GPU-Accelerated Symmetry Transform for Object Keypoints

Jason Owens

December 7, 2016

1   Introduction

Finding salient object candidates in arbitrary natural images is a problem in computer vision and robotic perception that has yet to be solved. There are many ways this challenge is currently approached:

- scan the entire image, at multiple scales, with some kind of object recognition algorithm, generating a heat map indicating the score or probability of a known, detected object (i.e. sliding window),

- propose a smaller set of likely object regions based on some engineered or learned features of the image (e.g. edge boxes, geodesic object proposals, objectness),

- compute a segmentation of the image to propose object-like regions,

- learn a saliency function to predict regions of the image that may contain objects.

In recent years and in combination with advanced deep learning systems for object recognition tasks, the use of object proposal algorithms has become almost a de-facto standard. As indicated, object proposal algorithms produce a significantly smaller of image regions to test than almost any other mechanism. The main idea is to either engineer or train a detector for what has been called the "objectness" of a region, i.e. how well that region reflects aspects of containing an object. Object aspects often include such properties as closed contours, convexity, and compactness. In the image domain, the contours can be reflected in the edges derived from the gradient image.

However, processing image contours without any additional information (e.g. learning an object contour prior) often yields undesirable results, with many proposals that do not represent an object. When we discuss objectness, we usually care about whole physical objeccts we can pick up (box of cereal, phone, pencil, flashlight, coffee mug) and not the aspects of an object's appearance that may also provide strong gradient edge responses.

In 1995, Reisfeld et al. proposed the use of a symmetry transform operator in the image domain for use as an attentional operator. Reisfeld showed that by extracting contours from the symmetry magnitude and selecting local maxima, it was possible to compute attention keypoints that indicated important features of the image. For example, figure ?? shows an example from [?] where selecting the maximum output of the radial symmetry transform produces keypoints for the face of Nixon, the forehead of Elvis, as well as several interesting points on the flags.
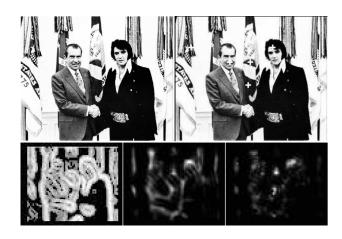
Figure 1:   Salient keypoint extraction using symmetry magnitude, from Reisfeld 1995

Figure 2: Symmetry keypoints naively extracted from a symmetry transform