Exploiting Protrusion Cues for Fast and Effective Shape Modeling via Ellipses

Alex Wong[†‡] Brian Taylor[†] Alan L. Yuille[‡]

University of California, Los Angeles[†] Johns Hopkins University[‡]

http://ccvl.jhu.edu/

motivation

Ellipses provide a **compact representation of complex objects** and their structure, as most **objects can be broken down rigid parts**. Given the ability of the state of the art semantic segmentation and motion boundary detection methods to produce reliable object masks, representing such as **ellipse models becomes a natural next step for understanding the segmented objects' structures**.

our contribution

We propose a novel ellipse fitting method **based on psychology and cognitive science studies** on shape decomposition and show that our shape **coverage compares well with the state of the art** methods, while **significantly outperforming them in run-time** by as much as 508 times in our evaluation of the methods on over 4000 2D shapes.

our approach

1. We leverage symmetry axis transform and protrusions as cues to partition the foreground pixels, joining regions whose protrusion strength (Eqn. 1) is less than a threshold τ.

$$f_p = \frac{\gamma(\mathbf{s}_h^{\mathbf{j}}) - r_{\mathbf{j}}}{|\mathbf{b}_1^h - \mathbf{b}_2^h|} \qquad (1)$$

2. We estimate a set of ellipses for each local region by minimizing Eqn. 2, which describes the coverage and complexity of the model.

$$C(\mathbf{R}^j, \mathbf{e}^j) = ||\mathbf{R}^j - \bigcup_{k=0}^{\kappa} I(\mathbf{e}_k^j)||^2 + \kappa \frac{||\mathbf{R}^j||}{\eta(\mathbf{s}^{\mathbf{j}}) + 1} \qquad (2)$$

3. We then minimize the cost of the entire model by comparing cost of pairs of adjacent ellipses to a single ellipse covering both region.

$$D_{uv} = C(\mathbf{Q}_u, \mathbf{E}_u) + C(\mathbf{Q}_v, \mathbf{E}_v) - C(\mathbf{U}_{uv}, \mathbf{E}'_{uv}) \qquad (3)$$

4. We select and merge the pair that gives the highest reduction on the model cost via Eqn. 3 until the cost cannot be reduced.

$$(u, v) = \arg\max_{u,v}(D_{u,v}) \qquad (4)$$

experimental results

	Intersection-Over-Union			
Method	LEMS	MPEG-7	SiSHA	PASCAL Horses
Ours	0.9634	0.9597	0.9638	0.9507
DEFA	0.9623	0.9615	0.9629	0.9545
AEFA	0.9711	0.9600	0.9654	0.9527
EMAR	0.9326	0.9219	0.9600	0.9470
	Run-time (seconds)			
Method	LEMS	MPEG-7	SiSHA	PASCAL Horses
Ours	3.88s	2.20s	2.19s	19.96s
DEFA	211.44s	255.87s	125.28s	2503.21s
AEFA	745.45s	1001.92s	528.36s	12088.02s
EMAR	120.57s	78.41s	63.23s	740.81s

We compared our method against three others: Augmentative Ellipse Fitting Algorithm (AEFA) and Decremental Ellipse Fitting Algorithm (DEFA) from [1] and an EM ellipse fitting algorithm (EMAR) [2].

We ran each model on over 4000 shapes given by four 2D shape benchmarks: LEMS, MPEG-7, SiSHA, and PASCAL Horses.

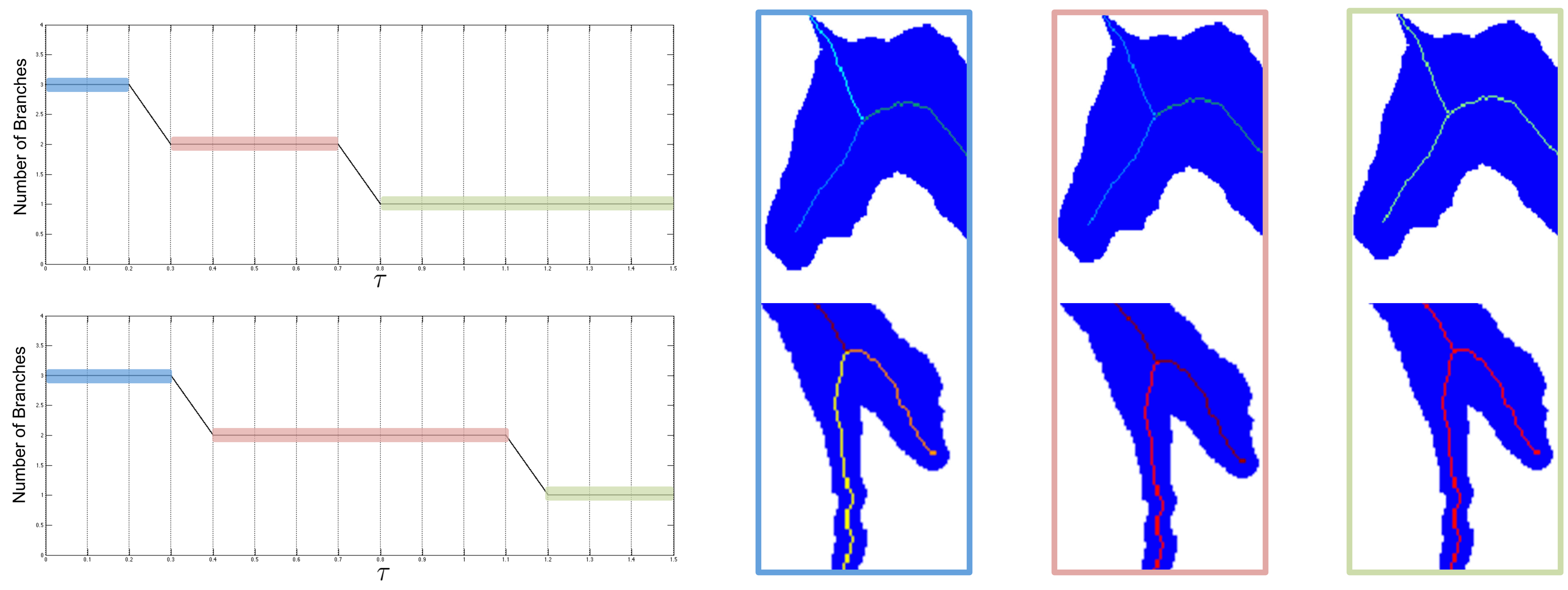
We find that our model:

- performs comparably to DEFA and AEFA (where the maximal difference in mean IOU is less than 0.01) and outperforms EMAR
- gains a 109-fold of mean run-time improvement over DEFA and 508-fold over AEFA

references

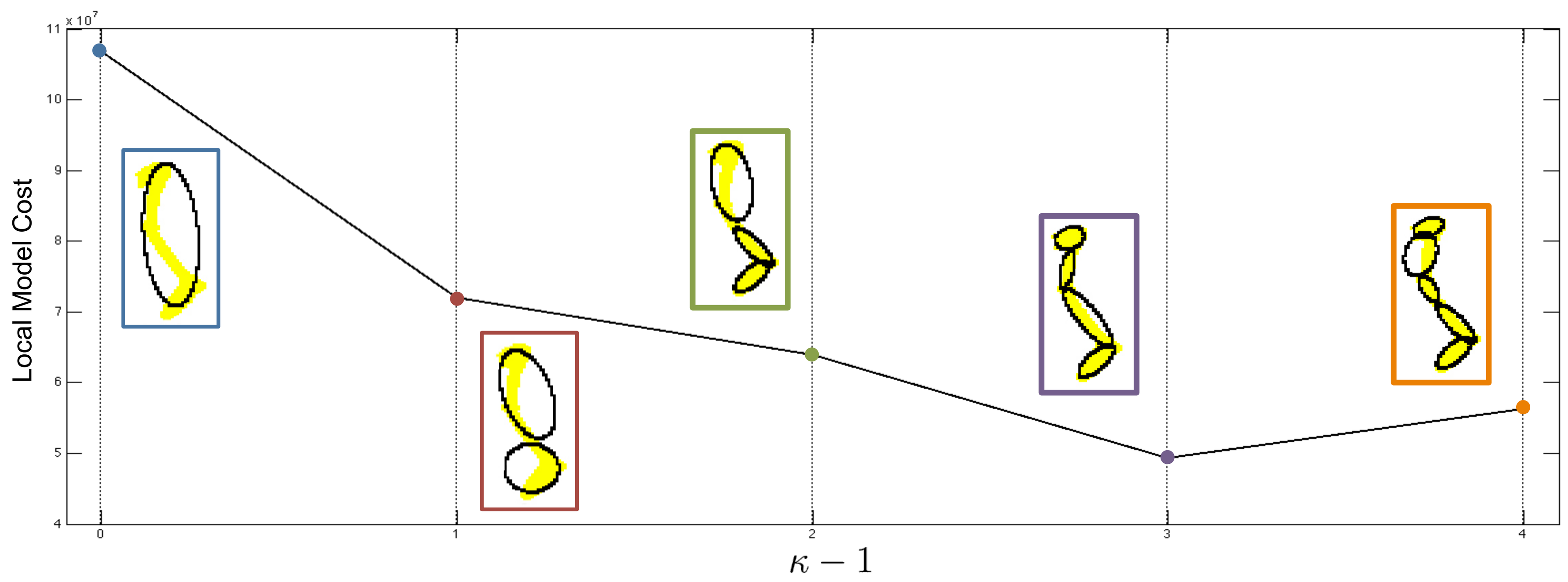
1. Costas Panagiotakis and Antonis Argyros. Parameter-free modelling of 2d shapes with ellipses. *Pattern Recognition*, 2015.
2. Da Xu, Richard Yi, and Michael Kemp. Fitting multiple connected ellipses to an image silhouette hierarchically. *Image Processing, IEEE Transactions on*, 19(7):1673–1682, 2010

symmetry axis and protrusion cues



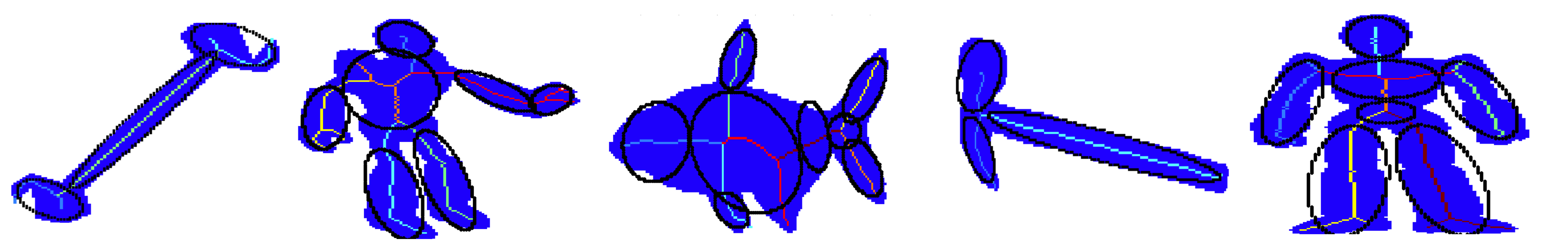
We approximate the candidate regions to fit a set of ellipses via cues from the symmetry axis and the protrusion strength (Eqn. 1) of its segments. A branch with $f_p < \tau$ is merged with the segment it protrudes from. We begin at the terminal segments and move inward.

local cost minimization



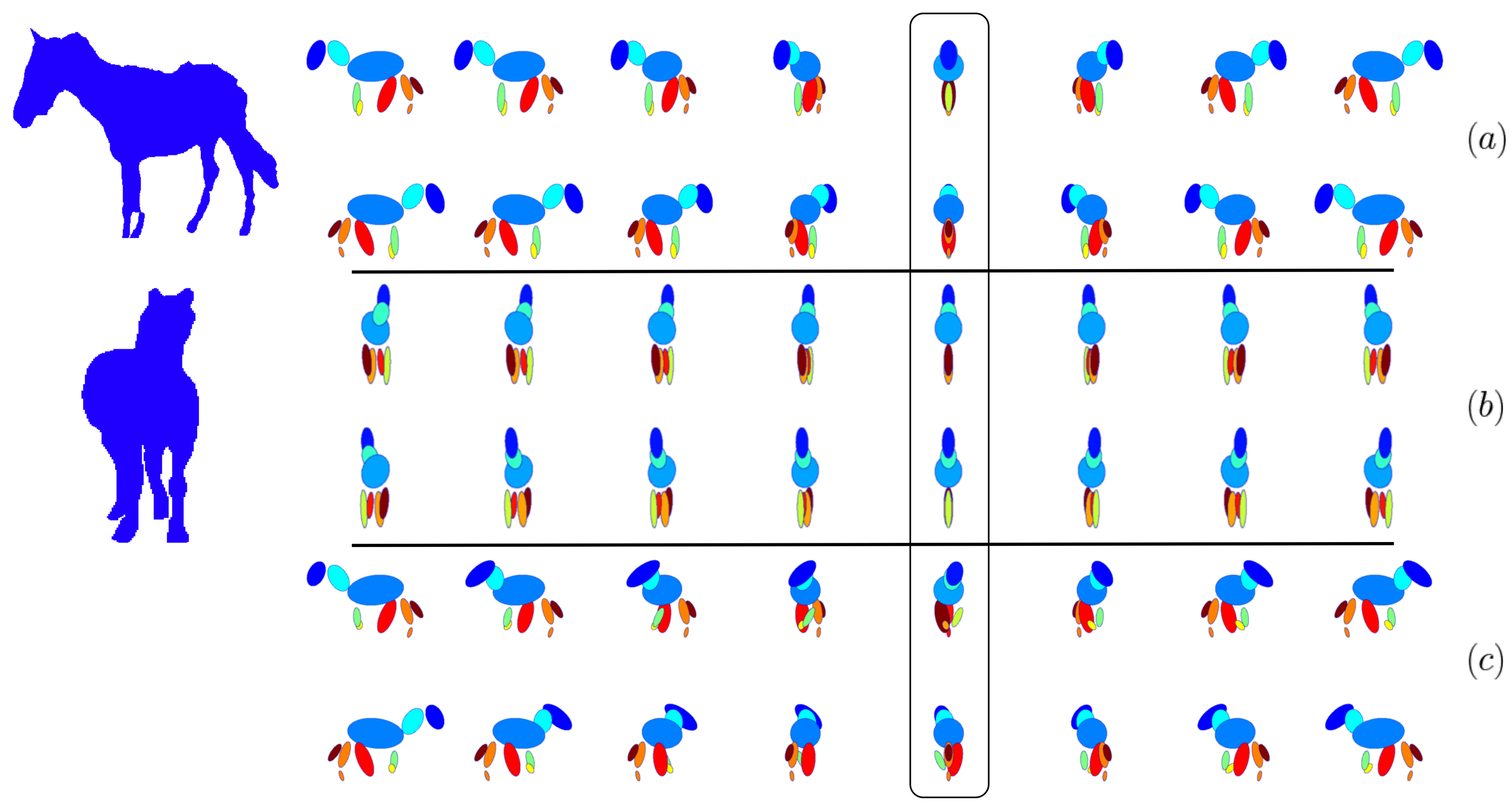
We partition a local region based on cuts along the symmetric axis determined by the curvature of the axis. We model the partitions with a set of ellipses and select the local model that best minimizes Eqn. 2.

sample results



Sample output from our method: the colored lines denote the symmetry axis and branch merging via protrusion cues. Each ellipse correspond to a semantically meaningful part (e.g. head, arm, leg, fin).

rendering a richer model



Pseudo-3D model built by applying our method to two images in PASCAL Horses. (a) denotes the model of the profile view of a horse. (b) denotes the model the frontal view of another horse. (c) denotes a model generated by combining (a) and (b). By combining multiple models we are able to hypothesize the 3D structure (see highlighted).

acknowledgements

We would like to thank Becky Li for her work on preparing the PASCAL Horses dataset and setting up the experiments. This work is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/ Interior Business Center (DoI/IBC) contract number D16PC00007 and NSF award CCF-1317376.