

Electronic Letters on Computer Vision and Image Analysis 7(4):62-72, 2009

Finding Kinematic Structure in Time Series Volume Data

Tomoyuki Mukasa*, Shohei Nobuhara*, Atsuto Maki* and Takashi Matsuyama*

* *Grad. School of Informatics, Kyoto University, Sakyo-ku Yoshida Honmachi, Kyoto, Japan*

Received 6th June 2007; accepted 25th March 2009

Abstract

This paper presents a new scheme for acquiring 3D kinematic structure and motion from time series volume data. Our basic strategy is to first represent the shape structure of the target in each frame by Reeb graph which we compute by using geodesic distance of target's surface, and then estimate the kinematic structure of the target which is consistent with these shape structures. Although the shape structures can be very different from frame to frame, we propose to derive a unique kinematic structure by way of clustering some nodes of graph, based on the fact that they are partly coherent to a certain extent of time series. Once we acquire a unique kinematic structure, we fit it to other Reeb graphs in the remaining frames, and describe the motion throughout the entire time series. The only assumption we make is that human body can be approximated by an articulated body with certain numbers of end-points and branches. We demonstrate the efficacy of the proposed scheme through some experiments.

Key Words: Vision-Based Motion Capture, Video and Image Sequence Analysis, Reeb graph, Motion Tracking and Analysis.

1 Introduction

The description of moving articulated objects, e.g. human figure, is important in many applications including technical analysis of sports and dance, and production of video contents. Motion-capture system is well-known and available for such purposes. However, it demands special markers and suits, which is sometimes burden to the actor. Moreover, the scope of motion description is limited to the case that the precise structure of the target is given. For these reasons, vision-based human motion capture is an active research area these days [1, 7]. In order to realize a scheme for describing arbitrary moving objects which can be approximated by articulated rigid body without using any special markers and suits, we employ time series volume data as the input which we compute from multiviewpoints videos.

1.1 Shape structure and kinematic structure

Two important notions in this paper are *kinematic structure* and *shape structure*. Let us consider a target object in motion, and assume that we can segment it to several primitives based on the characteristics of motion. The characteristics of the motion are defined as a set of the *motion class* of a target and the motion description under its motion class. Before describing or modeling a target in motion, it is useful to classify its motion. Huang classified motions of rigid/non-rigid objects [14]. In his classification, motions are first divided to rigid motion

Correspondence to: mks@vision.kuee.kyoto-u.ac.jp

Recommended for acceptance by Francisco Perales

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

and non rigid motion. Secondly, non-rigid motion is divided to two classes, general class, which consists of elastic and fluid motion, and restricted class. Restricted class consists of five types of motion, articulated, quasi-rigid, isometric, homothetic, and conformal. In our research, we assume that motion class of target object is "articulated" since target objects are mainly human body. When the target object can be approximated by articulated rigid body, the motion class of each primitive is rigid. What we call kinematic structure is comprised of these segmented parts and their time-invariant connectivity. That is, each segmented part is connected at its end points by joints that are not disconnected.

A target object can also be described by yet another sort of primitives whose representation is based on their mutual distance and their time-variant connectivity. If the target is represented by mesh for example, the mesh topology can be generally changed from frame to frame. We call this structure shape structure, which is typically computed directly from volume data.

Since the structure that is directly available by observation is normally shape structure, we can define our problem domain as that of interpreting a shape structure into combinations of a kinematic structure and its joints' motion.

1.2 Global description of shape structure

Conventional approaches for acquisition of kinematic structure from volume data are either top-down or bottom-up. Top-down approach uses a specific model, cylinder-model for example, to match with volume data. It is impossible to acquire a description of non-rigid motion by this approach. Moreover, in this approach, we need employ a somewhat elaborate model for each observation target. On the other hand, there are several bottom-up methods which acquire kinematic structure from observed motion without specific model [4, 5, 6].

In these methods, the unit of motion description is voxel or vertex on surface, and its motion is assumed relatively small compared to that of the whole body. Therefore, it is difficult to discriminate the difference between the motion that is derived from global rigid motion and the small scale non-rigid motion on the surface that is caused for example by body vibration. Hence, it is hard to describe the whole body by a kinematic structure in these approaches.

In order to overcome this difficulty, we design a global description of shape structure so as to eliminate the effects of small scale motion, and acquire an effective representation of kinematic structure accordingly.

1.3 Coping with topological variations of shape structure

Given a time series volume data of an articulated body in motion as an input sequence of shape structure, the topology of the shape structure naturally changes before and after touching surfaces of body segments (primitive segments of body). Thus, global topology of shape structure and kinematic structure will remain identical only if there is no contact between any body segments. In this respect, two types of time intervals can be generally considered in a given time series data; in one type of time intervals the shape structure and the kinematic structure have the same global topology, and it is not the case in the other type of time intervals. In the former, which we call *topologically coherent intervals*, we can acquire kinematic structure by editing the shape structure so that it has constant number (and sufficient number to describe the motion) of primitives. However, in the latter type of intervals which we call *topologically in-coherent intervals*, the same approach does not apply and it is the very issue that we are going to deal with in the sequel.

Our basic strategy to cope with the topological variations of shape structure is to first separate the entire time series data to topologically coherent/incoherent intervals, and acquire a kinematic structure only from the topologically coherent intervals in a bottom-up manners, We then describe the entire time series by using the acquired kinematic structure, in a top-down fashion.

2 Algorithm

We acquire the kinematic structure through an off-line process as shown below.

1. Acquire time series visual hull [2] from multiviewpoints video.
2. Obtain surface mesh by applying marching cubes method [13] to the visual hull acquired in each frame.
3. Throughout the sequence construct Reeb graphs [8, 12, 10, 11] based on geodesic distance on the surface.
4. Separate the time series of Reeb graphs to different time intervals which are topologically either coherent or incoherent.
5. For each topologically coherent interval, cluster some nodes of Reeb graph based on their motion, and acquire a kinematic structure. We call each of the structures as "piecewise kinematic structure".
6. Integrate the piecewise kinematic structures.
7. Describe entire time series by the integrated kinematic structure.

(1),(2),(3) are processes for acquiring a global shape structure. (4) is a process for finding intervals in which we can make correlations between shape structures and kinematic structures. This process enables us to acquire the kinematic structure in process (5). In process (6) we acquire a kinematic structure that reflects the diversity of motion in the input visual hull. Finally, in process (7), we acquire sets of a kinematic structure and its motion for the entire time series.

We explain the details of process (3) to (7) in the following sections.

2.1 Pseudo Endoskeleton Reeb Graph

We employ the Reeb graph [8, 12] for the global shape description. The first step to construct a Reeb graph is segmenting surface S of an object on the basis of a continuous function $\mu(\mathbf{v})$ (\mathbf{v} is an arbitrary vertex on the surface S), defined on the surface. The second step is representing each segmented surface by a node. And the final step is linking the nodes based on the connectivity between the segmented surfaces. Now input sequence we use is time series mesh, that are discrete surfaces. Therefore we use geodesic function extended to discrete surfaces [9] as $\mu(\mathbf{v})$. Now, the surface of the object is represented by a mesh model. The geodesic distance $g(\mathbf{v}, \mathbf{p})$ is defined by shortest path between \mathbf{v} and another vertex \mathbf{p} on the mesh. Then, function $\mu(\mathbf{v})$ can be represented as

$$\mu(\mathbf{v}) = \sum_{\mathbf{p} \in S} g(\mathbf{v}, \mathbf{p}). \quad (1)$$

This indicates the sum of distance from vertex \mathbf{v} to all vertices on S .

Here, we revise the acquired Reeb graphs to be used as a prototype of kinematic structure. As we explained in section 1.1, kinematic structure is comprised of primitives that can be roughly approximated by cylindrical objects. The cylindrical primitives are connected to each other at the ends along their axial direction, and the entire volume corresponding to each primitive follows the edge, i.e. its axis, when the kinematic structure moves. The motion of kinematic structure can be therefore represented by motions of these edges that are constrained by their connectivity. On the contrary, nodes and edges in a Reeb graph based on geodesic distance are not always inside the cylinders. This feature is not appropriate for representing connectivity of articulated primitives in a kinematic structure. For this reason, we revise Reeb graphs in which all edges and nodes are constrained to be inside the mesh by using multiresolution Reeb graph method [10, 11]. Multiresolution Reeb graph is a hierarchical collection of the Reeb graphs which are obtained at different resolutions. Each node of a low-resolution Reeb graph is associated to a set of nodes of Reeb graph in a higher-resolution. Using

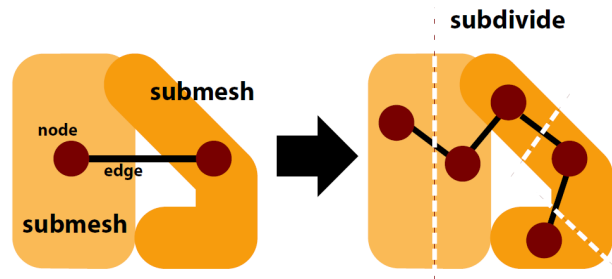


Figure 1: Generation of pERG.

this attribute, when a part of graph is outside of mesh, we replace the part with its associated part in a higher-resolution. If the highest-resolution is good enough, by repeating this replacement recursively we can acquire a revise Reeb graph in which all the edges and nodes are inside the mesh (Figure 1). We call this structure pseudo Endoskeleton Reeb graph (pERG).

2.2 Setting topologically coherent intervals

The pERG is stable when there are no contact between any body segments. However, while observing an identical object, the graph structure may easily vary due to a contact. Considering this problem, before acquiring a kinematic structure, we define topologically coherent intervals during which the coherence between the graph structure and the kinematic structure do not break. The term "topology" here indicates the numbers of end points and branch points, and their connectivity in a graph.

In this research we assume the target to be human or the like whose structure can be roughly approximated by an articulated rigid body. Therefore, the pERG in topologically coherent intervals should have just an appropriate number of branches but no loop. These properties are equivalent to the following conditions of graph structure.

- (i) "The graph has five end-points".
- (ii) "The graph has two branch-points, one has four branches, and the other has three branches", or "The graph has three branch-points each of which has three branches".

The latter is due to the fact that the pERG is not necessarily symmetric.

2.3 Acquiring piecewise kinematic structures

The process of acquiring a piecewise kinematic structure is summarized in two steps:

1. Adjust the number of graph's nodes in each topologically coherent intervals.
2. Cluster the nodes based on its motion in each topologically coherent intervals.

In process 1, we select an initial frame in each topologically coherent interval, and deform the pERG (we call it initial model) to the graphs in neighbouring frames. After this process, the number of graph nodes are kept constant in each interval, and thus we can define correspondence of nodes between different frames, which is useful for clustering the nodes in process 2.

The selection of initial frame is carried out by similarity of pERGs in each interval so that the pERG in the initial frame has the highest score in terms of the sum of correlations of all pERGs within the interval. For computing the similarity, we use augmented multiresolution Reeb graph method [11]. This method allows us to use various attributes to calculate the similarity, including shape and texture of a mesh segment and balk of volume, which are both associated to each node.

2.3.1 Deforming pERGs

We deform initial models so as to deal with the changes in the number of nodes of pERG.

In order to deform an initial model and fit it to other pERGs, we shift the nodes of initial model to the positions where the nodes of other pERG are, while maintaining connective relations of the nodes of initial model.

We take the following constraints into account for the movement of nodes.

- (a) Maintain the distance between nodes as much as possible.
- (b) Move a node of initial model to its nearest neighbor node in the target pERG.

We enforce these constraints to work on each node. (a) can be realized by elastic force between nodes (we call the force as internal force), and (b) by the external force to move these nodes. Given these forces working at each node, we can compute how initial model deforms its shape by solving Newtonian equation with backward Euler integration method.

As detailed later, it should be noted that on every deformation we select a pERG which has the highest correlation to the target pERG as a new initial model, to avoid the need to correlate a node to a very distant position.

External force : The nodes of the initial model should be moved to the neighborhood of the nearest nodes of the target pERG without any crossing in pathways of nodes' movement. We define the external force which urges such nodes' movement based on the point-set deformation algorithm [3].

We denote the pERG at time a and b by N and M respectively. First, we find nearest neighbor nodes in M for each node in N , and make a pair (m_a, m_b) . We also make another pair (n_a, n_b) for each node in M in a similar way. Note that (m_a, m_b) and (n_a, n_b) are not always identical generally. Now, we denote the distances from an arbitrary point $p(x, y)$ to node n_a and m_a as $d_a(n)$ and $d_b(m)$, respectively. We also denote the drift force of node n_a to n_b , node m_a to m_b by $f_a(n)$ and $f_b(m)$, respectively. Then, we can define the external force $f_{external}(x, y)$ working on $p(x, y)$ as

$$f_{external}(x, y) \equiv \frac{\sum_n \alpha_n f_a(n)}{\sum_n \alpha_n} + \frac{\sum_m \beta_m f_b(m)}{\sum_m \beta_m}, \quad (2)$$

$$\alpha_n = \exp\left(\frac{-d_a(n)}{\sigma}\right), \quad \beta_m = \exp\left(\frac{-d_b(m)}{\sigma}\right). \quad (3)$$

Internal force : We introduce the internal force, $F_{internal}(v)$, as

$$F_{internal}(v) \equiv \sum_1^n k_j (\|\mathbf{q}_{v_j} - \mathbf{q}_v\|) \frac{\mathbf{q}_{v_j} - \mathbf{q}}{\|\mathbf{q}_{v_j} - \mathbf{q}\|} - k'_j \mathbf{q}_v, \quad (4)$$

where k_j denotes a constant, v_j a node connected to v within 2 hops, \mathbf{q}_{v_j} its 3D position, and n the number of v_j , respectively. The internal force works so as to maintain the smoothness of pERG's shape.

Updating Initial Model Based on Correlation of pERG : The temporal sampling rate of the input volume data can often be too low to follow the rapid motion of the target object. That is, it is not always true that temporally neighboring graphs are strongly-correlated.

On every deformation, we select a frame in which the pERG has the highest correlation to the target pERG among the frames in which the pERG has already deformed, and use its deformed pERG as a new initial model. We utilize a correlation computing method of pERGs that is proposed in [11]. Acquisition of Kinematic Structure Based on the Motions of Nodes For each topologically coherent interval, we acquire kinematic structure by clustering nodes of pERG based on cross-correlation of motion between neighboring nodes (see Figure 2).

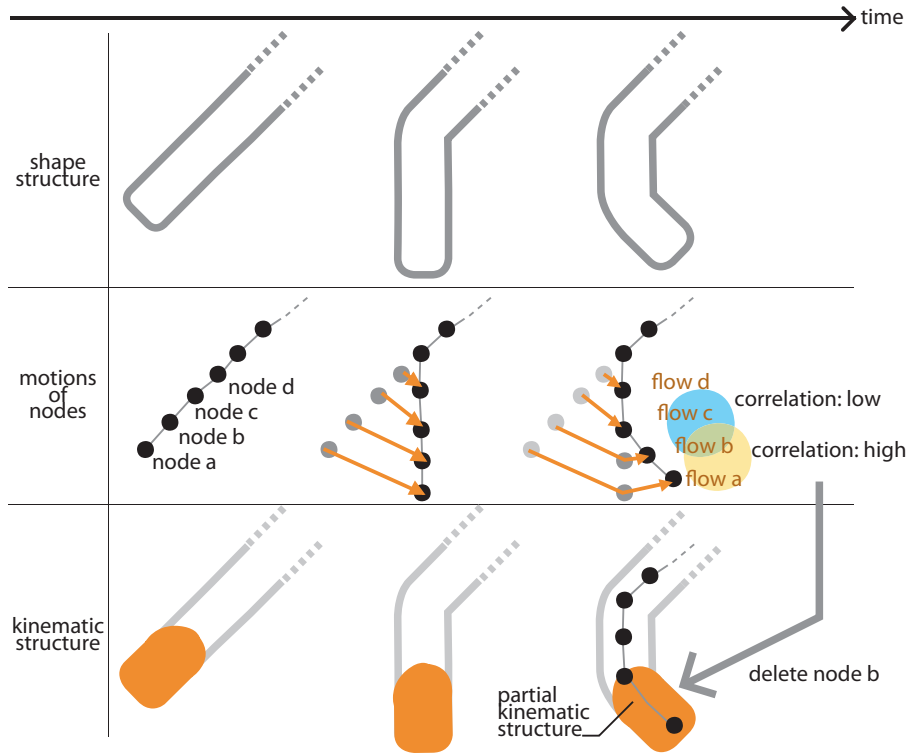


Figure 2: Clustering of nodes based on their motion

We call these kinematic structures piecewise kinematic structures. We denote the number of nodes in the initial model by n , the velocity of node p_i , ($i = 1, 2, \dots, n$) at time t by \mathbf{v}_t^i , and the cross-correlation function of nodes p_i, p_j at time interval $[t_b, t_e]$ as $r(i, j)$ can then be represented as

$$r(i, j) = \frac{\sum_{k=t_b}^{t_e} (\mathbf{v}_k^i - \mathbf{v}_{mean}^i)(\mathbf{v}_k^j - \mathbf{v}_{mean}^j)}{\sqrt{\sum_{k=t_b}^{t_e} (\mathbf{v}_k^i - \mathbf{v}_{mean}^i)^2} \sqrt{\sum_{k=t_b}^{t_e} (\mathbf{v}_k^j - \mathbf{v}_{mean}^j)^2}}, \quad (5)$$

where

$$\mathbf{v}_{mean}^i = \frac{\sum_{t=t_b}^{t_e} \mathbf{v}_t^i}{t_e - t_b}, \quad \mathbf{v}_{mean}^j = \frac{\sum_{t=t_b}^{t_e} \mathbf{v}_t^j}{t_e - t_b}. \quad (6)$$

Regarding a set of neighboring two nodes, the higher the cross-correlation is, the more probable that both of the nodes belong to the same rigid body.

2.4 Integration of Piecewise Kinematic Structure

After acquiring piecewise kinematic structures, we choose the most detailed partial kinematic structure for each branch, corresponding to arm, leg, or head among the piecewise kinematic structures. Finally, we deform the partial kinematic structures so as to fit them all over the time series pERGs, and integrate them (see Figure 3).

The integrated kinematic structure reflects various patterns of kinematic motion in the input volume data.

2.5 Motion description of entire time series data

Once we acquire the integrated kinematic structure through above processes, we match it to pERGs in topologically incoherent intervals, and describe the target's motion in the entire time series. This process may be

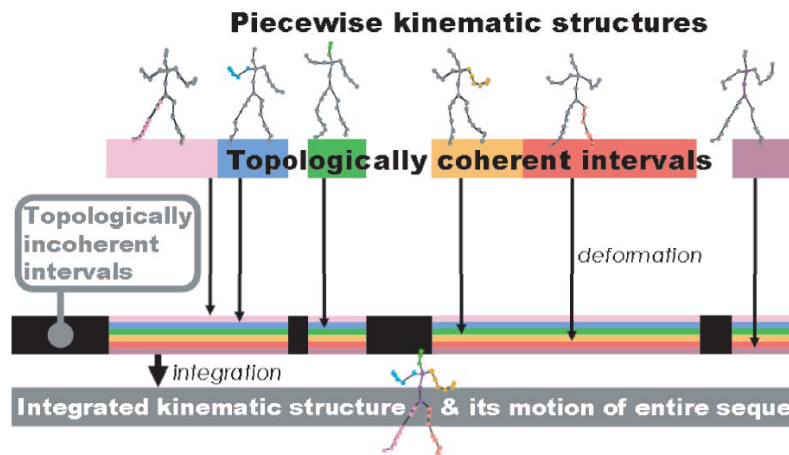


Figure 3: Integration of piecewise kinematic structures

regarded as a kind of simplified top-down approach in which a certain specific model is matched to the volume model.

In this process, we deform the integrated kinematic structure by shifting each node in the structure while maintaining the connectivity of its nodes. This is also implemented by the method explained in section 2.3.

2.6 Feasibility and limitation

In our method, target object is modeled by an articulated rigid body which is comprised of cylindrical objects that are connected with each other at the ends. The possibilities of contact of these cylindrical objects can be classified into three types. They are:

- Two cylindrical objects touch one another at their ends.
- An end of one cylindrical object touches the other's surface which is parallel to the axis.
- Two cylindrical objects touch one another on their side surfaces which are parallel to each of their axes, respectively.

In process 7, which we described in the previous section, we assume that once a contact occurs the touching parts of the body retain the way they touch. The proposed method properly works as long as there is no contact in the pERG, or the touching type remains the same even if it occurs.

However, it is not guaranteed that our method correctly works when the touching type changes, or areas of contact changes while maintaining their touching of type 3. Therefore, we consider such case as an exception that an observed person gets his or her arms tangled after an touching occurred.

3 Experiments

3.1 CG data

We generate CG scene of a walking human by applying a motion capture data * to a CG model (See Figure 4). We then shot the scene by 15 virtual cameras.

circumnavigating the CG object, and acquired its time series voxel data using a technique based on visual hull. Finally, we acquired kinematic structure from the time series visual hull by the proposed method.

*ICS Action Database, Copyright (C) ICS, The University of Tokyo, 2003-2004 (<http://www.ics.t.u-tokyo.ac.jp/>)

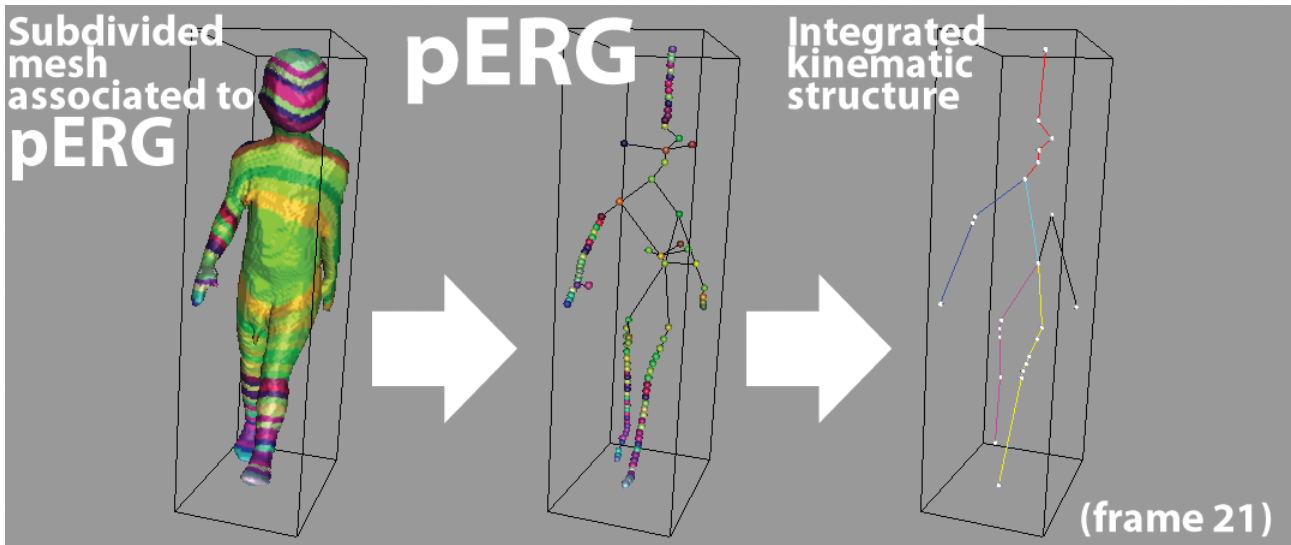


Figure 4: Acquisition process of the integrated kinematic structure.

We evaluated the acquired kinematic structure by comparing it with the original motion capture data. In this evaluation, as a criterion, we computed detection rate of joint positions. Given a pair of correctly detected node and its correlated joint position in motion data, they should meet the following conditions.

- The joint position is closest to the node.
- The node is also closest to the joint position.

We denote the motion capture data by M , acquired kinematic structure by N , joint points in structures M and N by m_i , ($i = 0, 1, \dots, i_{max}$) and n_j , ($j = 0, 1, \dots, j_{max}$), respectively. We denote the distance between m_i and n_j by $dist(m_i, n_j)$, and also denote the distance from a particular joint position in M to a node in N by $dist_{m_i}(n_j)$. When node n_j is a correctly detected node, n_j and its pair m_i meet the formula,

$$dist_{m_i}(n_j) = dist_{n_j}(m_i) = \min\{dist(m_i, n_j)\}. \quad (7)$$

Under these definitions, the detection rate, $R_{detection}$, can be defined as

$$R_{detection} = \frac{S}{i_{max}} \quad (8)$$

where S is the number of the pair of node m_i and joint position n_j which meet equation 7.

We also computed the mean value of distance between joint positions in the motion capture data to their nearest neighbor nodes in the acquired kinematic structure.

By using this, we can define above described criterion E_{dist} by

$$E_{dist} \equiv \frac{\sum_i dist(m_i, n_{i'})}{i_{max} + 1} \quad (9)$$

Figure 5 shows the evaluated values, $R_{detection}$ and E_{dist} throughout the input frames. Both $R_{detection}$ and E_{dist} are normalized by the target's height. The mean value of estimation error, E_{dist} , stays less than 0.075 within a small range of deviation. showing that the kinematic structure is stably acquired by the proposed method. On the other hand, there is a room for improvement in terms of estimation accuracy for practical use. For instance, the detection rate, $R_{detection}$, hovers around 60 percent, which can be ideally increased in the same framework.

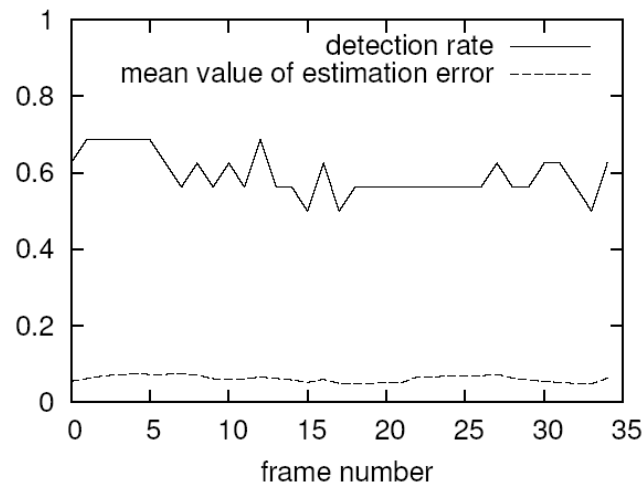


Figure 5: Detection rate ($R_{detection}$) and mean value of estimation error (E_{dist}).

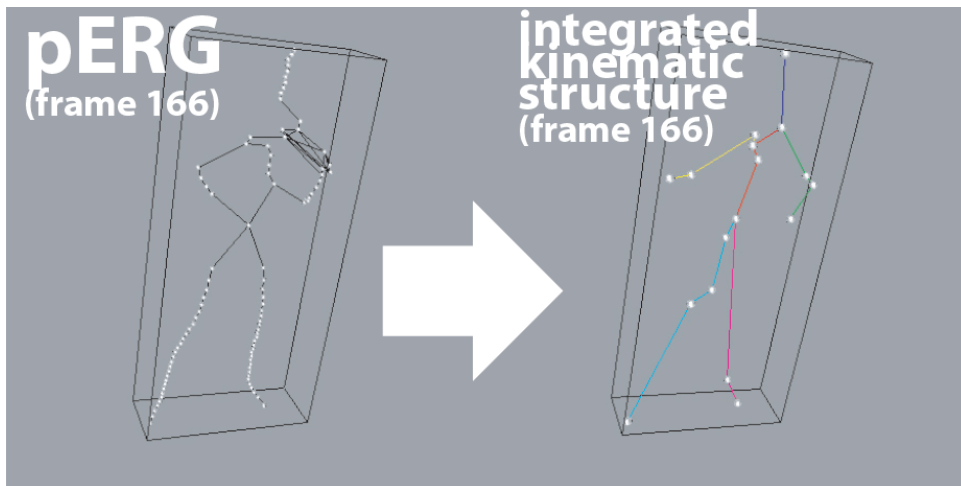


Figure 6: Modification of pERG at frame 166

3.2 Real data

We used nine cameras surrounding the target object (a dancing woman), and acquired time series visual hull from multiviewpoint videos. Applying the proposed method, we acquired the kinematic structure. Applying the proposed method, we acquired the kinematic structure from 330 frames of time series visual hull. This acquisition process is exemplified at frame 166 as in Figure 6 as an example. In this frame, the partial graph associated for the left hand touches the body, making a loop. On the contrary, the loop is removed in the acquired kinematic structure, and the placement of its end point is roughly correct. However, there is an obvious posture estimation error in the partial kinematic structure that is associated with the right hand. One of the reason causing this error was the simplicity of the graph-based model fitting in the last process, in which we shifted nodes of the initial model mainly on the basis of their Euclidean distance. Namely, it is because we cannot fully guarantee the edges to be inside the original mesh after the shifting. Another reason for the error was the noise in 3D surface reconstruction.

The key to overcome these defects, we will be to device the volumetric fitting method in this last process of our algorithm.

4 Conclusions

In this paper, we proposed a scheme for acquiring kinematic structure from time-series voxel data using the pERG for the global shape description. For the case of human figure in motion, we have presented our early results that show the stable characteristic of the proposed scheme.

However, it is not guaranteed that the kinematic structure which reflects exact structure of the observed object can be always acquired. The proposed algorithm assumes that at least one topologically coherent interval is contained in entire time series. If no such intervals are observed, our algorithm would fail. However, even in some topologically incoherent intervals, certain body parts could present useful motion which is not observed in topologically coherent intervals. In this case, it will be more appropriate to acquire partial kinematic structures for this kind of body parts in incoherent intervals. For these reasons, there are rooms for improvement in term of accuracy.

Therefore, our future work will incorporate the following concepts.

- Acquire each partial kinematic structure in an interval which shows its characteristic motion, and integrate them.
- Revise the fitting process from the current graph based method to the volumetric method with born and skin model.

Finally, since the proposed method works on time series voxel data, we plan to expand the scope of the scheme to deal with non-rigid body in motion.

Acknowledgements

This work has been supported by "Young leader-training program of Kyoto University Global COE Program, Informatics Education and Research Center for Knowledge-Circulating Society" and "Foundation of Technology Supporting the Creation of Digital Media Contents" project (CREST, JST). We would like to thank Takeshi Takai for his valuable support of the experiments.

References

- [1] A. Hilton, "A Survey of Advances in Vision-Based Human Motion Capture and Analysis", *International Journal on Computer Vision and Image Understanding*, Vol.104, No.2-3 :90-127, 2006.
- [2] A. Laurentini, "How Far 3D Shapes Can Be Understood from 2D Silhouettes" *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.17, No.2 :188-195, 1995.
- [3] D. J. Burr, "A Dynamic Model for Image Registration", *Computer Graphics and Image Processing* 15, :102-112, 1981.
- [4] K. M. Cheung, S. Baker and T. Kanade "Shape-From-Silhouette Across Time: PartII: Applications to Human Modeling and Markerless Motion Tracking" *International Journal of Computer Vision*, Vol. 63, No. 3 :225-245, 2005.
- [5] S. Nobuhara, T. Matsuyama, "Heterogeneous Deformation Model for 3D Shape and Motion Recovery from Multi-Viewpoint Images", 2nd International Symposium on *3D Data Processing, Visualization, and Transmission* :566-573, 2004.
- [6] I. Masaaki, K. Yoshinari, M. Michihiko, "Estimation of the Location of Joint Points of Human Body from Successive Volume Data" *ICPR2000*, :699-702, 2002.

- [7] T. Mukasa, S. Nobuhara, A. Maki, and T. Matsuyama: "Finding Articulated Body in Time-series Volume Data", *AMDO 2006*, :395-404, 2006.
- [8] G. Reeb, "Sur les point singuliers d'une forme de Pfaff completement integrable ou d'une fonction numerique" *Comptes Rendus Acad. Sciences Paris*, vol. 222, :847-849, 1946.
- [9] S. Biasotti, B. Falcidieno, and M. Spagnuolo, "Extende reeb graphs for surface understanding and classification", *Conf, Discrete Geometry for Computer Imagery*, :185-197, 2000.
- [10] M. Hilaga, Y. Shinagawa, T. Kohmura, T.L. Kunii, "Topology Matching for Fully Automatic Similarity Estimation of 3D Shapes", *Proc. of SIGGRAPH 2001*, :203-212, 2001.
- [11] T. Tung, F. Schmit, "The augmented multiresolution reeb graph approach for content-based retrieval of 3D shapes", *International Journal of Shape Modeling (IJSM)*, Vol.11, No.1, :91-120, June 2005.
- [12] N. Werghi, Y. Xiao, and J. P. Siebert, "A Functional-Based Segmentation of Human Body Scans in Arbitrary Postures" *IEEE Transactions on Systems, Man and Cybernetics: PartB*, Vol. 36, No.1 :153-165, 2006.
- [13] Y. Kenmochi, K. Kotani, A. Imiya, "Marching Cubes Method with Connectivity", *Proc. of International Conference on Image Processing*, :361-365, 1999.
- [14] T. S. Huang, "Modeling, Analysis, and Visualization of nonrigid object motion" *Proc. 10th ICPR*, :361-364, 1990.