Pattern Recognition 43 (2010) 3833–3845

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# Viewpoint independent object recognition in cluttered scenes exploiting ray-triangle intersection and SIFT algorithms

Georgios Kordelas, Petros Daras *

*Informatics & Telematics Institute, 1st km Thermi Panorama Road, 57001 Thermi, Thessaloniki, Greece*

ARTICLE INFO

*Article history:*
Received 9 April 2009
Received in revised form
29 March 2010
Accepted 22 May 2010

*Keywords:*
3D object recognition
Distance maps
Ray-triangle intersection
Clutter
Occlusion

ABSTRACT

Viewpoint independent recognition of free-form objects and estimation of their exact position are a complex procedure with applications in robotics, artificial intelligence, computer vision and many other scientific fields. In this paper a novel approach is presented that addresses recognition of objects lying in highly cluttered and occluded scenes. The proposed procedure relies on distance maps, which are extracted and stored off-line for each of the 3D objects that might be contained in the scene. During the on-line recognition procedure distance maps are extracted from the scene. Greyscale images, derived from scene's distance maps, are matched with those of the object under recognition by applying similarity measures to the descriptors that are extracted from the images. The similarity is then estimated from image patches, which are defined using the SIFT descriptor in an appropriate way. After finding the best similarities the position of the object in the scene is estimated. This process is repeated until all objects are successfully recognized. Multiple experiments, which were performed on both 2.5D synthetic and real scenes, proved that the proposed method is robust and highly efficient to a satisfactory degree of occlusion and clutter.

## 1. Introduction

In the recent years, significant progress has been made towards the recognition of free-form objects. The immediate objective of object recognition systems is to correctly identify an object in a scene of objects, in the presence of clutter and occlusion and to estimate its position and orientation. Those systems can be exploited in robotic applications where robots are required to navigate in crowded environments and use their equipment (i.e. range scanners, arms) to recognize and manipulate objects. Robots with advanced capabilities could be used to service elderly/impaired people or for surveillance in sensitive environments. Object recognition can be performed using 2D images, which is an affordable solution due to the wide availability of low cost cameras. Approaches exploiting cameras are fast and low cost, yet they are also very sensitive to illuminations, shadows and occlusions and do not provide accurate estimation of object's pose. Thus, the focus of the relevant scientific communities is on the development of 3D object recognition algorithms that overcome the aforementioned limitations.

The idea of recognizing objects in range data has already been investigated in several scientific studies. Campbell's and Flynn's survey [1] provides an extended overview of 3D object recognition techniques. However, a short complement to this survey and a report to recent methods is presented here for the sake of completeness. COSMOS [2], one of the earliest algorithms, is based on the computation of principal curvatures of the surface. This method is limited to objects with smooth surfaces and is applicable to just unoccluded views of an object. Chua and Jarvis [3] propose a point signature (PS) for 3D object recognition where a sphere centered at a given point is intersected with the surface and creates a 3D space curve on which a plane is fitted. Point signature was proved to be sensitive to noise and surface sampling [14]. Hetzel et al. [13] combine pixel depth, surface normals and curvature in a multidimensional histogram in order to directly model the probability distribution of different feature combinations. Their experiments proved the efficiency of this method; however, the database used includes only non-cluttered, self-occluded range images of 30 free-form objects. Johnson and Hebert propose the spin image method [7], which is vulnerable to sampling and resolution (level-of-detail) of the models and has low discriminative power. Additionally, this method is applied to every vertex of the object or the scene, therefore the number of the descriptors increases as the number of vertices does. When the number of descriptors is compressed, using principal component analysis (PCA), the average recognition rate decreases significantly (almost 10%). Nevertheless, spin images have been used in many applications such as parts-based 3D object classification [4] and for recognizing members of classes of

* Corresponding author.
*E-mail addresses:* kordelas@iti.gr (G. Kordelas), daras@iti.gr (P. Daras).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.05.030

3D shapes [5]. In [9], an enhancement of the spin images algorithm is presented by using vertex interpolation. Although these changes resolved sensitiveness of spin images to variations in resolution, descriptor's discriminative power was not improved significantly. Spherical harmonics [11] and locality-sensitive hashing [12] are exploited in [10] to perform efficient retrieval of shapes; the work is tested on 3D shape information obtained from laser range scanners. However, the approximate location of the shape to be retrieved is already known, thus algorithm's task is limited to identify database's correct shape. Mian et al. [14], recently proposed a tensor-based surface representation defined on pairs of oriented points. Their descriptors are 3D tensors that measure the variation of surface position. Correspondence between 3D Tensors is established using a voting process to find pairs of tensors with high overlap ratio.

A more recent approach is presented in [17], where the similarities between input 3D images are computed by matching their descriptors with a pyramid kernel function. The similarity matrix of the images is used to train support vector machines-based (SVM) classification [19], and new images can be recognized by comparison with the training set. The experiments were performed on the same database as in [13], and thus robustness with respect to clutter was not examined. In [18], an initial implementation of the distance map descriptor was presented; however, this approach was viewpoint dependent. The algorithm presented in [16] (an extension of work in [15]) calculates the local surface properties of patches, which are defined on the extracted feature points. By comparing local surface patches for a model and a test image, and casting votes for the models containing similar surface descriptors, the potential corresponding local surface patches and candidate models are hypothesized. The evaluation experiments were simple, since at most two objects existed in the scene. In [20], the generalized Hough transform is extended to detect instances of an object in laser range data, independently to the scale and orientation of the object. However, this method is restricted to simple objects that can be represented with few parameters, such as planes, spheres and cylinders.

The plethora of the existing algorithms [6,17,5,4] use spin images [7]. These methods either modify the spin image or integrate it with other components, so as to improve its performance. Moreover, the majority of the methods was tested on self-occluded scenes without presence of clutter [2,4,3,13]. Thus, there is a need for the development of novel methods that address the object recognition problem in a more efficient way.

In this paper a novel approach for recognition of 3D objects in range scenes, is presented. The primary step of the proposed algorithm is to place the 3D object in a proper position and then to form a coordinate basis used to extract distance maps for this object. During the 3D object's recognition procedure, distance maps are extracted for the scene according to a coordinate system, which allows keeping their total number very low. Matching between scene's and object's distance maps is established using the SIFT algorithm on greyscale images that are generated from the distance maps. The whole procedure is novel and provides a different insight in the "treatment" of the object recognition problem. A major difference to previous methods (i.e. [7]), where descriptors are extracted on the vertices of the reconstructed point cloud, lies on the extraction of scene's descriptors, which is based on a coordinate system that is formed according to scanning parameters.

The advantages of the proposed method are the following: the approach used to extract distance maps, especially for the scene, allows keeping their number low since it is independent of 3D object's number of vertices. Added to this, the employment of a simple 1D hash table allows significant acceleration of the execution time. Another advantage is its robustness with respect to objects' level-of-detail since, unlike spin images, it is not required the library objects to have similar resolution.

The results on synthetic scenes proved that the proposed algorithm is robust to a high degree of clutter and occlusion and experimental comparison with the spin image approach on real scenes verified the superiority of the proposed algorithm.

The rest of this paper is organized as follows. In Section 2, the off-line extraction of 3D object's distance maps along with the automatic extraction of scene's distance maps are presented. Section 3 introduces a similarity measure based on the SIFT algorithm. In Section 4, the performed experiments both on synthetic and real data are given, while conclusions are drawn in Section 5.

## 2. Computation of distance maps

### 2.1. Model's initial distance maps

The goal of this procedure is twofold: firstly to place the 3D model in a proper initial position and secondly to create a coordinate basis around the object, which is used to define object's initial distance maps in such a way that largest portion of object's surface will be described with the minimum number of descriptors.

### 2.1.1. Initial position of 3D model

Each model's vertices are stored in the matrix $\boldsymbol{V}_{model}$ (where $\boldsymbol{V}_{model}$ is a $N \times 3$ matrix of 3D coordinates). The PCA [8] on $\boldsymbol{V}_{model}$ is computed and the three orthogonal principal components are derived. The object is rotated around its center of mass, so that the first principal component becomes parallel to $z$-axis and the second principal component becomes parallel to $y$-axis. After rotation, the object is denoted as $\boldsymbol{V}_{PCA}$. The object is then translated by $\boldsymbol{V}_{final} = \boldsymbol{V}_{PCA} - [x_m, y_m, z_a]$, where $\boldsymbol{C}_m = [x_m, y_m, z_m]^T$ is $\boldsymbol{V}_{PCA}$'s center of mass and $\boldsymbol{P}_a = [x_a, y_a, z_a]^T$ is $\boldsymbol{V}_{PCA}$'s point with minimum $z$-coordinate. This procedure intends to place the object in such a position that $z$-axis passes centrally through object's volume since the coordinate basis used to extract object's initial distance maps is constructed around $z$-axis. The points of intersection between $\boldsymbol{V}_{final}$ and $z$-axis with minimum $z$-coordinate and maximum z-coordinate are $\boldsymbol{P}_{mim} = [x_{min}, y_{min}, z_{min}]^T$ and $\boldsymbol{P}_{max} = [x_{max}, y_{max}, z_{max}]^T$, respectively (Fig. 1(a.2) and (b.2)). Fig. 1 depicts two objects after estimation of their initial position.

### 2.1.2. Extraction of 3D object's initial distance map

*2.1.2.1. Circular sector formation.* Before advancing to the extraction of initial distance maps, a circular sector $S$ of $N$ points, indexed by variable $f$ ($f$=0,1,…,$N$), with radius $R$ (a global parameter used throughout this paper) and center $O = [0,0,0]^T$ is created on $xy$ plane. Circular sector's points are sampled uniformly on the circular disc by creating a centroidal Voronoi tessellation (CVT) [21] of points within the sector region. Since points are sampled uniformly its rather impossible that a point coincides with $O$; however, the circular sector's point that has the minimum Euclidean distance from $O$ is denoted as point $K$ and it is assumed to coincide with $O$. The distribution of points over a specific circular area, using a polar coordinate system and CVT is depicted on Fig. 2(a) and (b), respectively. This figure proves the efficiency of CVT to generate uniform points. $S$ is adapted around points of