Prime Shapes in Natural Images

Qi Wu
http://www.cs.bath.ac.uk/~qw219
Peter Hall
http://www.cs.bath.ac.uk/~pmh

Media Technology Research Centre
Department of Computer Science
University of Bath,
Bath, UK

Shape has been well studied in many disciplines, yet to the best of our knowledge the question as to whether there is a set of elementary planar shapes that appear commonly in the world around us has never been asked. If such a set exists, then the elemental shapes could play a similar role in shape analysis as the primary colours do in colour analysis. This paper uses a fully unsupervised framework to find out the 'primary shapes' in image segmentations. It conclude that the most common of those found are familiar enough to be named: shapes such as triangles, squares and circles (more exactly, these shapes up to affine transformation). We propose to use qualitative shapes as features in future applications. For example, hierarchies of qualitative shape can be used for cross-modal matching [2].

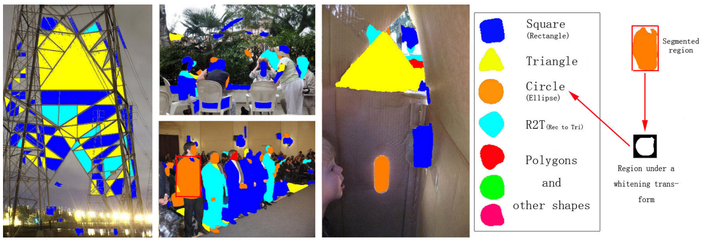

Figure 1: Segmented regions classified by prime shapes obtained from MIT database. Example: the segmented torso of a man is classified as an ellipse/circle.

Our proposition has its roots in Art, most particularly 20th century Western Art. Painters such as Picasso (*e.g.* Seated Woman with Wrist Watch), Leger (*e.g.* Card Players), and schools such as Italian Futurism, Tubism, and Orphism, depicted objects (and motions) as being composed of just a few basic geometric forms: cones, cylinders, bricks and so on. Additionally, it is very common for artists to make initial sketches using simple shapes to layout a scene, as any book on drawing instruction will testify. Empirical evidence that aligns with artistic intuition has existed since at least the 1970's, when psychologists such as Rosch [4] showed simple shapes (specifically triangles, squares, and circles) are easier for humans to recall that other shapes. This paper provides evidence that simple shapes are integral to what might be called 'the visual signal'. As Figure 1 shows, we can classify image regions into qualitative shape that have been learned without supervision.

Our experiment is designed to find out whether common simple shapes objectively exist in image segmentations. We wish to remove as much bias as we can, so supervised methods are ruled out and we have been sure to use a range of segmentation methods, shape descriptors, and databases. Importantly, *we will not define simple shapes in advance, rather they should be an emergent property based on image statistics*. Our approach is to automatically cluster regions that have been segmented from images, and compare these clusters with those created from a database of randomly created images; there is no human interaction at all.

Three main steps are included:

- Shape Production: we choose three types of segmentation methods to produce shapes, which are *Thresholding*, *MSER* [3] and *Berkeley Segmentor* [1].
- Shape Description: we opted for *Zernike Moments* [5] and *Chebyshev Moments* as the shape descriptor. Before the computing its description, we apply a whitening transfer that brings the region into the unit disc. This will map any triangle into equilateral form, any rectangle into a square, and any ellipse into a circle. However, scaling into the unit disc changes the effective sample rate. To make sure that this plays no role in moments competition, we resample the shapes into a 50^2 binary images.

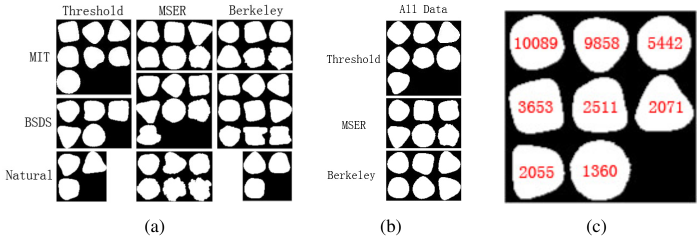

Figure 2: A matrix of final results. (a) Each entry shows the shape icons yielded by different databases, different segmentation methods. (b) Shape icons for different segmentation methods yielded by combining all three databases, different segmentation methods. (c): Final grouping result by combing all databases and segmentations. The number of each prime shape is plotted in each corresponding icon. The total number of segmented regions, classified or not, is 56992

- Shape Classification: this phase consists of two steps, mean-shift comes first and then an agglomerative clustering algorithm is applied to do the second times grouping. To locate statistically significant clusters in shapes drawn from an image database we count the total number of shapes in a cluster of a given size to get $p(m|D)$, which is the probability of observing a cluster of size m, given source $D \in \{\text{Image Database}, \text{Random}\}$. We keep only those clusters of size m for which $p(m|\text{Image Database}) > p(m|\text{Random})$.

Final shapes for each database and each segmentation method can be seen in Figure 2. The shapes tend to be be simple and nameable; shapes such as circle, square and triangle are common. There are some irregular looking shapes too, but these are not often observed compared to the regular shapes.

During the research, we also noticed that the priors on different prime shapes depends on the database used, and these contain different sorts of photograph. This suggests a scene classification application which can be found in our paper.

The main contribution of this paper is a discovery which is unique, so far as we know: regions in image segmentations naturally form classes that correspond to simple, easily recognisable shapes. In fact, more than half of the segmented regions in the datasets can be classified into prime shapes. We found this to be true no matter what segmentation algorithm we used, no matter what databased we used, an no matter how we described the shape of segmented regions. There are no arbitrary parameters in our clustering algorithm, which is fully unsupervised. In short, we have provided empirical evidence to suggest that *natural images contain simple shapes to a statistically significant degree*. And the results clearly show prime shapes emerging from segmentations: they are 'features in the signal', and as such may be of use to many applications in Computer Vision and maybe elsewhere, not just scene classification.

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2294 – 2301, June 2009.

[2] A. Balikai, P. Rosin, Y.-Z. Song, and P.M. Hall. Shapes fit for purpose. In *British Machine Vision Conference*, 2008.

[3] J Matas, O Chum, M Urban, and T Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761 – 767, 2004. ISSN 0262-8856.

[4] Eleanor H. Rosch. Natural categories. *Cognitive Psychology*, 4(3): 328 – 350, 1973. ISSN 0010-0285.

[5] M.R. Teague. Image analysis via the general theory of moments. *JOSA*, 70(8):920–930, 1980.