

Recognition of Simple 3D Geometrical Objects under Partial Occlusion

Alexandra Barchunova and Gerald Sommer

Bielefeld University, 33615 Bielefeld, Germany
{abarch}@cor-lab.uni-bielefeld.de

Abstract. In this paper we present a novel procedure for contour-based recognition of partially occluded three-dimensional objects. In our approach we use images of real and rendered objects whose contours have been deformed by a restricted change of the viewpoint. The preparatory part consists of contour extraction, preprocessing, local structure analysis and feature extraction. The main part deals with an extended construction and functionality of the classifier ensemble Adaptive Occlusion Classifier (AOC). It relies on a *hierarchical fragmenting* algorithm to perform a local structure analysis which is essential when dealing with occlusions. In the experimental part of this paper we present classification results for five classes of simple geometrical figures: prism, cylinder, half cylinder, a cube, and a bridge. We compare classification results for three classical feature extractors: Fourier descriptors, pseudo Zernike and Zernike moments.

1 Introduction

Contour-based recognition of partially occluded objects involves handling of several challenging issues. Contour acquisition and its quality improvement is the first task. In this work we have made use of some common techniques for noise cancelling like cautious Gaussian smoothing and B-Spline modelling [7]. Partial occlusion of shape poses a big challenge for algorithms with a global approach. An object is made invisible in a local environment. At the same time its shape in this environment is replaced by the shape of the occluding object. Recognition of occluded shapes by a human involves an analysis of the local structure, a search for a characteristic contour fragment allowing a clear assignment to the corresponding object class. Automatisation of the recognition process requires likewise a method that conducts such a local structural analysis of the object contours under partial occlusion. The algorithm that performs the local structure analysis in our work will be referred to as hierarchical fragmenting. For a given input contour this method generates several fragment levels, whereby the structural complexity of the fragments increases from level to level. The B-spline interpolation for noise cancelling and hierarchical fragmenting are both based on the segmentation that uses local maxima of the curvature function for extraction of the points of interest. The importance of such points in contour-based approaches has been investigated in [3].

In this paper we present classification results for two classical affine-invariant feature extractors based on Fourier descriptors and Zernike moments [10,9].

Within our experimental setup we allow a restricted perspective deformation of the contour. The scene is shot only from above while the camera is moving parallel to the surface. It has been shown in our work that it is possible to compensate a certain degree of perspective deformation by training. Experiments with rendered object images have shown that the larger the degree of perspective change of the contour the larger the classification error when using an affine-invariant feature extractor.

Multiple classifier systems have been employed in complex computer vision tasks starting in the eighties. Different aspects of development in this field have been discussed by T.K. Ho in [4]. In our approach each member of the ensemble specializes on its own degree of occlusion, defined by the corresponding level of the hierarchical fragmenting. The final hypothesis is generated by application of the class-related weighted average method. The determination of weights is formulated as a linear optimization problem being solved for an additional set of occluded shapes. The empirical results have shown that our approach delivers a considerable improvement of classification results compared to using a single classifier for recognition of partially occluded objects.

2 Contour Preprocessing and Feature Extraction

Factors like hardware noise, reflection or shadows result in an acquired contour that doesn't comply with the smooth shape of the original objects (see Figure 1(a)). Our goal is to improve the contour locally without losing the information about the global structure. In our approach we describe the global structure by a set of *points of interest* (POI), which we later interpolate with B-splines. Here we use a common way to define a point of interest as a local maximum of the curvature function [5,2]. Let x and y be functions defining a discrete contour in a parametric representation. Then the curvature in the point $p := (x(t), y(t))$, $t \in \underline{T}$ ¹ is defined as follows:

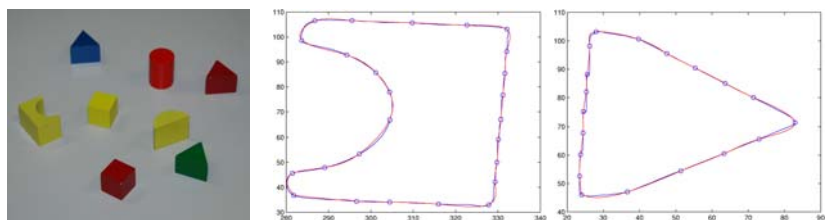
$$k(p) := \frac{x'(t)y''(t) - x''(t)y'(t)}{(x'^2(t) + y'^2(t))^{3/2}}, \quad (1)$$

where T denotes the number of discrete points in the contour. POI $p_0 = (x(t_0), y(t_0))$ is defined for $\epsilon \in \mathbb{N}$, $t_0 \in \underline{T}$ and an environment $U_0 = U(t_0, \epsilon)$ as follows:

$$|k(p_0)| = \max_{p \in P_u} |k(p)|, \quad (2)$$

where $P_u = \{(x(t), y(t)) | t \in U_0\}$. In order to obtain realistic values for $k(p)$, we first apply Gaussian smoothing with $\sigma = 5$ on the contour data. Then we calculate averaged derivatives that we use instead of regular discrete derivatives in evaluation of $k(p)$. Each derivative is built as an average over 5 neighbouring points, which reduces the influence of noise. In the second step we apply a procedure for POI extraction that can be schematically described as follows:

¹ Throughout this work $t \in \underline{T}$ denotes $t \in \{1, \dots, T\}$ for a $T \in \mathbb{N}$.



(a) Test objects in a typical scene. (b) Extracted POI (blue points); interpolated contour (red) and original contour (blue).

Fig. 1. Typical test objects, preprocessed and segmented contours

1. Sort $\{(p, k(p))\}$ in descending order according to the values $k(p)$
2. Select the first available point from the sorted list to be the next POI; prohibit selection of further points belonging to the local environment of the selected point
3. Go to 2 if there are points available in the list, otherwise emit the chosen POI

The value of the local environment parameter in 2 depends on the structural complexity of the objects. In our experiments we have used the value $1/10$ of the contour length. Finally, we conduct B-spline interpolation for the calculated POI. This results in an improvement of the local contour structure, while sustaining the global shape characteristics (see Figure 1(b)).

Contours of three-dimensional objects are rarely planar. Thus we can either try to reconstruct the three dimensional structure of the curve or we work with the two dimensional projection. Here we make use of the second option. In our tests (see Section 4) we have shown that it is possible to use affine-invariant feature extractors on such data and compensate the restricted perspective deformation by learning. For a given sampled contour we calculate a vector of normalised Fourier descriptors and (pseudo) Zernike moments. In our experiments we have used a constant number of points, $L = 64$, to represent any kind of contour data. The dependency between the dimensionality of the feature vector and the classification error will be described in Section 4.

3 Classifier Ensemble and Its Organisation

In the following sections we will describe the nature of Adaptive Occlusion Classifier (AOC) by looking at the following four main components: the data set, feature extractors, basis classifiers and the combination technique [6].

3.1 Data Set

The data pool consists of real camera images and rendered images² of the non-occluded test objects. In our experiments we investigate solely artificially

² In our experiments we have used POV-Ray [1] for rendering.

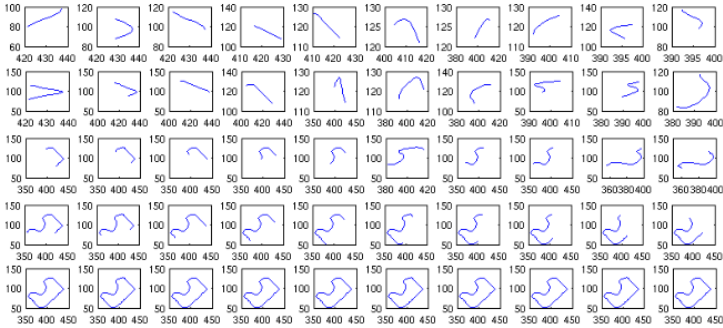


Fig. 2. Five data subsets of fragments generated by the hierarchical fragmenting of a bridge contour; rows correspond to the hierarchical levels

generated straight line boundary occlusion and make use of two methods for occlusion generation. The first method of *contour occlusion* simply deletes a given part of the contour data and connects the gap with a straight line. It is computationally efficient but can produce unrealistic contours when applied to non-convex shapes. The second method calculates the positions of the pixels within the contour that can be deleted to yield a linear occlusion of the object area. The later method of *area occlusion* is computationally more complex but delivers realistic linear occlusions independent of the shape convexity.

On the data level the members of the AOC are assigned to their personal data subsets. These are generated by the hierarchical fragmenting algorithm based on the set of POI (or control points) for the B-spline interpolation. Consider a contour fragment located between **three** neighbouring control points (see the example fragments in the first line of Figure 2). A set of all such fragments builds up the first hierarchical level or the first data subset. Note that the set of POI used in the algorithm contains a subset of structurally descriptive points, e.g. a corner of a prism. Analogously, the subset of the generated fragments contains a subset of local shape-descriptive fragments, e.g. a fragment, containing a corner of a prism. The hierarchical fragmenting algorithm on the i -th step connects two neighbouring segments of the $(i-1)$ -th level to a new one. This generates levels of contour fragments of growing structural complexity, defining different levels of partial occlusion. In Figure 2 you can see some example fragments of five hierarchical levels for a bridge contour segmentation.

3.2 Feature Extractors and Basis Classifiers

In a perfect scenario classifier training results in a model that completely covers the feature space of the partially occluded objects. Because the set of all possible occlusions is vast, we have to choose training data that allows an approximate solution. In our work we use the data subsets generated by the hierarchical fragmenting which the feature extractor transforms into feature vector subsets.

As mentioned above, we use normalized Fourier descriptors, Zernike and pseudo Zernike moments. On the classifier level we use Local Credibility Criterion (LCC) classifiers [8]. LCC classifiers consists of multiple hypersphere shaped models. Their feasibility or credibility is determined by the ratio of correct responses to the number of overall responses: $\gamma = R_c/R_t$. Both the set of models and their number is dynamic.

3.3 Combination Technique and Weight Vector Estimation

Let $F \subset \mathbb{R}^n$ be a feature space and $C := \{1, \dots, K\} \subset \mathbb{N}$ the set of class labels. Let $D := \{D_1, \dots, D_L\}$ be the classifier ensemble, where a classifier D_i for $i \in \underline{L}$ can be described by the following map:

$$D_i : F \rightarrow [0, 1]^K, \quad x \mapsto (d_{i1}(x) \dots d_{iK}(x)). \tag{3}$$

In this work we have used the *class-related weighted average* method in order to combine the responses of the AOC members. For each $j \in \underline{K}$, $i \in \underline{L}$ the class-related weighted average for a sample x is defined by:

$$\mu_j(x) := \sum_{i=1}^L w_{ij} d_{ij}(x), \tag{4}$$

where the w_{ij} 's denote the class specific weights.

AOC was designed to allow contour-based classification of objects with different degrees of occlusion. Each member of the ensemble specializes on its own degree of occlusion during training as well as during testing. The main task of the weight vector is to integrate the individual classifier class responses to a final hypothesis according to their classification performance. For this purpose we use an additional set of contours with a random uniform area occlusion up to 80 percent, denoted by $\gamma_{max}^w = 0.8$. Further let $X := \{(x_n, c_n) | n \in \underline{N}\}$ be a labeled sample set. The response matrix $R_l \in \mathbb{R}^{N \times K}$ of the l th classifier to the sample set X is given by:

$$R_l := \begin{pmatrix} d_{l1}(x_1) \dots d_{lK}(x_1) \\ \dots \dots \dots \\ d_{l1}(x_N) \dots d_{lK}(x_N) \end{pmatrix} \tag{5}$$

We define an auxiliary function that allows a correct building of a scalar product and adding up of columns of the matrix R_l , $l \in \underline{L}$ for correct solving of the minimization problem (see Eq. 8). For a $k \in \underline{K}$ we define:

$$f_k : \mathbb{R}^{N \times K} \rightarrow \mathbb{R}^{N \cdot K}, \quad R_l \mapsto (v_1, \dots, v_p)^T, \tag{6}$$

where $p = N \cdot K$ and for $i \in \underline{p}$, $n \in \underline{N}$:

$$v_i := \begin{cases} d_{lk}(x_n), & \text{if } i = N \cdot (k - 1) + n \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Let $r = L \cdot K$ be the dimension of the weight vector. We combine the responses of the classifier ensemble in the following matrix $\hat{R} \in \mathbb{R}^{p \times r}$:

$$\hat{R} := (f_1(R_1), f_2(R_1), \dots, f_j(R_i), \dots, f_K(R_L)).$$

For a given sample set X the weight vector $w \in \mathbb{R}^r$

$$w := (w_{11}, w_{12}, \dots, w_{1k}, \dots, w_{LK})^T,$$

can be calculated by minimizing the distance between the optimal response matrix R_{opt} and the weighted (see Eq. 4) response matrix \hat{R} :

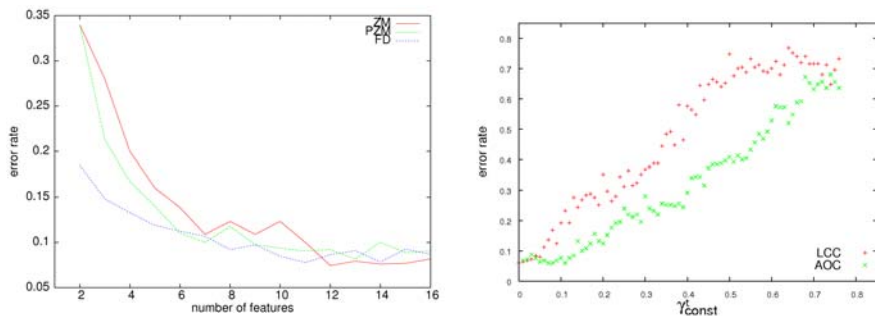
$$\min_{w \in \mathbb{R}^r} \|\hat{R}w - R_{opt}\|. \quad (8)$$

4 Experimental Results

The data pool consists of 1000 camera images and 1600 rendered non-occluded images. For every type of geometrical figure (bridge, cylinder, half cylinder, cube and prism) the data pool provides the same number of images. The objects have been recorded with a perspective deformation through a change of the viewpoint or the POV-Ray configuration.

In the first experiment we analyse the dependency between the dimensionality of the feature vector and the classification error rate. For the training of the ensemble members we have randomly selected 370 contours of non-occluded objects out of the data pool. For the calculation of the weight vector we have used 250 contours with generated random area occlusion up to 30 percent, $\gamma_{max}^w = 0.3$. In all algorithms for random occlusion generation we use uniform distribution on the interval $(0, \gamma_{max}]$. On average random uniform area occlusion up to 30 percent is approximately equivalent to random uniform contour occlusion up to 50 percent in our experimental setup. We have tested 250 contours with $\gamma_{max}^t = 0.3$. In the Figure 3(a) you can see the dimension of the feature vector on the x -axis and the average classification error rate on the y -axis. Zernike moments (ZM), pseudo Zernike moments (PZM) and Fourier descriptors (FD) yield approximately the same results. The lowest average error is about 7 percent which can be explained by the ambiguity of the contours with partial occlusion.

In the next experiment (see Figure 3(b)) we have compared the performance of a single LCC classifier vs. AOC ensemble. For the calculation of the weight vector we have used contours with a random uniform area occlusion up to 80 percent, $\gamma_{max}^w = 0.8$. As a representation of the contour data we have chosen to use a 14-dimensional normalized vector of FD. On the x -axis we have plotted the constant area occlusion parameter γ_{const}^t . The value of this parameter indicates for all objects in the sample set a constant but randomly placed area occlusion specified by the value γ_{const}^t . On the y -axis you can see the recognition error. For all levels of occlusion AOC classifies considerably better than a single LCC. Data with a constant area occlusion of more than 80 percent cannot be classified better than by random guessing. This can be explained by a high degree of ambiguity of strongly occluded object shapes.



(a) Classification results for different feature (b) Classification results: AOC vs. single extractors LCC

Fig. 3. Classification results for AOC ensemble and a single LCC classifier

Table 1. Comparison of error rates for different sets of synthetic data

	R-R-S	R-S-R	R-S-S	S-R-R	S-R-S	S-S-R	S-S-S
Error rate (Set 1)	0.28	0.20	0.22	0.21	0.18	0.24	0.15
Error rate (Set 2)	0.27	0.19	0.24	0.22	0.21	0.23	0.22

The third experiment was aimed at exploring the capacity of rendered data alone as well as its potential in combination with real camera images within our applications. We have trained AOC with 350 samples. For weight vector estimation we have used 250 contours with $\gamma_{max}^w = 0.5$. Table 1 shows the summary of the test results for different combinations of real (R) and synthetic (S) data in training, weight vector estimation and testing (R/S-R/S-R/S). Our test data consists of 100 samples with $\gamma_{max}^t = 0.5$. The in-plane translation of the objects with regard to the camera position within Set 2 is two times as large as within Set 1. The classification using real camera images (R-R-R) yielded an average error of about 17 percent. Consider the sixth column of the table corresponding to the S-S-R configuration. The test results are only about 5 percent worse compared to the case where only real data (R-R-R) has been used in training and weight estimation.

5 Conclusions

In this paper we have shown that the usage of the AOC ensemble for recognition of partially occluded shapes of three-dimensional objects considerably improves the results in comparison to a single LCC classifier. Classification for γ_{max}^t yielded an average error rate of about 17 percent. This can be explained by the ambiguity of the partially occluded data.

Our procedure of hierarchical fragmenting delivers levels or subsets of contour segments with growing structural complexity. By using this method in our application we have demonstrated that it is well suited for carrying out local structure analysis. In our tests we have compared the following classical affine-invariant feature extractors: Fourier descriptors, pseudo Zernike moments and Zernike moments. All three have yielded comparable results. It can be concluded that the usage of 12 to 14 dimensional feature vectors is sufficient for our application.

Tests with rendered images have revealed an automation potential of the ensemble training within our experimental setup. Note that in our tests the usage of rendered images in training resulted in an error rate increase of approximately 5 percent. More tests could be conducted in this area.

References

1. Persistence of vision raytracer, <http://www.povray.org/>
2. Ghosh, A., Petkov, N.: Robustness of shape descriptors to incomplete contour representations. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(11), 1793–1804 (2005)
3. Ghosh, A., Petkov, N.: Effect of high curvature point deletion on the performance of two contour based shape recognition algorithms. *International Journal of Pattern Recognition and Artificial Intelligence* 20(6), 913–924 (2006)
4. Ho, T.K.: Hybrid methods in pattern recognition. *Series in Machine Perception and Artificial Intelligence*, ch. 5, vol. 47, pp. 367–382. Springer, Heidelberg (2002)
5. Kindratenko, V.: Development and Application of Image Analysis Techniques for Identification and Classification of Microscopic Particles. PhD thesis, University of Antwerpen (1997)
6. Kuncheva, L.I.: *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, Hoboken (2004)
7. Lim, K.B., Du, T., Zheng, H.: 2d partially occluded object recognition using curve moments. In: *Proceedings of the Seventh IASTED International Conference Computer Graphics And Imaging*, Kauai, Hawaii, USA (2004)
8. Prehn, H., Sommer, G.: Incremental classifier based on a local credibility criterion. In: *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*, AIA 2007, Innsbruck, Austria, pp. 372–377. ACTA Press (2007)
9. Teh, C.-H., Chin, R.T.: On image analysis by the methods of moments. *IEEE Trans. Pattern Anal. Mach. Intell.* 10(4), 496–513 (1988)
10. Zhang, D., Lu, G.: Review of shape representation and description techniques. *Pattern Recognition* 37, 1–19 (2004)