Shading Cues for Object Class Detection

Michael Stark   Michael Goesele   Bernt Schiele
Computer Science Department, TU Darmstadt, Germany

Abstract

Recognition of object classes in natural images has made tremendous progress in recent years. Today's approaches often rely on powerful learning approaches as well as robust local 2D shape or appearance features. Exploiting 3D shape cues however has become unfashionable in recent literature. While shading cues play a major role in human perception of object shape, shape-from-shading techniques are seldom used today for object class detection. Drawing on ideas from the early days in object recognition this paper aims to revisit the concept of using shading primitives to support object class detection. We demonstrate and discuss the applicability of this approach to real world images of a standard benchmark data set. Experimental results suggest that our shading cues can be useful for object class detection.

1. Introduction

In recent years, impressive progress has been reported in the recognition of a wide variety of object classes. Object models based on robust local appearance features [20], in combination with bag-of-words [5], or more spatially constrained models [7] perform well on recognition benchmarks. More recently, 2D-shape-based approaches have also shown to yield comparable performance [8].

Interestingly, none of these 'modern' recognition approaches makes explicit use of 3D shape information provided by shading cues. This is in contrast to early approaches in object class recognition and also contrary to intuition, since humans make extensive use of shading information to assess object shape [15, 17], which is important for recognition. One might argue that at least part of the shading information is encoded implicitly by appearance features, and thus available to 'modern' recognition algorithms. This comes at a cost, however: in order to reliably separate possibly relevant shading information from background, these algorithms need to use statistics over large numbers of training samples. Explicitly modeling or learning shading information can remedy this problem, by en-
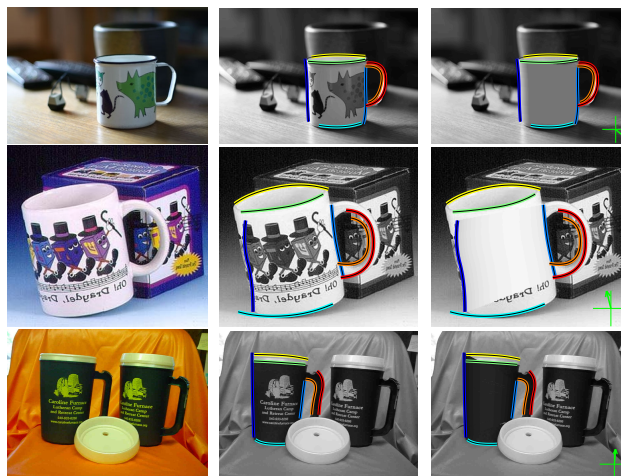

Figure 1. Shape-based object detections and shading cues on *ETHZ Mugs*. From left to right: (1) original image, (2) shape-based detection, (3) shading cue based on (2), back-projected into the image. The green arrow reflects estimated lighting direction, seen from above the scene.

coding relevant information into the model itself.

Inferring the shape of a surface from shading is unfortunately a difficult problem, and has long been a major focus of computer vision research. By nature, shape-from-shading (SFS) is highly ambiguous: without any prior knowledge, a given image of an observed scene could have been generated by an infinite number of different combinations of object surfaces present in the scene, their reflectance, and lighting conditions. As a consequence, SFS approaches are typically restricted to controlled environments, or reduce ambiguity by imposing strong assumptions on surface shapes, material, and lighting [14].

As a consequence, making direct use of SFS for object recognition in natural images has proven difficult although there have been many attempts. Worthington and Hancock [28] apply Shape-from-Shading techniques to object recognition, on the level of individual object instances (COIL-20 data set [22]). Their work builds upon a mid-level representation of surface topography based on local curvature and

shape-index [16] information, and uses histograms and region descriptors on top of this representation. Following a similar route, Lichtenauer et al. [19] suggest using orientation and curvature of isophotes (lines of equal brightness) as features in a classification framework for classifying image patches as face/non-face. Wu et al. [29] report improved performance for gender-classification of pose-aligned face images with needle-map features obtained via shape-from-shading. Nillius et al. [23] present generic shape detectors for cylinders and spheres, using model-based PCA, and a multi-scale sliding-window search over image regions. Mori et al. [21] describe shading on human limbs by prototypical, half-wave rectified gradient image patches, and use a similarity score in order to identify candidate limb image regions.

While these relatively recent approaches use SFS as bottom-up features, more than ten years ago, Haddon and Forsyth [13] suggested a promising alternative, by verifying given 3D shape hypotheses in a top-down fashion using shading cues. In line with Biedermann's theory of recognition-by-components [1], and similar in spirit to Weinshall [27], the authors suggest shading primitives as the basis for recognition. The recognition problem amounts to finding valid configurations of several primitives.

Borrowing from these ideas, we use a part-based object class model at the core of our approach. We explicitly model the 2D shape of individual parts, together with pairwise, semi-local symmetry relations, and the overall spatial layout. We then establish 3D shape hypotheses based on object parts and shading cues, and add them as additional cues to the final detection hypothesis. In particular, our paper makes the following contributions:

- We propose a shading model for cylindrical surface primitives, which we show to yield acceptable model fits on real world images, taken from a standard object detection benchmark [9], and analyze the failure cases.

- We present first results to integrate this shading model as an additional cue into an existing state-of-the-art shape-based object detection framework.

- We give quantitative experimental evidence that shading cues can indeed increase recognition performance.

The remainder of this paper is organized as follows. Section 2 introduces the shading model. Section 3 reviews the shape-based object detector. Section 4 gives experimental results, and Section 5 concludes with an outlook on future work.

2. Shading Model

Similar to the work of Haddon and Forsyth [13], our shading model follows the principle of hypothesis verification. Instead of recovering the 3D shape underlying an image area in a bottom-up fashion, it starts from a given 3D shape hypothesis and tries to verify this hypothesis based on image evidence. In particular, the observed image evidence must be consistent with the 3D shape, some estimated reflectance properties, and the estimated scene illumination. Proper regularization is required since the estimation is highly ambiguous — the same image can be generated by different combinations of surface shape, reflectance, and lighting.

2.1. A Shading Primitive

In the following, we present a concrete instantiation of this hypothesis verification framework for the case of cylindrical surfaces. Our model starts with the hypothesized occluding contours of a cylindrical shape (the cylinder sidewalls) in the image plane and tries to verify this hypothesis based on evidence from the pixels on the cylinder surface using a simple model for lighting and reflectance. (Figure 1 shows some successful examples on images from the *ETHZ Mugs* dataset.)

We assume that the directional lighting in the scene can be well approximated by a single point light source located far away from the surface of interest. In the limit, i.e., for infinite distance this corresponds to a directional light source. We model the remaining contribution as ambient illumination impinging on the surface uniformly from all directions. Both components of the model can be simply added due to the principle of superposition.

Regarding reflectance, we restrict ourselves to the simplest possible model and assume that the surface is diffuse (Lambertian) with a constant albedo [6]. Specular effects of surface texture are ignored and will be treated as outliers during parameter estimation. This model implies that barring occlusion effects the reflected radiance depends solely on the direction of incident radiance relative to the surface normal. All points with equal surface normals will exhibit equal brightness in the image.

Shading on cylindrical surfaces. Let us assume an orthographic projection of a cylindrical surface, with the viewing direction being perpendicular to the cylinder axis. We divide the surface into a set of circular cross-sections, such that the viewing direction is parallel to the corresponding sectional planes. A point on the observed half of a cross-section can then be described by the parameterization ϕ (see Figure 2). Due to orthographic projection, $s = \sin\phi$ can be used to parameterize the projection of the cross-section onto the image plane without introducing any distortions. We can now establish a functional dependency between s and the observed image values for the corresponding surface point $B(s)$.[1] Let ρ be the angle between the direction

---

[1] Note that we need to ensure that the image is in photometrically linear space. This typically requires applying an inverse gamma correction.
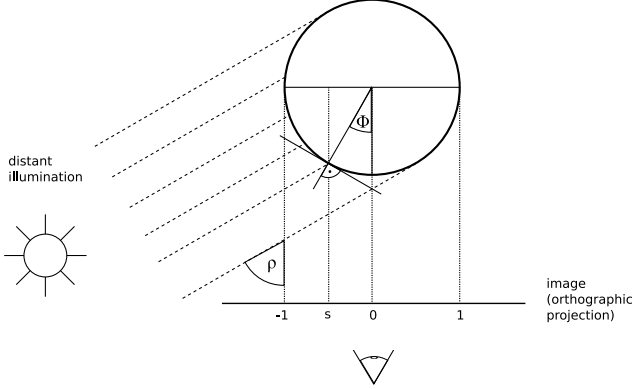
Figure 2. Geometry of visible cylinder (half) cross section, parameterized by $s = sin\phi$, viewing direction, and light source at infinity. ρ denotes the angle between illumination direction and (orthographic) viewing direction.

of incident light and the viewing direction, both projected on a plane perpendicular to the cylinder axis. The observed image value is then

$$B(s) = a + b * \max(0, \cos(\rho - \phi(s))). \quad (1)$$

The two scaling factors $a > 0$ and $b > 0$ determine the intensity of the ambient and the directional lighting, respectively, multiplied with the albedo. The maximum in Equation 1 ensures that surface points with normals pointing away from the directional light source, and which are therefore in shadow, do not contribute physically invalid, negative radiance.

Let us now assume viewing the cylinder from an elevated angle and/or rotating the camera around the viewing direction. The corresponding cross-sections are no longer perpendicular to the cylinder axis, and change their shape from circular to elliptical. As a consequence of both orthographic projection and directional lighting, these elliptical cross-sections can be transferred into equivalent circular cross-sections by sliding all constituent points along the cylinder's isophotes, parallel to the cylinder axis. Following this argumentation, Equation 1 can be proven valid for any cylinder cross-section without changing the parameterization s, as long as its projection on the image plane is a straight line connecting the two cylinder side-walls.

Implementation. In order to determine the model parameters a, b, and ρ we need to select a set of cross-section points

$$\{s_i\}_{i=1}^n, s_i \in [-1, 1] \quad (2)$$

and corresponding brightness values

$$\{b_i\}_{i=1}^n, b_i \in [0, 1]. \quad (3)$$

We obtain pairs of the form (s_i, b_i) by first sampling a fixed number of equidistant points on the two occluding contours

of a hypothesized cylindrical surface and then connecting corresponding pairs of points by straight lines. We finally sample pixel brightness values b_i by parameterizing each line by $s_i \in [-1, 1]$ using the Bresenham algorithm [3].

The parameters a, b, and ρ can now be determined using standard non-linear least squares optimization techniques, such as the Levenberg-Marquardt algorithm [18], by minimizing the sum of squared residuals

$$S(a, b, \rho) = \sum_{i=1}^n (b_i - B(s_i))^2. \quad (4)$$

In practice, we observe that the non-differentiable $\max(.)$ function does not pose any problems during optimization.

Since surface texture, specular reflections, and other unmodeled effects often yield a significant number of outliers (see, e.g., the textured mug in Figure 3(a)), we use RANSAC [10] to select a single consistent model. Invariance w.r.t. global brightness variations is achieved by selecting inliers according to a threshold on the squared residual $(\log b_i - \log B(s_i))^2$ in logarithmic space.

2.2. Example shading model fits

In order to demonstrate the validity of the proposed cylindrical shading model, we give qualitative as well as preliminary quantitative results on the *Mug* category of the ETHZ Shape Classes dataset [9]. Figure 3 visualizes exemplary shading model fits of varying quality on eight different images, starting from shape-based object detections (see Section 3). In particular, we select the single best true positive *Mug*-hypothesis per image, each consisting of seven parts (left and right side-walls, upper and lower rims, bottom, and two handle parts), and fit a cylindrical shading model between the side-walls of the *Mug*.

2.3. Discussion

Table 1 gives an assessment of the quality of obtained shading fits on *Mug* objects. It compares shading fits obtained by using detection hypotheses from the shape-based object detector with fits obtained from ground truth annotations of *Mug* side walls. Since we obtain ground truth side walls by marking actual shape features in images, these annotations are available only for 36 of 44 *Mug* images, due to imperfect shape feature detections. The table further gives estimates on the quality of the original shape fits, as these constitute the basis for shading fits. Since ground truth on the lighting conditions in which the images were taken is hard to acquire in retrospect, shading fit quality is assessed by visual inspection, and roughly categorized into *near perfect* (all parameters sensibly fit), *acceptable* (parameter estimates deviate slightly from human assessment), and *failure* (clearly erroneous parameter estimates).

We note the following observations: First, in 0.94 of the cases, an at least *acceptable* shading model can be fit from
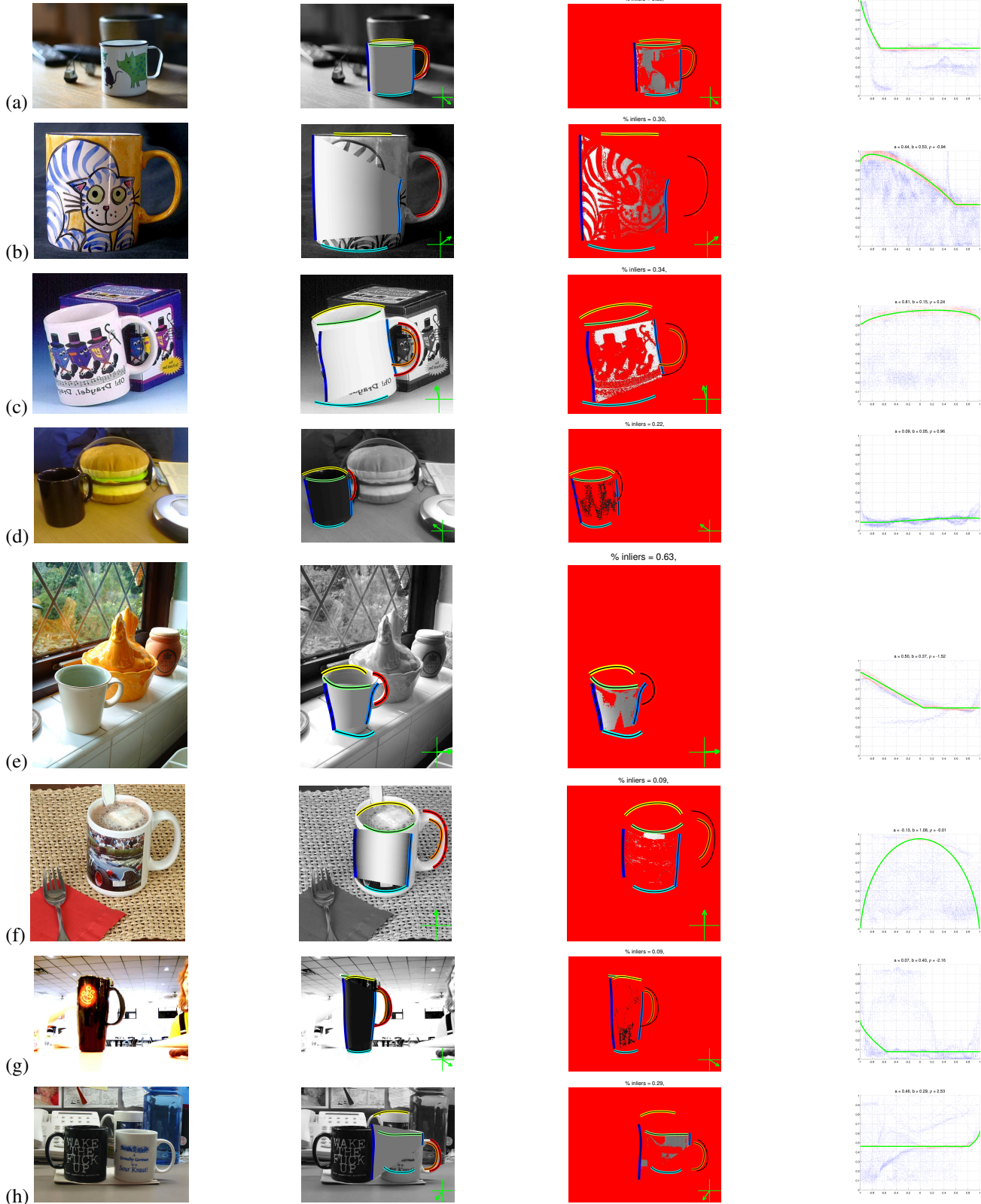
Figure 3. Example shading fits, based on shape-based object detection hypotheses (see Section 3). First column: original image. Second column: back-projected shading model. Third column: RANSAC inliers and estimated lighting direction, seen from above the scene. Fourth column: shading model fit with accepted samples (red) and outliers (blue). *Near perfect* fits: (a) - (d), *acceptable* fit: (e), *failures* ((f) - (h)) due to object texture (f), specularities (g), erroneous shape fits (h).

Fit quality	Shape	Shading on GT	Shading on shape fit
Failure	0.14	0.06	0.27
Acceptable	0.18	0.25	0.16
Near perfect	0.68	0.69	0.57
Non-failure	**0.86**	**0.94**	**0.73**

Table 1. Quality of shape and shading model fits. The last row summarizes the two preceding rows.

Failure case	Specularities	Shape fit	Lighting	Texture
Fraction	0.42	0.32	0.16	0.11

Table 2. Likely failure reasons of shading model fits.

available ground truth occluding contours. This indicates that the proposed shading model is in principle capable of modeling most shading artifacts present on the tested ETHZ *Mug* images, despite variations in shape, texture, material, and lighting. Second, despite the fact that this number decreases significantly if shape model fits are used as a basis an encouraging amount of 0.73 of the obtained shading models is still at least *acceptable*. These models correctly reflect the cylindrical 3D shape of the *Mug* objects, and can thus be beneficial for recognition.

Table 2 lists the most likely reasons for imperfect fits, again assessed by visual inspection. The most frequent likely reason (0.42) for failure is the presence of specularities and reflections, which are not explicitly included in the shading model, but possibly rejected as outliers by RANSAC. Figure 3(g) gives an example of an erroneous fit, caused by the highly specular, dark material of the mug.

The second most frequent reason for failure is the sometimes insufficient quality of shape model fits used as the basis for shading. According to Table 1, 0.14 of these shape fits are failures, resulting in erroneous support for the shading model. Figure 3(h) shows an example, where pixels on the mug and pixels from the background are wrongfully combined in the set of selected inliers.

Figure 3(e) gives an example of a still acceptable fit, showing a deviation in the estimated incident light direction from what one would expect: contrary to intuition, the incident light is estimated as coming strictly from the left, and not from the direction of the window. This is an instance of difficult lighting conditions, and attributed to 0.16 of the failure cases.

Surprisingly, texture is rarely a source of confusion (0.11 of the cases). Figure 3(f) shows one of the few examples where object texture (a photo printed onto the mug) is wrongfully picked up by the shading model (the shape fit for this example is also imperfect; the corresponding shading fit for ground truth side-walls is in fact *near perfect*). Figure 3(a) - (c) gives examples of successfully handled textures.

3. Shape Model

Our approach to integrate shading cues into object recognition is based on the shape-based object class detector presented by Stark et al. [25]. In particular, we use information from this model in two different ways: 1) the shading model described in Section 2 is used to verify hypotheses provided by the shape-based part-detections (cylinder side walls), and 2) a final score is calculated by combining the shape-based detection scores with the fitted shading model parameters (detailed in Section 4). We give a brief overview on the most important aspects of the shape-based detector, but refer the reader to [25] for details.

Local shape features. The detector is based on a novel flavor of local shape features, which constitute a discrete, over-complete representation of image contours. The shape features are based on the *Contour Segment Network* (CSN) of Ferrari et al. [9], and its associated local companions, k-Adjacent Segments (k-AS) [8]. Multiple k-AS of varying k are combined by fitting a parametric B-spline curve to all constituent edgel chains, using the exact same parameterization, independent of k. This offers the benefit of retaining the original *curvature* information and increasing the discriminative power of the features compared to the original k-AS represented by straight line approximations. Spline curves are fit against edgel chains, once they have been transformed into a translation and scale invariant space, using Procrustes analysis [4]. The resulting spline parameters are used as a low-dimensional local shape description.

Semi-local symmetry relations. In order to facilitate discrimination between object parts and background clutter, the detector of [25] includes another powerful perceptual cue, which relates pairs of local shape features by identifying and describing *symmetries* between them. In particular, it uses a B-spline-based implementation [24] of *Smoothed Local Symmetry* (SLS). SLS were originally proposed by Brady and Asada [2] in the context of planar shape analysis. Starting from the spline-based representation of SLS, [25] devises a semi-local symmetry descriptor, which captures both the shape of the symmetry axis and the lengths of lines connecting points on the axis with corresponding points on either of the symmetrical splines (the *symmetry lines*). The first is achieved by representing the axis as a local shape feature, as described in the previous paragraph. The second is achieved by recording a length profile of selected symmetry lines.

Probabilistic model. In [25] objects are modeled as an assembly of spatially arranged parts, in the spirit of the constellation model of Fergus et al. [7]. The corresponding probabilistic formulation subsumes individual part shape S,

binary symmetry relations B, relative part scales R, and their overall spatial layout X.

During detection, the goal is to find an *assignment* of all P model parts to local shape features, denoted the detection hypothesis $H = (h_1, \ldots, h_P)$. That is, h_p contains a local shape feature identifier assigned to part p. The detection problem can be formulated as a maximum a posteriori hypothesis search over the distribution $p(X, R, S, B, H|\theta)$, which is the joint posterior distribution of H and image evidence, given a learned model θ. It factors into separate likelihood contributions for local part shape, symmetry relations, spatial part layout, relative part scales, and a prior on hypotheses, as follows:

$$p(X, R, S, B, H|\theta) =$$
$$\underbrace{p(S|H,\theta)}_{Local\,Shape}\underbrace{p(B|H,\theta)}_{Symm.\,Rel.}\underbrace{p(X|H,\theta)}_{Layout}\underbrace{p(R|H,\theta)}_{Rel.\,Scale}\underbrace{p(H|\theta)}_{Prior}(5)$$

Learning and inference. Model parameters θ are learned using maximum likelihood estimation via supervised training. Supervision is provided by labeling contour segments in training images.

Inference aims at identifying the maximum a posteriori hypothesis $H_{\text{MAP}} = \arg\max_H p(H|X, R, S, B, \theta)$. This is equivalent to $\arg\max_H p(X, R, S, B, H|\theta)$. [25] approximates H_{MAP} by drawing samples from $p(X, R, S, B, H|\theta)$ using the Metropolis-Hastings (MH) algorithm [12]. The Single Component update variant of MH is employed, since it allows to separately update individual components of the target density, conditioned on the remaining portion of the current state of the Markov chain. This opens the possibility to guide the sampling towards high density regions by data-driven, bottom-up proposals [30, 26].

Multiple object instances are detected by running a number of independent Markov chains, and memorizing the per-chain highest-scoring hypotheses. The greedy non-maximum suppression described by Fritz and Schiele [11] is used to prune overlapping hypotheses.

Detection results. Figure 4 shows learned models and example detections for four categories *Giraffes*, *Swans*, *Mugs*, and *Bottles* of the ETHZ Shape Classes data set [9]. [25] reports competitive results for these four classes outperforming, e.g., [8, 11]. For further details on these results and the model we refer to [25].

4. Experiments

The following examines the potential benefit of adding our shading cue for object recognition. To integrate our shading cue into the probabilistic model of the shape-based object detector described in [25], we combine the outputs of both models in a discriminative framework (sometimes
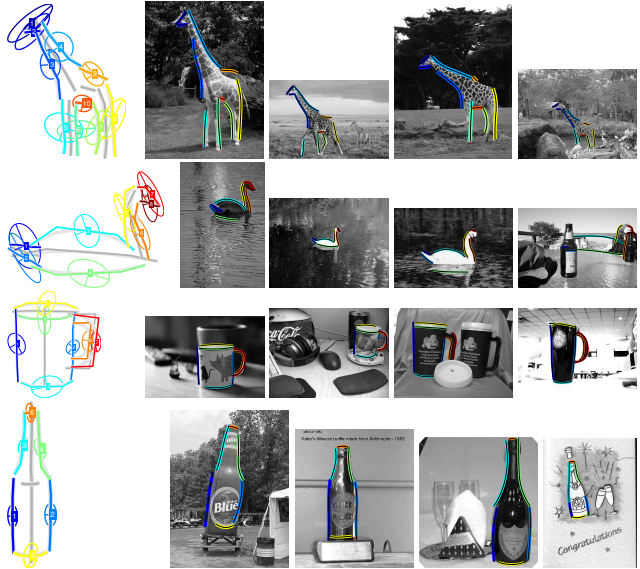

Figure 4. Learned ETHZ Shape Classes models (left) and example detections (right). For models, mean local part shapes and selected mean symmetry axes are drawn at mean positions and relative scales. Covariances of part positions are shown as ellipses. Reproduced from [25] with permission.

referred to in the literature as *late integration*). In particular, we train two linear SVM classifiers. The first is using the shading model parameters a, b, ρ, the fraction of inliers, and the mean squared residual on the inliers. The second additionally uses the shape-based detection score.

As in Section 2, we base our evaluation on the category *Mug* from the ETHZ Shape Classes data set [9]. We set up a binary classification task as follows: for each of the 251 images (44 *Mugs*, 207 non-*Mugs*) of the data set, we select the highest scoring detection hypotheses for the category *Mug*. We then either store it as a positive (in case it qualifies as a true positive detection according to an overlap criterion) or as a negative (in case it does not) training example. We then train and test classifiers on these examples using 5-fold cross validation, in order to have a reasonable amount of positive training examples available. Each model is individually optimized w.r.t. the maximum margin-training error minimization tradeoff parameter C of the linear SVM. Please note that this experiment is different from the original setup in [8] and therefore does not allow for direct comparison. However, as a first proof of concept and to understand the potential benefit of our shading cues for recognition we consider this experiment appropriate for the purpose of this paper.

Figure 5 plots precision and recall curves for the binary classification task described above. While the red curve is based on the shape-based detection score alone, the blue and green curves have been obtained by varying a threshold
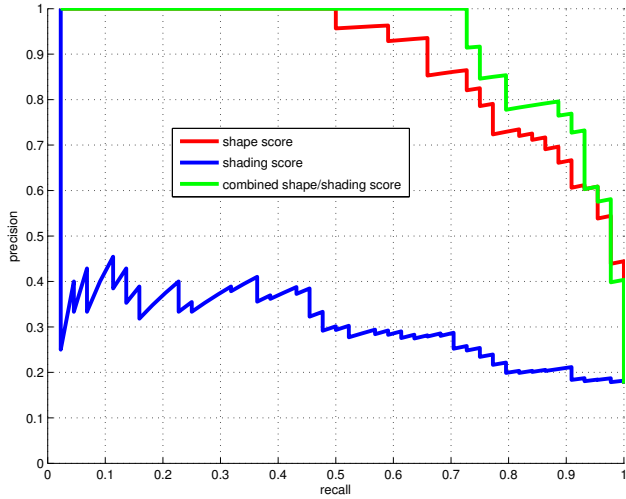
Figure 5. Precision/recall curves for classifying shape-based detection hypotheses into *Mugs*/non-*Mugs*, based on different scores.
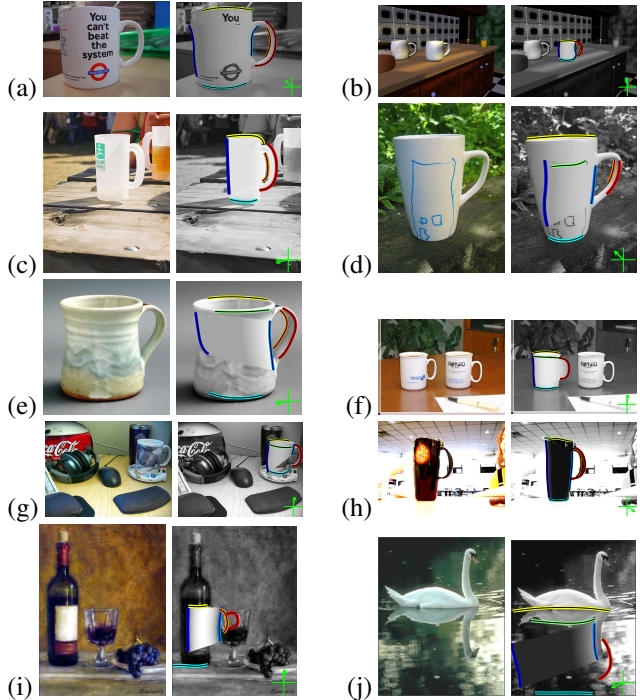

Figure 6. (a) - (h): complementary detections. (a) - (f): six high scoring *Mug*-hypotheses correctly classified by the combined shape/shading score, but missed by the pure shape score (precision level 1.0, highest recall). (g),(h) the only two hypotheses for the inverse case. (i), (j) Two false positives of the combined shape/shading score at EER.

on the corresponding SVM scores, obtained by 5-fold cross validation.

Observations. First, the classifier trained on shading model parameters alone essentially fails to discriminate between positive and negative examples. This is not surprising, since it fully neglects valuable information about the shape and spatial layout of *Mugs*, while concentrating fully on cylindrical shading (as, e.g., in Figure 6(i),(j)).

Second, the shape-based detection score shows good performance (Equal-Error-Rate (EER) of 77.3%), despite the negative examples being hard ones (since we picked the highest-scoring ones per image).

Third, combining the shape-based detection score and the shading model parameters yields a considerable improvement over the shape-based detection score. The shading cue improves recall at precision level 100% from 50% to 72.7%, and lifts EER from 77.3% to 79.6%. Figure 6(a) - (h) depicts complementary detections hit and missed by the two classifiers, respectively. It lists high scoring detections correctly classified by one, but mis-classified by the other classifier, at the highest achieved recall for precision 1.0. Apparently, the combined shape-shading classifier makes efficient use of available shading information, compensating weak shape model fits (Figure 6(c) - (f)). The two examples mis-classified by the shape-shading combination can be attributed to imperfect shading fits due to specularities and texture, respectively. Figure 6(i),(j) show two false positive classifications of the combined shape/shading score. While the bottle label is in fact an instance of cylindrical shading, the water surface underneath the swan is clearly an error.

5. Conclusions and Future Work

In this paper, we have introduced a shading model for cylindrical surface primitives, based on hypothesis verification, and demonstrated its validity on images of a standard data set for shape-based object detection. We have shown preliminary results of integrating this shading model as an additional cue into an existing, state-of-the-art, shape-based object detection framework, and obtained quantitative experimental evidence for its potential usefulness in recognition.

Based on these encouraging results, we consider the proper integration of the proposed shading cue into the Data-Driven Markov Chain Monte Carlo framework of [25] an obvious next step, as well as adding more 3D surface primitives.

References

[1] I. Biedermann. Recognition-by-components: a theory of human image understanding. In *Psychol Rev*, 1987.

[2] M. Brady and H. Asada. Smoothed local symmetries and their implementation. In *IJRR*, 1984.

[3] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 1965.

[4] T. Cootes. An introduction to active shape models, 2000.

[5] G. Csurka, C. Dance, L. Fan, J. Willarnowski, and C. Bray. Visual categorization with bags of keypoints. In *SLCV*, 2004.

[6] J. Dorsey, H. Rushmeier, and F. X. Sillion. *Digital Modeling of Material Appearance*. Morgan Kaufmann, 2007.

[7] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, 2003.

[8] V. Ferrari, F. Jurie, and C. Schmid. Accurate object detection with deformable shape models learnt from images. In *CVPR*, 2007.

[9] V. Ferrari, T. Tuytelaars, and L. J. V. Gool. Object detection by contour segment networks. In *ECCV*, 2006.

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[11] M. Fritz and B. Schiele. Decomposition, discovery and detection of visual categories using topic models. In *CVPR*, 2008.

[12] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo In Practice*. 1996.

[13] J. Haddon and D. Forsyth. Shape representations from shading primitives. In *ECCV*, 1998.

[14] B. Horn and M. Brooks. *Shape from Shading*. 1989.

[15] D. A. Kleffner and V. S. Ramachandran. On the perception of shape from shading. *Perception and Psychophysics*, 52:18–36, 1992.

[16] J. Koenderink and A. van Doorn. Surface shape and curvature scales. 1992.

[17] J. Koenderink, A. Van Doorn, C. Christou, and J. Lappin. Perturbation study of shading in pictures. *Perception*, 1996.

[18] K. Levenberg. A method for the solution of certain nonlinear problems in least squares. *The Quarterly of Applied Mathematics*, 1944.

[19] J. Lichtenauer, E. Hendriks, and M. Reinders. Isophote properties as features for object detection. In *CVPR*, 2005.

[20] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[21] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering human body configurations: Combining segmentation and recognition. 2004.

[22] S. A. Nene, S. K. Nayar, and H. Murase. Columbia object image library (coil-20), 1996.

[23] P. Nillius, J. Sullivan, and A. Argyros. Shading models for illumination and reflectance invariant shape detectors. In *CVPR*, 2008.

[24] P. Saint-Marc, H. Rom, and G. Medioni. B-spline contour representation and symmetry detection. *PAMI*, 15(11):1191–97, 1993.

[25] M. Stark, M. Goesele, and B. Schiele. A shape-based object class model for knowledge transfer. In *Twelfth IEEE International Conference on Computer Vision (ICCV)*, 2009.

[26] Z. Tu, X. Chen, A. Yuille, and S. Zhu. Image parsing: Unifying segmentation, detection and recognition. *ICJV*, 2005.

[27] D. Weinshall. Local shape approximation from shading. In *CVPR*, 1992.

[28] P. L. Worthington and E. R. Hancock. Object recognition using shape-from-shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(5):535–542, 2001.

[29] J. Wu, W. Smith, and E. Hancock. Gender classification using shape from shading. In *BMVC*, 2007.

[30] S.-C. Zhu, R. Zhang, and Z. Tu. Integrating bottom-up/top-down for object recognition by data driven markov chain monte carlo. In *CVPR*, 2000.