidance of trivial solutions in comparison with graph cut and other segmentation algorithms. All those advantages consolidate the chose of random walks algorithm in our approach.

4. RESULTS AND DISCUSSION

We have implemented our automatic foreground extraction approach in C++ programming language using openCV, boost , and eigen libraries, on a PC with Intel(R) Core(TM) i5 CPU, 2.40 GHZ, 4 Go in RAM and Windows 7 operating system. The method was tested on several videos, and the figure (3) illustrate the obtained result on walk video sequence with $\lambda = 0.1$, and $\beta = 0.01$, $\alpha 1 = 60$, $\alpha 2 = 1$, $s1 = 20$, $s2 = 3$ graph edges is construct by V4 neighbors pixels and frame size is 160×120. the complexity algorithmic of our approach requires $O(n)$ operations, where n is the number of the frame pixels. to efficiently resolve the linear system of equations (13) matrices as represented as a sparse matrix, so instead of storing the entire

matrix, a vector of weights may be stored to decrease the computational time. finally, GPU-based implementation is encouraged for real-time processing like behavior analysis and recognition, motion analysis, event detection.

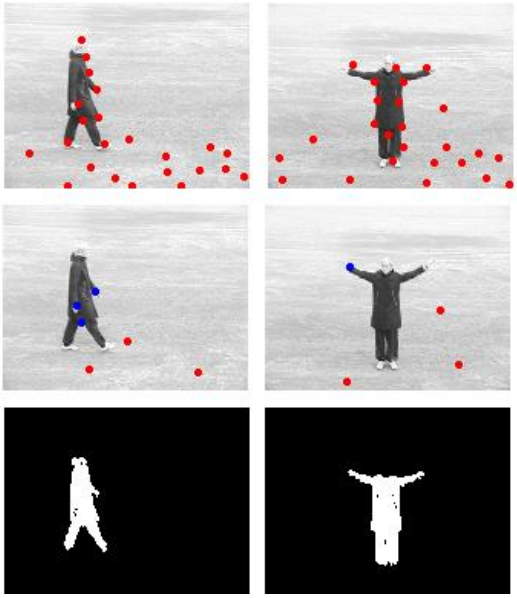

Figure 3. Experimental results using our approach on the walk video sequences, the first row illustrate the good feature to track, the second row represent the initial seeds(blue are in motion and red are stationary) and the third row display final segmentation

5. CONCLUSION

In our motion detection approach, we have presented a spatiotemporal video segmentation, by formulating the desired automatic foreground extraction as a graph based problem. In addition to spatial coherence used in image segmentation, we have applied the temporal information by introducing a likelihood term in the energy function, this term penalizes the similarity between adjacent pixels in the current frame and next frame. Like much other motion segmentation, the random walks algorithm was applied to minimize the defined energy function, and resolving our labeling problem to get the final pixels classification. The interest of our method in addition to his performance that not required any human interaction, they can be used in real time application. Our future work will improve the efficiency of our approach further and make the result more accurate.

REFERENCES

[1] Qingnan Fan, and al., "JumpCut : Non-Successive Mask Transfer and Interpolation for Video Cutout," *ACM Transactions on Graphics,* 2015; 34(6).
[2] Alireza Fathi, and al., "Combining Self Training and Active Learning for Video Segmentation," *BMVC,* 2011; 1-11.
[3] Tinghuai Wang and Bo Han and John Collomosse, "TouchCut: Fast Image and Video Segmentation using Single-Touch Interaction," *Computer Vision and Image Understanding,* 2014; 120; 14-30.
[4] Alon Faktor and Michal Irani, "Video Segmentation by Non-Local Consensus Voting," *BMVC,* 2014; 1-12.
[5] Fabio Galasso, Roberto Cipolla and Bernt Schiele, "Video Segmentation with Superpixels," *ACCV,* 2013; 760-774.
[6] Chetan Bhole and Christopher Pal, "Fully Automatic Person Segmentation in Unconstrained Video using Spatio-temporal Conditional Random Fields," *Image and Vision Computing,* 2016; 51.
[7] Jianbo Shi and Jitendra Malik, "Motion Segmentation and Tracking Using Normalized Cuts," *ICCV,* 1998; 1154-1160.
[8] Matthias Grundmann, et al., "Efficient Hierarchical Graph-Based Video Segmentation," CVPR, 2010; 2141-2148.
[9] Qian Zhang and King Ngi Ngan, "Multi-view Video Based Multiple Objects Segmentation using Graph Cut and Spatiotemporal Projections," *Journal of Visual Communication and Image Representation,* 2013; 21; 453-461.
[10] Anestis Papazoglou and Vittorio Ferrari, "Fast object segmentation in unconstrained video," *ICCV,* 2013; 1777-1784.

[11] Johan Vertens and Abhinav Valada and Wolfram Burgard, “SMSnet: Semantic motion segmentation using deep convolutional neural networks," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017; 582-589.
[12] Jianbo Shi and Carlo Tomasi, “Good Features to Track," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, 1994; 593-600.
[13] Leo Grady, “Random walks for Image segmentation," *Ieee Transaction On Pattern Analysis and Machine Intelligence*, 2006; 28(11).
[14] Chenliang Xu and Caiming XiongJason and J. Corso, “Streaming Hierarchical Video Segmentation," *ECCV*, 2012; 626-639.
[15] Zhengyang Wu, et al., “Robust video segment proposals with painless occlusion handling," CVPR, 2015.
[16] Anna Khoreva, et al., “Classifier Based Graph Construction for Video Segmentation," CVPR, 2015.
[17] Carsten Rother and Vladimir Kolmogorov and Andrew Blake, “"GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts," *ACM Trans. Graphics*, 2004; 23.
[18] Bruce D. Lucas and Takeo Kanade, “An iterative image registration technique with an application to stereo vision," *IJCAI'81 Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981; 674-679.
[19] Grady, “Multilabel RandomWalker Image Segmentation Using Prior Models," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005; 763-770.
[20] D. Zhang and O. Javed and M. Shah, “Video Object Segmentation Through Spatially Accurate and Temporally Dense Extraction of Primary Object Regions," *Computer Vision and Pattern Recognition*, 2013; 628-635.
[21] J. Shi and J. Malik, “Motion Segmentation and Tracking Using Normalized Cuts," *ICCV*, 1998; 1154-1160.
[22] Sutikno and H. Wibawa and P. Yusuf Budiarto, “Classification of Road Damage from Digital Image Using Backpropagation Neural Network," *IAES International Journal of Artificial Intelligence*, 2017; 6(4); 159-165.
[23] C. Mishra and D. L. Gupta,“Deep Machine Learning and Neural Networks: An Overview," *IAES International Journal of Artificial Intelligence*, 2017; 6(2); 66-73.
[24] M. Lalaoui and A. El Afia and R. Chiheb,“A Self-Tuned Simulated Annealing Algorithm using Hidden Markov Model," *International Journal of Electrical and Computer Engineering(IJECE)*, 2018; 8(1), 291-.298.