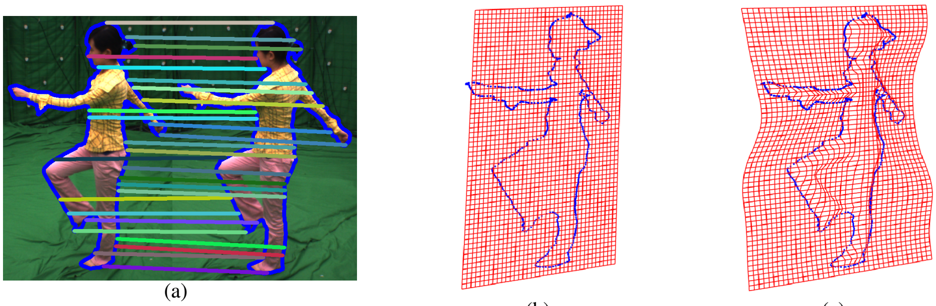
Vision, Modeling, and Visualization (2013)
Michael Bronstein, Jean Favre, and Kai Hormann (Eds.)

Poster

Dynamic Time Warping Based 3D Contours

Simone Croci, Aljoscha Smolic, Oliver Wang

Disney Research Zurich



(a) (b) (c)

Figure 1: *Our method computes an alignment of two contours (a), which allows us to position and orient a planar billboard (b) and to create a non-planar 3D billboard (c). The input figures in (a) are from [YQW10].*

Abstract

In this work, we present a method for computing 3D contours from the silhouettes of objects in multiple views. Our approach unwraps contours into 1D signals and computes an efficient, globally optimal alignment using a modified dynamic time warping technique. Using our method, we can approximate real model geometry by interpolating the 3D contour on the billboard plane, thereby reducing the stereo problem to 1D and allowing for much more efficient and robust computation methods.

Categories and Subject Descriptors (according to ACM CCS): I.4.9 [Image Processing and Computer Vision]: Applications—

1. Introduction

Virtual view synthesis (VVS) describes the process where novel views of a scene are generated from virtual cameras. One difficulty of VVS is efficiently acquiring accurate scene geometry. Planar billboards are commonly used as a simple geometric proxy. Objects are first segmented from an input image and then projected onto planar billboards (see for example [HS06]). In this work we introduce a method to efficiently compute non-planar 3D billboards. Another important contribution of this work is the application of dynamic time warping (DTW), originally introduced in [BK59], to the stereo correspondence problem.

First, our method takes two input images from different stereo views and an object segmentation. From the segmented images we extract the 2D contours (a list of points on the boundary of the silhouette) and we compute the alignment (correspondences) using a modified DTW method (see Figure 1(a)). From the alignment, we triangulate the 2D contour points obtaining a 3D contour, and then we fit a surface through it obtaining a 3D billboard (see Figure 1(b) and 1(c)).

2. Alignment

Dynamic Time Warping DTW is a method based on dynamic programming for finding the best alignment between two 1D signals. We use DTW to align two 2D contours

DOI: 10.2312/PE.VMV.VMV13.229-230

from different views: $\mathbf{c}_1(i_1) \in \mathbb{R}^2$ with $i_1 = 1,\ldots,N_1$ and $\mathbf{c}_2(i_2) \in \mathbb{R}^2$ with $i_2 = 1,\ldots,N_2$. For each alignment a cost is defined and the goal is to find the alignment with the smallest cost. The cost of an alignment is defined as the sum of the distances $\Phi(i,j)$ between the pairs of matched contour points $\mathbf{c}_1(i)$ and $\mathbf{c}_2(j)$. Given two contour points $\mathbf{c}_1(i)$ and $\mathbf{c}_2(j)$ we propose the following distance:

$$\Phi(i,j) = d_{patch}(i,j) + \lambda\, d_{epipolar}(i,j) \tag{1}$$

$d_{patch}(i,j)$ is defined as the sum of the color differences between two windows centered at the contour points $\mathbf{c}_1(i)$ and $\mathbf{c}_2(j)$. $d_{epipolar}(i,j)$ is based on the distances between point $\mathbf{c}_2(j)$ and the epipolar line of $\mathbf{c}_1(i)$ and vice versa.

DTW fills a N_1xN_2 matrix $\mathbf{X}$ in scan-line order, where each entry $\mathbf{X}(u,v)$ stores the minimum cost of aligning the contour points $\mathbf{c}_1(i_1)$ with $i_1 = 1,\ldots,u$ and $\mathbf{c}_2(i_2)$ with $i_2 = 1,\ldots,v$. Each entry (u,v) of the matrix $\mathbf{X}$ is computed by evaluating Φ at (u,v) and by considering only already computed neighboring entries. The cost of the best alignment is in the entry $\mathbf{X}(N_1,N_2)$, and the optimal alignment can be obtained from $\mathbf{X}$. The running time of DTW is then in $O(N_1N_2)$.

Modified Dynamic Time Warping A limitation of DTW is that the starting points of the two 2D contours must be known. Since this information is not available, we have developed a modified DTW where we wrap the second contour around the first twice, and compute the best alignment over all starting points simultaneously. For this purpose a matrix $\mathbf{X}'$ of size $2N_1 \times N_2$ is computed similarly to $\mathbf{X}$.

Optimization Most of the running time of the modified DTW is spent in constructing the matrix $\mathbf{X}'$ and more specifically in evaluating Φ. In order to reduce the running time we have implemented three improvements. First, Φ is saved when evaluated in order to avoid to compute it twice. Second, we keep an upper bound of the alignment cost in order to prune the propagation of partial contours above this bound. Third, we have developed a special criteria that avoids to propagate unlikely partial contours.

Alignment filtering A filter has been specifically designed to cope with alignment errors that occur when a 2D contour region in one view is not present in the second view.

3. 3D Billboard

Given two corresponding 2D contour points the 3D point is obtained by intersecting rays in world space. We can use this set of 3D points to help with several VVS tasks.

Billboard location and orientation For VVS, billboards must be correctly placed in 3D world coordinates, to allow for rendering novel views. One possibility would be to consider the mean of the 3D points. In the case of reconstructing standing people, we determine billboard locations by considering only the leg contour points (established as points near the ground whose neighbors are vertical), which we found to give robust billboard placement locations.

Besides the position of the billboard we can also compute the orientation of the billboard around a vertical axis by applying principal component analysis (PCA) to the projection of the 3D contour points onto an horizontal plane (see Figure 1(b)). The first PCA component defines the major axis, and will be the best approximation for a planar or near planar billboard orientation.

3D Billboards Once we have defined the position and orientation of the planar billboard, we can use the 3D contour points to create an offset from the planar billboard. We interpolate all positions within the contour using normalized convolution. For standing people we additionally inflate the interior of the 3D billboard: the more interior a point is, the more inflation it gets until a certain limit (see Figure 1(c)).

4. Conclusion

In this work, we have presented a simple and efficient way to enhance billboard rendering accuracy by integrating some 3D information. A modified DTW method has also been proposed for solving a 1D version of the stereo problem with unknown starting correspondences, and several efficiency optimizations.

However, our method is not without limitations. Our assumption that the contour points match in two views can be violated when camera positions are too far apart. In this case, we have proposed a filtering approach to remove these contour points.

Future work would be to investigate combining additional cameras to reduce fitting ambiguities, using more complicated geometric representations driven by our 3D contours, and exploring in greater detail what the theoretical bounds are on VVS for different objects given semi-planar geometric representations.

For more details about our work see [CSW13].

References

[BK59] Bellman R., Kalaba R.: On adaptive control processes. *Automatic Control, IRE Transactions on 4*, 2 (1959), 1–9. 1

[CSW13] Croci S., Smolic A., Wang O.: *Dynamic Time Warping Based 3D Contours*. Tech. rep., 2013. URL: http://zurich.disneyresearch.com/videodata/vmv/3DContourTechReport.pdf. 2

[HS06] Hayashi K., Saito H.: Synthesizing free-viewpoint images from multiple view videos in soccer stadium. In *Computer Graphics, Imaging and Visualisation, 2006 International Conference on* (2006), IEEE, pp. 220–225. 1

[YQW10] Yebin L., Qionghai D., Wenli X.: A point-cloud-based multi-view stereo algorithm for free-viewpoint video. *IEEE trans. Visualization and Computer Graphics 16*, 3 (2010), 407–418. 1