IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **1(2)** (2011), pp. 152–157

Complex Networks applied to Dynamic Vision Systems

Regino Criado[1], Miguel Romance[1], Angel Sánchez[2,†] and Pedro D. Suárez[2]

[1] *Department of Applied Mathematics, Rey Juan Carlos University, 28933 Móstoles, Madrid (Spain)*

[2] *Department of Computing Sciences, Rey Juan Carlos University, 28933 Móstoles, Madrid (Spain)*

Abstract. We presented an novel application of complex network theory to the visual tracking problem in videos considering multiple targets. The implemented system has been tested successfully with different videos of different spatial resolutions producing accurate targets' detection and tracking at a real-time framerate.

Keywords: Visual tracking, object detection, complex network, centrality measure.
MSC 2000: 05C90, 68T45, 05C82, 68R10.

† **Corresponding author:** angel.sanchez@urjc.es
Received: October 14, 2011
Published: October 24, 2011

1. Introduction

The purpose of video tracking is to automatically determine the positions of one or more moving objects (i.e. targets) at each of the frames in a video sequence and using only one static camera. This is one interesting topic in Computer Vision since the solution to this problem presents multiple applications to areas like: visual surveillance, traffic monitoring, facial expression analysis from video, medical imaging, among many others [7]. The tracking goal is to accurately determine the positions of the involved targets along the frames and also to efficiently perform this task (i.e. ideally at real-time frame rate). Different situations could difficult the tracking process like: variations in illumination conditions in videos, possible partial occlusions of some tracked objects in the scene change of the shape of these objects due to the camera perspective. Two major components are present in an automatic visual tracking system: target detection and tracking, respectively. Target detection is dependent on how the objects in the scene are represented (i.e. as blobs, as contours, etc). Object tracking usually involves aspects like including in the model some prior information about the scene and/or objects, considering the objects dynamics and evaluating the adjustment between the predicted targets' positions by the movement model and corresponding measured positions of these target in the frames. Some commonly used probabilistic algorithms for visual tracking [7] are Kalman filters and particle filters.

2. Complex network modeling and feature extraction from images

In the proposed visual tracking method, the first stage is to create the complex network for each single frame of the video. For the sake of efficiency, the construction of the network for a complete frame is only carried out for the first video frame. This is required in our approach to estimate initially the positions of involved targets to be tracked. For the subsequent images, where the positions of targets to be tracked have been estimated (from the previous frame), the complex network is built only for the reduced regions of pixels called regions-of-interest (ROI) which surround the position where each target was previously located. This fact reduces significantly the computational cost of the method.

The complex network modeling the scene for a color image (or a subimage) is built as follows. A graylevel version of the frame I with $N \times N$ pixels is created by averaging the corresponding pixels for the three color channels, such that for pixel $p \in I$ we have its intensity value $f(p) \in [0, 255]$. Next, we compute an watershed-based segmentation $R = R(I) = \{r_1, \ldots, r_k\}$ of I and we choose a set of pixels $X(R) = \{p_1, \ldots p_k\} \subseteq I$ such that for every

$1 \leq j \leq k$, $p_j \in r_j$. There are several methods for selecting these pixels from the segmentation R (for example, by choosing the centroid of each region, at random, and many others), but the results obtained are similar since all the pixels in a given region have similar intensity. By using these region-centroid pixels $X(R) = \{p_1, \ldots p_k\}$ as nodes we construct a weighted, sparse and spatial network $G(I) = (X(R), E)$ by defining each link weight $w(p_i, p_j)$ as follows

$$w(p_i, p_j) = \begin{cases} \|RGB(p_i) - RGB(p_j)\|_2, & \text{if } r_i \text{ and } r_j \text{ are adjacent regions,} \\ 0, & \text{otherwise,} \end{cases}$$

where $\|\cdot\|_2$ denotes the Euclidean norm and $RGB(p) = (r(p), g(p), b(p))$ is the RGB vector of pixel p. Note that in this case the weighted associated network is sparse and its number of nodes $k \ll N$, which makes that the computations on such networks be much more efficient than those previously stated in the associated networks. It is easy to check that the networks introduced in [4] or [5] can be also defined by using this model, simply by considering the appropriate *feature vector* describing some *visual* properties of each region r_j.

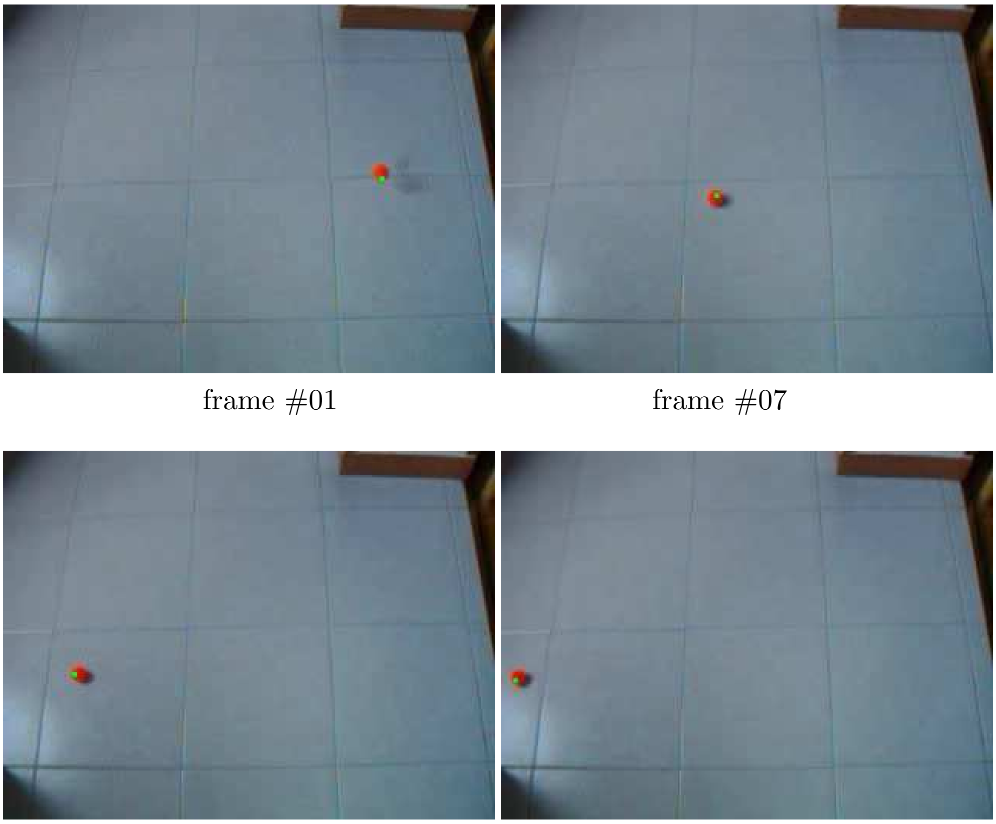
Next, we use the centrality measure proposed by the authors [2, 3], to determine the interest points in the image. This way, the interest of a graph node (i.e. image pixel) is computed as the sum of the weight links which are incident to this node. The nodes having higher interest (i.e. one node per target) are selected to be tracked in the successive frames of the video analyzed.

3. Visual tracking approach using complex networks

We describe a new approach for determining the positions of one or multiple targets at all the frames of a video sequence by using the framework of complex network analysis [1]. First, the object(s) to be tracked are selected in the initial frame of the video by the positions of the highest-values interest points [6] by applying the method described in the previous section [2]. Later, the corresponding positions of these located targets are iteratively projected to the next frame. This is done using the corresponding ROI window around each target position in previous image. In the current frame, new complex network graphs [2] are built only for each of the targets' ROI. Now, we search separately for in these reduced "ROI graphs" the most similar centrality values of interest points with respect to the corresponding ones located in the previous frame (i.e. those similar points below a distance threshold).

This way, the new position of each target in the current frame is determined. Our method is computationally efficient since each tracked interest point (where each point corresponds to one different moving object) is searched only in the corresponding ROI region and not in the whole frame. Due to the classical mathematical definition of the interest points which is of local nature, we use in this paper some local measures of the associated geometrical complex network graphs.

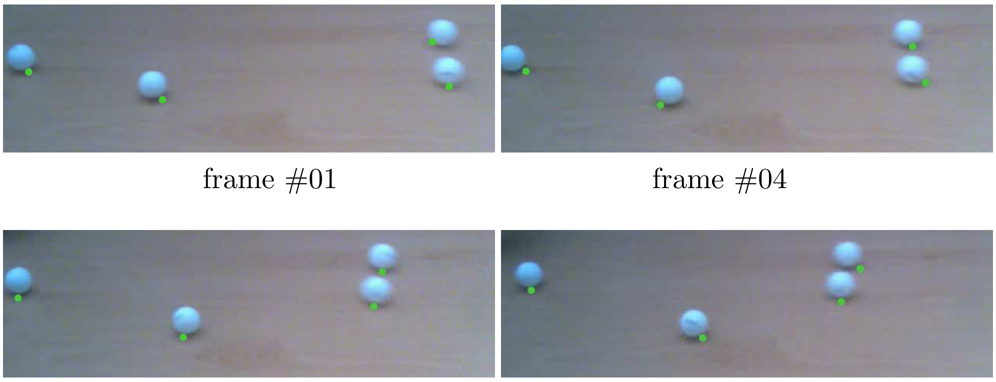
Figures 1 and 2 respectively show the tracking results for video sequences with one and multiple targets (i.e. small balls) being tracked. It is possible to check visually the accuracy of the automatically-detected targets positions for different frames in both videos.



frame #01 frame #07

frame #14 frame #21

Figure 1: Tracking results corresponding to a 320×240 video with only one tracked target. The red point close to the small ball shows the position of the target estimated by our system in the current frame.



frame #01 frame #04

frame #07 frame #10

Figure 2: Tracking results corresponding to a 640×192 video with four tracked targets. Each red point close to each small ball shows the position of the corresponding target estimated by our system in the current frame.

Our approach shows in the first video frame, the interest points above an interest value. If this interest threshold is properly chosen, these filtered points are candidates to describe the initial positions of the targets to be tracked. This selection can be done automatically by successive refinement of the threshold to guarantee that each point correspond to an target in the scene or can be done manually. Then, these corresponding targets are automatically tracked along the rest of the frames.

All the system has been implemented in MATLAB in a standard Pentium 4 without code optimization. Although our system can work at a real-time framerate for the considered video resolutions, its tracking accuracy is affected by factors like: the scene illumination conditions, the contrast of the targets with respect to the background, the velocity of each target and the frame rate of the video. Some parameters of the algorithm like the size of the ROI corresponding to each target must be tuned according to each specific video characteristics. In general, for a video with a 320×240 spatial frame resolution as the corresponding the single object sequence, a ROI size of 25-30 pixels is suitable.

4. Conclusions and future work

We presented an application of complex network theory to the visual tracking problem. Our approach models the scene as a weighted graph where nodes correspond to uniform regions (obtained by a watershed procedure) and edges represent the intensity differences between adjacent regions. Next, a local centrality measure is computed for each node defining its importance (or probability to be considered as tracked target) in the first frame. Node positions with higher importance values are considered as the targets whose positions are efficiently projected along the subsequent frames. The system has been tested successfully with different videos presenting different number of targets. However, this approach is still a prototype with several limitations. One of these is that the number of tracked targets must remain constant along the video. As future work we will work on correcting the previous restriction and also on the automatic adjustment of system parameters for each video considered. Moreover, the approach will be test on more difficult videos (i.e. outdoor scenes).

Acknowledgements

This work has been partially supported by the Spanish Government Projects TIN2008-06890-C02-02 and MTM2009-13848.

References

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and HD.U. Hwang, *Phys.Rep.* **424**, 175–308 (2006).

[2] R. Criado, M. Romance, M. and A. Sanchez, *J.Comput.Appl.Math.*, in press (2011).

[3] R. Criado, M. Romance, M. and A. Sanchez, *Int.J.Bif.Chaos*, in press (2011).

[4] L. da F. Costa, *Complex Networks, Simple Vision*, physics/0403346 (2004).

[5] L. da F. Costa, *Visual Saliency and Attention as Random Walks on Complex Networks*, physics/0603025 (2007).

[6] K. Mikolajczyk and C. Schmid, *J.Comput.Vision* **60**(1), 63–86 (2004).

[7] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer-Verlag (2010).