NYU TANDON SCHOOL OF ENGINEERING

Automatic Object Segmentation

Duc Tran, Andrew Weng, Russell Wustenberg, Ye Xu

Abstract

Motion analysis is a powerful tool for identifying objects within video sequences. While learning-based segmentation methods are dominant today, previous research has shown that a ground-up approach exists for image segmentation based on motion analysis. We build upon previous work done by the joint research between Computer Vision Group at Universität Freiburg and the University of California, Berkeley. We implement their from-principles segmentation pipeline in a naive approach to better understand the analytical methods by which they achieved their results.

Keywords: segmentation, optical flow, moving image, trajectory analysis, replication study

Prior research

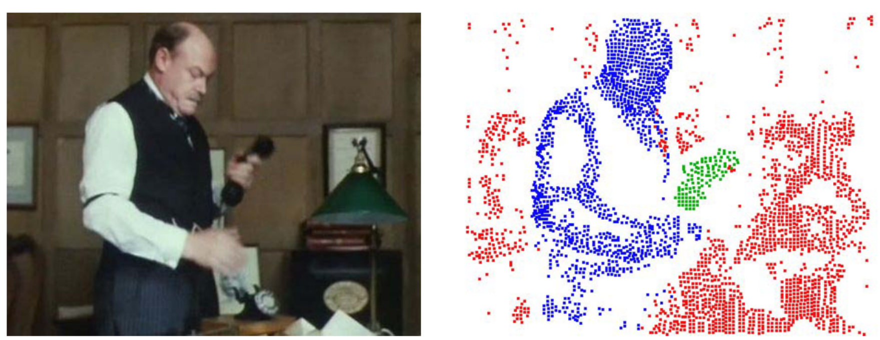

Fig. 1. Successful Object segmentation taken from [1]

The Freiburg-Berkeley authors took biological vision as their model. Biological vision does not rely upon a learned visual model for the task of segmentation [3]. Instead, cues are taken from textures, shape and, most notably, motion [4]. The informal name for this strategy is LTMA - long term motion analysis. In this study, we attempt an 'unpacking' of the LTMA algorithm and a from-principles replication of several of its components.

Picking Initial Points

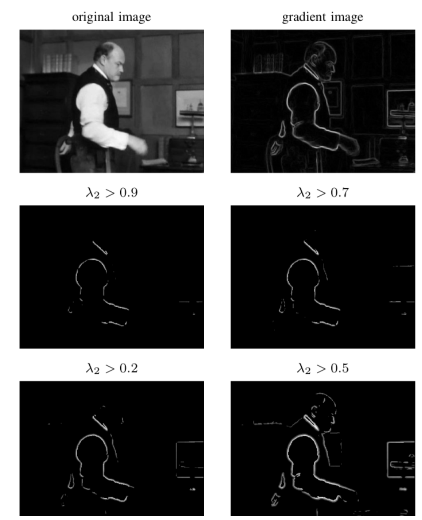

Fig. 2. Results of eigen-thresholding at various levels on a frame taken from Miss Marple, drawn from BMS-26 [2]

Since we employ Fareneback optical flow, which is a dense technique, we take measures to limit the number of points of interest. We apply eigen-thresholding on the input image to limit the data environment.

A few techniques of edge and structure detection were considered and implemented in the course of our experimentation, but eigen-thresholding (**Fig. 2**) yielded the best results and stayed true to the authors' implementation.

Creating Trajectories

Once the image has been reduced to only trackable points, their movement between frames can be estimated using optical flow. We define a class of Track objects which collect a history of pixel coordinates over their lifespan.

Occlusion Detection

We took steps to prevent the tracking of points that became occluded by passing behind an object, passing out of the frame, or otherwise becoming non-existent. Then, we compute inverse optical flow by bilinear interpolation and only track points that produce forward and backward flow values that fall within a tolerance and are not on a boundary.

Determining Similarity between Trajectories

The definition of 'similarity' between trajectories and the involved calculations are ultimately what create values that can be used for clustering and segmentation.

Affinity

We calculate 'affinity' - the measure of the joint similarity between motion and spatial location of two trajectories. Affinity is calculated pairwise for every trajectory relationship and the values are sampled from the frame in which the motion vectors have the greatest distance. An 'affinity matrix' is then constructed.

Clustering and Classification

Spectral Clustering is applied to our affinity matrix to generate a final label clustering of trajectories. We chose to implement a home-grown spectral clustering algorithm. Spectral Clustering is not limited by many of the factors that limit K-means clustering, but is significantly more involved to implement. Below are the output images from our classification.

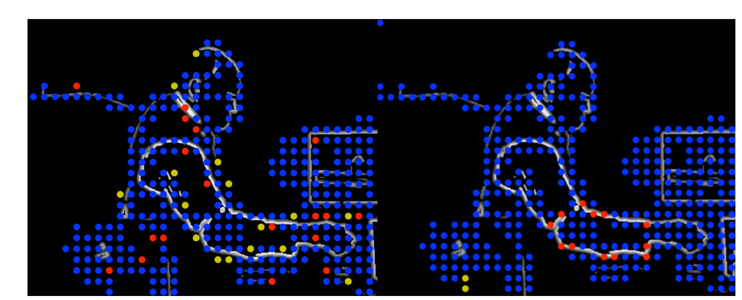

Fig. 3. Clustered and classified trajectories from our implementation.

[1] Peter Ochs, Jitendra Malik, and Thomas Brox. "Segmentation of moving objects by long term video analysis". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 33.3 (2011), pp. 500–513.

[2] Thomas Brox and Jitendra Malik. "Object segmentation by long term analysis of point trajectories". In:Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).Vol. 6315 LNCS. PART 5. 2010, pp. 282–295. ISBN : 3642155545.DOI: 10.1007/978-3-642-15555-0{\}21.

[3] Elizabeth S. Spelke. "Principles of Object Perception".In:Cognitive Science14.1 (Jan. 1990), pp. 29–56.DOI:10.1207/s15516709cog1401{\}3.

[4] Yuri Ostrovsky, Ethan Meyers, Suma Ganesh, et al."Visual parsing after recovery from blindness". In:Psychological Science20.12 (2009),pp. 1484–1491.ISSN: 09567976.DOI:10.1111 / j.1467 - 9280 . 2009 .02471.x.