JOINT BUNDLED CAMERA PATHS FOR STEREOSCOPIC VIDEO STABILIZATION

Heng Guo, Shuaicheng Liu, Shuyuan Zhu, Bing Zeng

School of Electronic Engineering
University of Electronic Science and Technology of China, Chengdu, China

ABSTRACT

This paper presents a method to stabilize shaky stereoscopic videos captured by hand-held devices. Directly applying traditional monocular video stabilization techniques to two views independently is problematic as it often brings undesirable vertical disparities and produces inaccurate horizontal disparities, which violate original stereoscopic disparity constraints, leading to erroneous depth perception. In this paper, we show that monocular video stabilization methods, such as the bundled camera paths stabilization, can be extended for stereoscopic videos by taking additional disparity constraints during the stabilization. In particular, we first estimate disparities between two views. Then, we compute camera motions as meshes of bundled paths for each view. Next, we smooth paths of two views separately and iteratively. During each iteration, we adjust the meshes of one view by our proposed 'Joint Disparity and Stability mesh Warp (**JDSW**)'. The final result is generated after several iterations of paths smoothing and meshes adjusting, in which temporal stability and correct depth perception are achieved simultaneously. We evaluate our method by various challenging stereoscopic videos with different camera motions and scene types. The experiments demonstrate the effectiveness of our method.

***Index Terms*—** Stereoscopic, video stabilization, bundled paths, disparity, mesh warp

1. INTRODUCTION

In recent years, a variety of stereoscopic cameras and displays became available. Thanks to the success of 3D movies which accelerates the development of stereoscopic techniques, some of which has been explored in the research community, including stereoscopic cloning [1, 2, 3], warping [4, 5], inpainting [6, 7], panorama [8, 9] and retargeting [10, 11]. In this work, we focus on stereoscopic video stabilization [12]. Videos captured by hand-held cameras often appear remarkably shaky. Video stabilization aims to remove camera jitters, synthesizing videos with smoothed camera motions. Stabilization techniques recover camera trajectories, either in 2D [13, 14] or in 3D [15, 16], to represent the camera motions, which are then smoothed via low-pass filtering [17, 18].

However, the traditional monocular video stabilization methods cannot be applied to stereoscopic videos directly. The disparity constraints between left and right views of a stereoscopic video will be damaged in the stabilized frames due to the ignoring of the disparity constraints, which generates problematic depth perception, leading to 3D fatigue to the viewers. Liu *et al* [5] proposes a method for stereoscopic image warping, which preserves disparities during the image transformation. Meanwhile, bundled camera path stabilization [19] can handle videos with scene parallax under various challenging camera motions. In this work, we propose a video stabilization technique that combines the merits of two approaches to stabilize stereoscopic videos. Specifically, we begin by calculating disparities between two views. Then we compute camera motions as meshes [19] and smooth them iteratively. During each iteration, the disparities are warped [5] and the meshes are adjusted by our proposed 'Joint Disparity and Stability mesh Warp (**JDSW**)'. We demonstrate the performance of our approach through many challenging casual captured videos. Please refer the project page for videos. [1]

2. RELATED WORK

Video stabilization can be roughly categorized as 3D [15, 16], 2D [13, 18, 19], and 2.5D [20, 21] methods according to their adopted motion models. The 2D methods estimate 2D transformations (e.g., affines or homographies) to represent the camera motion and smooth them for stable videos. 3D methods reconstruct 3D camera paths as well as 3D scene structures and smooth the 3D camera trajectory. The stabilized video is synthesized along the smoothed camera trajectory guided by the 3D scene structures. If 3D reconstructions are applicable for videos, the 3D methods often produce superior results compared with other methods. However, 3D reconstruction is fragile, especially for consumer-level videos. To be more practical, the 2.5D methods relax the requirement of full 3D reconstruction to some partial 3D constraints embedded in long feature tracks, such as epipolar [21] constraint or subspace constraint [20]. Liu *et al.* [12] extended the subspace stabilization to handle stereoscopic videos. They show that the low-rank subspace constraint for monocular video also holds for stereoscopic video and design a smoothing strategy to smooth feature tracks from two views.

[1] http://www.liushuaicheng.org/ICIP2016/stabilization/index.html

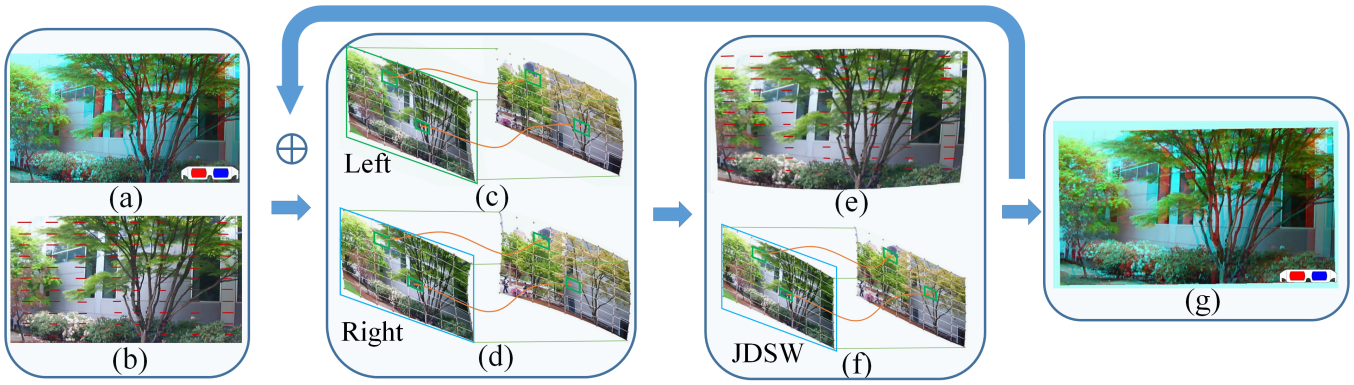


Fig. 1. Our system pipeline. (a) The input stereoscopic video. (b) Estimated disparities sampled for illustration. (c) and (d) are left and right videos, which are smoothed by bundled paths method [19] separately. (e) The disparities are warped according to [5]. (f) Without loss of generality, we begin by warp the right meshes using the JDSW. The process of (c,d,e,f) are iterated until both videos are stabilized enough. During each iteration, one view is warped according to JDSW, the other view is warped directly by stabilization. (g) shows the final result.

However, they rely on the long feature tracks which are hard to obtain when camera undergoes quick motions (e.g., fast panning, quick zooming). In this work, we adopt bundled camera paths approach [19], which is a 2D method that only requires feature matches between neighboring frames. It can achieve stabilization effects similar to the 3D methods while retains the efficiency and robustness of 2D methods. It handles scene parallax by dividing frames into several mesh grids, yielding multiple camera paths. We show that it can be successfully extended for stereoscopic videos by stabilizing two views jointly.

Stereoscopic disparity manipulation and maintain is crucial for high quality stereo image/video editing. Recent advances on the analysis of stereoscopic images have paved the way for our research. Image cloning methods [1, 2] copied the stereo content somewhere else and pasted into a new 3D scene compatibly. Lang *et al.* [22] and Lee *et al.* [11] exploited a locally adaptive algorithm and a disparity histogram for nonlinear disparity mapping, respectively. Wang *et al.* [6] developed a stereoscopic inpainting system for simultaneous color and depth recovery. Niu *et al.* [5] extended 2D image warping to 3D stereoscopic image warping. Du *et al.* [4] developed a method to change the views perspectively for stereo images. Didyk *et al.* [23] introduced a perceptual model of disparity for computer graphics and related applications. For depth coherence, we adopt method of [5] to warp disparities.

3. STEREOSCOPIC STABILIZATION

Figure 1 shows our pipeline. We calculate disparities between left and right videos and estimate bundled camera motions between neighboring frames within each video. The pipeline involves three types of operation, traditional monocular bundled paths stabilization, disparity warp and JDSW warp. For the clarification and completeness, we begin by briefly revisiting disparity warp [5] and bundled stabilization [19], following which we describe JDSW warp and finish the pipeline.

3.1. Disparity

Estimation. Per-pixel dense disparity estimation methods are well documented in [24], which is still a challenging vision problem [25]. We estimate disparities using sparse features [26]. We exclude outliers by homography-based RANSAC [27]. We further calculate dense optical flow [28] as disparity and sample them uniformly (every five pixels) to cover the textureless regions.

Warp disparities. The disparities are warped by minimizing the following energy [5]:

$$\sum_{d_i} \sum_{d_j \in N(d_i)} \|(\hat{d}_i - \hat{d}_j) - s_i(d_i - d_j)\|^2 \qquad (1)$$

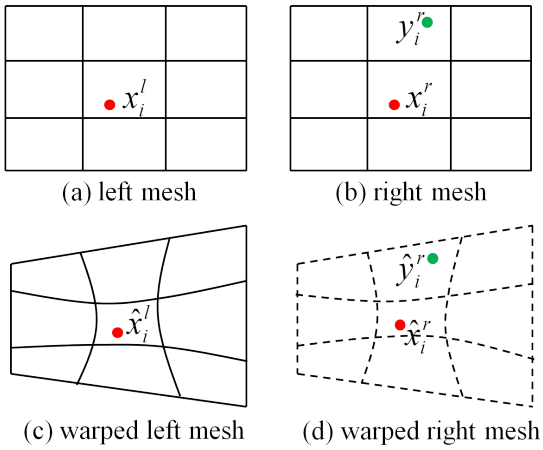
where d_i and $\hat{d}_i$ are original and warped disparities of a point i, respectively [2]. $N(\cdot)$ represents the neighboring pixels. The s_i is a scaling factor. It is obtained from a similarity transform H_{s_i} that fitted from neighboring pixels (3x3) of i before and after warping. The boundary condition is set to $\hat{d}_{min} = sd_{min}$, where d_{min} has the minimum magnitude. Notably, the warping function described in [5] is a user-specified warp. In our scenario, the warping comes from stabilization. For more details, please refer to [5].

3.2. Bundled-paths stabilization

Smooth a single path.

A single homography $F(t)$ is estimated between neighboring frames in the original video. The camera path is defined as consecutive multiplications of these homographies:

[2] We use $\hat{(\cdot)}$ to represent disparities or points in the warped coordinates.



(a) left mesh (b) right mesh (c) warped left mesh (d) warped right mesh

Fig. 2. The red dot and green dot denote disparity point and motion point, respectively. (c) the left mesh is warped from (a) by bundled stabilization. (d) the right mesh is warped from (b) according to our JDSW warping.

$C(t) = F(t)F(t-1)...F(1)F(0), F(0) = I$. Given the original path $\mathbf{C} = \{C(t)\}$, the smoothed path $\mathbf{P} = \{P(t)\}$ is obtained by minimizing the following energy:

$$\mathcal{O}(\{P(t)\}) = \sum_t \|P(t) - C(t)\|^2 + \sum_t (\lambda_t \sum_{r \in \Omega_t} \omega_{t,r}(C) \cdot \|P(t) - P(r)\|^2) \quad (2)$$

where Ω_t denotes the neighborhood at frame t. The strength of smoothing is controlled by λ_t. The smoothing kernel $w_{t,r}$ is a bilateral smoothing weight. The output video is obtained by applying a transform $B(t)$ to the input video, which is defined as $B(t) = C^{-1}(t)P(t)$.

Smooth bundled paths. Each frame is divided into a grid of 16×16 cells and camera paths are estimated for each cell. The estimation is based on the mesh warping, which warps the frame t to frame $t-1$. All Paths are smoothed simultaneously by a space-time optimization:

$$\sum_k \mathcal{O}(\{P_k(t)\}) + \sum_t \sum_{h \in N(k)} \|P_k(t) - P_h(t)\|^2, \quad (3)$$

where $N(k)$ includes eight neighbors of the grid cell k. This produces a warping transform $B_k(t)$ for each cell, which brings the original mesh to its desired position.

3.3. Notations

After bundled paths smoothing, we obtain $B_k^l(t)$ and $B_k^r(t)$ for both views. Let us omit the index t for simplicity. The left mesh can be warped using B_k^l as shown in Fig. 2 (a) and (c). However, we cannot directly warp right mesh using $B_k^r(t)$, which will damage the disparities between two views. An original disparity point pair is defined as (x_i^l, x_i^r), with disparity of d_i, shown as red dots in (a) and (b) of Fig. 2, where $d_i = [d_i, 0]^T$. Notably, We always use k,h to index mesh cells and i,j to index points. We further denote $\hat{x}_i^l$ as transformed point of x_i^l using corresponding B_k^l ($\hat{x}_i^l = B_k^l * x_i^l$). The disparity point $\hat{x}_i^r$ (Fig. 2 (d)) can be obtained using $\hat{x}_i^l$ and corresponding warped disparities $\hat{d}_i$ ($\hat{x}_i^r = \hat{x}_i^l + \hat{d}_i$). We further denote motion points as y_i^r, which are uniformly sampled (skip five pixels) in the right frame (green dot in Fig. 2 (b)). The warped motion point of right view $\hat{y}_k^r$ (Fig. 2 (d)) can be obtained using B_k^r as $\hat{y}_i^r = B_k^r * y_i^r$.

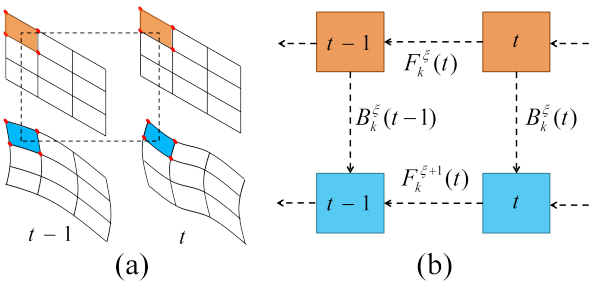


Fig. 3. Mesh configuration and motion relationship. (a) meshes (left or right) at $t-1$ and t before and after warp. (b) The unknown motions $F_k^{\xi+1}(t)$ can be derived as $(B_k^\xi(t))^{-1} F_k^\xi(t) B_k^\xi(t-1)$.

The green dot $\hat{y}_i^r$ indicates stabilized position that the right mesh should move to, while the red dot $\hat{x}_i^r$ encodes the correct disparities which should be satisfied. We seek for a mesh warping that best satisfies two requirements (dotted mesh in Fig. 2 (d)). Note that there are many disparity points and motion points, we only show one of each in Fig. 2 for illustration.

3.4. Joint Disparity and Stability mesh Warp (JDSW)

The joint mesh warp achieves both stability and disparity coherence. Let V denotes mesh vertices for the input mesh. We warp the mesh by optimizing over the following energy:

$$E(\hat{V}) = \lambda_1 E_d(\hat{V}) + \lambda_2 E_s(\hat{V}) + \lambda_3 E_r(\hat{V}) \quad (4)$$

where $\hat{V}$ are the unknown mesh vertices. $E_d(\hat{V})$, $E_s(\hat{V})$ and $E_r(\hat{V})$ account for disparity term, stability term and similarity term respectively, with λ_1, λ_2 and λ_3 being the associated weights. We set $\lambda_1 = 5, \lambda_2 = 1, \lambda_3 = 1$ in our system.

Disparity term. We can obtain pair of point constraint $(x_i^r, \hat{x}_i^r)$ through initial disparity correspondences (x_i^l, x_i^r), the warped disparity $\hat{d}_i$ and warping transform B_k^l, where $\hat{x}_i^r$ is calculated as $\hat{x}_i^r = B_k^l * x_i^l + \hat{d}_i$. We represent each point x_i^r by its 2D bilinear interpolation of four vertexes $V_i = [v_i^1, v_i^2, v_i^3, v_i^4]$ of enclosing grid cell: $x_i^r = V_i w_i$, where $w_i = [w_i^1, w_i^2, w_i^3, w_i^4]^T$ are the interpolation weights that

Fig. 4. Our results on various challenging videos with different scene types and camera motion.

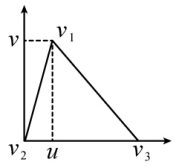
sum to 1. For the warped vertexes $\hat{V}_i = [\hat{v}_i^1, \hat{v}_i^2, \hat{v}_i^3, \hat{v}_i^4]$, we hope the same weights can be used to represent $\hat{x}_k^r$ after warping. The disparity term is defined as:

$$E_d(\hat{V}) = \sum_i \|\hat{V}_i w_i - B_k^l * x_i^l - \hat{d}_i\| \quad (5)$$

Stability term. Similarly, the motion points y_i^r are represented by their corresponding enclosing grid cell as $y_i^r = V_i w_i$. The stability term is then defined as:

$$E_s(\hat{V}) = \sum_i \|\hat{V}_i w_i - B_k^r * y_i^r\| \quad (6)$$

Rigidity term. It enforces spatial smoothness during mesh deformation. Each grid cell is divided into two triangles. For each triangle, v_1 can be represented by the other two vertices v_2 and v_3 in a local coordinate system. Let (u, v) be the local coordinates of v_1. We encourage the $\hat{v}_1$ to be still represented by $\hat{v}_2$ and$\hat{v}_3$ under the same local coordinates after warping. The following distance should be minimized respect to $\hat{v_1}, \hat{v_2}$ and $\hat{v_3}$ [15, 29]:



$$\|\hat{v}_1 - (\hat{v}_2 + u(\hat{v}_3 - \hat{v}_2) + vR_{90}(\hat{v}_3 - \hat{v}_2))\|^2, \quad (7)$$

where u, v are the same values computed before warping, $R_{90} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. $E_r(\hat{V})$ consists of all such cost from all triangles collected from every grid cells. The Equ. 4 is quadratic and can be minimized by a sparse linear system.

3.5. Iterations

We obtain warped meshes for two views in previous step, with left meshes obtained by bundled smoothing and right meshes generated by JDSW. Now, we need to apply them to the next iteration as described in Fig. 1 (c,d,e,f). As shown in Fig. 3, the motions between neighboring frame at iteration $\xi + 1$ of cell k can be computed as: $F_k^{\xi+1}(t) = B_k^{\xi}(t)^{-1} F_k^{\xi}(t) B_k^{\xi}(t-1)$. It is then applied to Equ. 3 for bundled smoothing of next iteration. We also update the disparity In the implementation, JDSW is applied to the left and right meshes alternately. (e.g., at iteration ξ, JDSW is applied to the right meshes, at $\xi + 1$ to the left meshes). Empirically, our method converges in 5 iterations.

4. RESULTS

We run our method on an Intel i5 3.1GHZ Quad-Core machine with 8G RAM. For a stereoscopic video of 640×480 resolution, our algorithm can achieve the speed of 1.5fps. Fig. 4 shows several results produced by our method. Please refer to the videos in the supplementary file. The video is best viewed in 3D eye glasses. Some of these videos have challenging camera motions,(e.g., quick rotation) and scene types, (e.g., large depth variations, dynamic scenes). Thanks to the advantages of bundled stabilization, we can stabilize them successfully. We also show a comparison with subspace stereoscopic video stabilization method [12]. As we do not require long feature tracks, we can stabilize videos with fast camera motions, where the length of the feature tracks drops quickly, which challenges the subspace method. We also show that our method can maintain more visual contents compared with [12] in supplementary video.

5. CONCLUSION

We have presented a method for stereoscopic video stabilization, which combines the merits of bundled paths stabilization [19] and disparity preserving warping [5]. We proposed a novel warping method JDSW which jointly considers disparities and stabilities in mesh warping during the stabilization. We validate our method on various casual captured videos.

6. ACKNOWLEDGE

This work has been supported by National Natural Science Foundation of China (61502079, 61370148 and 61300091).

References

[1] S. Luo, I. Shen, B. Chen, W. Cheng, and Y. Chuang, "Perspective-aware warping for seamless stereoscopic image cloning," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 182:1–182:8, 2012.

[2] W. Lo, J. Baar, C. Knaus, M. Zwicker, and M. Gross, "Stereoscopic 3d copy & paste," vol. 29, no. 6, pp. 147, 2010.

[3] R. Tong, Y. Zhang, and K. Cheng, "Stereopasting: interactive composition in stereoscopic images," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 8, pp. 1375–1385, 2013.

[4] S. Du, S. Hu, and R. Martin, "Changing perspective in stereoscopic images," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 8, pp. 1288–1297, 2013.

[5] Y. Niu, W. Feng, and F. Liu, "Enabling warping on stereoscopic images," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 183, 2012.

[6] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images," in *Proc. CVPR*, 2008, pp. 1–8.

[7] T. Mu, J. Wang, S. Du, and S. Hu, "Stereoscopic image completion and depth recovery," *The Vis. Comput.*, vol. 30, no. 6-8, pp. 833–843, 2014.

[8] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic stereo imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 279–290, 2001.

[9] F. Zhang and F. Liu, "Casual stereoscopic panorama stitching," in *Proc. CVPR*, 2015.

[10] T. Basha, Y. Moses, and S. Avidan, "Geometrically consistent stereo seam carving," in *Proc. ICCV*, 2011, pp. 1816–1823.

[11] K. Lee, C. Chung, and Y. Chuang, "Scene warping: Layer-based stereoscopic image resizing," in *Proc. CVPR*, 2012, pp. 49–56.

[12] F. Liu, Y. Niu, and H. Jin, "Joint subspace stabilization for stereoscopic video," in *Proc. ICCV*, 2013, pp. 73–80.

[13] M. Grundmann, V. Kwatra, and I. Essa, "Auto-directed video stabilization with robust l1 optimal camera paths," in *Proc. CVPR*, 2011, pp. 225–232.

[14] S. Liu, L. Yuan, P. Tan, and J. Sun, "Steadyflow: Spatially smooth optical flow for video stabilization," in *Proc. CVPR*, 2014, pp. 4209–4216.

[15] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3d video stabilization," *ACM Trans. Graph.*, vol. 28, pp. 44, 2009.

[16] S. Liu, Y. Wang, L. Yuan, J. Bu, P. Tan, and J. Sun, "Video stabilization with a depth camera," in *Proc. CVPR*, 2012, pp. 89–95.

[17] M. Gleicher and F. Liu, "Re-cinematography: improving the camera dynamics of casual video," in *ACM Conf. Mult.*, 2007, pp. 27–36.

[18] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H. Shum, "Full-frame video stabilization with motion inpainting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 1150–1163, 2006.

[19] S. Liu, L. Yuan, P. Tan, and J. Sun, "Bundled camera paths for video stabilization," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 78, 2013.

[20] F. Liu, M. Gleicher, J. Wang, H. Jin, and A. Agarwala, "Subspace video stabilization," *ACM Trans. Graph.*, vol. 30, no. 1, pp. 4, 2011.

[21] A. Goldstein and R. Fattal, "Video stabilization using epipolar geometry," *ACM Trans. Graph.*, vol. 31, no. 5, 2012.

[22] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3d," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 75, 2010.

[23] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H. Seidel, "A perceptual model for disparity," vol. 30, no. 4, pp. 96, 2011.

[24] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision (IJCV)*, vol. 47, no. 1-3, pp. 7–42, 2002.

[25] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1582–1599, 2009.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004.

[27] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA, 2 edition, 2003.

[28] C. Liu, *Beyond pixels: exploring new representations and applications for motion analysis*, Ph.D. thesis, MIT, 2009.

[29] T. Igarashi, T. Moscovich, and J. Hughes, "As-rigid-as-possible shape manipulation," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 1134–1141, 2005.