## Back to the Future: Learning Shape Models from 3D CAD Data





Figure 1: Learning shape models from 3D CAD data. (a) Collection of 3D CAD models composed of semantic parts, (b) viewpoint-dependent, non-photorealistic renderings, (c) learned spatial part layouts, (d) multi-view detection results.

In the 70's and 80's the predominant approach to recognition was based on 3D representations of object classes [2, 6, 7, 8, 9]. While being based on an intriguing paradigm, these approaches showed only limited success when applied to real-world images. This was due to both the difficulty to robustly extract 2D image features and the inherent ambiguity when matching them to 3D models. Today, the predominant paradigm to recognition relies on robust features and powerful machine learning techniques. While enabling impressive results, these methods have at least two inherent limitations: they are typically limited to a single viewpoint, and rely on the existence of sufficient representative real-world image training data, limiting their generality and scalability.

The starting-point of this paper is therefore to go back to the idea of learning object class models from 3D computer aided design (CAD) models (Fig. 1(a)), not using any real-world training images of the object class of interest. In contrast to early approaches, we draw from a multitude of advancements in both object class recognition and 3D modeling, which we use as tools for designing highly performant object class models. The first tool is an abstract shape representation that establishes the link between 3D models and natural images. It is based on the viewpointdependent, non-photorealistic rendering of 3D CAD model edges (silhouettes, creases, and part boundaries, Fig. 1(b)). The second tool is a collection of discriminatively trained part detectors [1], based on robust dense shape feature descriptors on top of this representation (Fig. 1(b)). The third tool is a powerful probabilistic model governing the spatial layout of object parts in each viewpoint (Fig. 1(c)), capable of representing the full covariance matrix of all part locations, reminiscent of the constellation model [3]. All three tools aim at capturing representative object class statistics from a collection of 3D CAD models, increasing the robustness of the resulting object class model.

For a given viewpoint, following [10], we formulate recognition as a MAP search over constellations H, given the shape of individual parts S, their relative scales R, and their overall spatial layout X.

$$p(X, R, S, H|\theta) = \underbrace{p(S|H, \theta)}_{Local Shape} \underbrace{p(X|H, \theta)}_{Layout} \underbrace{p(R|H, \theta)}_{Relative Scale} \underbrace{p(H|\theta)}_{Prior}$$
(1)

The MAP solution is approximated by efficient data-driven MCMC sampling [12], using part detector responses as proposals. Multi-view recognition is achieved by training individual detectors for a discrete set of viewpoints (bank-of-detectors), and combining all detections via nonmaximum suppression.

Fig. 2 (first row) gives example car detections for the 3D object classes data set [11] using a bank of 8 detectors, together with a quantitative comparison of multi-view recognition performance with related work (second row, left) and a viewpoint confusion matrix (second row, right). Achieving an average precision (AP) of 81.0% (green curve), our bank of 8 detectors clearly outperforms all three related approaches (APs 55.3% [11]

- <sup>1</sup> MPI Informatics,
- Saarbrücken, Germany
- <sup>2</sup> Computer Science Department, TU Darmstadt, Germany



Figure 2: Multi-view object class detection results. Example detections (first row). Comparison to state-of-the-art ([4, 5, 11], second row, left), confusion matrix for viewpoint classification (second row, right).

(cyan curve), 72.6% [4] (blue curve), and 76.7% [5] (magenta curve)). Notably, all related approaches use real-world training images either as the sole source of information [4, 11] or in combination with 3D models [5]. Relying solely on 3D CAD data further allows as to train models for arbitrary viewpoints. We thus analyze both the sensitivity to viewpoint variation and the required density of sampled viewpoints in our experiments, improving the performance of our approach further by increasing the number of detectors from 8 to 36 in a 10-degree spacing (Fig. 2, red curve, AP 89.9%).

- [1] M. Andriluka, S. Roth, and B. Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In *CVPR*, 2009.
- [2] R.A. Brooks, R. Creiner, and T.O. Binford. The acronym modelbased vision system. In *Intern. Joint Conference on Artiticial Intelligence*, pages 105–113, 1979.
- [3] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, 2003.
- [4] G. Gill and M. Levine. Multi-view object detection based on spatial consistency in a low dimensional space. In *DAGM*, 2009.
- [5] J. Liebelt and C. Schmid. Multi-view object class detection with a 3D geometric model. In *CVPR*, 2010.
- [6] D.G. Lowe. Three-dimensional object recognition from single twodimensional images. *Artificial Intelligence*, 31:355–395, 1987.
- [7] D. Marr and H.K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Roy. Soc. London B 200*, pages 269–194, 1978.
- [8] R. Nevatia and T.O. Binford. Description and recognition of curved objects. Artificial Intelligence, 8:77–98, 1977.
- [9] A.P. Pentland. Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28:293–331, 1986.
- [10] M. Stark, M. Goesele, and B. Schiele. A shape-based object class model for knowledge transfer. In *ICCV*, 2009.
- [11] H. Su, M. Sun, L. Fei-Fei, and S. Savarese. Learning a dense multiview representation for detection, viewpoint classification and synthesis of object categories. In *ICCV*, 2009.
- [12] 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.