Video Contour Tracking

Zhao Weng 304946606
Yao Xie 804946717
Honglin Zheng 304947026
Wenshan Li 105026914

Abstract

Both object tracking in video and contour extraction in image are active research in computer vision. However, state-of-the-art algorithms in both areas have limitations to some extent. For object tracking algorithms, most of them only capture the position and the general size of the moving object, which are represented as rectangle bounding box. For contour extraction models, most of them are applied on a single image. In this project, we propose a method to perform video contour tracking that incorporates two models to identify both the moving objects and their fine-grained contours in a video effectively.

Models and Method

Phase 1: Object Detection and Tracking

Object tracking has been an active research topic. We are going to use the datasets below to fine tune an existing model, so that we can extract the coordinates and size information for each moving objects, which can be used later for fine grained contour extraction.

Phase 2: Contour Extraction

OpenCV has an existing implementation for contour extraction based on image mask of identified objects. Topological structural analysis of digitized binary images by border following is used to do border following to extract contours for objects in each frame of the video.

Datasets

http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm#object

We plan to experiment our method on the following three, but not limited to, datasets.

YouTube-BoundingBoxes Dataset

YouTube-BoundingBoxes is a large-scale data set of video URLs with densely-sampled high-quality single-object bounding box annotations. The data set consists of approximately 380,000 15-20s video segments extracted from 240,000 different publicly visible YouTube videos, automatically selected to feature objects in natural settings without editing or post-processing. All these video segments were human-annotated with high precision classifications and bounding boxes at 1 frame per second.

DAVIS: Densely Annotated Video Segmentation
The dataset consists of fifty high quality, Full HD video sequences, spanning multiple occurrences of common video object segmentation challenges such as occlusions, motion-blur and appearance changes. Each video is accompanied by densely annotated, pixel-accurate and per-frame ground truth segmentation.

UCLA Aerial Event Dataset
The events in this dataset was collected with scripts involving the interactions between humans and objects at two different sites. After camera calibration and frame registration, there are totally 27 videos in the dataset, the length of which ranges from 2 minutes to 5 minutes. The hierarchical semantic information of objects, roles, events and groups in the videos was annotated. The annotation in the dataset includes individuals, objects, groups, events. human roles and goals (destinations).