ACOUSTIC RANGE IMAGE SEGMENTATION BY EFFECTIVE MEAN SHIFT

U. Castellani, M. Cristani, V. Murino

Dipartimento di Informatica, University of Verona
Strada le Grazie 15, 37134 Verona - Italy

ABSTRACT

Image perception in underwater environment is a difficult task for a human operator, and data segmentation becomes a crucial step toward an higher level interpretation and recognition of the observing scenarios. This paper contributes to the related state of the art, by fitting the mean shift clustering paradigm to the segmentation of acoustical range images, providing a segmentation approach in which whatever parameter tuning is absent. Moreover, the method exploits actively the connectivity information provided by the range map, by using reverse projection as acceleration technique. Therefore, the method is able to produce, starting from raw range data, meaningful segmented clouds of points in a fully automatic and efficient fashion.

1. INTRODUCTION

Automatic segmentation of three-dimensional (3D) data is still an open research field, that can be considered as bridge between the classical image segmentation and the more general clustering of multi-dimensional data. In specific, the 3D segmentation is the focus of a vast literature and several surveys, reporting interesting approaches for different data representations such as unorganized points, range image, or 3D polygonal meshes [1].

In this paper, we focus on the segmentation of range acoustic images in underwater environments, for which the problem becomes more challenging because of the very noisy nature of acquired data. In this framework, we propose a new clustering-based 3D segmentation method by introducing a non parametric density estimation approach, based on the mean shift paradigm [2]. The mean shift (MS) clustering operates by shifting a fixed size estimation window from each data point towards the direction of maximal density, and converging into a basin of attraction, that represents a local mode. The points converging to the same centroid belong to the same region.

Although the mean shift has shown to be a powerful technique for several fields of research such as image and video segmentation [2, 3], tracking [4], clustering, and data mining [5], very few works have been addressed to it within the context of 3D data segmentation [6, 7] and, for the best of our known, none of them is related to range images. Furthermore, all these approaches rely on the tuning of several parameters, where the kernel is empirically specified.

In this paper, the mean shift paradigm has been extended to range images. Each point of the range data lives in a 7-dimensional *joint space*, formed by three subspaces, describing respectively the 3D coordinates, the normal and the curvature of that point. In this framework, a multi-dimensional mean shift clustering operation is performed; the granularity of this operation is determined by some parameters, i.e. the kernel bandwidths, one for each subspace, that, together, form a multi-dimensional kernel bandwidth. Large bandwidths lead to global but coarse separations, whereas small bandwidths better identify local modes, however risking over-partition.

According to the concept of *stable* segmentation [8], for each subspace, we find out the bandwidth value providing the most robust partition, using the MS clustering on that subspace. Thus, we fuse all the best bandwidth values, so as to form a multidimensional kernel which is an adapted to the characteristics.

Furthermore, as observed in [6], when the dimension of the space increases, as well as the number of points involved in the computation, the search for neighbors in feature-space is a key component, affecting the efficiency and feasibility of the algorithm. In order to treat this issue, a speed-up technique has been proposed. The main idea consists in the implementation of the *reverse* projection paradigm, that exploits connectivity properties of range data, explained in the following.

2. SOURCE DATA

Three-dimensional acoustic data are obtained with a high resolution acoustic camera, the *Echoscope* 1600 [9]. The scene is insonified by a high-frequency acoustic pulse, and a two-dimensional array of transducers gathers the backscattered signals. The whole set of raw signals is then processed in order to form computed signals whose profiles depend on echoes coming from fixed steering directions (called *beam signals*), while those coming from other directions are attenuated. Successively, the distance of a 3D point can be measured by detecting the time instant at which the maximum peak occurs in the beam signal [9]. According to the spherical scanning technology, range values are measured from each steering direction (u, v), where u and v are indices related

to the elevation *(tilt)* and azimuth *(pan)* angles respectively. Fig. 1 shows a projection of the acquiring volume to the ZX (or ZY) plane, on which the sector associated to the central beam is marked.

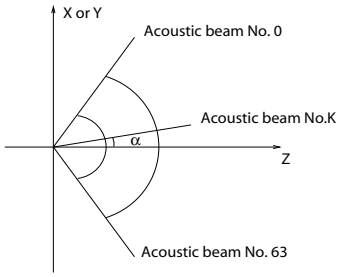


Fig. 1. Subdivision of the beams onto the acquiring volume. Each beam is associated to a (u, v) coordinate of the range image.

Going into details, the Echoscope carries out 64 measures for both tilt and pan by defining a 64×64 range image $\mathbf{r}_{u,v}$. Spherical coordinates are converted to usual Cartesian coordinates, referring to a coordinate system centered at the camera, by the use of the following equations [9]:

$$x = \frac{r_{u,v} \tan(vs_\alpha + U_{OFF})}{\sqrt{1 + \tan^2(us_\alpha + U_{OFF}) + \tan^2(vs_\beta + V_{OFF})}} \quad (1)$$

$$y = \frac{r_{u,v} \tan(vs_\beta + V_{OFF})}{\sqrt{1 + \tan^2(us_\alpha + U_{OFF}) + \tan^2(vs_\beta + V_{OFF})}} \quad (2)$$

$$z = r_{u,v} \sqrt{\tan^2(us_\alpha + U_{OFF}) + \tan^2(vs_\beta + V_{OFF})} \quad (3)$$

where s_α and s_β are elevation and azimuth increments respectively and U_{OFF}, V_{OFF} are offsets. These parameters are fixed by the acquisition sensor, determining the aperture of the acquisition (i.e., field of view and resolution). The result is a cloud of 3D points in *x,y,z* coordinates, each of them refers to an entry of a 64×64 matrix.

Therefore, in order to reverse the process, the projection of a 3D point (x, y, z) onto the range image is specified by the following equation:

$$u = \frac{\alpha - U_{OFF}}{s_\alpha}; \qquad v = \frac{\beta - V_{OFF}}{s_\beta} \quad (4)$$

where $\alpha = arctg(y/z)$ and $\beta = arctg(x/z)$.

3. MEAN SHIFT

The mean shift procedure is an old non-parametric density estimation technique [8, 2]; the theoretical framework of the mean shift arises from the Parzen Windows technique, that in particular hypotheses of regularity of the input space (independency among dimensions, see [2] for further details) estimates the density at point $\mathbf{x}$ as:

$$\hat{f}_{h,k}(\mathbf{x}) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2\right) \quad (5)$$

where $c_{k,d}$ is a normalizing constant, n is the number of points available, and $k(\cdot)$ the kernel profile, that models how strongly the points are taken into account for the estimation, in dependence with their distance h to $\mathbf{x}$.

Mean shift extends this "static" expression, differentiating (5) and obtaining the density gradient estimator

$$\hat{\nabla} f_{h,k}(\mathbf{x}) = \frac{2c_{k,d}}{nh^d} \left[\sum_{i=1}^{n} g\left(\left\| \frac{\mathbf{x}_i - \mathbf{x}}{h} \right\|^2\right)\right] \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\| \frac{\mathbf{x}_i - \mathbf{x}}{h} \right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\| \frac{\mathbf{x}_i - \mathbf{x}}{h} \right\|^2\right)} - \mathbf{x}\right] \quad (6)$$

where $g(x) = k'(x)$; this quantity is composed by three terms: the second one is *proportional* to the normalized density gradient obtained with the kernel profile k, the third one is the *mean shift* vector, that is guaranteed to point towards the direction of maximum increase in the density. Therefore, starting from a point $\mathbf{x}_i$ in the feature space, the mean shift produces iteratively a trajectory that converges in a stationary point $\mathbf{y}_i$, representing a mode of the whole feature space.

4. THE PROPOSED METHOD

In this paper, we consider each point $\mathbf{x}_i$ of the source data as a 7-dimensional entity, living in a *joint domain*. In specific, $\mathbf{x}_i = [\mathbf{x}_{i,s}, \mathbf{x}_{i,n}, \mathbf{x}_{i,c}]'$, where each component identifies the 3D (x, y, z) *spatial*, the 3D *normal* and the 1D *curvature* subdomain. The curvature is modelled by the *curvedness* index [1]; for each sub-domain we assume Euclidian metric.
In order to explore the joint domain, a multivariate kernel is used [2], that is:

$$K_{h_s,h_n,h_c}(\mathbf{x}) = \frac{C}{h_s^3 h_n^3 h_c} \prod_{u \in \{s,n,c\}} k\left(\left\| \frac{\mathbf{x}_u}{h_u} \right\|^2\right) \quad (7)$$

where C is a normalization constant, and h_s,h_n,h_c are the kernel bandwidths for each sub-domain. As intra-subspace kernel $k(\cdot)$, we adopt the Epanechnikov kernel [2], that differentiated leads to the uniform kernel $g(\cdot)$, i.e., a d-dimensional unit sphere.

Therefore, aiming at automatically estimating the kernel bandwidth dimension, we propose a task-oriented selection technique, that exploits decomposition stability criteria, composed by three steps.

1. *Standardization*: we rearrange each sub-domain as a hypercube, where the length of the side is fixed as the value of the largest dimension of that subspace, i.e. $R_{j \in \{s,n,c\}}$.
2. *Separate choice of the best bandwidth*: we divide uniformly the range of each subspace in $2N_{\max}$ values, and we consider those $N_{\max}$ values falling in the range $[R_{j \in \{s,n,c\}}/2N_{\max}, R_{j \in \{s,n,c\}}/2]$, enumerating them as $\{h^{(v)}_{j \in \{s,n,c\}}\}, v = 1, \ldots, N_{\max}$. With these values, we

perform *separately* for each sub-domain mean shift clustering. After these trials, we choose as best bandwidth value $h_j^{(v_{\text{best}})}$, where $v_{\text{best}} = (v_{\max} - v_{\min})/2$ indicates the center of the largest operating range $[h_j^{(v_{\min})}, h_j^{(v_{\max})}]$ (i.e., a plateau) over which the same number of partitions are obtained for the given data.

3. *Final clustering*: we perform again the mean shift clustering in the joint domain by using the kernel formed by concatenating the optimal sub-domain bandwidth values (see Eq. 7)).

This method individuates separately for each sub-domain, that we suppose to be independent from the other, the bandwidth most stable, in the sense claimed by [8], p.541. Putting together the best bandwidth values in a unique composite bandwidth corresponds to define a kernel that has the form of Eq. 7, leading to a mean shift vector equal to

$$m(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i \prod_{u \in \{s,n,c\}} g\left(\left\| \frac{\mathbf{x}_{i,u} - \mathbf{x}_u}{h_u} \right\|^2 \right)}{\sum_{i=1}^{n} \prod_{u \in \{s,n,c\}} g\left(\left\| \frac{\mathbf{x}_{i,u} - \mathbf{x}_u}{h_u} \right\|^2 \right)} - \mathbf{x} \qquad (8)$$

The speed-up technique is introduced in order to deal with several range images, all of them acquired by the same sensor (i.e., both the range of acquisition and the density of the points are similar for all the images). In such a situation, the optimal parameters can be calculated only on a single image, using the speed up technique to perform the segmentation on the remaining images. The proposed technique consists in reorganizing the range image $\mathbf{r}_{u,v}$ by adding the normal components and the curvature for each of its entry. Indeed, let us consider a point of the feature space $\mathbf{x}_l^t$ at the step t. By using Eq. 4 the point is re-projected onto the range image at the position (u, v) (Fig. 2, step 1). Then, the range connectivity information is used and a set ζ^t of 'potential' neighbors are selected by fixing a squared window W of size d centered at (u, v)(Fig. 2, step 2). Therefore, the next position $\mathbf{x}_l^{t+1}$ is computed by applying Eq. 8 where, instead of using the whole set of points, the sum is carried out only among the points $\mathbf{x}_i \in \zeta^t$ (Step 3, Fig. 2). Note that the window size d should be considered as a coarse approximation of the correct bandwidth of the spatial sub-space. Thus, its value is easy to estimate by adopting a conservative approach since it influences the speed of the processing while not affecting the accuracy of the segmentation.

5. EXPERIMENTS

The proposed method has been tested using a P4 3Ghz (Matlab code) on both synthetic and real acoustic data. The normals and the principal curvatures are computed by using classical quadric fitting estimation [1]. After the standardization of the data, we select the best bandwidth values for each subdomain, using $N_{\max} = 10$. The speed up technique has been applied by using a window size $d = 10$.

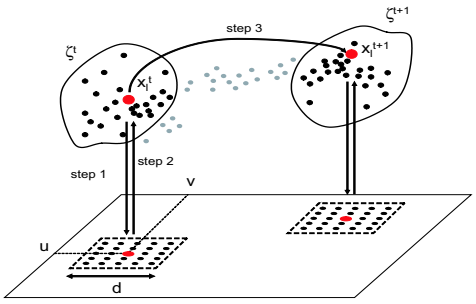
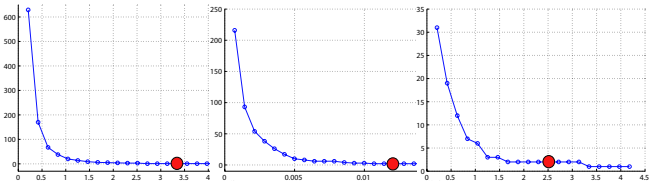


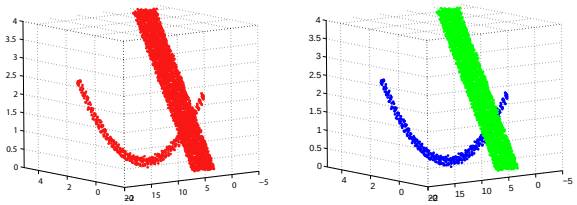
Fig. 2. Speed up technique



(a) (b) (c)

Fig. 3. Best bandwidths selection: (a) spatial coordinates, (b) surface normals, (c) curvatures. The lighter point is the best value, in a stability sense.

The first experiment (synthetic) shows the efficacy of the automatic bandwidths estimation. The scene consists of a plane trespassed by a gauge (Fig. 4.a) and gaussian noise has been added to data. The best bandwidth values are automatically estimated for each subspaces. Fig. 3 shows the progress of the bandwidth evaluation for the *spatial* (Fig. 3.a), the *normal* (Fig. 3.b) and the *curvature* (Fig. 3.c) subspaces. Each graph represents the number of clusters obtained using increasing bandwidth values. In all the graphs, is easy to note the largest plateuses, in the middle of them the best bandwidth is selected (that appears with a lighter marker). Therefore, these values are merged by using the multidimensional kernel of Eq. 7 and the final segmentation is obtained (Fig. 4.b), where the plane and the gauge are correctly separated.



(a) (b)

Fig. 4. Experiment 1: sampled points (a) and results of segmentation

In the second experiment (real) the scene is composed of a single pipe on a flat bottom (Fig. 5.a). Also for this experiment the best bandwidths are recovered for all the three subspaces and the final segmentation is obtained (Fig. 5.b). As

expected, the bottom and the pipe are correctly segmented. In the third experiment (real) the scene is more complex and

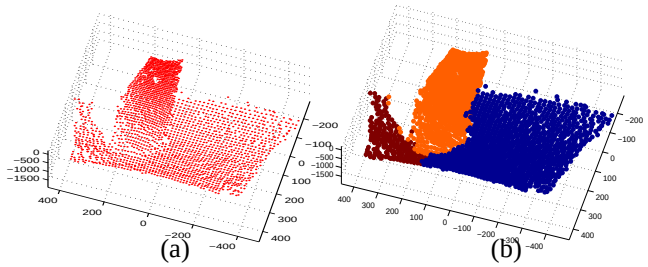


Fig. 5. Experiment 2: source data (a) and result of the segmentation (b)

it consists of a big pillar on the left, the seabottom, and two pipes on the right (Fig. 6, 1st row). The data are very noisy and the objects on the scene are very little recognizable. The best kernel estimation obtained from the previous experiment has been used for this experiment as well. The recovered segmentation is fully convincing since the four objects are correctly separated and the perception of the scene is improved (Fig. 6, 2nd row).

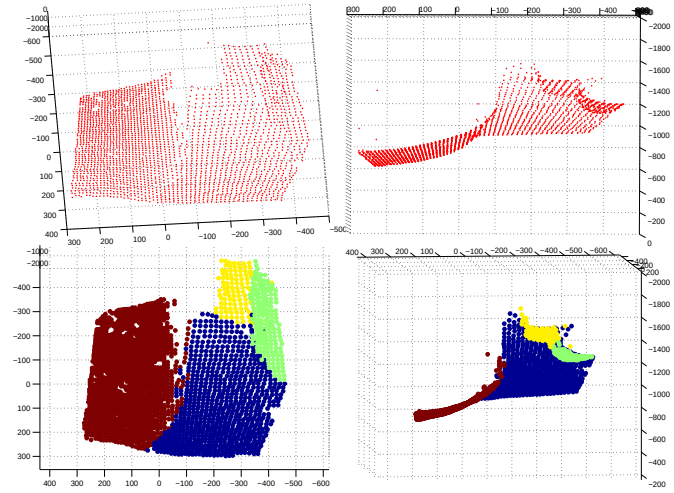


Fig. 6. Experiment 3: in the 1st row, front view (a) and top view (b) of the source data. In the 2st row, our results

Experiment	N. points	Non-Optimized (sec.)	Optimized (sec.)
Real 1	2399	58.0469	9.6094
Real 2	2835	188.06525	16.6406

Table 1. Performance of the MS segmentation for the real experiments

Finally, in Tab. 1 a performance evaluation is reported. The speed of the MS segmentation is drastically reduced for both the real experiments, when the proposed optimized approach is carried out. Note that the improvement of the proposed method is stronger in the second real experiment, when the number of points is increased. An exhaustive evaluation of the performance will be exploited for future works.

6. CONCLUSIONS

In this paper a new method for acoustic image segmentation is proposed. The mean shift paradigm has been applied effectively to the 3D range images by modelling correctly both the geometric properties of the source data and the information coming from the range connectivities. With respect to the current mean shift-based 3D segmentation methods our approach improves the automatism of the kernel bandwidth estimation, basing on a stability principle, and the speed of the algorithm, resorting to a reverse projection approach. Results are satisfying in terms of accuracy of segmentation and speed.

7. REFERENCES

[1] Sylvain Petitjean, "A survey of methods for recovering quadrics in triangle meshes," *ACM Comput. Surv.*, vol. 34, no. 2, pp. 211–262, 2002.

[2] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[3] J. Wang, B. Thiesson, Y. Xu, and M. Cohen, "Image and video segmentation by anisotropic kernel mean shift," in *ECCV (2)*, 2004, pp. 238–249.

[4] R.T. Collins, "Mean-shift blob tracking through scale space," in *CVPR (2)*, 2003, pp. 234–240.

[5] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: A texture classification example," in *ICCV*, 2003, pp. 456–463.

[6] A. Shamir, "Geodesic mean shift," in *Proceedings of the 5th Korea Israel conference on Geometric Modeling and Computer Graphics*,, 2004, pp. 51–56,.

[7] H. Yamauchi, S. Lee, Y. Lee, Y. Ohtake, A. Belyaev, and H.P. Seidel, "Feature sensitive mesh segmentation with mean shift," in *Shape Modeling International 2005*, 2005, pp. 236–243.

[8] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, second edition, 1990.

[9] R.K. Hansen and P.A. Andersen, "A 3d underwater acoustic camera - properties and applications," in *Acoustical Imaging*, P.Tortoli and L.Masotti, Eds., pp. 607–611. 1996.