Application of the Reeb Graph Technique to Vehicle Occupant’s Head Detection in Low-resolution Range Images

Pandu RangaRao Devarakota, Marta Castillo-Franco, Romuald Ginhoux, Bruno Mirbach
IEE S.A.,
ZAE Weiergewan, 11, rue Edmond Reuter,
L-5326, Contern, Luxembourg
{pdv,mca,rgi,bmi}@iee.lu

Björn Ottersten
Signal Processing group, School of Electrical Engineering,
Royal Institute of Technology (KTH)
SE-100 44, Stockholm, Sweden
bjorn.ottersten@ee.kth.se

Abstract

In [3], a low-resolution range sensor was investigated for an occupant classification system that distinguish person from child seats or an empty seat. The optimal deployment of vehicle airbags for maximum protection moreover requires information about the occupant’s size and position. The detection of occupant’s position involves the detection and localization of occupant’s head. This is a challenging problem as the approaches based on local shape analysis (in 2D or 3D) alone are not robust enough as other parts of the person’s body like shoulders, knee may have similar shapes as the head. This paper discusses and investigate the potential of a Reeb graph approach to describe the topology of vehicle occupants in terms of a skeleton. The essence of the proposed approach is that an occupant sitting in a vehicle has a typical topology which leads to different branches of a Reeb Graph and the possible location of the occupant’s head are thus the end points of the Reeb graph. The proposed method is applied on real 3D range images and is compared to Ground truth information. Results show the feasibility of using topological information to identify the position of occupant’s head.

1. Introduction

During the last decades, many research projects have investigated solutions in the field of intelligent transportation systems (ITS) to provide cheap and reliable systems to many safety applications, including occupant classification for "smart" airbag deployment [5, 7, 12]. The feasibility of an occupant passenger classification system using low-resolution range sensor was investigated in [3]. The aim of this system is to detect the occupancy of passenger seat and classify it into one of the following four classes:

1. Empty seat
2. Rearward facing infant seat (RFIS)
3. Forward facing child seat (FFCS)
4. Adult (P) (see Figure 1).

A low-resolution range sensor which is based on time–of–flight principle was used. This range sensor is advantageous since it provides directly a dense range image, independent of the illumination conditions and object textures. In [7, 4], an occupant classification problem is addressed however based on low-resolution infrared images which were computed from a stereo camera system.

The system in [3] can also be used to detect the size, position, and posture of the occupant. Especially the head position of the occupant gives a valuable input to the the airbag control unit to deploy the airbag at an optimal level for the occupant position, that is providing optimal protection while reducing the injury risk due to an airbag deployment. In case of a so-called out-of-position where the head is too close to the airbag module, the airbag should not be triggered (see Figure 2) or must be de-powered.
A robust head detection in such situation is a challenging problem due to the constraints of the application. Typically, head detection is based on analyzing the local shape of head–like objects in the data. The typical model of the head with an elliptical shape is often considered for this purpose [6, 12]. The classification is then based on the shape features extracted from each head-like shapes detected in the data. However, such a head detection is not very robust due to the following reasons: On one side, the shape of a human head varies a lot due to hair style, hat, etc.,. On the other side, there are many regions in a scene such as shoulder, elbow, of the occupant that have locally the shape of the head model. In addition, there might be other objects, occupants can carry, for example a balloon, that fit even better to the head model than the occupant head itself.

The only chance to overcome these situations where possibly false detection may occur is to take not only local shape features into account but the topology of the whole scene. The basic idea is that an occupant of a vehicle has a typical topology and can be represented in terms of a skeleton or graph. The important parts of the scene can be represented using the branches of the skeleton. For instance, the skeleton of the scene where the passenger seat is occupied, represents two branches, one corresponds to the backrest and other corresponds to the upper torso of the person. In this way, the possible head candidate position can be identified as the end points of the skeleton.

This paper presents a general topological analysis framework for low resolution range images that offers a systematic way to detect the vehicle occupant’s head. There are several representations available to code the topology and our approach is based on Reeb Graph (RG) technique which was originally proposed to extract a skeleton of closed-surface 3D objects [10, 1]. The object of the paper is not to show the Reeb graph method is superior to other alternative skeletonisation methods but to show how topology information is useful to localize the position of head of the occupant. The topology representations other than the Reeb graph can also be used under the similar conditions. The traditional Reeb graph approach requires a dense 3D representation of the closed object surface, as, e.g. a triangulated mesh representation. However, this is not the case for the 3D representation extracted from a range image. There, one obtains a cloud of 3D points representing the part of the object surface that is visible from the camera point of view. This kind of data has not the same connectivity as dense volumetric data and rather be seen as 2D1/2 data. A second property of this data is that the density of data points is not constant but varies with the distance of the object surface points to the camera, as is the case for a wide–angle camera surveilling passenger seat inside the car. Moreover, data drawn from real images are subject to noise in the range measurement.

We propose here an adaptation of the Reeb graph method to this kind of 2D1/2 data, which is motivated by the work of Xiao et al. [14] who proposed a discrete Reeb graph approach to 3D point-cloud data obtained by a full scan of the object. They explored the connectivity based on a distances between the 3D points. For our data we propose to use rather neighborhood connectivity of voxels for building a Reeb graph. The benefit of using a voxel representation of the 2D1/2 data is that by choosing an appropriate voxel activation level for each voxel, depending on its distance to the camera, one can cope with the large variation of the density of data and eliminate outliers which can occur due to noise in the data. Moreover, the voxel connectivity is computationally more efficient than calculating pairwise distances, which is important for the development of an automotive application and is also subject to cost constraints.

A novel 3D low-resolution range camera which is based on a time–of–flight principle is used [8]. The advantage of this camera is that it provides directly a dense range image, independent of the ambient illumination conditions and object textures. The scene is broadly illuminated by a modulated infrared light beam. Then, the modulated beam which is reflected by an object, is detected by the receiver. Due to the time of flight of the light to and from the target, the detected beam has a phase shift compared to the phase of the modulation signal in the illumination. This phase difference \( \Delta \phi \) can be calculated by sampling at four temporal
Figure 3. A example range image and its preprocessed image are shown in fig. 3(a) and fig. 3(b) in a false color representation, respectively. Blue pixels (dark gray) correspond to points closer to the camera and red pixels (light gray) correspond to a large distance. In fig. 3(c), intensity image of the same scene is shown which is acquired by a high-resolution 2D camera to provide a reference.

3. Topology Coding

The topology of an object represented in a skeletal structure proved to be very useful in various applications of computer vision, such as, for instance, 3D indexing [13], classification [2], or 3D object reconstruction [10]. The Reeb graph is one method to represent the topology as a skeleton that shows how different parts of an object are connected to each other. [9].

Reeb graphs are based on Morse theory which says that the change of the topology of a manifold can be detected by identifying the critical points of a smooth function defined on the manifold. A point is a critical point if and only if its derivative with respect to a local coordinate system on the manifold vanish. Reeb graph shows graphically the configuration of critical points and their topological relationships. Using the knowledge of the critical points, the way a manifold is embedded in the 3D space can be coded in terms of a Reeb graph.

One of the main concerns in the Reeb graph extraction is its invariance. The extracted Reeb graph should be invariant under translation, rotation and scaling based on the application. For example, in object recognition applications, one prefer to have a Reeb graph invariant under different realistic possible variations that may occur on the object. By defining a proper Morse function this can be achieved. There are several functions studied in the literature. Each function has different properties and suitability to the type of application. The height function introduced in [10] is suitable for models spread along the vertical axis such as human being representation. The function returns the value of the coordinate z if it is defined on a 2-D manifold. The height function is easy to compute but it is variant to object rotations. In [1], a Morse distance function was proposed and it was shown therein that it is invariant to translation, scale and rotation. Due to its simplicity and invariance properties, we consider the distance function in our method as a Morse function.

4. Method for Head Detection from Reeb Graph

4.1. Reeb Graph extraction on Range data

In the past, the Reeb graph extraction has been studied [10, 2, 1] for 2D and 3D images. A study of the Reeb graph extraction from range data was reported in [13] and is used for the task of 3D indexing. These studies are based on the assumption that 3D objects which are modelled as
explicit surfaces such as triangulated meshes or parametric surfaces, are readily available at hand, which are otherwise time consuming to compute. In addition the possible definitions of critical points generally suffer from stability since small perturbations of the vertex coordinates (in case of triangulated mesh representation) may result in rather different configurations. In [2], an extended Reeb graph (ERG) approach to a discrete surface is proposed, where the data is described with a polygonal contours. These representations, however, are not ideal for potentially noisy and incomplete data such as 3D points produced by a range sensor. In addition, the direct application of Reeb graph on range images is not possible as the data does not comply with the assumption of a smooth closed manifold.

The construction of Reeb graph for a 3D dense range data obtained by a full scan around the object was proposed in [14] which was referred therein as Discrete Reeb graph (DRG). The connectivity between any two points is defined based on a threshold on the value of absolute difference between them.

Our method is motivated by the approach in [14], we have, however, chosen the concept of voxel neighborhood to establish connectivity in the graph. The voxel representation of the 2D1/2 data is particularly advantageous as variations of density of the data can be taken into account by assigning to each voxel a activation level depending on the distance to the camera (see below). Moreover, the voxel representation allows to remove outliers that occur due to noise in the range measurement.

4.2. Voxel Neighborhood

For all experiments the range image which is represented in Cartesian coordinates is used. That is, the range image is transformed to Cartesian coordinates in the vehicle coordinate system where a triple (x,y,z) is assigned to each pixel. For more details about this transformation and how it is done, please refer to [3]. Figure 4(a) shows an example of a range image and a scatter plot of the corresponding coordinate points is shown in Figure 4(b).

First step of our method consists of transformation of these 3D data points into a volume representation. For this purpose, 3D grids of equally-sized cube volume elements (voxels) are used. That is, the 3D coordinate space is sampled using a regularly spaced array of voxel elements. The size of the voxel is thereby a parameter that is adapted to the data and the application. That is, the voxel size is chosen larger than the minimum resolution of the camera, but small enough to resolve details of the topology we are interested, i.e. the typical size of a persons head. We have chosen an equal voxel size of 10cm which corresponds to the typical head radius of a person. Figure 4(c) shows the voxel representation of the scene shown in Figure 4(b).

Each coordinate points in the image is assigned uniquely to one voxel, but each voxel can contain several pixels. A voxel is said to be activated only if it contains sufficient number of pixels, that is the number of pixels larger than the activation level given by $C/d^2$. Here $d$ is the mean value of all pixels the voxel element contain and constant $C$ is chosen such that the activation level varies from 1 pixel at a largest distance to 10 pixels at a mean distance of the camera to occupant. That is the activation level is chosen to cope with the outliers which occurs due to the noisy nature of the data over the distance.

The extraction of Reeb graph is now based on the notion of connectivity described in the following definitions:

**Definition 1** (Concept of voxel neighborhood) Two voxel elements $v$ and $w$ are said to be 6-adjacent if they share a face, 18-adjacent if they share a face or an edge, and 26 adjacent if they share a face, or an edge, or a vertex. [11]

**Definition 2** (Connectivity of points) Let $p_1(x_1, y_1, z_1)$ and $p_2(x_2, y_2, z_2)$ are two data points in the range image. Two points are defined as connected if their corresponding voxel elements are said to be 6-adjacent.

**Definition 3** (Connectivity of point sets) Two point sets $P = \{p_i\}, i = 1, \cdots , m$ and $Q = \{q_j\}, j = 1, \cdots , n$ are defined as connected if $\exists p_i \in P$ and $q_j \in Q$ such that their corresponding voxel elements are said to be 6-adjacent.

4.3. The Algorithm

The aim of the algorithm is first to describe the topology of a scene corresponds to a vehicle occupant using the Reeb graph extraction. The goal is that the resulting Reeb graph will have three branches, one corresponds to the backrest of the seat, other two correspond to the upper and lower torso of the occupant (in case of a situation where the passenger seated in a normal position). For this purpose one need to identify and extract first the critical points of the Reeb graph where the change in topology occurs. Branches of the graph should then be allocated to represent similar topology regions correspond to different parts of the scene. This graph information can be used to identify the position of the occupant’s head. The end points of the Reeb graph are useful for this purpose.

The algorithm is as follows: As a Morse function, we choose the radial distance function which was proposed in [1]. This function is advantageous over other functions since it is rotationally invariant. It was also shown in [1] that an invariant Reeb graph can be obtained under scale and translation.

Our algorithm starts with the identification of an origin $o$ of the scene from where the distance function grows gradually in $K$ steps of a fixed size, till all data are covered. Hence the step size play as a role of resolution of the Reeb graph. In [1], the centroid of the object is considered as a
starting point. Since our goal is to obtain a graph invariant under rotation of the upper torso of the person, in our application, we have chosen the point close to the hip of the passenger (see Figure 4(b)) as origin. This point is localized as follows: We first determine the range $[X_{\text{min}}, X_{\text{max}}]$ of values in which voxels are activated and choose $X$-coordinate of the centroid $(X_c = (X_{\text{min}} + X_{\text{max}})/2)$. Among all voxels with this $X$-value, the one with the minimum $Z$-value is chosen as a starting point.

The level sets of the distance function are concentric spheres. In each of the $K$ steps, we find the corresponding data points that lie in a shell of radius $[r_k, r_{k+1}]$ ($k = 1, \ldots, K - 1$). The connectivity between points are then defined based on the definition of connectivity of points described in the previous section and then accordingly, the corresponding connected point sets are identified. Next, each connected point set is assigned with a cluster node at its centroid. At the end the connectivity between all point sets are detected based on the definition of connectivity of point sets and all connected point sets are joined by an edge segment. Along this procedure, a cluster node will be identified as a saddle critical point if it has at least two branches attached to it. Similarly a cluster node will be identified as one of the extremal critical point (minimum or maximum) if it is the end point of the branch. Finally, the end points of the graph can be identified as the region where the occupant’s head lies.

5. Experimental Results and Discussion

We conducted experiments to evaluate the performance of the method, in particular, the ability of Reeb graph of a scene to detect the position of occupant’s head. If the passenger sits in a normal position, one would expect a Reeb graph with three branches though, in practice different situations of the scene are possible which introduce variations in the Reeb graph. Some examples of different situations of the scene are as follows: passenger head lies close to the head-rest, passenger bend forward, or passenger carrying any head-like objects. A database of several sequences were recorded with two different passenger subjects and they were asked to perform several tasks to take into account the mentioned situations. We first consider few snapshots of these sequences to study the possibility of different variations in the Reeb graph on the mentioned situations. Next, the head detection algorithm explained in Section 4.3 is evaluated on all sequences and results are compared with ground-truth information to see whether there is always a branch leading to the head of the passenger.

5.1. Ground Truth Information

To show that the Reeb graph can be used to localize the occupant’s head position, results of the head detection algorithm are compared against ground-truth data which is acquired by 3D string sensors. The 3D string sensor consists of three variable-length cables that are attached to the potentiometers placed on distinct locations on the dashboard and in the ceiling of a car. The free extremities of the cables are linked on to each other with a belt. During the recordings, occupants were asked to wear this belt on their forehead. The 3D string sensor gives directly a measure of coordinates of the crossing point, i.e., the $X$, $Y$, and $Z$ values of the crossing point at the forehead where three sensor cables were linked. Since this value gives a measure of coordinates of forehead point, the coordinates of the center of occupant’s head can be obtained by adding a fixed offset to it. This measure is then taken as a ground-truth data to compare with the coordinates of possible head location given by the Reeb graph approach.

5.2. Performance

Figure 5 shows three examples of range images of different situations of the scene. Figure 5(a) depicts the situation where the passenger head lies close to the head-rest, Figure 5(b) shows the situation of the passenger bent for-
ward, and Figure 5(c) shows the situation where passenger carries a balloon. The scatter plots of corresponding coordinate points are shown in Figure 6. We selected the step size in which the distance function grows gradually, equal to the typical size of the occupant’s head. Remember that the notion of connectivity of point sets is based on the voxel neighborhood where the size of voxel is adapted to the typical size of the occupant’s head. Therefore it would be natural to choose the resolution level equal to the typical size of the occupant’s head. The Reeb graph of the situation as in Figure 5(a) is shown in Figure 7(a) and the corresponding end points are marked in square symbol. In this particular situation Reeb graph produce only two end points since the passenger head lies close to the head-rest, they are considered as topologically connected and thus represented with only one branch. The other branch corresponds to the legs of the passenger. Figure 7(b) shows the Reeb graphs of the situation where passenger bend forward as in Figure 5(b). As expected Reeb graph produce three end points, one corresponds to the head-rest, one corresponds to the head of the passenger, and third one corresponds to the legs of the passenger. In the situation where the passenger bent forward and looking for something under the seat, Reeb graph produce only one branch correspond to the legs and head of the passenger.

It is now interesting to see the topology of a particular scene where the passenger is holding a ball as in Figure 5(c). Figure 7(c) shows the Reeb graphs of such a situation. Reeb graph in this case has four end points representing the back-rest, passenger’s head, ball, and lower torso of the occupant, respectively. In the situation where the balloon is on lap, Reeb graph produce only one branch correspond to the balloon and legs of the passenger since they are considered as topologically connected. Clearly in all three situations, Reeb graph always produce a branch leading to the passenger’s head.

In order to see the effect of number of steps $K$ on the Reeb graph, the Reeb graph extraction is carried out at three different resolution levels. The Reeb graphs of the situations as in Figure 6 are shown in Figures 8 to 10 which are evaluated at different resolution levels and the corresponding end points are marked in square symbol. In all three situations, similar Reeb graphs with the same topology as before are obtained. Therefore the resolution parameter $K$ does not have any influence on the number of Reeb graph branches. However, the position of the end-points may change according to the selected resolution level and hence it may play a role on the precision of the head detection results. Since the goal is to localize the head, for all further experiments the voxel size is chosen as the resolution level of the Reeb graph.

Now the head detection algorithm is evaluated on all sequences and results are then compared with the ground-truth measurements. A total of 10 sequences consists of 3225 frames of images were acquired for the evaluation. The sequences where passenger holding head-like objects were not included in the evaluation as in these cases, ground-truth information could not be obtained; objects would obstruct the string sensors. In order to verify if there is always a head candidate close to the ground-truth position, the notion of successful detection is considered as follows: If the Reeb graph of a particular scene produce $n$ endpoints, $E_1, E_2, \ldots, E_n$, respectively, and $E_k (1 < k < n)$ is the closest end-point to the ground-truth measurement, then the point $E_k$ is said to be successfully detected only if it is within 15cm from the ground truth measurement. Since the ground-truth gives a measure of forehead point, for the comparison to be fair the following procedure is considered: Following the discussion at end of the section 4.3, the end-points of the Reeb graph are computed as the most
forward point of the connected point sets instead of the centroid of the connected point sets. This most forward point is achieved by computing the most forward point of the ellipse fitted to the connected point sets.

![Figure 8. Reeb Graphs for an in-position occupant as in fig. 5(a) at different resolution levels (a) $K = 2$ (b) $K = 4$ (c) $K = 8$. The shown square marking points are end points of the graph.](image8)

![Figure 9. Reeb Graphs for an out-of-position occupant as in fig. 5(b) at different resolution levels (a) (a) $K = 2$ (b) $K = 4$ (c) $K = 8$. The shown square marking points are end points of the graph.](image9)

![Figure 10. Reeb Graphs for an occupant holding a balloon as in fig. 5(c) at different resolution levels (a) (a) $K = 2$ (b) $K = 4$ (c) $K = 8$. The shown square marking points are end points of the graph.](image10)

It is now interesting to see the number of end-points present in each graph since if there was a large number of end-points, the use of Reeb graph would appear less interesting. In Figure 11(b), the average number of head candidates i.e. the average number of end points of the graph over all frames of images are plotted. Clearly, a maximum of 4 end-points are present in the graph. Typically there are 2 or 3 end-points. Results shows that the topology information of the scene successfully detects the position of the occupant’s head while the number of possible head candidates is small.

### 6. Conclusions and Future Work

This paper presents a general topological analysis framework for low-resolution range images that offers a systematic way to detect the vehicle occupant’s head. The Reeb graph technique is considered for this purpose. However, alternative topology representations other than the Reeb graph could also be useful. The previous approaches to Reeb graph extraction are applied to a dense 3D representation of a closed object surface, such as, e.g. triangulated mesh representation. Here, we adapted the Reeb graph concept to low-resolution range images. This work is motivated by the work of Xiao et al [14] who proposed a discrete Reeb graph (DRG) approach to unorganized cloud of 3D points. They explored the use of connectivity notions defined by calculating the distance between 3D points. To our data, we proposed a voxel neighborhood connectivity notion for building the Reeb graph. The voxel neighborhood connectivity of range data has the following features: 1) By choosing an appropriate voxel activation level for each voxel, depending on its distance to the camera, one can cope with large variation of the density of data. In this process, the influence of outliers due to the noise in the data can also be reduced. 2) The voxel connectivity is computationally more efficient than the calculating pairwise distances. This is particularly important in the automotive application which is also a subject in the cost constraints.

The approach is applied to several sequences which were recorded to take into account the different variations of the scene. Results on this data showed that the Reeb graph detects in most of the cases successfully the head. On this way, it was shown that an average of only 2.5 head candidates are resulted for each image. What is missing, and what will be investigated therefore in the future is a method to select the correct head candidate out of the Reeb graph end-points and to combine this method with a data association and tracking algorithm.

### Acknowledgments

The authors would like to thank Prof. Hamid Krim and Dr. Sajjad Baloch from North Carolina State University,
Number of frames successfully detected | %Correct | Average number of end-points in each image
--- | --- | ---
3114/3225 | 96.56% | 2.5

Table 1. Successful head detection results

![Histogram of difference error in X (cm)](image1)

![Histogram of number of head candidates](image2)

Figure 11. a) Histogram of the difference between ground truth measurement and its closest end-point of the Reeb graph over all frames. b) Histogram of number of end-points over all frames.

For fruitful discussions on the Reeb graph approach. This project is funded by IEE S.A., Luxembourg and Luxembourg International Advanced Studies in Information Technology (LIASIT), Luxembourg.

References