

## Point Cloud Transport

Hozuma Nakajima, Yasushi Makihara, Hsu Hsu, Ikuhisa Mitsugami, Mitsuru Nakazawa,  
Hirotake Yamazoe, Hitoshi Habe, Yasushi Yagi

*ISIR, Osaka University, Japan*

{*nakajima, makihara, hsu, mitsugami, nakazawa, yamazoe, habe, yagi*}@am.sanken.osaka-u.ac.jp

### Abstract

*In this paper we propose a method for temporal interpolation of a point cloud undergoing occlusions and topological changes. The point cloud is first merged into fine clusters, which are then further merged into coarse clusters for each source and target shape. In conjunction with trash box bins to cope with occlusions, a coarse correspondence between a source and a target shape is found that minimizes the transportation cost in the earth mover's distance framework. Subsequently, a fine correspondence is found in a similar way based on the coarse correspondence constraint to suppress locally isolated motion. Finally, the source and target point clouds are transported based on the fine correspondence. Experiments with point cloud sequences captured by a Kinect range finder show promising results.*

### 1. Introduction

Recently, range data have been used in a wide range of fields including 3-D face recognition [3], 3-D face modeling [1], human pose estimation [12], and cultural heritage modeling [9]. To capture dynamic objects (e.g., facial expressions and human poses), range data with higher frame-rates are preferable in several applications for better visualization and performance.

One of the ways of generating such range data is temporal interpolation by shape morphing techniques, which fall mainly into two categories: (1) surface-based approaches, and (2) volume-based approaches.

The surface-based approaches usually start by finding correspondences between the source and target surfaces, which is also an important process for non-rigid surface registration, otherwise known as deformable object registration [11, 6]. Most of the surface-based approaches, however, cannot successfully cope with topologically different shapes. As an exception, Bronstein et al. [5] proposed a topology-invariant similarity comprising both intrinsic and extrinsic similarities. In the

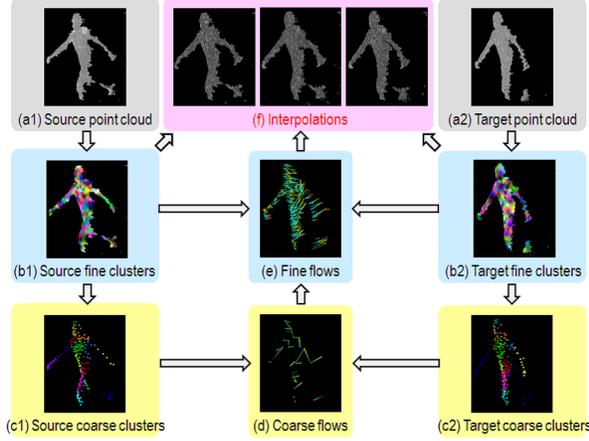
method, a shape is modeled as a metric space with a two-dimensional smooth compact connected surface embedded in the three-dimensional Euclidean space [4]. Acquiring such smooth compact connected surfaces, however, could be difficult if the range data contain significant noise and also discontinuities due to occlusions.

The volume-based approaches treat topologically different shapes in a more natural way by interpolating signed distance fields for source and target shapes (e.g., a positive value for the inside shape and a negative value for the outside shape) [10] or by applying a level-set method to the signed distance fields [2]. However, they naturally require volume data with a closed surface and hence cannot be applied to range data with an open surface observed from a single view.

In this paper, we propose an interpolation method for range data with open surfaces containing noise, discontinuities due to occlusions, and topological changes. Because it is difficult to reconstruct smooth surfaces from such range data, we formulate this interpolation as *point cloud transport* rather than non-rigid surface deformation. In this formulation, a point cloud in a source shape is directly warped into that of a target shape beyond topological changes.

The proposed method was inspired by topology-free volume-based 2D shape morphing [8], which, however, suffers from incorrect correspondences due to occlusions, as well as from locally isolated motion due to the absence of constraints on the smoothness of adjacent movements.

To overcome such problems, we introduce (1) trash box bins, and (2) a coarse-to-fine framework. Trash box bins realize *appearing* and *disappearing* processes for points with no plausible correspondences due to occlusions. Such points in the source and target shapes go to or come from the trash box bins equipped with the target and source shapes, respectively. In the coarse-to-fine framework, the coarse correspondence is employed as a global constraint for the fine correspondence, thereby suppressing undesirable locally isolated motion.



**Figure 1. Overview.** In (b) and (c), each point is depicted by an alpha-blended color based on membership to each cluster. In (d) and (e), flows from source to target clusters are depicted by a color transition from yellow to cyan, while a color transition from/to red indicates the appearance/disappearance by trash box bins. In (f), interpolated point clouds are shown with a transition rate of 0.25, 0.5, and 0.75 from left to right.

## 2. Proposed framework

### 2.1. Overview

An overview of the proposed framework is given in this section (see Fig. 1). A point cloud is first merged into fine clusters, which are then further merged into coarse clusters for each source and target shape. A coarse correspondence between a source and a target shape is found in conjunction with the trash box bins to cope with occlusions. Subsequently, a fine correspondence is found in conjunction with the coarse correspondence constraint to suppress locally isolated motion. Finally, source and target point clouds are transported based on the fine correspondence. Details of each of these steps are given in the following subsections.

### 2.2. Clustering

Given a point cloud composed of  $N$  3D data points  $\{\mathbf{x}_i\}$  ( $i = 0, \dots, N-1$ ),  $N_F$  fine clusters are obtained by fuzzy C-means (FCM) clustering [7]. We define the  $j$ -th fine cluster's mean  $\bar{\mathbf{x}}_{Fj}$  and weight  $w_{Fj}$  as

$$\bar{\mathbf{x}}_{Fj} = \frac{\sum_{i=0}^{N-1} m_{Fij} \mathbf{x}_i}{\sum_{i=0}^{N-1} m_{Fij}}, \quad w_{Fj} = \sum_{i=0}^{N-1} m_{Fij}, \quad (1)$$

where  $m_{Fij}$  is the membership of the  $i$ -th 3D data point to the  $j$ -th fine cluster, which holds  $\sum_{j=0}^{N_F-1} m_{Fij} = 1 \forall i$ .

Furthermore,  $N_F$  fine clusters are merged into  $N_C$  coarse clusters in the same way. We define the  $l$ -th coarse cluster's mean  $\bar{\mathbf{x}}_{Cl}$  and weight  $w_{Cl}$  as

$$\bar{\mathbf{x}}_{Cl} = \frac{\sum_{j=0}^{N_F-1} m_{Cjl} w_{Fj} \bar{\mathbf{x}}_{Fj}}{\sum_{j=0}^{N_F-1} m_{Cjl} w_{Fj}}, \quad w_{Cl} = \sum_{j=0}^{N_F-1} m_{Cjl} w_{Fj}, \quad (2)$$

where  $m_{Cjl}$  is the membership of the  $j$ -th fine cluster to the  $l$ -th coarse cluster, which satisfies  $\sum_{l=0}^{N_C-1} m_{Cjl} = 1 \forall j$ . Figures 1(b) and (c) illustrate examples of the membership of the fine and coarse clusters.

### 2.3. Coarse correspondence

The second step involves acquiring a coarse correspondence so as to minimize the transportation cost from the source to target shapes in a similar way to that given in [8]. Let the sets of means and weights for the source coarse clusters be  $\bar{X}_C^s = \{\bar{\mathbf{x}}_{Cl}^s\}$  and  $w_C^s = \{w_{Cl}^s\}$  ( $l = 0, \dots, N_C^s - 1$ ) and those for the target coarse clusters be  $\bar{X}_C^t = \{\bar{\mathbf{x}}_{Cl}^t\}$  and  $w_C^t = \{w_{Cl}^t\}$  ( $m = 0, \dots, N_C^t - 1$ ), where  $N_C^s$  and  $N_C^t$  are the number of source and target coarse clusters, respectively. Note that superscripts  $s$  and  $t$  denote source and target, respectively.

Moreover, an extra cluster for the trash box bin is added for each source and target shape. As a result, the numbers of source and target coarse clusters are incremented to  $N_C^s + 1$  and  $N_C^t + 1$ , respectively. For notation convenience, we assign such extra clusters to the  $N_C^s$ -th source and  $N_C^t$ -th target coarse clusters, respectively. Weights  $w_{C N_C^s}^s$  for the  $N_C^s$ -th source coarse cluster and  $w_{C N_C^t}^t$  for the  $N_C^t$ -th target coarse cluster are then set to the sum of the weights of the opposite-side coarse clusters excluding the trash box bins; more specifically,  $w_{C N_C^s}^s = \sum_{m=0}^{N_C^s-1} w_{Cm}^s$  and  $w_{C N_C^t}^t = \sum_{l=0}^{N_C^t-1} w_{Cl}^t$ , respectively. Note that this trash box weight setting enables all the clusters to be accommodated by the trash box bins in the transport stage; that is, entirely appearing or disappearing in an extreme case.

The transportation cost and flow (transportation amount) from the  $l$ -th coarse cluster of the source shape to the  $m$ -th coarse cluster of the target shape including the trash box bins are denoted as  $d_{Clm}$  and  $f_{Clm}$  ( $l = 0, \dots, N_C^s, m = 0, \dots, N_C^t$ ), respectively. The transportation cost is then defined as

$$d_{Clm} = \begin{cases} \|\bar{\mathbf{x}}_{Cm}^t - \bar{\mathbf{x}}_{Cl}^s - \bar{\mathbf{v}}\|^2 & (l < N_C^s \text{ and } m < N_C^t) \\ 0 & (l = N_C^s \text{ and } m = N_C^t), \\ d_{Ctb}^2 & \text{otherwise} \end{cases}, \quad (3)$$

where  $\bar{\mathbf{v}}$  is a translation vector between the gravity centers of the source and target point clouds, and  $d_{Ctb}$  is the trash box cost for coarse clusters, which is equivalent to the threshold for appearing or disappearing in the case of occlusion. Therefore, the trash box cost must be set so that it is larger than the maximum Euclidean distance for plausible point cloud transportation, but smaller than the Euclidean distance by incorrect correspondence.

Finally, the earth mover's distance (EMD) flows are optimized so as to minimize the transportation cost by the Hungarian algorithm.

$$\{f_{Clm}^*\} = \arg \min_{\{f_{Clm}\}} \sum_{l=0}^{N_C^s} \sum_{m=0}^{N_C^t} f_{Clm} d_{Clm}, \quad (4)$$

which is subject to  $\sum_{l=0}^{N_C^s} f_{Clm} = w_{Cm}^t \quad \forall m$ ,  $\sum_{m=0}^{N_C^t} f_{Clm} = w_{Cl}^s \quad \forall l$ , and  $f_{Clm} \geq 0 \quad \forall l, m$ . Now, we can regard the obtained  $\{f_{Clm}^*\}$  as the cluster-based many-to-many correspondence weight coefficients between the source and target shapes. Figure 1(d) shows an example of the coarse flows.

## 2.4. Fine correspondence

The third step involves acquiring a fine correspondence. To suppress locally isolated motion, we consider the constraint that the fine clusters' flows are close to the dominant coarse flows. For this purpose, the dominant flow  $v_{Cl}^s$  from the  $l$ -th source coarse cluster is first obtained as

$$v_{Cl}^s = \begin{cases} \bar{x}_{Cm'}^t - \bar{x}_{Cl}^s & (m' < N_C^t) \\ \bar{v} & (m' = N_C^t) \end{cases}, \quad m' = \arg \max_m f_{Cml}^*. \quad (5)$$

Note that a translation vector  $\bar{v}$  between the gravity centers of the source and target point clouds is substituted as above when the flow to the trash box is dominant ( $m' = N_C^t$ ).

The dominant flow  $v_{Fj}^s$  from the  $j$ -th source fine cluster is then interpolated based on the coarse dominant flow  $\{v_{Cl}^s\}$  and membership  $\{m_{Cjl}^s\}$  of the source fine clusters to the coarse clusters is given as

$$v_{Fj}^s = \frac{\sum_{l=0}^{N_C^s-1} m_{Cjl}^s v_{Cl}^s}{\sum_{l=0}^{N_C^s-1} m_{Cjl}^s}. \quad (6)$$

The dominant flow  $v_{Fk}^t$  to the  $k$ -th target fine cluster is obtained in a similar way.

The transportation cost and flow from the  $j$ -th source fine cluster to the  $k$ -th target fine cluster are denoted as  $d_{Fjk}$  and  $f_{Fjk}$  ( $j = 0, \dots, N_F^s, k = 0, \dots, N_F^t$ ), respectively. The total transportation cost is then defined as

$$d_{Fjk} = \begin{cases} \min\{\|\bar{x}_{Fk}^t - \bar{x}_{Fj}^s - v_{Fj}^s\|^2, \\ \|\bar{x}_{Fk}^t - \bar{x}_{Fj}^s - v_{Fk}^t\|^2\} & (j < N_F^s \text{ and } k < N_F^t) \\ 0 & (j = N_F^s \text{ and } k = N_F^t) \\ d_{Ftb}^2 & \text{otherwise} \end{cases} \quad (7)$$

where  $d_{Ftb}$  is the trash box cost for fine clusters. Note that the above transportation cost is defined so that it is small as the correspondence approaches either of the interpolated dominant flows  $v_{Fj}^s$  and  $v_{Fk}^t$  from the source and to the target, respectively. In addition, the trash box cost  $d_{Ftb}$  for fine clusters should be smaller than that for coarse clusters  $d_{Ctb}$  to avoid locally isolated motion.

Finally, the EMD flows are optimized so as to minimize the transportation cost in the same way as for coarse clusters and the obtained flows are denoted as  $\{f_{Fjk}^*\}$ . Figure 1(e) illustrates an example of the fine flows.



(a) Ball handling

(b) Sitting

**Figure 2. Range data. Left and right images show the source and target.**

## 2.5. Transport

Point clouds of the source and target shapes are transported according to a transition rate  $\alpha$ . From a *transport* point of view, an intermediate point cloud is expressed as a set of multiple weighted data points, because there are multiple point clouds that come from both the source and target shapes, and which are derived from multiple memberships of each 3D data point to the clusters as well as many-to-many correspondences as a result of the EMD framework. In fact, the  $i$ -th 3D data point  $x_i^s$  in the source point cloud is transported to  $x_{ijk}^s$  ( $j = 0, \dots, N_F^s - 1$ ) by the flow  $f_{Fjk}^*$  from the  $j$ -th source fine cluster to the  $k$ -th target fine cluster by weight  $w_{ijk}^s$ , expressed as

$$x_{ijk}^s = \begin{cases} x_i^s + \alpha(\bar{x}_{Fk}^t - \bar{x}_{Fj}^s) & (k < N_F^t) \\ x_i^s + \alpha v_{Fj}^s & \text{otherwise} \end{cases} \quad (8)$$

$$w_{ijk}^s = (1 - \alpha) m_{Fij}^s f_{Fjk}^*, \quad (9)$$

The  $i$ -th 3D data point in the target point cloud is transported to  $x_{ijk}^t$  ( $k = 0, \dots, N_F^t - 1$ ) by weight  $w_{ijk}^t$  in the same way, expressed as

$$x_{ijk}^t = \begin{cases} x_i^t - (1 - \alpha)(\bar{x}_{Fk}^t - \bar{x}_{Fj}^s) & (j < N_F^s) \\ x_i^t - (1 - \alpha) v_{Fk}^t & \text{otherwise} \end{cases} \quad (10)$$

$$w_{ijk}^t = \alpha m_{Fij}^t f_{Fjk}^*. \quad (11)$$

Figure 1(f) shows examples of the transition of point cloud transport.

## 3. Experiments

### 3.1. Setup

To evaluate the proposed method, range data for several human actions were captured by a Kinect<sup>TM</sup> range finder, as shown in Fig. 2. Foreground range data were extracted by background subtraction and a point cloud was acquired by projecting the foreground range data into 3D space based on the intrinsic camera parameters, calibrated in advance. The numbers of fine and coarse clusters were set to  $N_F = 200$  and  $N_C = 20$ , respectively. The trash box costs for fine and coarse correspondence were, respectively, set to  $d_{Ftb} = 100$  [mm] and  $d_{Ctb} = 200$  [mm] experimentally. The proposed method was compared with topology-free shape morphing [8] as a benchmark.

### 3.2. Results

Results of the point cloud transport transitions are shown in Fig. 3. Regarding the ball-handling scene, because the face is partially occluded by the ball in the

target shape (green circle in the first row of Fig. 3), the face in the source and the ball in the target partially correspond with each other in the benchmark method, which results in undesirable artifacts (red circles in the first row of Fig. 3).

On the other hand, the proposed method can appropriately handle occlusions as appearing or disappearing processes using the trash box bins and also suppress locally isolated motion with the coarse-to-fine framework. Therefore, the proposed method successfully suppresses such artifacts and realizes more natural point cloud transitions as shown from the second to fourth rows of Fig. 3.

#### 4. Conclusion

This paper described a point cloud transport method for temporal interpolation of range data. To overcome incorrect correspondence due to occlusions, we introduced appearing and disappearing processes using trash box bins. In addition, because locally isolated motion often occurs due to the lack of a constraint for the smoothness of adjacent motion, a coarse-to-fine framework was included in the proposed method. Both coarse and fine correspondences were found so as to minimize the transportation cost in the earth mover's distance framework. The experimental results show that the proposed method works well with occlusions and topological changes.

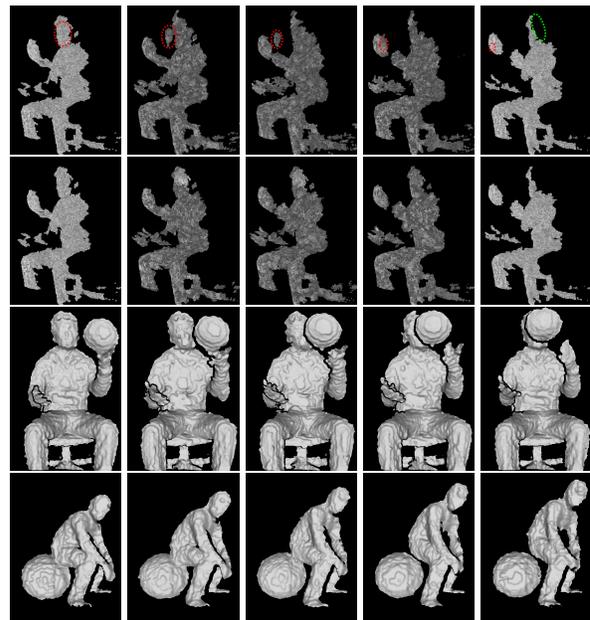
#### Acknowledgment

This work was supported by CREST (Core Research for Evaluation Science and Technology) of JST (Japan Science and Technology).

#### References

- [1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proc. of the 26th Annual Conf. on Computer Graphics and Interactive Techniques*, 1999.
- [2] D. Breen and R. Whitaker. A level-set approach for the metamorphosis of solid models. *IEEE Trans. on Visualization and Computer Graphics*, 7(2):172–192, Apr. 2001.
- [3] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *Int. Journal of Computer Vision*, 64(1):5–30, 2005.
- [4] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. *Numerical Geometry of Non-Rigid Shapes*. Springer, 2008.
- [5] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Topology-invariant similarity of nonrigid shapes. *Int. Journal of Computer Vision*, 81(3):281–301, 2009.
- [6] K. Fujiwara, K. Nishino, J. Takamatsu, B. Zheng, and K. Ikeuchi. Locally rigid globally non-rigid surface registration. In *Proc. the 13th IEEE Int. Conf. on Computer Vision*, pages 1527–1534, Nov. 2011.

- [7] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis*. John Wiley and Sons, 1999.
- [8] Y. Makihara and Y. Yagi. Earth mover's morphing: Topology-free shape morphing using cluster-based emd flows. In *Proc. of the 10th Asian Conf. on Computer Vision*, pages 2302–2315, Queenstown, New Zealand, Nov. 2010.
- [9] D. Miyazaki, T. Ooishi, T. Nishikawa, R. Sagawa, K. Nishino, T. Tomomatsu, Y. Takase, and K. Ikeuchi. The great buddha project: Modelling cultural heritage through observation. In *Proc. the 6th Int. Conf. on Virtual Systems and MultiMedia*, pages 138–145, Gifu, 2000.
- [10] B. Payne and A. Toga. Distance field manipulation of surface models. *IEEE Computer Graphics and Applications*, 12(1):65–71, 1992.
- [11] R. Sagawa, K. Akasaka, Y. Yagi, H. Hamer, and L. V. Gool. Elastic convolved icp for the registration of deformable objects. In *Proc. the 2009 IEEE Int. Workshop on 3-D Digital Imaging and Modeling*, pages 1558–1565, Kyoto, Japan, Oct. 2009.
- [12] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proc. the 22th IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1297–1304, Jun. 2011.



**Figure 3. Results of point cloud transport. Transition rates  $\alpha$  are 0.0, 0.25, 0.5, 0.75, and 1.0 from left to right, respectively. The first and second rows: benchmark method [8] and the proposed method for ball handling. The third and fourth rows: reconstructed surface mesh for ball handling and sitting by the proposed method.**