3D Skeleton Extraction from Volume Data Based on Normalized Gradient Vector Flow

Sang Min Yoon
GRIS, TU Darmstadt
Rundeturmstrasse 10
Darmstadt
64283 Germany
Sangmin.yoon@zgdv.de

Cornelius Malerczyk
ZGDV
Rundeturmstrasse 10
Darmstadt
64283 Germany
Cornelius.malerczyk@zgdv.de

Holger Graf
ZGDV
Rundeturmstrasse 10
Darmstadt
64283 Germany
Holger.graf@zgdv.de

ABSTRACT

Skeleton extraction and visualization of 3D reconstructed target objects from multiple views continues to be a major challenge in terms of providing intuitive and uncluttered images that allow the users to understand their data. This paper presents a three-dimensional skeleton extraction technique of deformable objects based on a normalized gradient vector flow in order to analyze and visualize its characteristics. 3D deformable objects are reconstructed by an image based visual hull technique from known extrinsic and intrinsic camera parameters and silhouettes which are extracted from each camera. Our 3D skeleton extraction methodology employs the normalized gradient vector flow which is a vector diffusion approach based on partial differential equations. The euclidean distance of the magnitude of a normalized gradient vector flow is used to extract the medial axis of volume data. A markerless 3D skeletonization of reconstructed objects from multiple images might be applied to retrieve the 3D model or correct the 3D motion of the target objects.

Keywords

3D Skeleton extraction, 3D reconstruction, and Normalized Gradient Vector Flow(NGVF)

1. INTRODUCTION

The acquisition of three-dimensional real world objects from a set of input images is an important topic in computer graphics as well as computer vision. Most techniques that have been developed during last two decades have focused on how to visualize and render the 3D motion of deformable objects in an arbitrary viewpoint. The most common representations for such objects are boundary meshes or point-sets. However, applications such as editing, animation, morphing or shape matching often need a higher level understanding of the shape and its structure. Such an understanding can be conveyed through the use of a skeleton representation of the object because a representation of deformable objects efficiently show their characteristics based on low level data.

Traditional 3D motion capture and skeleton extraction systems are mainly based on two approaches. One is to attach many sensors to the joints of a target object, and the other is to analyze a video sequences by feature detection, searching correspondence between the features from multiple views, recovering the 3D skeleton extraction and connection from feature correspondences. Especially, human skeleton extraction and sensor based motion capture systems are already widespread within comprehensive applications for the analysis of users' performances, medical diagnosis, surveillance, and 3D model retrieval systems. However, sensor based 3D skeleton extraction has many constraints in terms of user mobility and experimental environment even if it is robust and fast to understand thee 3D skeleton of target objects. On the other hand, markerless 3D skeleton extraction gives users convenience in moving, but it is difficult due to the fact that the quality of a 3D skeleton is dependent on the methodology of how to reconstruct the target object and whether the reconstructed objects include complex local topology, large missing data, and noise. This in turn requires a robust and accurate interpretation process.

In this paper, we propose a simple and efficient skeletonization algorithm, which employs image-

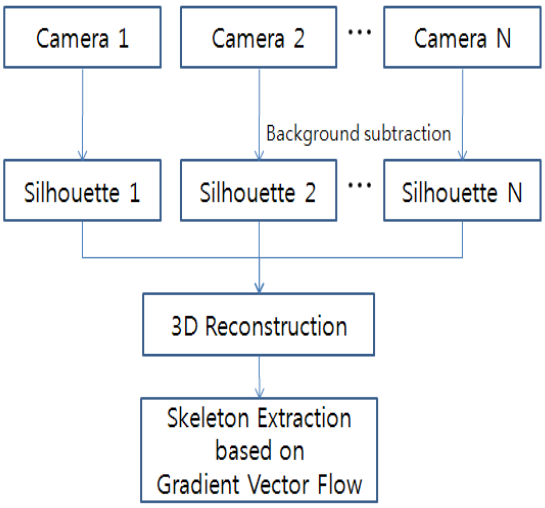


Figure 1. Overview of our proposed system

based 3D reconstruction techniques and extracts a 3D skeleton based on a normalized gradient vector flow technique, a vector diffusion approach based on partial differential equations. In order to be able to extract the 3D skeleton we need to know the cameras extrinsic and intrinsic parameters of the multiple camera alignment. Figure 1 shows the overview of our proposed methodology to extract the 3D skeleton of deformable volume data with normalized gradient vector flow from multiple images, silhouette extraction using background subtraction followed by the 3D reconstruction.

Configuration of the paper

This paper is organized as follows: Section 2 describes the previous research about 3D reconstruction from multiple images, background subtraction to extract the silhouette of deformable objects and 3D curve-skeleton extraction techniques. Section 3 explains the 3D reconstruction of target objects via image based visual hulls using the silhouettes which are extracted by background subtraction from multiple camera images. Such a reconstruction utilizes the known intrinsic and extrinsic camera parameters. Section 4 describes methods for the computation of a 3D gradient vector flow and the skeleton extraction of deformable objects from 3D volume data. In section 5, we will show our experimental results that qualify the performance of the proposed approach. Finally, we conclude and discuss our methodology in section 6.

2. PREVIOUS RESEARCH

Multi-view 3D Reconstruction

The topic of 3D scene reconstruction of deformable objects based on multiple images has been investigated during the last 20 years and produced numerous results in the area of computer graphics and computer vision. Especially, real-time 3D reconstruction of target objects within a GPU environment has been one of hot issues nowadays. The 3D reconstruction research starts from a stereo vision based reconstruction [MP79]. Okutomi et al. [OK93] extended the conventional two-view stereo reconstruction into a multiple camera environment. Kang et al. [KSC01] developed a method of multi-view stereo reconstruction from images to overcome the large occlusions. These methods are designed to reconstruct depth maps from particular viewpoints. Hence, they are not suitable for a full 3D scene reconstruction from images obtained from multiple surrounding cameras.

Image based visual hull reconstruction [MBG00] is a real-time 3D scene reconstruction technique from multiple view images. The algorithm does not need to solve the corresponding problem. Instead, it simply calculates the convex hull of silhouettes in all view images. While the visual hull method works robustly when cameras surround the object, a concave object cannot be reconstructed using the silhouette alone. This problem was solved by a voxel coloring method presented by Seitz et al. [SD97].

Background Subtraction

The principle of a background subtraction is to detect moving objects by building the difference between the current frame and a reference frame. A comprehensive overview and indepth literature review on background subtraction techniques can be found in Picacardi et al. [Pic04]. Several methods for performing background subtraction try to effectively estimate the background model from temporally trained sequences of images. Wren et al. [WAD97] has proposed to model the background independently at each pixel which is based on a Gaussian probability density function. Stauffer et al. [SG99] extended the uni-modal background subtraction approach by using an adaptive multi-modal subtraction method that modelled the pixel color as a mixture of Gaussians. Oliver et al. [ORP00] used an eigen-space model for background subtraction. Recent techniques which combine multiple cues such as color and depth maps are also used for video surveillance and monitoring system [BLL03].

3D Skeleton extraction

3D skeleton extraction can be largely classified into three categories according to [CSYB05]: voxel topology, computational geometry, and continuous implicit. The computation of skeleton extraction by voxel topology is derived by topological thinning [GS99] through iteratively removing its simple points from the boundary of a voxel set. The medial axis of a 3D shape by geometry is extracted using its own distance field [WML03] or a refined geodesic

(a) Input color images from multi-views (b) Extracted silhouettes by a kernel density based background subtraction technique

Figure 2. Input color images and extracted silhouettes of a target object from multi-view

field [DS06]. Implicit technique compute the skeleton from the ridge points of 3D fields such as fast marching [ZT99] or active contours [GG00].

3. 3D RECONSTRUCTION OF DEFORMABLE OBJECTS

In this section, we will explain how we extract the silhouette of target objects based on background subtraction and reconstruct the deformable object from the assumption that we know the intrinsic and extrinsic camera parameters by camera calibration. We first extract the silhouette of the target object with background subtraction technique.

Kernel Density Estimation based Silhouette Extraction

There exist many approaches to extract and segment the target objects with the lowest possible false alarm rates. Background subtraction is a method typically used to detect the deformable objects in the scene by comparing each new frame to a model of the scene background. We use a non-parametric technique for background modeling and foreground extraction. Our approach is based on kernel density estimation of the probability density function of the intensity of each pixel within each image. Kernel density estimation based background modeling aims at capturing and storing recent information about the image sequence, continuously updating this information in order to capture fast changes in the scene background [HCD04]. The intensity distribution of a pixel can change quickly. So we can estimate the density function of this distribution at any moment of time given only very recent history information if we want to obtain a sensitive detection. Using the recent pixel information, the probability density function of each pixel will have intensity value $I(x,y)$ at time t and can be non-parametric estimated using the kernel, K as

$$pdf\ (I_t) = \frac{1}{N}\sum_{i=1}^{N} K\ (I_t - I_i) \qquad (1)$$

where N is the recent pixel information for comparing the current image's pixel information. If we choose our kernel estimation function to be a Gaussain kernel for color image, then the density ca be estimated as

$$pdf\ (I_t) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{3}\frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1(I_{ij}-I_{ij})^2}{2\sigma_j^2}} \qquad (2)$$

where j is number of channel and σ is the standard deviation of Gaussain kernel. The foreground area of an image is segmented by an adequate threshold of equation (2).

Figure 2 shows the input color images and extracted target object. The origin of the world coordinate system which is defined by camera calibration is also displayed by a red(X), green(Y), and blue(Z) line. Thus, in order to extract the silhouette of our target objects, we use this kernel density estimation based background subtraction technique.

3D Reconstruction with Image-based Visual Hulls

In this section, we explain how we reconstruct the target object from multiple images. The image based visual hull methodology [MBG00] is usually computed with respect to a finite number of silhouettes. The image-based visual hull is defined by the camera's intrinsic and extrinsic parameters and silhouettes from each view. Generally, it is the

maximal volume whose projections onto multiple image planes result in a set of observed silhouettes of an object. One efficient technique for generating the 3D reconstructed object by a visual hull computes the intersection of the viewing ray from each designated viewpoint with each pixel in that viewpoint's image.

In order to reconstruct the visual hull surface the first intersection point of the ray traversing the box with the visual hull must be found. A point on the ray is in the visual hull if its projection lies within the silhouette in all view images. A simple approach to this problem is a ray matching algorithm: The ray is sampled at regular intervals and each resulting point is projected onto all views using the camera calibration data. The necessary small steps for a good approximation of the surface yield to a high processing cost and a bad performance.

Figure 3 displays the 3D reconstructed target object by image-based visual hulls in an arbitrary viewpoint.

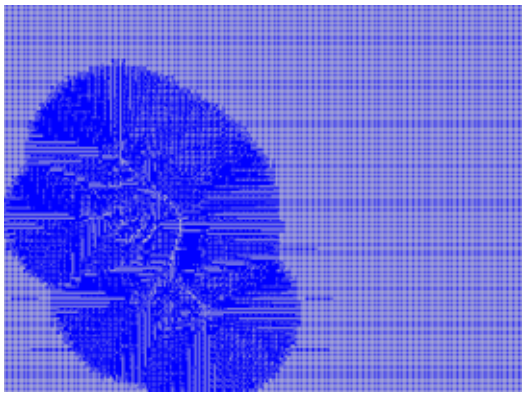
Figure 3. 3D Reconstructed target object and rendering in an arbitrary viewpoint

4. 3D SKELETON EXTRACTION BASED ON GRADIENT VECTOR FLOW

Gradient Vector flow (GVF) [XP98] begins defining the edge map of volume data as $f(x, y, z)$ derived from the original volume data. The edge map should have the property that $f(x, y, z)$ is large near the image boundaries and small within the homogeneous regions. The edge map of the original volume data is defined as

$$f(x, y, z) = -\| \nabla I(x, y, z) \|^2 \quad (3)$$

The basic premise of the energy minimizing formulation of deformable objects is to find a parameterized curve that minimizes the weighted sum of energy.



(a) 3D NGVF of reconstructed objects

(b) Skeleton extraction of a deformable object based on a 3D reconstruction

Figure 4. Normalized gradient vector flow of the reconstructed object and its skeleton from equation (5)

The GVF is the vector flow $V(x)$ that minimizes the following functional,

$$E(V) = \iiint \mu |\nabla V|^2 + |\nabla f|^2 |V - \nabla f|^2 \, dx \quad (4)$$

where $x = (x, y, z)$, μ is a regularization parameter. This variational formula consists of two terms. The first term, the sum of the squares of the partial derivatives of the vector field, makes the resulting vector flow smoothly. The second term stands for the difference between the vector flow and its initial status. Thus minimizing this energy will force $V(x)$ nearly equal to the gradient of the edge map where $\| \nabla f(x, y, z) \|$ is large.

The typical GVF methods cannot efficiently to extract the medial axis when a weak vector makes a very little impact on its neighbors that have much stronger magnitudes.

The normalized gradient vector flow technique (NGVF) [YB02] can tremendously affect a strong vector, both on its magnitude and on its orientation. One of the important properties of the $\| V(x) \|$ over

the Euclidean distance is that it does not form medial surfaces for 3D objects because only one boundary voxel contributes to the computation of distance [HF07]. The 3D skeleton is extracted from the medialness whose strength is controlled by the field strength. q

$$\lambda(x,y,z) = 1 - \left(\frac{|V(x,y,z)| - \min|V|}{\max|V| - \min|V|}\right)^{q}, 0 > q > 1 \quad (5)$$

Figure 4 shows the extracted NGVF from the reconstructed object and the extracted skeleton from medialness function of equation (5).

5. EXPERIMENTS

We implemented our proposed 3D skeleton extraction of deformable objects from multiple images and conducted some experiments on a standard PC with Pentium 4 2.2GHz CPU. Multiple images from 4 cameras consist of a color image which has 640x480resolution. For background subtraction, we trained 20 background images per each camera. The voxel size of target object is 128x128x128. Figure 5 shows the 3D reconstructed object by an image-based visual hull and its extracted skeleton. Those first experiments, showed the robustness and efficiency of our proposed skeleton extraction methodology.

6. CONCLUSION AND DISCUSSION

This paper presents a novel framework for computing markerless 3D skeletons based on an extraction from 3D reconstructed volumetric objects. Both the efficiency and robustness of the proposed framework have been validated within a controlled environment as well as reconstructing different deformable objects. The NGVF based 3D skeleton extraction methodology provides a medial axis of the 3D deformable objects which are reconstructed by image-based visual hulls. We need to benchmark our system within the next steps of research in order to precisely define the parameters and boundary conditions for motion analysis and its applications.

7. REFERENCES

[ACK01] Amenda, N., Choi, S., and Kolluri, R. The Power Crust, In proceeding of the ACM Symposium on Solid Modeling and Applications, pp249-260, 2001.

[BLL03] Barotti, S., Lombardi, L., Lombarid, P.,Multi-module Switching and Fusion for Robust Video Surveillance, In proceeding of Image analysis and processing, 2003.

[BKS01] Bitter, I., Kaufman, A.E., and Sato, M. Penalized Distance Volumetric Skeleton Algorithm, IEEE Transaction on Visualizationi and Computer Graphics, 7(3), pp.195-206, 2001.

[Blu67] Blum, H. A transformation for new descriptors of shapes, MIT Press, pp.362-380, 1967.

[CSYB05] Cornea, N., Silver D., Yuan, X., and Balasubramanian R., Curve-skeleton applications, In IEEE Visualization pp.95-102, 2005.

[DS06] Dey T.K., Sun, J., Defining curve-skeletons with medial geodesic function, In proceeding of SGP, pp.143-152, 2006.

[GG00] Golland P., Grimson, W., Fixed topology skeleton, In proceeding of CVPR, pp.1010-1017, 2000.

[GS99] Gagvani N., and Silver D., Parameter-controlled volume thinning, Graph Models and Image Processing, 61, 3, pp.149-164, 1999.

[HCD04] Han, B., Comaniciu, D., Davis, L., Sequential kernel density approximation through mode propagation, In proceeding of ECCV, 2004.

[HF07] Hassaouna M.S, and Farag A.A., On the Extraction of Curve Skeletons using Gradient Vector Flow, In proceeding of ICCV, pp.1-8, 2007.

[KSC01] Kang, S.B., Szeliski, R., and Chai, j. Handling occlusions in dense multi-view stereo, in proceeding of Computer Vision and Pattern Recognition, pp.103-110, 2001.

[MBG00] Matusik, W., Buehler, C., Gortler, S.J., and McMillan, L. Image-based Visual Hulls, In proceeding of ACM SIGGRAPH, 2000.

[MP79] Marr, D.C., and Poggio. T. A computation theory of human stereo vision, In proceeding of the Royal Society of London, B204, pp.301-328, 1979.

[OK93] Okutomi. M, and Kanade,T. A multiple-based stereo, IEEE Transaction on PAMI 15, pp.353-363, 1993.

[ORP00] Oliver, M.M., Rosario, B., Pentland, A.P., A Bayesian computer vision system for modeling human interactions, IEEE Transaction on PAMI, 2000.

[Pic04] Piccardi, M., Background subtraction techniques: a review, In proceeding of IEEE International Conference on System, Man, and Cybernetics.

[SD97] Saitz, S.M., and Dyer, C.M. Photorealistic scene reconstruction by voxel carving, In proceeding of Computer Vision and Pattern Recognition, pp.1067-1073, 1997. [SG99] Stauffer, C., Grimson, W., Adaptive background mixture models for real-time tracking,, In proceeding of CVPR, 1999.

[WAD97] Wren, C., Azarbayejani, A., Darrel, T., Pentland, A.P., Pfinder: real-time tracking of the human body, IEEE PAMI, 1997.

[WML03] Wu, F.C., Ma W.C., Liou, P.C., Laing, R.H., Ouhyoung M., Skeleton extraction of 3D objects with visible repulsive force, In proceeding of Pacific Graphics, pp.409-413, 2003.

[WP02] Wade, L., and Parent, R.E. Automated Generation of Control Skeletons for use in Animation. The Visual Computer, 18(2), pp.97-110,2002.

[XP98] Xu, C., Prince, J.L. Snake, shapes, and gradient vector flow. IEEE Transaction Image Processing, 7(3), pp.359-369, 1998.

[YB02] Yu, Z., Bajaj, R., Normalized gradient vector diffusion and image segmentation, In proceeding of ECCV, pp.517-530, 2002.

[ZT99] Zhou, Y., and Toga, A. Efficient Skeletonization of Volumetric Objects, IEEE Transaction on Visualization and Computer Graphics, 5(3), pp.195-206, 1999.

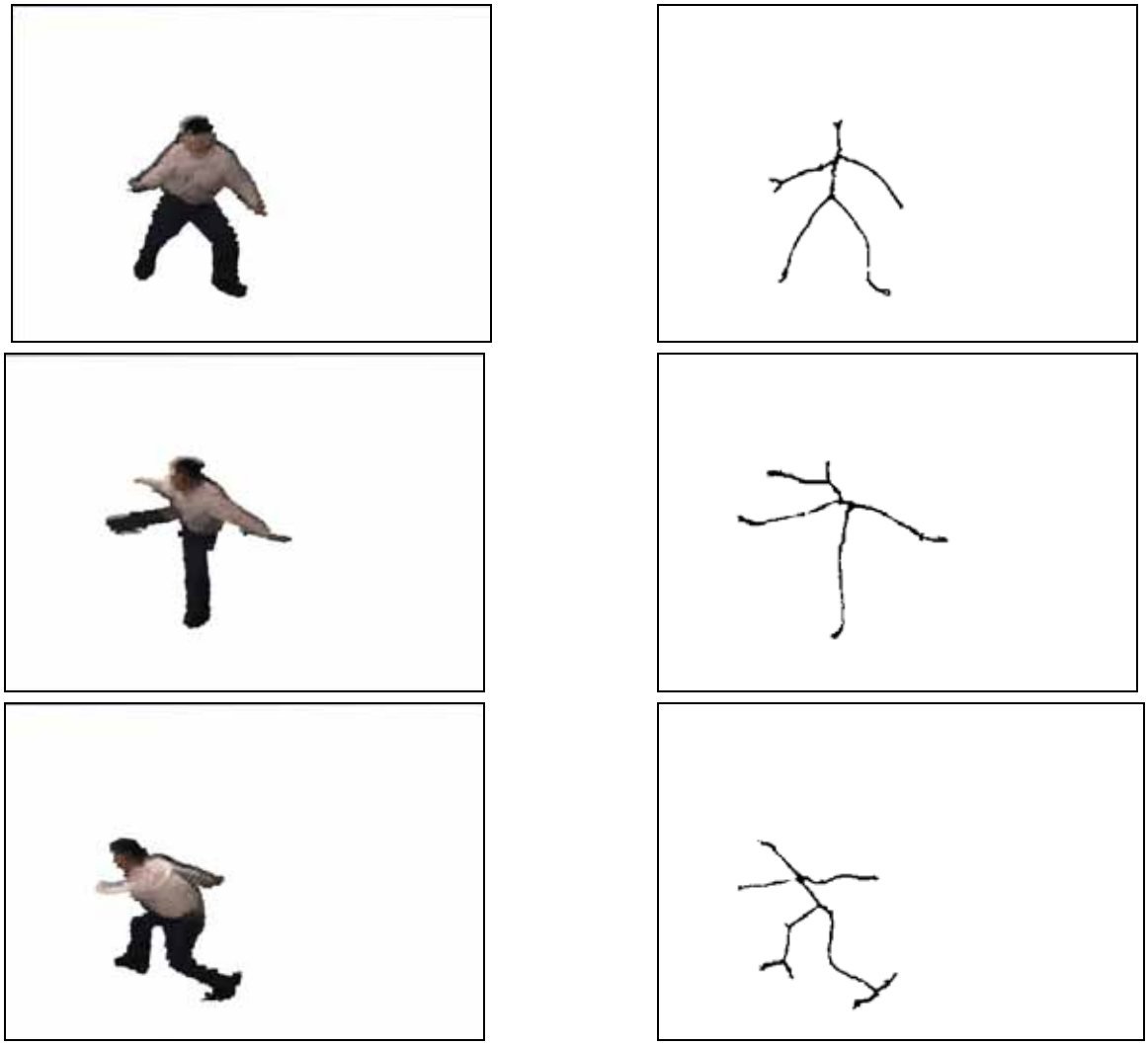


Figure 5. 3D reconstruction of deformable objects and their extracted 3D skeleton with our proposed methodology.