

# GPU-Accelerated Symmetry Transform for Object Keypoints

Jason Owens

January 5, 2017

## 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 set 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 often be reflected in the edges derived from the image gradient.

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 objects 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, e.g. logos or pictures on a cereal box.

In 1995, Reisfeld et al. proposed the use of a symmetry transform operator in the image domain for use as an attentional operator. Since symmetry is considered a strong indicator of an object with shape [3, 1], it's reasonable to conclude that regions exhibiting strong symmetry are likely to contain objects of interest.

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 1 shows an example from [4] 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.

Need to discuss Kootstra use of the symmetry in an object saliency framework (plus the limitations).

Like [2]. Extends the 2D appearance-based symmetry into a 3D approach on depth maps, and yields better detection results (check this out).



Figure 1: Salient keypoint extraction using symmetry magnitude, from Reisfeld 1995

## 2 Mathematical Approach

Reisfeld’s main idea is to use the gradient image to compute symmetry magnitude and direction for every pixel in the image. Given the magnitude image, one can then use non-maximum suppression to select local maxima as salient points in order to direction attentional processing. For convenience, we restate Reisfeld’s mathematical formulation to provide background for the computational approach discussed in the next section.

Let  $\mathcal{I} : \Omega \rightarrow [0, 1]$  be a grayscale image with domain  $\Omega \subset \mathbb{Z}^2$ . Then  $p_k \in \Omega$  represents some pixel coordinate in the image, and  $\nabla(p_k) = \left( \frac{\partial}{\partial x} \mathcal{I}(p_k), \frac{\partial}{\partial y} \mathcal{I}(p_k) \right)$  is the gradient at that coordinate. Reisfeld then computes a 2D polar coordinate for each pixel,  $(r_k, \theta_k)$ , where  $r_k = \log(1 + \|\nabla(p_k)\|)$ , and  $\theta_k = \text{atan2}(\frac{\partial}{\partial y} \mathcal{I}(p_k), \frac{\partial}{\partial x} \mathcal{I}(p_k))$ . Let  $l_{ij}$  be the line passing through two points  $p_i$  and  $p_j$ , and let  $\alpha_{ij}$  be the angle  $l_{ij}$  makes with the horizontal ( $x$ ) axis. For any pixel  $p_k$ , define the set  $\Gamma(p_k) = \{(i, j) \mid \frac{p_i + p_j}{2} = p_k\}$ , i.e. the set of pixel index pairs such that  $p_k$  resides on the center of the separating line  $l$ . A distance,  $D_\sigma(i, j)$ , and phase,  $P(i, j)$  function are used to determine the contribution  $C(i, j)$  for each point pair in  $\Gamma(p_k)$ , defined as follows:

$$D_\sigma(i, j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\|p_i - p_j\|}{2\sigma}} \quad (1)$$

$$P(i, j) = (1 - \cos(\theta_i + \theta_j - 2\alpha_{ij})) (1 - \cos(\theta_i - \theta_j)) \quad (2)$$

$$C(i, j) = D_\sigma(i, j) P(i, j) r_i r_j \quad (3)$$

Finally, the *symmetry magnitude* for a point  $p$  is defined as:

$$M_\sigma(p) = \sum_{(i, j) \in \Gamma(p)} C(i, j) \quad (4)$$

which simply sums the weighted contributions over the entire “symmetric pixel” neighborhood of  $p$ . A direction contribution function for each pixel,  $\psi(i, j)$ , used to compute the *symmetry direction*,  $\phi(p)$  for  $p$ , is defined as follows:

$$\psi(i, j) = \frac{\theta_i + \theta_j}{2} \quad (5)$$

$$\phi(p) = \psi(i^*, j^*) \quad \text{where} \quad (i^*, j^*) = \underset{(i, j) \in \Gamma(p)}{\operatorname{argmax}} C(i, j) \quad (6)$$

He provides no justification for scaling the gradient magnitude; presumably this is to better match the range of the distance and phase functions he defines

### 3 Computational Approach

### 4 Results

### 5 Future work

### 6 Conclusion

### 7 References

#### References

- [1] Yunfeng Li et al. “Symmetry Is the Sine qua Non of Shape”. In: *Shape Perception in Human and Computer Vision*. Ed. by Sven J. Dickinson and Zygmunt Pizlo. London: Springer London, 2013, pp. 21–40. ISBN: 978-1-4471-5194-4 978-1-4471-5195-1. URL: [http://link.springer.com/10.1007/978-1-4471-5195-1\\_2](http://link.springer.com/10.1007/978-1-4471-5195-1_2) (visited on 12/04/2016).
- [2] Ekaterina Potapova, Michael Zillich, and Markus Vincze. “Local 3d Symmetry for Visual Saliency in 2.5 D Point Clouds”. In: *Asian Conference on Computer Vision*. Springer, 2012, pp. 434–445. URL: [http://link.springer.com/chapter/10.1007/978-3-642-37331-2\\_33](http://link.springer.com/chapter/10.1007/978-3-642-37331-2_33) (visited on 12/04/2016).
- [3] D. Reissfeld and Y. Yeshurun. “Robust Detection of Facial Features by Generalized Symmetry”. In: IEEE Comput. Soc. Press, 1992, pp. 117–120. ISBN: 978-0-8186-2910-5. DOI: 10.1109/ICPR.1992.201521. URL: <http://ieeexplore.ieee.org/document/201521/> (visited on 11/15/2016).
- [4] Daniel Reissfeld, Haim Wolfson, and Yehezkel Yeshurun. “Context Free Attentional Operators: The Generalized Symmetry Transform”. In: *Int. J. Comput. Vision* 14 (1995), pp. 119–130.

