Eurographics Italian Chapter Conference (2006)
G. Gallo and S. Battiato and F. Stanco (Editors)

# 3D data segmentation using a non-parametric density estimation approach

U. Castellani, M. Cristani, V. Murino

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

**Abstract**

*In this paper, a new segmentation approach for sets of 3D unorganized points is proposed. The method is based on a clustering procedure that separates the modes of a non-parametric multimodal density, following the mean-shift paradigm. The main idea consists in projecting the source 3D data into a set of independent sub-spaces, forming a joint multidimensional space. Each sub-space describes a geometric aspect of the data set, such as the normals and principal curvatures, so as a dense region in a particular sub-space indicates a set of 3D points sharing a similar value of that feature. A non-parametric clustering method is applied in this joint space by using a multidimensional kernel. This kernel smoothly takes into account for all the subspaces, moving towards high density regions in the joint space, separating them and providing "natural" clusters of 3D points. The algorithm can be implemented very easily and only few parameters need to be freely tuned. Experiments are reported, both on synthetic and real data, assessing the quality of the proposed approach and promoting further developments.*

Categories and Subject Descriptors (according to ACM CCS): I.4.6 {Image Processing and Computer Vision}{Segmentation}[Partitioning]

## 1. Introduction

Segmentation is a vast and complex domain, both in terms of problem formulation and resolution techniques. It consists in formally translating the delicate visual notions of homogeneity and similarity, and defining criteria which allow their efficient implementation [Pet02]. The goal is to partition the source data into meaningful pieces, i.e. those parts corresponding to the different entities, in the physical and semantical sense of the application envisioned. In the realm of 3D data, the segmentation comprises numerous applications, for example texture atlas generation [LPRM02, SWG*03], shape simplification [CSAD04, KT96], shape recognition and matching [AA93, CF01], and shape modelling and retrieval [FKS*04,KT03,FK04]. The literature is vast and several surveys report interesting approaches for different data representations such as unorganized points, range image, or 3D polygonal meshes [Pet02, HJBJ*96, AA93]. Roughly speaking, the segmentation methods can be categorized into two main classes: *edge-based* and *region-based* [Pet02]. In the former, features corresponding to part boundaries are first detected and then regions are built, each one formed by sets of points delimited by the same boundary. In the latter, points sharing the same similarity property are grouped together. In particular, three are the most popular approaches to region-based segmentation: *split-and-merge* methods, identified by a top-down paradigm; *region-growing* methods, that adopt a bottom-up paradigm, and *clustering-based* methods, based on the projection of the points onto a higher dimensional space where the clusters (i.e., segments) are recovered by defining some particular distance functions [JMF99]. In this paper a new *clustering-based* method for 3D unorganized points segmentation is proposed. The segmentation is obtained from the source data by introducing a non-parametric density estimation approach based on the mean shift paradigm [CM].
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 [CM, WTXC04], tracking [Col03], clustering, and data mining [GSM03], very few works have been addressed to it within the context of three-dimensional data segmentation. In [HS03], the mean shift has been applied on the histogram of the principal curvatures in order to recover geometric primitives from range images. In summary, instead of classifying the surface patches on the basis of the mean and the gaussian curvatures values combinations [AA93], the authors have proposed to detect the type of primitives by analyzing the clusters obtained onto the principal curvatures space. In [Sha04], the 3D polygonal mesh is projected onto a 2D image by preserving the geodesic distances among the points and by coloring each pixel according to the curvature value of its corresponding 3D point. Therefore, the segmentation is obtained by adopting the standard mean shift approach for 2D image segmentation [CM]. In [YLL*05], the 3D polygonal mesh is directly analyzed and the authors proposed to apply the mean shift algorithm on the surface normals space (i.e., the feature space) as a pre-processing stage. The segmentation is then completed by adopting a classical region-growing approach.

In this paper, the source data are unorganized 3D points. For each point, both the normal (such as in [YLL*05]) and principal curvatures (such as in [Sha04]) are extracted and are joined onto the feature space.

The contribution of the paper is twofold: first, we assess the effectiveness of a non-parametric clustering paradigm as 3D segmentation approach for sets of unorganized data points; second, we show that in such a framework, the enrichment of the set of features makes the segmentation more accurate, without the necessity of special pre- or post-processing.

The rest of the paper is organized as follow. After describing an overview of the mean shift procedure in Section 2, the details of the proposed method are reported in Section 3. Results of the algorithm on synthetic and real images are shown in Section 4, and finally, in Section 5 conclusions are drawn and future perspectives are envisaged.

## 2. Mean Shift

The mean shift procedure is an old non-parametric density estimation technique [Fuk90]. The main underlying idea is that the data feature space is regarded as an empirical probability density function to estimate: therefore, a big concentration of points that fall near the location $\mathbf{x}$ indicates a big density near $\mathbf{x}$.
The theoretical framework of the mean shift arises from the parzen windows [DHS01] basic expression, i.e. the kernel density estimator, that is

$$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x}-\mathbf{x}_i) \quad (1)$$

where $\hat{f}(\mathbf{x})$ represents the approximated density calculated in the $d$-dimensional location $\mathbf{x}$, $n$ is the number of available points and

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x}). \quad (2)$$

Here above, $K_{\mathbf{H}}$ can be imagined as a weighted window used to estimate the density, dependent on the kernel $K$ and the symmetric positive definite $d \times d$ bandwidth matrix $\mathbf{H}$. The function $K$ is a bounded function with compact support (for full details, see [CM]); the bandwidth matrix codifies the uncertainty associated to the whole feature space.

In the case of particular radial symmetric kernels (see [CM]), $K$ can be specified using only a 1-dimensional function, the *profile* $k(\cdot)$, the same for each dimension. Moreover, if we assume independence among the feature dimensions and equal uncertainty over them, the bandwidth matrix can be rewritten as proportional to the identity matrix $\mathbf{H} = h^2\mathbf{I}$. Under such hypotheses, Eq. 2 can be rewritten as

$$\hat{f}_{h,K}(\mathbf{x}) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \quad (3)$$

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

Mean shift extends this "static" expression, differentiating (3) and obtaining the density gradient estimator

$$\hat{\nabla} f_{h,K}(\mathbf{x}) = \frac{2c_{k,d}}{nh^d}\left[\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n}\mathbf{x}_i g\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}}{h}\right\|^2\right)}-\mathbf{x}\right] \quad (4)$$

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 $K$.
The third one is the *mean shift* vector, that is guaranteed to point towards the direction of maximum increase in the density (see [CM]). 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.

## 3. The proposed method

Our segmentation method can be thought as a clustering process, derived from the approach proposed in [CM]. Briefly speaking, the first step of such process is made by applying the mean shift procedure to all the points $\{\mathbf{x}_i\}$, producing several convergency points $\{\mathbf{y}_i\}$. A consistent number of close convergency locations, $\{\mathbf{y}_i\}_l$, indicates a mode $\mu_l$. The labeling consists in marking the corresponding points $\{\mathbf{x}_i\}_l$ that produces the set $\{\mathbf{y}_i\}_l$ with the label $l$. This happens for all the convergency location $l = 1,2,\ldots,L$.

In this paper, we consider each point of the cloud as a $d$-dimensional entity, living in a *joint domain*. In specific, each $\mathbf{x}_i$ is formed by the $x_1$,$x_2$,$x_3$ spatial coordinates relative to the

$x$,$y$,$z$ axes (the *spatial sub-domain*), the $n_1$,$n_2$,$n_3$ normal coordinates (the *normal sub-domain*) and the $c$ curvature (the *curvature sub-domain*). In particular, the curvature is modelled by the so called *curvedness-index* [Pet02]:

$$CurvnInd(k_1,k_2) = \frac{2}{\pi} ln \sqrt{\frac{k_1^2 + k_2^2}{2}} \quad (5)$$

where $k_1$ and $k_2$ are the principal curvatures [AA93]. For each sub-domain we assume Euclidian metric.
In order to explore the joint domain, a multivariate kernel is used [CM, WTXC04], that has the form

$$K_{h_s,h_n,h_c}(\mathbf{x_i}) = \frac{C}{h_s^3, h_n^3, h_c} k\left(\left\|\frac{\mathbf{x}_{i,s}}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}_{i,n}}{h_n}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}_{i,c}}{h_c}\right\|^2\right) \quad (6)$$

where $\mathbf{x}_{i,s}$ indicates the spatial coordinates of the $i-$th point and so on for $\mathbf{x}_{i,n}$ and $\mathbf{x}_{i,c}$; C is a normalization constant, and $h_s$,$h_n$,$h_c$ are the kernel bandwidths for each sub-domain. These values give to each feature domain the intuitive concept of "importance": strictly speaking, the bigger is the the related kernel bandwidth, the less important is that feature. In other words, a big amplitude of the kernel tends to agglomerate points in few convergence locations, while a small kernel highlights better local modes, encouraging cluster separations.
In this paper we use the Epanechnikov kernel [CM], that can be described by the profile

$$k(x) = \begin{cases} 1-x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

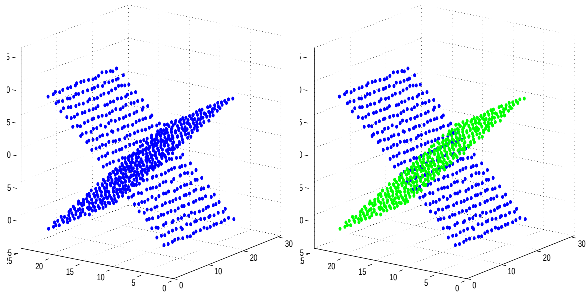
that differentiated leads to the uniform kernel, i.e. a $d$-dimensional unit sphere.

## 4. Experiments

The proposed method has been tested first on synthetic data, in order to give deep and clear insight to how it works. Gaussian noise has been added to each synthetic test in order to improve the realism of the experiments. Then, the method has been applied on real noisy data acquired with different 3D sensors. In all the experiments, we perform segmentation on clouds of 7-dimensional points, as explained in Sec.3, individuating 3 separated sub-domains (i.e., the spatial coordinates, the normals and the curvatures sub-domains). The normals and the principal curvatures are computed by using classical quadric fitting estimation [Pet02] and the curvedness index is recovered from eq. (5). From the source data, points having a degenerated value in computing the quadric fitting have been discarded. The kernel bandwidth values have been chosen easily, by following the principle explained above in Sec. 3. The current implementation of the proposed method is working under the Matlab environment.

### 4.1. Synthetic data

The first test has been run on two intersecting planes, each one opportunely sampled with 200 equally distanced points. Figure 1.a shows the image before the segmentation. Since

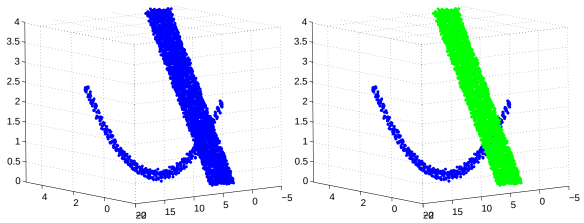


(a) (b)

**Figure 1:** *Planes: sampled points (a) and segmentation obtained using our method (b)*

the two planes are intersecting each other, they form a single connected component, therefore the spatial domain is not enough for segmenting correctly the scene. Thus, by introducing the information about the points directions (i.e., the normals) the joined space reveals two well separated clusters and the correct segmentation is obtained (Figure 1.b). It is worth noting that, for this experiment, the curvature information is not useful since all the points have the same flat curvature value.

The second experiment is performed on a plane trespassed by a gauge (Figure 2.a). For this experiment the normals are not enough for separating the two objects since they are not discriminative for the identification of the gauge. In fact, the gauge is composed of points having normals towards all the directions. Therefore, by introducing the information on the curvature the two objects are better distinguishable. In Figure 2.b the final result is proposed.

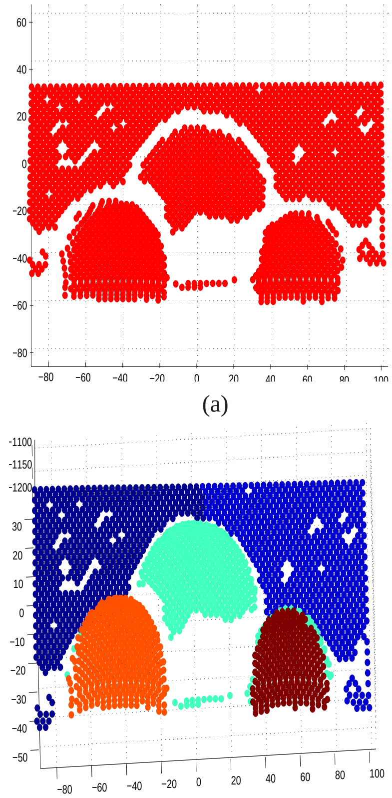


(a) (b)

**Figure 2:** *Plane and gauge: sampled points (a) and segmentation obtained using our method (b)*

### 4.2. Real data

The proposed method has been tested on two sets of range images acquired with different kind of sensors. The first set is from the *Minolta* database [CF98]. The scene consists of three spheres on a plane. Even if the scanner is a high resolution sensor, the image is quite noisy, especially onto the areas among the spheres (Figure 3.a). After the segmentation the result is reasonable since the three spheres are correctly separated (Figure 3.b). It worth noting that the plane on the background has been split into two parts. This is correct since the points on the upper central parts are attracted toward the points on their side. Furthermore, the sphere in the middle has been split into some small disconnected pieces. This is because in the proposed method no information about the points connectivities have been take into account. We are addressing this issue for the future works by using the connectivities information coming from either the range image or the polygonal mesh.

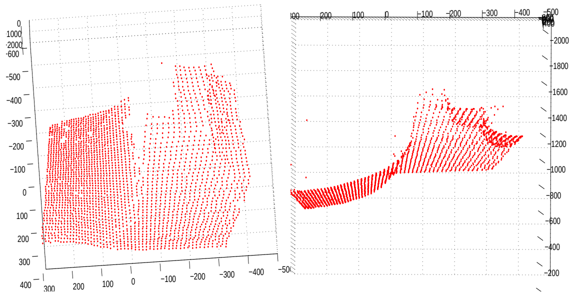


(a) (b)

**Figure 3:** *Spheres: sampled points (a) and segmentation obtained using our method (b)*

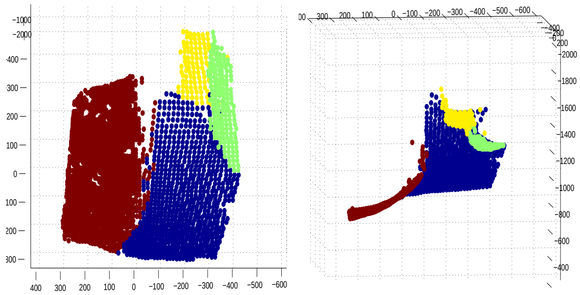
The second set is from the *EchoScope* [HA96] which is a 3D *acoustic* camera for underwater scene acquisition. The scene consists of a big pillar on the left, the seabottom, and two pipes on the right (Figure 4). The data are very noisy and the objects on the scene are very little recognizable. After the segmentation the result is fully convincing since the four objects are correctly separated (Figure 5).

It is worth noting that for all the proposed experiments, the parameters for the kernel definition have been manually tuned by using trials and errors. Since each sub-space is well bounded, the estimation of such parameters is not really tricky. Future works will address the automatic estimation of the kernel dimensions by exploiting some machine learning techniques, as suggested in [GSM03].



(a) (b)

**Figure 4:** *Underwater acoustic scenario: front view (a) and top view(b) of the sampled points*



(a) (b)

**Figure 5:** *Underwater acoustic scenario: front view (a) and top view (b) of the segmentation obtained using our method*

## 5. Conclusions

In this paper, we introduce a 3D segmentation technique derived by the mean shift procedure. The geometric attributes of the source data are extracted and modelled onto a set of subspaces that define the global feature space, on which the clusters are recovered. The main aim is to show that the non-parametric paradigm derived from the MS strategy well behaves with 3D segmentation issues. Moreover, we show that in such a framework, the enrichment of the set of features makes the segmentation more accurate, without the necessity of special pre- or post-processing. The proposed method considers as a distinct object the cloud of points exploiting jointly high similarity in the different feature spaces, where the importance of each feature can be easily modulated by the user. This permits to avoid the fitting of rigid parametric models with the data; actually, such an operation needs heavy tuning steps, that our approach in fact disregards. The results are promising, both on the synthetic and real data cases, characterized by different levels of noise. Further research is currently under study, specially devoted to make automatic the phase of kernel selection. Furthermore, a general framework for the extension of the proposed approach to polygonal meshes and range images is in progress.

## References


[AA93] ARMAN F., AGGARWAL J. K.: Model-based object recognition in dense-range images: a review. *ACM Comput. Surv. 25*, 1 (1993), 5–43.

[CF98] CAMPBELL R., FLYNN P.: A www-accessible 3d image and model database for computer vision research. In *Empirical Evaluation Methods in Computer Vision* (1998), Bowyer K., Phillips P., (Eds.), IEEE Computer Society Press, pp. 148–154.

[CF01] CAMPBELL R. J., FLYNN P. J.: A survey of free-form object representation and recognition techniques. *Comput. Vis. Image Underst. 81*, 2 (2001), 166–210.

[CM] COMANICIU D., MEER P.: Mean Shift: A Robust Approach Toward Feature Space Analysis.

[Col03] COLLINS R.: Mean-shift blob tracking through scale space. In *CVPR (2)* (2003), pp. 234–240.

[CSAD04] COHEN-STEINER D., ALLIEZ P., DESBRUN M.: Variational shape approximation. *ACM Trans. Graph. 23*, 3 (2004), 905–914.

[DHS01] DUDA R., HART P., STORK D.: *Pattern Classification*, second ed. John Wiley and Sons, 2001.

[FK04] FUNKHOUSER T., KAZHDAN M.: Shape-based retrieval and analysis of 3d models. In *GRAPH '04: Proceedings of the conference on SIGGRAPH 2004 course notes* (New York, NY, USA, 2004), ACM Press, p. 16.

[FKS*04] FUNKHOUSER T., KAZHDAN M., SHILANE P., MIN P., KIEFER W., TAL A., RUSINKIEWICZ S., DOBKIN D.: Modeling by example. *ACM Trans. Graph. 23*, 3 (2004), 652–663.

[Fuk90] FUKUNAGA K.: *Statistical Pattern Recognition*, second ed. Academic Press, 1990.

[GSM03] GEORGESCU B., SHIMSHONI I., MEER P.: Mean shift based clustering in high dimensions: A texture classification example. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision* (2003), pp. 456–463.

[HA96] HANSEN R. K., ANDERSEN P. A.: A 3-D underwater acoustic camera - properties and applications. In *Acoustical Imaging*, P.Tortoli, L.Masotti, (Eds.). Plenum Press, 1996, pp. 607–611.

[HJBJ*96] HOOVER A., JEAN-BAPTISTE G., JIANG X., FLYNN P., BUNKE H., GOLDGOF D., BOWYER K., EGGERT D., FITZGIBBON A., FISHER R.: An experimental comparison of range segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence 7*, 18 (1996), 673–689.

[HS03] HAMEIRI E., SHIMSHONI I.: Estimating the principal curvatures and the darboux frame from real 3-d range data. *IEEE Transactions on Systems, Man and Cybernetics 33*, 4 (2003), 626–637.

[JMF99] JAIN A. K., MURTY M. N., FLYNN P. J.: Data clustering: a review. *ACM Comput. Surv. 31*, 3 (1999), 264–323.

[KT96] KALVIN A. D., TAYLOR R. H.: Superfaces: Polygonal mesh simplification with bounded error. *IEEE Computer Graphics and Applications 16*, 3 (1996), 64–77.

[KT03] KATZ S., TAL A.: Hierarchical mesh decomposition using fuzzy clustering and cuts. *ACM Trans. Graph. 22*, 3 (2003), 954–961.

[LPRM02] LÉVY B., PETITJEAN S., RAY N., MAILLOT J.: Least squares conformal maps for automatic texture atlas generation. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 2002), ACM Press, pp. 362–371.

[Pet02] PETITJEAN S.: A survey of methods for recovering quadrics in triangle meshes. *ACM Comput. Surv. 34*, 2 (2002), 211–262.

[Sha04] SHAMIR A.: Geodesic mean shift. In *Proceedings of the 5th Korea Israel conference on Geometric Modeling and Computer Graphics,* (2004), pp. 51–56,.

[SWG*03] SANDER P. V., WOOD Z. J., GORTLER S. J., SNYDER J., HOPPE H.: Multi-chart geometry images. In *SGP '03: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing* (Aire-la-Ville, Switzerland, Switzerland, 2003), Eurographics Association, pp. 146–155.

[WTXC04] WANG J., THIESSON B., XU Y., COHEN M.: Image and video segmentation by anisotropic kernel mean shift. In *ECCV (2)* (2004), pp. 238–249.

[YLL*05] YAMAUCHI H., LEE S., LEE Y., OHTAKE Y., BELYAEV A., SEIDEL H.: Feature sensitive mesh segmentation with mean shift. In *Shape Modeling International 2005* (2005), pp. 236–243.