Eurographics Italian Chapter Conference (2007)
Raffaele De Amicis and Giuseppe Conti (Editors)

# 3D Objects Face Clustering using Unsupervised Mean Shift

M. Farenzena, M. Cristani, U. Castellani and A. Fusiello

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

**Abstract**

*In this paper, a novel approach to face clustering is proposed. The aim is the completely unsupervised extraction of planes in a polygonal a mesh, obtained from a 3D reconstruction process. In this context, 3D coordinates points are inevitably affected by error, therefore resiliency is a primal concern in the analysis. The method is based on the Mean Shift clustering paradigm, devoted to separating modes of a multimodal non-parametric density, by using a kernel-based technique. A critical parameter, the kernel bandwidth size, is here automatically detected by following a well-accepted partition stability criterion. Experimental and comparative results on synthetic and real data validate the approach.*

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computational Geometry and Object Modeling]:

## 1. Introduction

Polygonal meshes remain a preferred representation for surface data, because of their ability to efficiently approximate complex shapes and their data-structural simplicity. In particular, triangular meshes are widely used in many engineering, medical, and entertainment applications. Anyway, in order to process a three-dimensional (3D) object in the form of a triangular mesh for further and more sophisticated analysis, extracting and opportunely organizing higher order features represents a fundamental step. In this study, we focus on *face clustering*: this operation is useful, for example, in a Computer Graphic context, for shape simplification [DCSD04, MGH01, KT96], shape modelling and retrieval [TF04, DCSD04], or to accelerate the face culling process.

In the context of Computer Vision, instead, face clustering strategies could be useful for image-based modelling applications. Here, an important operation is the automatic extraction and division of a mesh object (acquired from real images) into consistent sets of informative portions. In particular, *planes* can be extracted and organized into different entities, depending on their orientation and position. This operation should be considered as a first step for further refinements of the 3D structure as in [MFD06], and in general for a higher level analysis and processing of the mesh object. However, this context requires to see the problem from a perspective slightly different from a pure Computer Graphic problem: if a mesh derives from a 3D reconstruction process, in fact, the location of the 3D points is inevitably affected by error, and so error resiliency should be a leading factor for a face clustering algorithm on these data.

In this paper, we propose a noise-resilient, fully automatic method for face clustering, able to effectively partition meshes affected by noise. This approach relies on the Mean Shift (MS) clustering paradigm [CM02], which is a powerful general purpose procedure for non-parametric scattered data. The main underlying idea of such approach is that the data space is regarded as an empirical probability density function to estimate. In short, the MS procedure operates by shifting a fixed size estimation window, *the kernel*, from each data point towards a local mode, characterized by a high concentration of data points. The points converging to the same mode are included in the same cluster.

MS has shown to be a powerful technique for several research fields such as image and video segmentation, tracking, clustering and data mining [CM02, Col03, GSM03]. In the context of face clustering, instead, MS clustering has been applied to surface normals only as a pre-processing step to mesh segmentation. In [HY05], for example, it has been applied to perform a sort of local smoothing aimed at removing noise from data. Other state-of-the-art approaches to face clustering [SWG*03, CSAD04] mainly use region growing methods, and the focus is not on noise resiliency but only on the visual quality of the results. Moreover, most methods rely on manual tuning of several parameters.

This work, instead, builds on [MCV06], where a method for automatic selection of kernel parameters in MS algorithm is developed. [MCV06] Faces the problem of 3D segmentation on unorganized 3D points; here, the method is applied to extract planes from noisy 3D mesh objects. In summary, our method firstly groups triangles' normals, organizing together triangles with similar normals. Then, for each cluster, a further clustering operation based on the distances of the triangles from the origin is carried out, thus permitting to separate triangles belonging to parallel planes.

In literature, approaches for automatic estimation of MS parameters are present: a recent and important theoretical framework has been proposed by Comaniciu in [Com03], but it is based on the assumption that data are locally distributed with a Gaussian distribution, and corrupted by Gaussian noise. While commonly accepted and motivated in a pure pattern recognition context, this assumption does not hold in general for data characterizing rigid geometrical data: for example, punctual information which characterize corners and spikes, such as normals and spatial positions, are far from being characterized by a local Gaussian configuration. Therefore, we do not want to impose Gaussian assumptions: we accept every data configuration, only assuming that the clutter affecting the data is bounded, with a uniform distribution holding inside the bound. In order to sensibly validate our assumptions, an extensive performance comparison between our approach and [Com03] has been performed, showing better performances of our method in this context.

The rest of the paper is organized as follows. Sec. 2 describes an overview of the Mean Shift procedure, while Sec. 3 illustrates the automatic estimation of the bandwidth parameter. Sec. 4 depicts the proposed method and Sec. 5 shows the experimental results, on synthetic and real data. Finally, the conclusions are in Sec. 6.

## 2. Mean Shift

The Mean Shift procedure is a non-parametric density estimation technique [Fuk90, CM02]. The theoretical framework of the Mean Shift arises from the Parzen Windows technique, that, in particular hypotheses of regularity of the input space (such as indipendency among dimensions [CM02]), estimates the density at point $\mathbf{x}$ as:

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

where $c_{k,m}$ is a normalizing constant, $n$ is the number of points available, and $k(\cdot)$ is the kernel profile, that models how strongly points are taken into account for the estimation, according to the *bandwith* $h$, that establish a threshold on their distance from $\mathbf{x}$.

Mean Shift extends this "static" expression, by differentiating (1) and obtaining the gradient of the density as:

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

where $g(x)$ is the gradient of $k(x)$. In the above equation, the first term in square brackets is *proportional* to the normalized density gradient, and the second term is the *Mean Shift* vector, that is guaranteed to point towards the direction of maximum increase in the density [CM02]. Therefore, starting from a point $\mathbf{x}_i$ in the feature space, the Mean Shift iteratively produces a trajectory that converges in a stationary point $\mathbf{y}_i$, representing a mode of the whole feature space.

## 3. Bandwidth automatic estimation

The bandwidth parameter $h$ defines the level of detail of the analysis. Large bandwidth values lead to global but course separation, whereas small bandwidth values better identify local modes, but at the risk of over-partitioning the data space. Good segmentation results could be obtained after an accurate parameters tuning. In line with the concept of stable segmentation [Fuk90] we exploit the same strategy developed in [MCV06]. We single out extreme values $h_{min}$ and $h_{max}$ for $h$ and we uniformly divide the range $[h_{min}, h_{max}]$. Then, for each value of $h$ we perform a Mean Shift clustering. After these trials, we consider the graph of the number of clusters as a function of $h$ and we choose as the best bandwidth value the centre of the largest plateau (see Figure 1).

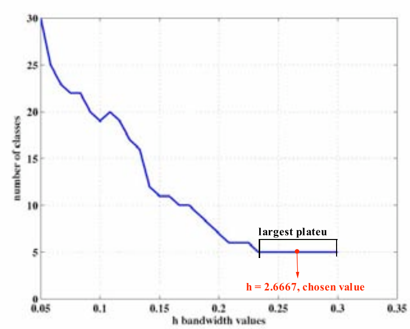



**Figure 1:** *Example of automatic bandwidth selection: on the graph of the number of clusters obtained from each trial with a h value, the centre of the largest plateau is selected.*

## 4. Proposed method

The proposed technique is composed by a two-step, hierarchical strategy. Firstly, normals of every mesh triangle are organized in separate clusters. Then, each cluster is further partitioned, in order to separate parallel planes.

As a reminder, a plane is expressed by the equation

$$ax+by+cz+d=0 \tag{3}$$

The parameters $a$, $b$ and $c$ are the coordinates of the plane's normal, while $d$ represents the plane's distance from the origin. Two triangles belonging to different parallel planes differ from $d$ only, so after the first clustering operation these triangles are in the same cluster. A second partitioning based on the parameter $d$ permits to separate them.

Specifically, for the former clustering operation, data are the points $\mathbf{x}_i = \mathbf{n}_i$, where $\mathbf{n}_i$ is the 3D normal of the $i$-th mesh triangle.

The adopted kernel is [CM02]:

$$K_{h_n}(\mathbf{x}) = \frac{C}{h_n^3} k\left(\left\|\frac{\mathbf{x}}{h_n}\right\|^2\right) \tag{4}$$

where $C$ is the normalization constant and $h_n$ the kernel bandwidth. Assuming a bounded error on 3D points, and a uniform distribution inside the bound (these assumptions are appropriate for this kind of applications, see [AJ94]), the appropriate $k$ employed is the Epanechnikov kernel [CM02]:

$$k(x) = \begin{cases} 1-x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

the function $k$, when differentiated, leads to the uniform kernel $g(\cdot)$, i.e. a multi-dimensional unit sphere.

The latter clustering operation is performed on data $x_i = d_{ic}$, where $d_{ic}$ is the plane's distance from the origin (3) of the $i$-th triangle in $c$ cluster. The kernel used is the same of the previous partition process, with bandwidth value $h_d$.

Finally, in order to automatically select the parameters $h_n$ and $h_d$ as described in Sec. 3, we need to single out their range of variability. We fixed $h_n \in [0.05, 0.2]$, whereas $h_d$ is adaptively computed so as to reflect the intrinsic scale of the problem. The lower bound of the range is $\min(d_i)$, whereas the upper bound is 10% of the median of the maximum distances between every pair of 3D points.

## 5. Experimental Results

The proposed method has been tested on both synthetic and real data.

As for the synthetic experiments, we used the four mesh objects in Fig. 2. As previously said, we assume that the 3D mesh points are affected by an uniform (bounded) error. So, the 3D points had been perturbed varying the bound's width and uniformly generating their positions inside the bound. For each object, we considered as bound's width $t$ values from 0.5% up to 3.0% of the median among the maximum distances between every pair of mesh points (see Figure 3). This related $t$ to the intrinsic scale of the object. For each $t$, we performed 50 independent trials. The mean percentages of mismatches wrt the ground truth are detailed in Table 1. Examples of the detected planes are depicted in Figs. 4 and 5.

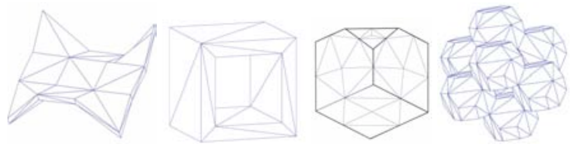


**Figure 2:** *The four objects used for the synthetic experiments, here referred as (from the left)* test*,* boxwhole*,* cutcube *and* foursix

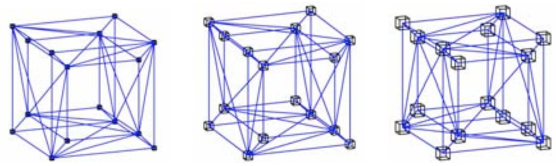
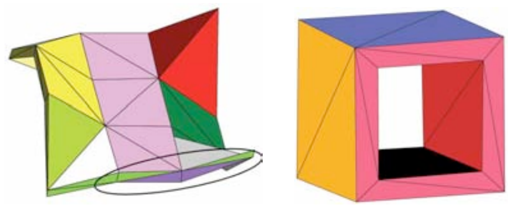


**Figure 3:** *Synthetic experiments for* boxwhole*, changing the error bound's width for different values of $t$. On the left, $t = 1$, i.e. the cubes width is 1% of the median among the maximum distances between every pair of mesh points; in the center, $t = 2$; on the right, $t = 3$.*

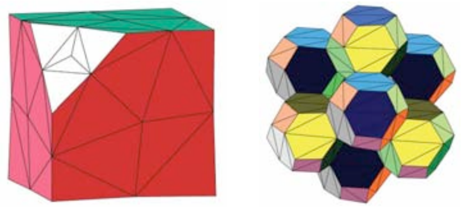


2 mismatches 0 mismatches

**Figure 4:** *Examples of detected planes for t = 1.5, with the number of mismatches. For the* test *object on the left, the mismatches are the two highlighted triangles.*

As the reader can notice, the algorithm works remarkably well, with a very little percentage of mismatches.

In the experimental comparative evaluation, we applied the same hierarchical clustering process, employing instead the bandwidth selection theorem proposed by [Com03]. The theorem implies a different MS formulation, giving to each data point a particular bandwidth value, instead of choosing a fixed bandwidth value for all the data space. Such bandwidth value is the one that maximize the module of the normalized MS vector, that from each location in the data space points towards the nearest mode. As previously said, the bandwidth selection theorem works well when the data is locally distributed as a Gaussian distribution, corrupted by



0 mismatches 0 mismatches

**Figure 5:** *Examples of detected planes for t = 1.5, with zero mismatches.*

| t | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|
| *test* | 0 | 0.12 | 1.35 | 3.97 | 6.88 | 9.67 |
| *boxwhole* | 0 | 0 | 0 | 0.09 | 0.36 | 3.09 |
| *cutcube* | 0 | 0 | 0 | 0.02 | 0.23 | 0.78 |
| *foursix* | 0 | 0 | 0 | 0 | 0.02 | 0.08 |

**Table 1:** *Synthetic experiments results: mean percentage of mismatches wrt the number of mesh triangles vs the bound's width $t$.*

Gaussian noise. The results are reported in Table 2. As expected, our algorithm shows up better performances.

| t | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|
| *test* | 1.53 | 2.63 | 4.60 | 6.07 | 6.97 | 8.33 |
| *boxwhole* | 0.06 | 0.44 | 1.19 | 2.19 | 2.88 | 3.00 |
| *cutcube* | 0.03 | 0.43 | 0.60 | 1.30 | 1.40 | 2.27 |
| *foursix* | 5.19 | 6.14 | 7.99 | 9.09 | 10.03 | 10.59 |

**Table 2:** *Synthetic experiments results obtained with the automatic bandwidth selection developed in [Com03]: mean percentage of mismatches wrt the number of mesh triangles vs the bound's width $t$.*

We tested the approach on real cases as well, using models obtained from an image-based reconstruction process [HZ00]. The *church* model is composed by 43 points and 93 triangles. The planes extracted by our algorithm are 38, with 6 triangles wrongly grouped, four of which derives from a wrong normals classification, the others from a wrong parallel planes separation (see Fig. 6).

The *tribuna* model is composed by 272 points and 364 triangles. The planes extracted are 52, showing an intuitive planes-organization of the object. Only 20 triangles were erroneously clustered. In this case, the errors equally derive from normals and parallel planes clustering.

## 6. Conclusions

In this paper, we propose a noise-resilient and fully automatic approach to face clustering based on the Mean Shift algorithm. The aim is the extraction of planes of a mesh acquired from a 3D reconstruction process. Resiliency to noise is proved by both synthetic and real experimental results. Concerning the automatic choice of the bandwidth value, in this paper we prove that, in case of rigid geometrical structures described by punctual measurements such as face normals, methods relying on Gaussian assumptions perform poorly. Instead, methods of bandwidth selection based on more general principles such as the stability of the partition, gives better results.

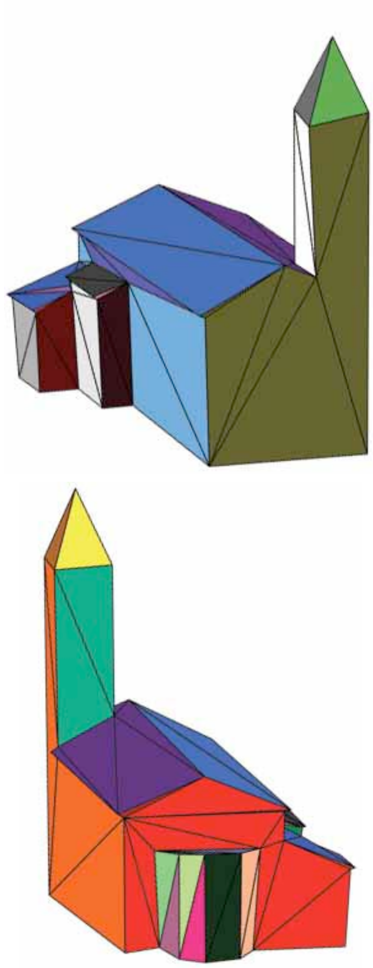


**Figure 6:** *Two views of* church *model, with the planes extracted by our algorithm.*

## References

[CSAD04] COHEN-STEINER D., ALLIEZ P., DESBRUN M.: Variational shape approximation. *ACM Transactions on Graphics 23*, 3 (2004), 905–914. 1

[HY05] H. YAMAUCHI S. LEE Y. L. E. A.: Feature sensi-

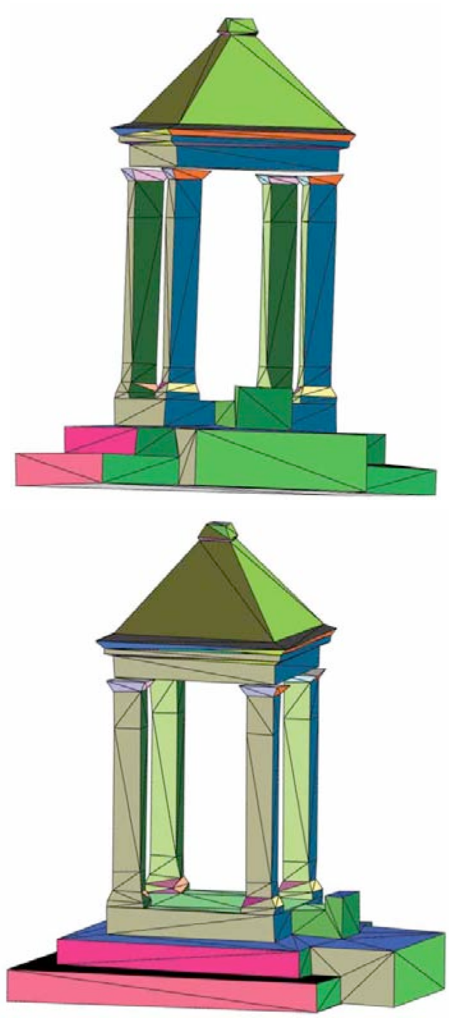


**Figure 7:** *Two views of* tribuna *model, with the planes extracted by our algorithm.*


tive mesh segmentation with mean shift. *Shape Modeling International* (2005), 236–243. 1

[KT96] KALVIN A., TAYLOR R.: Surfaces: Polygonal mesh semplification with bounded error. *IEEE Computer Graphics and Applications 16* (1996), 64–77. 1

[MCV06] M. CRISTANI U. CASTELLANI, V.MURINO: Adaptive feature integration for segmentation of 3d data by unsupervised density estimation. In *International Conference on Pattern Recognition ICPR 2006* (August 2006), vol. 4, pp. 21–24. 2

[MFD06] M. FARENZENA A. FUSIELLO, DOVIER A.: Reconstruction with interval constraints propagation. In *Computer Vision and Pattern Recognition* (2006), pp. 1185–1190. 1

[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. 1

[TF04] T. FUNKHOUSER M. KAZHDAN P. S. E. A.: Modeling by example. *ACM Transactions on Graphics 23* (2004), 652–663. 1

[DCSD04] D. COHEN-STEINER P. ALLIEZ, DESBREUN M.: Variation shape approximation. *ACM Transactions on Graphics 23* (August 2004), 905–914. 1

[MGH01] M. GARLAND A. WILLMOTT , HECKBERT P.: Hierarchcal face clustering on polygonal surfaces. In *Proceedings Symposium on Interactive 3D Graphics* (2001), pp. 49–58. 1

[Col03] R. COLLINS: Mean-shift blob tracking through scale space. In *CVPR (2)*, pages 234–240, 2003. 1

[Com03] D. COMANICIU: An algorithm for data-driven bandwidth selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):281–288, 2003. 2, 3, 4

[CM02] D. COMANICIU AND P. MEER: Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002. 1, 2, 3

[Fuk90] K. FUKUNAGA: *Statistical Pattern Recognition*. Academic Press, second edition, 1990. 2

[GSM03] B. GEORGESCU, I. SHIMSHONI, AND P. MEER: Mean shift based clustering in high dimensions: A texture classification example. In *ICCV*, pages 456–463, 2003. 1

[AJ94] T.D. ALTER AND D.W. JACOBS: Error Propagation in 3D-from-2D Recognition: Scaled-Orthographic and Perspective Projections. In *ARPA94*, pages 973–980, 1994. 3

[HZ00] R. HARTLEY AND A. ZISSERMAN: *Multiple view geometry in computer vision*. Cambridge University Press, 2000. 4