A comparison of Non-Euclidean Image-Based Rendering

Dana Cobzaş, Martin Jägersand
CS, U. of Alberta, Edmonton, AB, T6G 2E8, CANADA
http://www.cs.ualberta.cs/~{dana,jag}

Abstract

In image based rendering methods, viewing geometry is used to ensure a physically valid image transforms. We compare and contrast methods based on affine and projective viewing geometry and with and without an explicit 3D object structure representation.

1 Introduction

Image based rendering computes new views based on example views using geometric image-to-image transforms. Modeling, on the other hand, attempts to recover 3D structure, and render by texture mapping. In a blend of the two, non-Euclidean structure is used as an intermediate stage, but rendering is from image-to-image transforms. We compare two methods, 2D view morphing [5] and 3D affine reprojection [3], where the affine structure is also recovered from the image sequence [1, 6].

2 Theory

Assume a sequence of input images $\mathbf{I}_k, k = 1 \dots N$, for N different viewpoints. Let $\mathbf{x}_k, \mathbf{y}_k$ be row vectors of the image row and column projection of M corresponding physical points $\mathbf{q}$ tracked through the image sequence[1].

2.1 2D View Modeling

For a camera is moving parallel to its image plane, views interpolated from corresponding image points are geometrically-correct transitions between the original images. For solving the general case of non-parallel views (see Figure 1), two input images $\mathbf{I}_1$ and $\mathbf{I}_2$ have to first be *prewarped* in order to align the image planes without changing the optical centers of the two cameras. The prewarped images $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$ are then *interpolated* to generate an intermediate view $\hat{\mathbf{I}}_s$ that can be *postwarped* into $\mathbf{I}_s$ that has the desired camera orientation.

Prewarping Assuming perspective projection, we can compute the prewarping transformations $\mathbf{H}_1, \mathbf{H}_2$ that will align the image planes from the fundamental matrix $\mathbf{F}$. The fundamental matrix can be estimated from a set of 8 or more corresponding points between the two images. In order to have the views parallel, $\mathbf{F}$ should have the form:

$$\hat{\mathbf{F}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Then the image planes can be aligned by choosing any two projective transformations $\mathbf{H}_1, \mathbf{H}_2$ such that $(\mathbf{H}_2^{-1})^T \mathbf{F} \mathbf{H}_1^{-1} = \hat{\mathbf{F}}$. As described in [5], this can be achieved by applying a 3D rotation that makes the image planes parallel, followed by an affine warp to align the scanlines.

[1] We write matrices capitalized, vectors bold and scalars plain, row concatenation (x,y) and column concatenation (x;y).

Interpolation To interpolate the rectified images, form corresponding triangles from the tracked points. First interpolate the vertices of the triangular patches, then form the intermediate image by warping the interior of the triangles from both reference images. Here it is assumed that the triangles represent physical planes.

Postwarping The interpolated image is aligned with the rectified image planes. To reorient it for a natural transition between the original images, we apply a homography defined by the desired position of at least four points. We choose the first and last image of $\mathbf{I}_k$ for doing the interpolation and we used the intermediate positions of the control points to compute the postwarp.

2.2 Affine Modeling

Following [1], let $m_{kx} = \frac{1}{M}\sum_{i=0}^{M} x_i, m_{ky} = \frac{1}{M}\sum_{i=0}^{M} y_i$ be the centroid of the tracked points in frame k. Compose measurement matrices $W = (\mathbf{x}_k; \mathbf{y}_k)$, and zero mean $\bar{W} = (\mathbf{x}_k - m_{kx}; \mathbf{y}_k - m_{ky})$. Let $[U, S, V] = \text{svd}(\bar{W})$ be the singular value decomposition of $\bar{W}$.

Rank theorem Under orthography $\text{rank}(\bar{W}) = 3$. Under most viewing conditions with a real camera the effective rank is 3. Proof: see [1].

Interpret $\tilde{V} = S_{1\dots3,1\dots3} V_{1\dots3,1\dots N}$ as the affine coordinates of $\mathbf{q}$ and $\tilde{U} = U_{1\dots N,1\dots3}$ as the image projection matrices, where each view k is represented by the rows $P_k = (\tilde{U}_k; \tilde{U}_{k+M})$, and the columns of P represent the projections of the affine basis vectors. Hence rotation and scaling is represented by P and image plane translation by $\mathbf{m} = (m_x; m_y)$ and $W \approx \hat{W} = \tilde{U}\tilde{V} + \mathbf{m}$.

Reprojection property Given the image projections of a new basis $(\mathbf{m}, P) = (\mathbf{m}, (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3))$, all the affine points can be reprojected by $\begin{bmatrix} x \\ y \end{bmatrix} = \hat{W} = PV + m$.

Constructing P to correspond to a real physical viewpoint change requires additional knowledge. Published approaches include knowing the camera calibration or the metric coordinates of the affine basis[1, 6]. For small viewpoint changes (≤ 10 Deg) we approximate physical rotations by a linear combination of the closest example poses P_k. For arbitrary large changes we require metric knowledge X of at least 4 object features, and solve for a matrix Q s.t. $QX = V$. Then the rotated reprojection $\hat{W} = PQRQ^{-1}V + m$ can be done with a standard Euclidean rotation matrix R.

Rendering The affine model relates the coordinate projections W of the example images I, and the desired new pose $\hat{W}$. To render the new pose we arrange the points in $\hat{W}$ to represent destination quadrilaterals $\hat{\mathbf{y}}_d = (\mathbf{x}, \mathbf{y})$, and source quadrilaterals $\tilde{\mathbf{y}}_s$ in the closest example image. The rendering maps physical planes correctly under a projective camera model. Hence, we chose the points in W so that each object plane is represented by at least four points.

3 Experimental evaluation

A sequence of 15 images was created from different viewpoints of a toy house. Using XVision [2], 10 corresponding points were tracked (See the crosses in Fig. 2).

2D View Modeling We compute the fundamental matrix required by the prewarping step using *image-matching* software [4]. The errors from the estimated epipolar geometry affect the rectification process. In our case, the errors in the orientation of rectified epipolar lines are less than one pixel. We generate a movie [*www-video-1*] with the intermediate frames reconstructed from the first and last image of the original sequence. Most of the visible shape distortion errors are caused by 3-step the subsamplig process (rectification, interpolation, postwarping) and by the tracking errors that affect the postwarping transformation. The residual errors for the 10 tracked points recomputed through the 8D postwarping transformation are about one pixel per point.

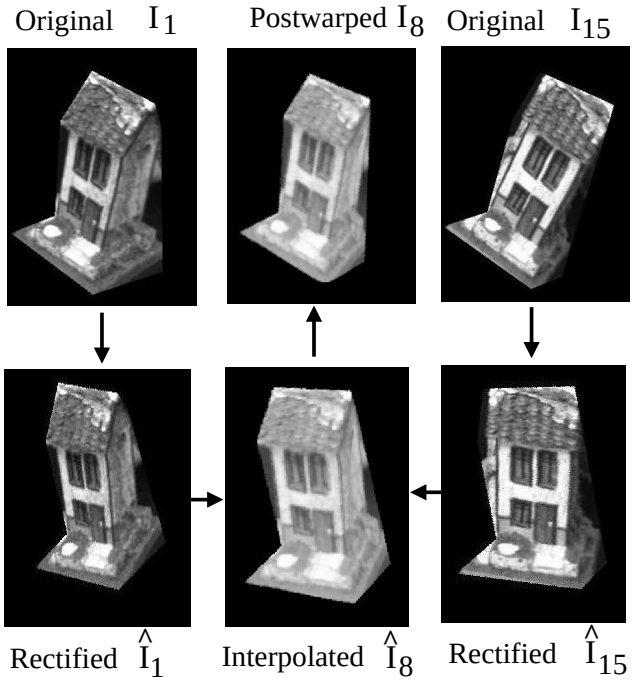

Figure 1: Steps in view interpolation

Affine Modeling We construct a 30×10 measurement matrix W from tracking the 10 points shown by crosses in Fig. 2. Constraining the point projections to the rank 3 affine subspace reduces tracking error from approximately 10 pixels to a few pixels. A movie [*www-video-2*] of smooth pose changes was generated by linearly interpolating $(P, m) = \sum_k a_k (P_k, m_k)$ for varying a_k's, and using the basis projections from the training images. The basis projections were close enough to not cause visible errors from approximating rotation arcs with lines. Subjectively, there is some unevenness in the motion. This is most likely caused by the errors in the estimated (P, m) due to tracking. The effect is minimized by sourcing the texture from the training image I_k corresponding to the closest (P_k, m_k), and hence canceling the errors from the largest a_k. We noted that the affine basis (plotted in [*www-video-2*]) over-rotated significantly, but for the rendering this didn't matter since the point reprojections were accurate to within a pixel. We used a standard SGI O2 cam, for which the linear camera assumption used in deriving the affine geometry interpretation does not hold precisely, but the rank 3 subspace still performs accurate data fitting.

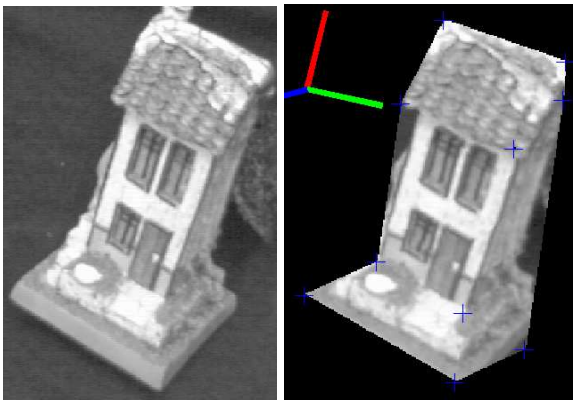

Figure 2: Left: Original house image. Right: House rendered from affine shape in a similar pose.

4 Discussion

We have shown two methods which can render new poses using only example images and image-to-image transforms. Two fundamental differences are: 1. The affine method uses a 3D non-Euclidean structure as an intermediate step, while the 2D view modeling only uses images and epipolar geometry, but no 3D object structure. 2. In the affine case information from a whole sequence of views is factored into stable pose and structure estimates, while the non-linear projective model for the 2D method is based on pairs of images. Since neither method recovers Euclidean viewing geometry, reorientations cannot be based on metrics and exact angles. Reorientations instead are done by interpolating example poses. Here the 2D projective model can more accurately deal with large differences between the example poses compared to the linear affine model. For both methods it is critical to extract corresponding points, a known hard problem. This is easier in a closely spaces image sequence, where tracking[2], or after rectification, disparity can be used.

References

[*www-video-*] On-line videos of the experiments are available on http://www.cs.ualberta.ca/~jag/ImRend/

[1] Tomasi C. and Kanade T. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9:137–154, 1992.

[2] K. Toyama G. Hager. The xvision system: A general-purpose substrate for portable real-time vision applications. *Computer Vision and Image Understanding*, 69(1):23–37, 1998.

[3] K. N. Kutulakos and J. R. Vallino. Calibration-free augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 4(1):1–20, 1998.

[4] Image Matching Software. http://www-sop.inria.fr/robotvis/personnel/zzhang/softwares.html.

[5] S. M. Seitz and C. R. Dyer. View morphing. In *Computer Graphics (SIGGRAPH'96)*, pages 21–30, 1996.

[6] D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *In Proc. International Conference on Computer Vision (ICCV'93)*, pages 675–682, 1993.