Object Matching Using Boundary Descriptors

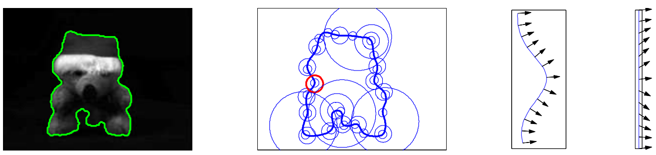
Ognjen Arandjelović
ognjen.arandjelovic@gmail.com

Swansea University, UK

The problem of recognizing 3D objects from images has been one of the most active areas of computer vision research in the last decade. This is a consequence not only of the high practical potential of automatic object recognition systems but also significant breakthroughs which have facilitated the development of fast and reliable solutions. These mainly centre around the detection of robust and salient image loci (keypoints) or regions, and the characterization of their appearance (local descriptors). While highly successful in the recognition of textured objects even in the presence of significant viewpoint and scale changes, these methods fail when applied on texturally smooth (i.e. nearly textureless) objects. Unlike textured objects, smooth objects inherently do not exhibit appearance from which well localized keypoints can be extracted.

Since their texture is not informative, characteristic discriminative information of smooth objects must be extracted from shape instead. Considering that it is not possible to formulate a meaningful prior which would allow for the reconstruction of an accurate depth map for the general class of smooth 3D objects, the problem becomes that of matching apparent shape as observed in images. This is a most challenging task because apparent shape is greatly affected by out of plane rotation of the object. What is more, the extracted shape is likely to contain errors when the object is automatically segmented out from realistic, cluttered images. The bag of boundaries (BoB) method of Arandjelović and Zisserman was the first to address this problem explicitly [1].

Boundary keypoint detection The problem of detecting characteristic image loci is well researched and a number of effective methods have been described in the literature. When dealing with keypoints in images, the meaning of saliency naturally emerges as a property of appearance (pixel intensity) which is directly measured. This is not the case when dealing with curves for which saliency has to be defined by means of higher order variability which is computed rather than directly measured. In this paper we detect characteristic boundary loci as points of local curvature maxima, computed at different scales. Starting from the finest scale after localizing the corresponding keypoints, Gaussian smoothing is applied to the boundary which is then downsampled for the processing at a coarser scale. Having experimented with a range of factors for scale-space steps, we found that little benefit was gained by decreasing the step size from 2 (i.e. by downsampling finer than one octave at a time). We estimate the curvature at a vertex as the curvature of the circular arc fitted to three consecutive boundary vertices: the vertex of interest, its predecessor, and its successor. An example of a boundary contour and the corresponding interest point loci are shown respectively in Figures 1(a) and 1(b). The method used to perform Gaussian smoothing of the boundary is explained next.



(a) Contour (b) Keypoints (c) Descriptor

Figure 1: (a) Original image of an object overlaid with the object boundary (green line), (b) the corresponding boundary keypoints detected using the proposed method and (c) an illustration of a local boundary descriptor based on the profile of boundary normals' directions (the interest point is shown in red in (b)).

Boundary curve smoothing. The most straightforward approach to smoothing a curve such as the object boundary is to replace each of its vertices by a Gaussian-weighted sum of vectors corresponding to its neighbours. However, this method introduces an undesirable artefact which is demonstrated as a gradual shrinkage of the boundary. In the limit, repeated smoothing results in the collapse to a point – the centre of gravity of the initial curve. We solve this problem by applying two smoothing operations, with the second update to the boundary vertices being applied in the "negative" direction and weighted by a constant such that in the limit repeated smoothing does not change the circumference of the boundary.

Local boundary descriptor Following the detection of boundary keypoints, our goal is to describe the local shape of the boundary. After experimenting with a variety of descriptors based on local curvatures, angles and normals, using histogram and order preserving representations, we found that the best results are achieved using a local profile of boundary normals' directions.

To extract a descriptor, we sample the boundary around a keypoint's neighbourhood (at the characteristic scale of the keypoint) at n_s equidistant points and estimate the boundary normals' directions at the sampling loci. This is illustrated in Figure 1(c). For each sampling point, a circular arc is fitted to the closest boundary vertex and its two neighbours, after which the desired normal is approximated by the corresponding normal of the arc, computed analytically. The normals are scaled to unit length and concatenated into the final descriptor with $2n_s$ dimensions. After experimenting with different numbers of samples, from as few as 4 up to 36, we found that our method exhibited little sensitivity to the exact value of this parameter. For the experiments in this paper we used $n_s = 13$.

We apply this descriptor in the same way as Arandjelović and Zisserman did theirs, or indeed a number of authors before them using local texture descriptors. The set of training descriptors is first clustered, the centre of each cluster defining the corresponding descriptor word. An object is then represented by a histogram of its descriptor words. Since we too do not encode any explicit geometric information between individual descriptors we refer to our representation as a bag of normals (BoN).

Evaluation. To evaluate the effectiveness of the proposed method we used the publicly available *Amsterdam Library of Object Images* and performed three experiments:

- We compared the BoB and BoN representations in terms of their robustness to viewpoint change. The representations of all 1000 objects learnt from a single view were matched against the representations extracted from viewpoints at 5–85° yaw difference. Each object image was used as a query in turn.
- We compared the BoB and BoN representations in terms of their robustness to segmentation errors. The representations of all 1000 objects learnt from a single view were matched against the representations extracted from the same view but using distorted segmentation masks. In this experiment we distorted the segmentation mask by morphological erosion using a 3×3 'matrix of ones' structuring element. As before, each object image was used as a query in turn.
- We compared the BoB and BoN representations in terms of their robustness to segmentation errors. This time we distorted the segmentation mask by morphological dilation using a 3×3 'matrix of ones' structuring element. As before, each object image was used as a query in turn.

Overall, the performance of the BoB and BoN representations was found to be similar. Some advantage of the BoN was observed in rank-1 matching accuracy: each 5° change in yaw can be estimated to decrease the BoB performance by approximately 12% and the BoN performance by approximately 10%. In the second and third experiments the superiority of the proposed BoN representation was more significant. For example, the distortion of the segmentation mask by two erosions reduces the rank-1 matching rate of the BoB by 30% and that of the BoN by half that i.e. 15%. The negative effects of dilation of the mask were less significant for both representations but qualitatively similar: repeated twice, dilation reduces the rank-1 matching rate of the BoB by 25% and that of the BoN by only 10%.

[1] R. Arandjelović and A. Zisserman. Smooth object retrieval using a bag of boundaries. *In Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 375–382, November 2011.