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

Abstract

This paper provides evidence that about half of all the regions in segmented images can be classified as one a few simple shapes. Using three segmentation algorithms, three different image databases, and two shape descriptors, we empirically show that shapes such as triangles, squares, and circles are observed, up to an affine transform and at a much higher rate than random shapes. This result has potential value in applications such as scene understanding, visual object classification, and matching because qualitative shapes can be used as features. We show an application in scene categorisation based on what might be called 'bag of shapes' .

1 Introduction

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 [5].

The literature studying planar shape covers many areas. We focus on just those areas of direct relevance to this paper — Image Processing and Computer vision — which alone is large and growing larger. Morphology is a standard topic in lecture room texts, such as [10], and decomposing shapes into parts is a future interest of ours. Here though, we are more interested in describing planar shapes for classification. Boundary descriptors permit a scale based representation [22], Fourier descriptors are a common example [25, 28, 31]. Shape skeletons are the dual of shape boundary, and also have been used as a descriptor [14, 29, 33]; modern techniques reduce noise sensitivity [27]. Region based descriptors are robust to noise when compared with either boundary or skeletal descriptors. Typical descriptors include geometric moments such as Hu [11], Zernike [13], and Chebyshev [26]. Alternative region based descriptors also exist, such as [8]. Shape is put to use in many Computer Vision tasks not limited to matching [32], classification [9, 18, 35], and retrieval [1, 3, 6].

It is not possible, nor is it our purpose, to review the extensive shape related literature here; the above is a small but representative sample. What is important to this paper is that

 http://dx.doi.org/10.5244/C.26.45

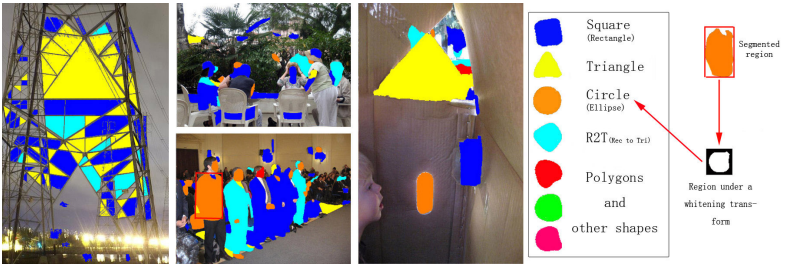


Figure 1: Segmented regions classified by prime shapes obtained from MIT database (see section 2.5). Example: the segmented torso of a man is classified as an ellipse/circle.

none of the literature we know of asks as we do: *" is there a set of shapes commonly present in natural images?"* Intuitively we would expect such shapes to be simple, but most of the existing literature — especially recent publications — use relatively complex silhouettes of real objects (cups, horses, hands *etc*). This paper tests the proposition that simple (prime) shapes exist in natural images — that they are part of 'the signal'. Since this appears to be a unique question, and since its affirmation or denial should not depend on details of shape description *etc.*, we curtail our review of the Computer Vision literature in favour of motivating our work a little more.

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 [30] showed simple shapes (specifically triangles, squares, and circles) are easier for humans to recall than other shapes. It is interesting to speculate that this may explain why humans have words to describe these shapes, and it is interesting also that our experiments show it is exactly these shapes that occur in natural images with a frequency which is well above that expected by chance alone. In other words, 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. On the right-side of the figure also shows an example that how a noisy segmented region can be classified into circle or ellipse.

2 Experimental Method

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. Figure 2

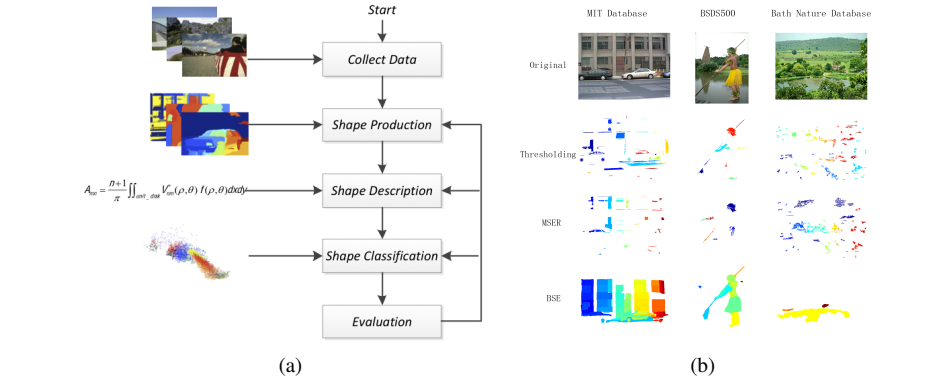


Figure 2: *(a)* Experimental Framework *(b)* Regions produced by different segmentation algorithms over different image databases.

gives a more detailed illustration of our experiment.

No matter how we configured our experiment, we concluded that simple shapes such as triangles, squares, and circles (up to an affine transform) are the most common classes of shape to emerge from natural images. However, the ratio of these shapes is not constant across the databases, which leads to an immediate application in scene categorisation, explained in Section 3.

2.1 Three Image Databases, and a Random Generator

Our experiments are based on three images databases, chosen because they offer a diverse set of content. We also used a databases of shapes created at random. Typical images from these database can be seen in Figure 2.

MIT Database is publicly available [12]. Designed for eye tracking experiments, the database contains 1003 images, including street scenes, buildings, animals and natural landscapes *etc*. We randomly choose 200 images as training data, and 100 for test data.

Berkeley Segmentation Database (BSDS500) [19] is a well known, publicly available data base often used for experiments in contour detection and image segmentation. It includes 500 pictures, most of them are natural images, but also includes faces and animals. We randomly choose 200 images as training data, and 100 for test data.

Bath Nature Database has 50 pictures of outdoors; forest, field, seascapes *etc*, there are few is any man-made objects. Most of the images in this database contain no obvious basic shape that we (as humans) can perceive. Also, there is no obvious foreground or background.

Random Shapes. In addition to fixed databases we generate random shapes, one per image, as in Figure 3. To create random shapes we developed the following algorithm: (i) Create an $N \times N$ image of independent random numbers. (ii) Choose the central pixel to be the current shape. (iii) Mask all pixels that form the outer border of the current 4-connected shape, and add the masked pixel with the highest random number to the shape. (iv) continue until the shape has the required number of pixels. The halting number is randomly drawn from a uniform distribution over $[100, 600]$.

Figure 3: Typical random shapes.

2.2 Three Segmentation Algorithms

To offset bias regarding any particular segmentation algorithm we used three; one very simple, one popular, one state of art. In each case any region that touched a picture boundary or which contained less than 100 pixels was removed from further consideration. The first was to remove any bias introduced by straight boundary edges, the second to remove noise — needed primarily for thresholding. Typical segmentation output can be seen in Fig 2 *(b)*.

Thresholding is perhaps the simplest methods for image segmentation. A grayscale, I image maps to a binary image: $b = I > \tau$, for threshold τ. Assuming gray values in $[0,1]$, we set $\tau = 1/2$. We used 4-connected regions, both black and white, as shapes.

Maximally Stable Extremal Regions (MSER) are regions found by analysis of successive threshold images [20]. MSER yields a hierarchy of regions and we only use the regions from the last two layers.

Berkeley Segmentation Engine(BSE) is based on the probability of boundary (Pb) maps introduced by Arbelaez *et al* [2]. It is considered as one of the most successful segmentation technique because it compares very well against human produced grounding truth using the Berkeley Segmentation Dataset (BSDS-500).

2.3 Two Shape /Region Descriptors

There are many shape descriptors to choose from; we opted for Zernike moments [34] and Chebyshev moments [23]. These moments operate over sets of points, so are useful for describing solid regions of the kind produced by typical image segmentation algorithms. They are fast to compute, again useful when faced with many regions in an image segmentation. They are invariant to rotation and robust to noise, and have been used for classification by shape [17]. This paper presents results using Zernike moments only, almost identical results are obtained using Chebshev moments; these results are available in the supplementary material.

2.3.1 A Whitening (Affine) Transform and Re-sampling

We normalise each region (shape) before computing its description. We apply a whitening transform that brings the region into the unit disc, as follows. Let $X = \{x_i\}$ be points of a region, with $\bar{x}$ their centroid and $C = ULU^T$ their covariance. Then $y_i = L^{-1/2}U^T(x_i - \bar{x})$ is a whitening transform, which applies an affine transform to the shape by centering it at the origin, rotating it to a canonical frame and differential scaling over each eigenaxis. This will map any triangle into equilateral form, any rectangle into a square, and any ellipse into a circle. The new shape will have a unit covariance in each eigendirection, so we scale by the point most distant from the origin to map the shape into the unit disc.

Scaling into the unit disc changes the effective sample rate. To make sure that this plays no role in moment computations, we resample the shapes into a 50^2 binary image.

2.3.2 Zernike moments

Zernike moments [34] are constructed using a set of complex polynomials which form a complete orthogonal basis set defined on the unit disk. They are parameterised by two integers; $n \geq 0$, and m such that $|m| \leq n$ and $n - |m|$ is even. In polar coordinates, (ρ, θ), the $(n,m)^{th}$ Zernike basis function, $V_{nm}(\rho,\theta)$, defined over the unit disk is

$$V_{nm}(\rho,\theta) = R_{nm}(\rho)\exp(jm\theta), \;\; \rho \leq 1, \tag{1}$$

in which $j = \sqrt{-1}$. The Zernike radial polynomials, $R_{nm}(\rho)$, are defined as:

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \times \frac{(n-s)!}{s!((n+|m|)/2-s)!((n-|m|)/2-s)!}\rho^{n-2s} \tag{2}$$

For a binary image $f(x,y)$, the mn^{th} Zernike moment is

$$Z_{mn} = \frac{n+1}{\pi}\sum_x\sum_y f(x,y)V^*_{nm}(\rho,\theta), \;\; x^2+y^2 \leq 1. \tag{3}$$

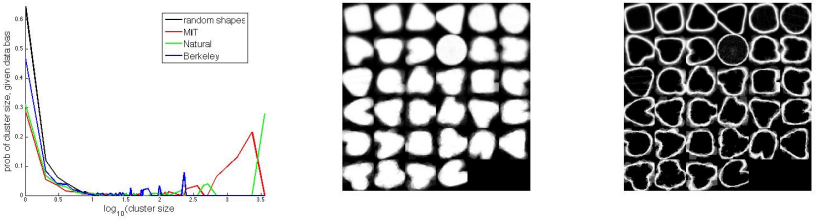
We use all moments up to $n = 6$, and $m \in [-n,n]$, giving $(n+1)(n+2)/2$ basis functions. Before clustering (see below) we normalise the Zernike moments as follows: (i) we use the absolute value, so that the moments are invariant to rotation; (ii) we divide by the zeroth order moment, so that the moments are invariant to pixel area.

2.4 Clustering

The problem now is to find clusters in a collection of shapes, in a *fully unsupervised* way. We use two steps to balance computational efficiency and accuracy, both are explained below, after an overview. First we use mean shift clustering, which is fast, well known, and above all is non-parametric; in particular the number of clusters is not specified. We locate mean shift clusters that are statistically significant, typically about 30 to 40 clusters, each with an associated iconic shape. However, we have found mean-shift tends to produce too many clusters in that some iconic shapes which should obviously be grouped are not. Fixing this inside mean-shift is very difficult, we found it much easier to switch to an agglomerative clustering based on shape correlation. This latter algorithm does not use any moment description at all, and in principle could be used on all (whitened) shapes. However, agglomerative clustering depends on pair-wise interactions — mean-shift reduces the number of pairs from about $(10^4)^2$ to a more manageable 35^2, approximately.

2.4.1 Mean Shift Clustering

To use mean-shift we project all of the descriptors into a deflated eigen-space (aka. principal component analysis) to leave a descriptor of about 17 dimensions. A whitening transform ensures the data exhibits a unit standard deviation in each eigen-direction: now a single number can now control the bandwidth of a mean shift clustering algorithm, because the data are equally spread in all directions. The bandwidth is automatically set to be $(1/3)(\text{vol}/N)^{1/n}$ in which vol is the hyper-volume of the bounding box enclosing the N data points, which exist in a n dimensional space. This is the characteristic radius of a hyper-sphere surrounding each datum, assuming they are uniformly distributed, but scaled because we found most of the points were clustered in about $1/3$ of the hyper-volume, in each direction.



(a) (b) (c)

Figure 4: *(a):*The probability of number of shapes in a group of a given size for different databases using threshold segmentation and Zernike moments. Random shapes do not create clusters of more than about 10 shapes. *(b):*Average shapes from mean shift clusters . *(c):*Error images over mean shift clusters. These shapes come from the MIT database, and are ordered so that the largest sized classes are top-left and smallest in bottom-right.

Mean shift yields many clusters of different sizes. Some clusters contain hundreds or even thousands of shapes, others contain just one or two. In order to decide which clusters are statistically significant we produce 10^4 random shapes, see 2.1. We clustered the random shapes using mean shift, and found that no cluster size exceeded about 10, and the vast majority were singletons, see Figure 4. This result suggests that the dimensions that remain after deflation contain enough information to discriminate random shapes: we are not finding shapes because our descriptor lacks resolution. The c in figure 4 shows clusters lose of regions of similar shapes.

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 $p(m|\text{Image Database}) > p(m|\text{Random})$. There are typically around 30 to 40 such clusters, which together contain between 30% and 80% of all shapes, depending on the image database and segmentation method, see Table 1. This is enough to have confidence that our hypothesis is valid — that simple shapes do exist to a statistically significant degree. But repetition of classes is undesirable, so these clusters are passed to the next clustering step.

2.4.2 Agglomerative Clustering

Mean-shift yields a few tens of clustered shapes, that number can be reduced to less than ten by agglomerative clustering, using a method developed by ourselves. We begin by rotating all (whitened) shapes in a cluster to the first. Now $s(x,y,i,j)$ denotes a point in the i^{th} aligned shape in the j^{th} cluster. It is now easy to compute the mean shape, and to estimate the spatial error distribution:

$$m(x,y,j) = \frac{1}{N_j}\sum_{i=1}^{N_j} s(x,y,i,j), \tag{4}$$

$$e(x,y,j) = \left(\frac{1}{N_j}\sum_{i=1}^{N_j}(s(x,y,i,j)-m(x,y,j))^2\right)^{1/2}, \tag{5}$$

where N_j is the number of shapes in class j. The error image, e, locates where the class varies most — which invariably is at the boundary. We normalise e so the image sums to unity. The thresholded mean shape $\mathsf{icon}(x,y,j) = m(x,y,j) \geq \text{mean}[m(x,y,j)]$ acts as an *icon* for the binary shapes in the class. Typical mean shapes and error images coming from mean shift can be seen in Figure 4, which informally suggests regions of similar shape form clusters. We must now combine classes, so need a class descriptor.

Our class descriptor uses the boundaries pixels of the icons, call these $b(x,y,j)$. We rotate a boundary image about its centre and at each angle, θ, computing its *similarity* to the error image in the same class using

$$\phi(\theta, j) = \sum_{xy} e(x,y,j) b(x,y,j,\theta). \tag{6}$$

$\phi(.,.)$ is now a one-dimensional signal that characterises an icons rotational symmetry against its own error set. Next, we compute the maximum of the normalised cross correlation between pairs of classes, to obtain an inter-class similarity score:

$$c(j,k) = \max \frac{\sum_\theta (\phi(\theta,j) - \bar{\phi}(\theta,j))(\phi(\theta-\alpha,k) - \bar{\phi}(\theta,k))}{(\sum_\theta (\phi(\theta,j) - \bar{\phi}(\theta,j))^2)(\sum_\theta)(\phi(\theta-\alpha,k) - \bar{\phi}(\theta,k))}. \tag{7}$$

This is not a symmetric function, so that $c(j,k) \neq c(k,j)$ in general. We set $c(i,i) = 0$. We merge classes j and k only if their inter-class scores are such that they share a mutually maximal class:

$$\left(\arg\max_i c(j,i) = \arg\max_i c(i,j) \right) = \left(\arg\max_i c(k,i) = \arg\max_i c(i,k) \right). \tag{8}$$

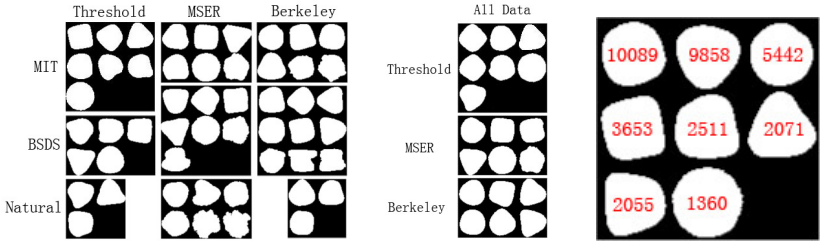
This ensures the pair of classes are tightly bound. In practice, we can group several clusters simultaneously because a single icon may be mutually maximal with several others, so that this form of agglomerative clustering is very efficient. All shape classes within a single group are bundled into one, aligned, and a new mean and error image is computed by weighted sums. For example $m(x,y,j') = \sum_j N_j m(x,y,j) / \sum_j N_j$, similarly for error images. Agglomerative clustering halts when there is no change in the number of shape classes.

2.5 Results

Final shapes for each database and each segmentation method can be seen in Figure 5. The shapes tend to be be simple — and nameable shapes such as circles, squares and triangles are common. In some cases we also see a square under a homography, which lies between square and triangle in feature space, and a simple shape lies between the square and circle, we conjecture it is a composite of higher order regular polygons. There are some irregular looking shapes too, but these are not often observed compared to the regular shapes. The fractional number of these 'prime' shapes depends on segmentation and database, but is consistently high; as Table 1 shows, over 1/2 of all segmented regions fall into one of the discovered categories.

3 An Application

Having found prime shapes, we now make use of them. One obvious application is to classify regions in a new segmentation. To do this we construct a Gaussian mixture model (GMM)



(a) (b) (c)

Figure 5: Matrices of final results. In each matrix element the shapes are ordered by descending frequency from top-left to bottom-right. (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

	Thresholding	MSER	Berkeley
MIT	66.65%	67.12%	81.45%
BSDS	59.92%	54.60%	80.65%
Natural	62.64%	36.24%	60.36%

Classified Shapes	37039
Un-Classified Shapes	19953
Classification Fraction	64.99%

Table 1: *Left*: The percentage of 'prime shapes' detected as statistically significant amongst total shapes from each database, using different segmentation algorithms. *Right:* The number of classified and un-classified shapes from all three databases, and the fraction of classified shapes. See Figure 5 to the total in each shape class.

of the density of shape moments for each class that it output by the clustering algorithm. In addition, we construct a GMM over all the random shapes. Using random shape classes allows for the possibility that a given segmented region is not classified as a prime shape. For a shape class S, let $\Omega = (N, \{\mu_i, C_i\})$ be a GMM with means μ_i and covariance matrices C_i.

$$p(x|S) = \sum_{i=1} N p(x|\Omega_i) p(\Omega) \tag{9}$$

The posterior that shape x belongs to class S follows from Baye's rule

$$p(S|x) = \frac{p(x|S)p(S)}{\sum_{T \in \mathcal{S} \cup R} p(x|T)p(T)} \tag{10}$$

where $\mathcal{S}$ is the set of prime shape class, and R the random shape class. the priors $p(T)$ are the proportion of shapes clustered into class T from the image data base being used; typically $p(T = R) \approx 0.5$ but since the random shapes are thinly distributed the likelihood $p(x|S = R)$ is often small, so segmented regions are classified into prime shape categories. Typical output can be seen in Figure 1

We noticed that the priors on different prime shapes depends on the database used, and these contain different sorts of photograph. The MIT database, for example contains street scenes, where as our natural database is exclusively landscapes, forests, coastal scenes *etc.*

(a) Coast (b) Forest (c) Highway (d) Inside City

(e) Mountain (f) Open Country (g) Street (h) Tall Building

Figure 6: Typical pictures from Oliva's database [24].

This suggests a scene classification appliaction, which is a popular research area, for example [7, 15, 16, 24]. Our classifiers learn the ratio of prime shapes (and random shapes) associated with a given category of scene. We assume the ratio of priors for a given category is Dirichlet distributed, because the ratios for any given image sum to unity and are a multinomial distribution. That is, if z is the ratio of priors then

$$p(z) \sim D(\beta_1, ..., \beta_K) = \frac{\Gamma(\Sigma_k \beta_k)}{\prod_k^K \Gamma(\beta_k)} \prod_k^K z_k^{\beta_k - 1}, \tag{11}$$

where Γ denotes the gamma function, and β are the Dirichlet parameters found by fitting [21]. Each distinct scene category, C has a distinct vector of β values. Given a new scene it is then easy to compute its ratio of prime shapes (and random), z, and hence compute $p(z|C)$ and therefore the posterior $p(C|z)$.

We used Oliva's database [24], partitioning the data into 800 training images and 800 test images. The test set images have given ground-truth categories, so we could produce the confusion matrix seen in Table 2 — next to Oliva's own results for comparison, which is representative of state of the art. Our result is not quite as strong as Oliva's, but strong nonetheless; broad classes such as 'Urban' and 'Natural' are very well classified. Given our approach is a simpler algorithm than any state of the art alternative, we found this to be a surprising result. We can explain the ambiguous cases in our table, but space prevents us, and scene classification is just one example. Our purpose here is to show that prime shapes have value in real applications; cross modal matching [4, 5] is a particular interest of ours.

	T	I	S	H	C	O	M	F
tal	**80**	0	0	11	0	0	0	9
ins	0	**85**	3	5	0	5	2	0
str	2	**42**	**20**	15	3	0	17	1
hig	7	1	9	**83**	0	0	0	0
coa	3	0	3	19	**75**	0	0	0
ope	17	0	4	0	17	**2**	**47**	13
mou	1	0	0	0	9	0	**90**	0
for	11	0	1	0	4	1	31	**52**

	T	I	S	H	C	O	M	F
tal	**82**	9	2	0	0	0	5	1
ins	3	**90**	3	1	0	1	0	0
str	1	5	**89**	2	0	1	2	1
hig	0	3	2	**87**	4	4	1	0
coa	0	0	0	8	**79**	12	1	0
ope	0	0	2	5	13	**71**	6	3
mou	1	0	2	2	2	5	**81**	7
for	1	0	0	0	0	1	6	**91**

Table 2: Confusion Matrix. *(Left):Our Proposed Method (Green: Urban Scene, Yellow: Natural Scene), (Right):Spatial Envelope* [24]

4 Conclusion

This is an experimental paper. Its main contribution 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*.

Clustering and other details such as alignment and noise handling can no doubt be improved, perhaps to sharpen the output icons. Yet 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.

References

[1] N. Alajlan, M.S. Kamel, and G.H. Freeman. Geometry-based image retrieval in binary image databases. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6):1003 –1013, june 2008. ISSN 0162-8828.

[2] 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.

[3] X. Bai, X. Wang, W. Liu, and Z.Tu. Co-transduction for shape retrieval. In *European Conference on Computer Vision*, volume 6313 of *Lecture Notes in Computer Science*, pages 328–341, 2010. ISBN 978-3-642-15557-4.

[4] A. Balikai and P.M. Hall. Depiction invariant object matching. In *British Machine Vision Conference*, 2012.

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

[6] Mohammad Reza Daliri and Vincent Torre. Robust symbolic representation for shape recognition and retrieval. *Pattern Recognition*, 41(5):1782 – 1798, 2008. ISSN 0031-3203.

[7] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 524 – 531 vol. 2, june 2005.

[8] P.-E. Forssen and D.G. Lowe. Shape descriptors for maximally stable extremal regions. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1 –8, oct. 2007.

[9] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt. Shape representation and classification using the poisson equation. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–61 – II–67 Vol.2, june-2 july 2004.

[10] Robert M. Haralick, Stanley R. Sternberg, and Xinhua Zhuang. Image analysis using mathematical morphology. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9(4):532 –550, july 1987. ISSN 0162-8828.

[11] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179 –187, february 1962. ISSN 0096-1000.

[12] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2106 –2113, 29 2009-oct. 2 2009.

[13] A. Khotanzad and Y.H. Hong. Invariant image recognition by zernike moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(5):489 –497, may 1990. ISSN 0162-8828.

[14] L.J. Latecki, R. Lakamper, and T. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 424 –429 vol.1, 2000.

[15] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169 – 2178, 2006.

[16] Heping Li, Fangyuan Wang, and Shuwu Zhang. Global and local features based topic model for scene recognition. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 532 –537, oct. 2011.

[17] S.X. Liao and M. Pawlak. On image analysis by moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(3):254 –266, mar 1996. ISSN 0162-8828.

[18] Haibin Ling and D.W. Jacobs. Shape classification using the inner-distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(2):286 –299, feb. 2007. ISSN 0162-8828.

[19] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416 –423 vol.2, 2001.

[20] 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.

[21] T.P. Minka. Estimating a dirichlet distribution. Technical Report 8, 2003.

[22] F. Mokhtarian and A.K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(8):789 –805, aug 1992. ISSN 0162-8828.

[23] R. Mukundan, SH Ong, and PA Lee. Discrete orthognal moment features using chebyshev polynomials. 2000.

[24] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[25] Eric Persoon and King-Sun Fu. Shape discrimination using fourier descriptors. *Systems, Man and Cybernetics, IEEE Transactions on*, 7(3):170 –179, march 1977. ISSN 0018-9472.

[26] Z.L. Ping, R.G. Wu, and Y.L. Sheng. Image description with chebyshev-fourier moments. *JOSA A*, 19(9):1748–1754, 2002.

[27] I. Pitas and A.N. Venetsanopoulos. *Nonlinear digital filters: principles and applications*, volume 84. Springer, 1990.

[28] A.P. Reeves, R.J. Prokop, S.E. Andrews, and F.P. Kuhl. Three-dimensional shape analysis using moments and fourier descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(6):937 –943, nov 1988. ISSN 0162-8828.

[29] H. Rom and G. Medioni. Hierarchical decomposition and axial shape description. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(10):973–981, Oct 1993. ISSN 0162-8828.

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

[31] Y. Rui, A.C. She, and T.S. Huang. Modified fourier descriptors for shape representation-a practical approach. In *Proceedings First Int'l Workshop Image Databases and Multi Media Search*, volume 22, page 23, 1996.

[32] S.G. Salve and K.C. Jondhale. Shape matching and object recognition using shape contexts. 9:471 –474, july 2010.

[33] H. Sundar, D. Silver, N. Gagvani, and S. Dickinson. Skeleton based shape matching and retrieval. In *Shape Modeling International, 2003*, pages 130 – 139, may 2003.

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

[35] A. Temlyakov, B.C. Munsell, J.W. Waggoner, and Song Wang. Two perceptually motivated strategies for shape classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2289 –2296, june 2010.