DELAUNAY TRIANGLES MODEL FOR IMAGE-BASED MOTION RETARGETING

Dong Hoon Lee and Soon Ki Jung
Department of Computer Engineering, Kyungpook National University 1370 Sankyuk-dong, Puk-ku, Taegu 702-701, Korea

Key words: vision based motion capture, delaunay triangulation, motion estimation, character animation and image-based modelling and rendering

Abstract: We present an automatic system for retargeting a human body extracted from an image sequence into a new character in a still image. In contrast to analysing the articulated motion of its skeleton in the previous vision-based human body tracking and posture recognition system, we use direct 2-D image warping based on a silhouette. At first, we represent the performer's silhouette with the Delaunay Triangles Model (DTM) of which the boundary points are the critical points of the silhouette. We then use a set of affine transformations of Delaunay triangles for the human body motion, which is applied to a new character for the deformation of the subject's DTM. The final animation of the subject is texture mapped using backward Radial Basis Functions (RBFs). Although our algorithm presented in this paper is not applicable to the human body with self-occluded motion, it allows believable photo-realistic motion retargeting.

1. INTRODUCTION

The pursuit of photo-realism is of major interest in computer graphics and virtual reality. In particular, realistic animations of avatar or an autonomous agent are important essential elements for constructing a virtual environment. Many researchers have studied realistic motion and expression generation and focused on motion control of geometric human body and facial models. Motion capture or motion retargeting is one of these efforts.

N. Magnenat-Thalmann et al. (eds.), *Deformable Avatars*

In this paper, we will focus on vision-based human body tracking and posture recognition systems.

Systems based on computer vision capture the motion parameters from images. These systems are called kinematics analysis systems [11, 6, 2, 8]. With this approach, the human body model is composed of a number of parts that allows movement among them so that reconstruction of the human body motion should calculate shape and joint angle parameters. The main difficulties are related to modelling humans and to the expensive search procedure for the recovery of shape and motion parameters.

Our method is based on understanding low-level features without an exhaustive search of high-level parameters. Previous similar approaches [1, 13, 7, 3] involve the extraction of the motion flow field and then segmenting it into piecewise smooth surfaces. These surfaces are then grouped and recognized as human parts, maybe using various types of features. Unfortunately, optical flow segmentation methods are rarely sufficiently general, and the recognition process may involve prohibitive search procedures.

Most existing approaches require a skeleton model of human motion. In contrast, we model the human body as a Delaunay Triangles Model (DTM), which has a set of 2-D Delaunay triangles approximating the silhouette of the human body and affine motion for each Delaunay triangle. In order to get a reliable DTM, we perform the following three steps:

First, we extract a set of feature points from the silhouette using the critical point detection (CPD) algorithm described by Zhu and Chirlian [14]. Second, we build a polygon with the set of points. We then tessellate the polygon using Delaunay triangulation [9]. The consistency of Delaunay triangulation between frames is preserved by a model based 2-D tracking of each triangle with piecewise constant affine motion and the adaptation of DTM by managing the critical value of each candidate critical point. Affine motion of each triangle for the subject's DTM is applied to the DTM of the new character for motion retargeting. The motion of the inner part of each triangle is interpolated using backward Radial Basis Functions (RBFs) for image warping of the subject.

We capture the performer's animation footage and the subject's still image (we can also use any picture of famous movie stars) as input data. The subject must have a similar posture to the performer's to ease the correspondence problem. Our algorithm assumes that the person in the image stands facing the fixed camera and does not have self-occluded motion. We assume this because we cannot extract 3-D motion from just one camera. These assumptions show our study is not intended for a challenging case of analysing complicated human motion and animating synthetic modelled character, but rather for rendering a real human appearance.

The overall algorithm is summarized in Figure 1.

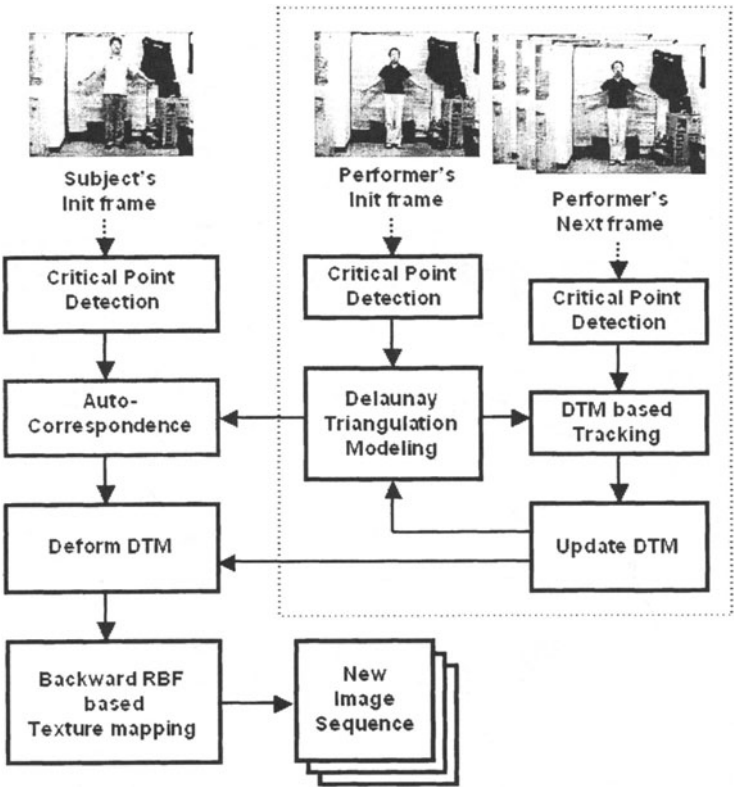


Figure 1. System block diagram.

The rest of this paper is organized as follows. We describe the details of the automatic critical point detection algorithm and the DTM-based tracking of the triangle model in Sections 2 and 3, respectively. Section 4 includes the algorithm of motion retargeting using backward RBFs. We show some results in Section 5. We present conclusions in Section 6.

2. DELAUNAY TRIANGLES MODEL

2.1 Human Body Extraction

To construct a Delaunay Triangles Model (DTM), we first have to extract a human silhouette. Video stream is nearly impossible to get a human

silhouette using a manual method because it has a large number of image data, so we design a background modelling strategy for automatic human extraction.

In the first stage, we made a model of a background image with several image frames that do not contain a person. To eliminate the effect of the luminance and shadow, we model each pixel as a mixture of Gaussians in Equation (1). The Gaussian distributions of the adaptive mixture model are then evaluated to determine which are most likely to result from a background process. Based on the persistence and the variance of each of the Gaussians of the mixture, we determine which Gaussians may correspond to background colours. Pixel values that do not fit the background distributions are considered foreground.

$$f(X) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\|X-u\|^2}{2\sigma^2}} \tag{1}$$

where X is the vector value of each pixel colour element R, G, B, and u is the mean of each pixel about background image.

2.2 Critical Point Detection

The human body is approximately modelled by Delaunay triangles, which consists of triangles connected together by feature points. To generate feature points, we used the critical point detection(CPD) algorithm described by Zhu and Chirlian [14].

2.2.1 Pseudo Critical Points

We assume the human silhouette is a simple closed contour without any interior holes. In a simple closed contour obtained by human body extraction in the previous step and simple border tracing technique, each pixel p_i has two neighbours p_{i-1} and p_{i+1}. We then transfer the contour to polar coordinates because it is easier to handle rotation and scaling changes in polar coordinates than in rectangular coordinates. When the centroid of the shape is used as the origin, the representation of the shape becomes very simple. The 2-D contour can be decomposed into two 1-D curves: $\rho(i)$ and $\theta(i)$, the local maxima and minima (zero crossing points) are more important in describing the curve character than the other points. Therefore, these points are selected as candidate of critical points. We select zero crossing points as pseudo critical points.

2.2.2 Critical Value

The pseudo critical points are just candidates of critical points. Some pseudo critical points must be deleted. The CPD algorithm assigns a critical value to each point on the boundary that is simply the area of the triangle constructed from the given point and its two immediate neighbours. The height of the triangle reflects the information of directional change providing the support region is a constant and the bottom of the triangle reflects the information of feature size. The critical value in each point represents the possibility of becoming a critical point. Thus, a larger critical value has a higher probability to be chosen as a critical point.

An iterative decimation process is used which removes the point with the smallest critical value, recomputes the critical value of the immediate neighbours of the point which has just been deleted and reidentifies the point with the smallest critical value. The process terminates when the remaining smallest critical value is above some threshold set by the user.

2.3 Delaunay Triangles Model (DTM)

The set of feature points from the previous step is tessellated by Delaunay triangulation. The silhouette information is insufficient to model human motion. That is the reason we construct triangle structures for human modelling. Each triangle has the information of a set of feature points with posture data and affine transformation elements. The transformation elements represent a rotation, translation, scale and shear between frames. Each affine motion of Delaunay triangles represents the complicated nonrigid human motion. Figure 2 shows the real human figure and its DTM. The result of tessellation includes the background region, so we should eliminate the triangles outside of human to get the final result.

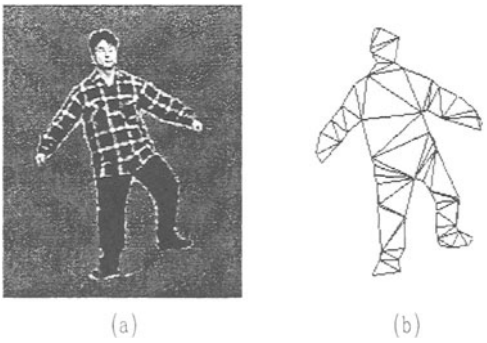


Figure 2. Delaunay Triangles Model (DTM).

3. DTM-BASED TRACKING

3.1 Tracking

In this section, we describe DTM-based feature tracking using predicted measurements. The human motion is assumed as a 2-D piecewise constant affine motion of each triangle which comprise a DTM. The 2-D motion of a triangle is a general 2-D affine transformation, representing a combination of rotation, translation, scale and shearing.

We predict the posture of each 2-D triangle with affine parameters calculated from a previous triangle and select real measurements from the set of candidate critical points on the silhouette. The candidate critical points are a set of pseudo critical points, not all point on the silhouette because the pseudo critical points have compact features that represent the character of a human silhouette. We can model the error function for selecting real measurements with the weighted nearest neighbour method as below (Equation 2). The critical point with the smallest error value is selected as the real measurement.

$$E(\hat{p}, p_i) = \frac{\lambda \times |\hat{p} - p_i|}{I(p_i)} , \tag{2}$$

where $I(p_i)$ is a critical value of each pseudo critical point, p is a predicted measurement, λ adjusts the weight between the distance and critical value, and the domain of candidate points, p_i, is determined by the user. .

3.2 Update DTM using Critical Value

A DTM constructed from an initial posture does not guarantee to approximate the silhouette after arbitrary human motion such as the bending motion of arms. In this case, some critical points newly appear. The DTM is dynamically updated through missed features, which are not tracked during the tracking step, with a high critical value in the feature extraction step as shown in Figure 3. The update process of the proposed algorithm can be summarized as follows:

1) DTM-based tracking.
2) Calculate the critical value with the tracked real measurement and a critical point with a high critical value.

3) Find the point with the lowest critical value besides the critical value calculated from a triangle that is composed of three real measurements.
4) Compare the lowest critical value with the specified critical value, I. If the lowest critical value is smaller than I, delete the critical point with this critical value, then go to step 2. Otherwise, stop the recursion.
5) Update the DTM.

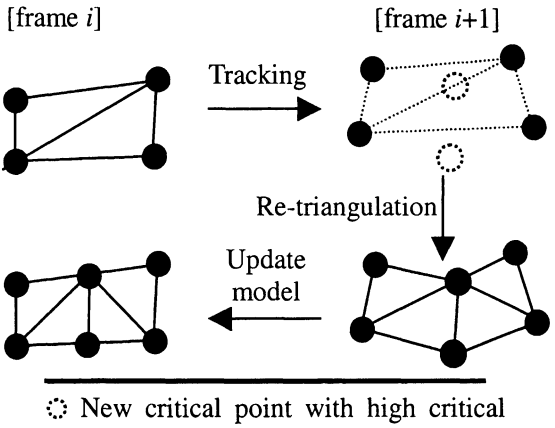


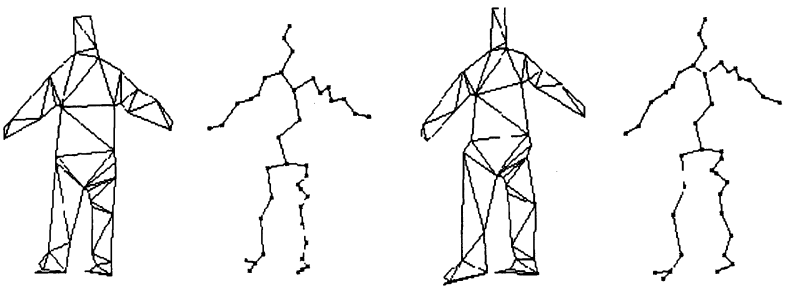
Figure 3. Update object model.

4. MOTION RETARGETING

4.1 DTM-based Motion Retargeting

As mentioned before, we would like to animate the subject from the performer's motion using a set of affine transformations for Delaunay triangles, which compose of the human body. For this purpose, we split the affine motion into several primitive components such as translation, rotation, scale, and shear. First, we extract the stick figure that forms a graph of which the nodes are the center points of triangles and the edges mean the adjacencies between triangles in DTM. We then calculate the translation of the whole body from the translation motion of a specific node, which has the maximum degree of the graph and the maximum area of the triangle. We call the node as the root of the graph. The rotation of each triangle can be extracted by the orientation motion of the corresponding node with respect

to the root. The scale factor is calculated from the variation of the distance between the node and the root. Then the remained terms are for the local affine motion of each triangle. The DTM-based motion retargeting is shown in Figure 4.



(a) Performer's DTM and graph (b) Subject's retargeted DTM and graph

Figure 4. This figure shows the result of retargeting the performer's DTM into the subject's DTM. The variations of the split affine motion between frames in the performer's graph model are applied to the subject's.

4.2 RBF based Texture Mapping

Classical approximation theory solves the problem of approximating or interpolating a continuous multivariate function by an approximation function with the appropriate choice of a parameter set. Finding a parameter set is often referred to as learning or training in the neural network sense. In the training stage, a goal is to figure out given an approximation function and a set of training examples that will provide the best approximation of F [10]. Radial Basis Functions are often chosen as approximating high dimensional smooth surfaces. Examples of RBFs are Gaussian functions, multi-quadrics and thin plate splines with linear terms added. The RBF training equation is expressed as Equation (3):

$$\Phi w = d, \tag{3}$$

where d is defined as the matrix of the coordinate of the performer's feature points, Φ is the matrix of $\{\varphi(\|x - x_i\|) \mid i = 1,2,...,N\}$ which is a set of N radial-basis functions which are made with the feature coordinates of the performer's next frame and w is unknown coefficients (weights).

Then, the set of weights, w, is given by

$$w = \Phi^{-1}d. \tag{4}$$

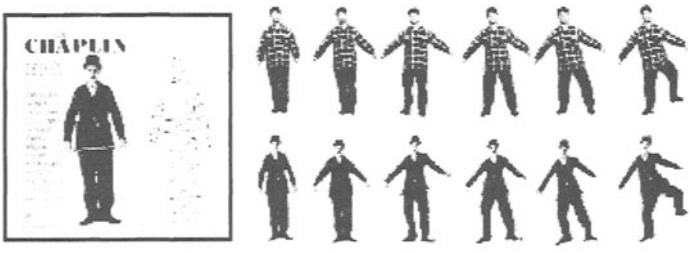
The general mapping function can be given in two forms: either relating the output coordinate system to that of the input, or vice versa. These functions are known as forward and inverse mapping. In this paper, we use inverse mapping because it guarantees that all output pixels are computed unlike with the forward mapping scheme.

We already have the spatial transformation parameter, w, from the previous step, so we can get the forward warped image easily by instituting a subject feature in the initial frame into the radial centre. However, we cannot get the initial subject's coordinates for inverse mapping because this information is contained in the radial basis function ϕ.

To operate inverse mapping, we simply obtain the weights of the RBF with the set of subject's feature point in frame i+1 as input x and those in frame i as desired output d in the training step. Then every pixel of subject image in frame i is scanned as input and each output pixel mapped back onto the input via the spatial transformation mapping function.

5. RESULTS AND EXPERIMENTS

The performer's motion is recorded using one video recorder and then captured off-line while playing back frame-by-frame. The system uses a SONY DCR-TRV310 video recorder and we implement the algorithm on a Pentium II PC which has a 350MHz CPU and 128Mbytes memory. The subject must have a similar posture to the performer's to ease the correspondence problem. The person in the experiments stands facing the fixed camera and does not have self-occluded motion. We create a variety of motions of retargeted objects by choosing different motions of source film footage. Figure 5 shows example of results of our proposed scheme.



(a) (b)

Figure 5. The example of results of the proposed scheme. The segmentation and feature extraction of subject image is done manually in left image and right image shows its DTM in (a). (b) shows the performer's image sequences (upper) and the result of motion retargeting using Charlie Chaplin still image (below).

6. CONCLUSIONS

In this paper, we have presented an automatic system for retargeting a human body motion extracted from an image sequence into a new character in a still image. In contrast to analysing the articulated motion of its skeleton in the previous vision-based human body tracking and posture recognition system, we have modelled the human body as a Delaunay Triangles Model (DTM), which has a set of 2-D Delaunay triangles approximating the silhouette of the human body and affine motion for each Delaunay triangle. We have used a set of affine transformation of Delaunay triangles for the human body motion that was applied to a new character for the deformation of the subject's DTM. The final animation of the subject was texture mapped using the backward Radial Basis Functions (RBFs).

Our study was not intended to challenge the case of analysing complicated human motion and animating a synthetic modelled character but rather for rendering real human appearance. Therefore, although our algorithm presented in this paper is not applicable to the human body with self-occluded motion, it allows believable photo-realistic motion retargeting. Our system can be utilized in the field of entertainment for the purpose of generation of the motion with old celebrity's image or mimicking the celebrity's motion such as Charlie Chaplin's. Furthermore, this will allow the generation of realistic avatars from widely available video clips.

ACKNOWLEDGEMENT

This work was supported by the Korea Science and Engineering Foundation(KOSEF) through the Virtual Reality Research Center at KAIST and by Korea Research Foundation Grant(KRF-99-041-E00294).

REFERENCES

[1] A. Bottino, A. Laurentini, and P. Zuccone, Toward non-intrusive motion capture, *Computer Vision-ACCV98*, pages 416-423, 1998.

[2] D. M. Gavrila and L. S. Davis, 3-d model-based tracking of humans in action: a multi-view approach, *In IEEE Conf. on CVPR*, pages 73-80, San Francisco, USA, 1996.

[3] I. Haritauglu, D. Harwood, and L. Davis, w^4s: A real-time system for detecting and tracking people in 2½d, *Computer Vision-ECCV98*, pages 877-892, 1998.

[4] Simon Haykin, *Neural networks - A comprehensive foundation*, 2nd Edition, Prentice-Hall, 1999.

[5] Andrew Hill, Chris J. Taylor, and Alan D. Brett, A framework for automatic landmark identification using a new method of nonrigid correspondence, *IEEE Trans. on PAMI*, 22(3), pages 241-251, March 2000.

[6] E. A. Hunter, P. H. Kelly, and R. C. Jain, Estimation of articulated motion using kinematically constrained mixture densities, *In Proceedings of IEEE Non-Rigid and Articulated Motion Workshop*, pages 10-17, Puerto Rico, USA, 1997.

[7] S. Ju, M. Black, and Y. Yacoob, Cardboard People: A parameterized model of articulated image motion. *In 2nd International Conference on Face and Gesture Analysis*, pages 38-44, Vermont, USA, 1996.

[8] I. A. Kakadiaris and D. Metaxas, Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection, *In IEEE Conf. on CVPR*, pages 81-87, San Francisco, CA, USA, 1996.

[9] Kazuhiro Nakahashi and Dmitri Sharov, Direct surface triangulation using the advancing front method, *AIAA-95-1686-CP*, pages 442-451, 1995.

[10] Mark J. L. Orr, Introduction to radial basis function networks, technical report, Centre for Cognitive Science, University of Edinburgh, April 1996.

[11] S. Wachter and H. –H. Nagel, Tracking of persons in monocular image sequences, *In Proceedings of IEEE Non-Rigid and Articulated Motion Workshop*, pages 2-9, Puerto Rico, USA, 1997.

[12] George Wolberg, *Digital Image Warping*, IEEE Computer Society Press, 1990.

[13] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, Pfinder: Real-time tracking of the human body, *IEEE trans. on PAMI*, 19(7) pages 780-785, July 1997.

[14] Pengfei Zhu and Paul M. Chirlian, On critical point detection of digital shapes, *IEEE trans. on PAMI*, 17(8), pages 737-748, August 1995.