3DVFX: 3D Video Editing using Non-Rigid Structure-from-Motion

Parashar Shaifali, Bartoli Adrien

EUROGRAPHICS 2019 / P. Alliez and F. Pellacini
(Guest Editors)

Volume 38 (2019), Number 2

3DVFX: 3D Video Editing using Non-Rigid Structure-from-Motion

S. Parashar[†1] and A. Bartoli[‡1]

[1]EnCoV, Institut Pascal, Université Clermont-Auvergne, France

a) Original images

b) Two red blobs added using 3DVFX

Figure 1: *3DVFX use case. The user provides the object to be added and marks the desired location in one of the images. The rest of the images are modified by transporting the texture across the 3D meshes of the reconstructed surfaces and reprojecting them to the images.*

Abstract

Numerous video post-processing techniques can add or remove objects to the observed scene in the video. Most of these techniques rely on 2D image points to perform the desired changes. Structure-from-Motion (SfM) has allowed the use of 3D points, however only for the objects that remain rigid in the scene. We propose to use both 2D image points and 3D points to modify the scene's deformable objects using Non-Rigid Structure-from-Motion (NRSfM). We rely on a recent effective NRSfM solution to develop a complete pipeline including manual 3D editing of an image and automatic 3D transfer of the edits. We perform object manipulation tasks such as retexturing a real deforming object.

CCS Concepts

• ***Computing methodologies*** → ***Reconstruction; Texturing; Mixed / augmented reality;***

1. Introduction

The entertainment industry increasingly relies on VFX, which adds special effects in video post-processing. For rigid scenes, VFX techniques can use the 2D image data and the 3D scene structure which can be reliably computed using SfM [HZ00]. SfM exploits the inter-image visual motion between monocular images taken from different viewpoints. It has been studied in computer vision for at least the past four decades and evolved to mature solutions. However, SfM breaks for non-rigid or deformable objects. Almost two decades ago, it was generalized to NRSfM [BHB00, ASKK09, GM11], which exploits deformation constraints and reconstructs multiple 3D structures, one per input image. Until recently, NRSfM methods were not as reliable as SfM ones in terms of speed, accuracy and stability, which limited their use in video post-processing. Hence, when dealing with deformable objects in the scene, current video post-processing methods either rely only on the 2D image points or use SfM along with mesh-editing techniques to develop a work-around, thus adding to the cost of VFX.

† shaifali.parashar@gmail.com
‡ adrien.bartoli@gmail.com

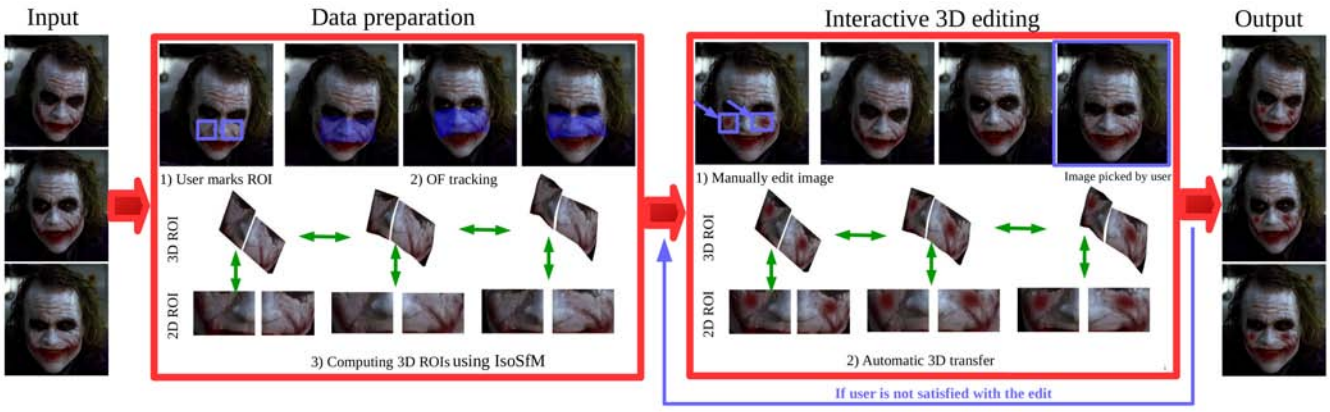

Figure 2: *3DVFX pipeline. First we prepare 2D and 3D data for the ROI marked by the user in one of the images. Then, the user edits one of the images in 2D or 3D. The edit is always reflected in the corresponding 3D. The manual 3D edit is automatically transferred to the rest of the 3D point clouds. OF stands for optical flow.*

We recently proposed an NRSfM solution [PPB18], called IsoSfM, for the objects deforming isometrically, i.e., whose geodesic distances are preserved throughout the deformation. A piece of paper or cloth are typical examples. Many objects in nature undergo a near-isometric deformation and therefore, they can be studied under isometry. IsoSfM uses a local framework which exploits the first and the second-order properties of the surfaces. It obtains a local solution to NRSfM by solving a system of bivariate cubic polynomials. It has been shown to overcome most of the shortcomings of other NRSfM methods. It is significantly more accurate and applicable to both short and wide-baseline images. It has a linear complexity in terms of number of images and number of points and therefore it is extremely fast compared to other methods. It requires very few images to perform a stable reconstruction and can handle a large number of images as well. Since it is a local method, it works with missing data.

We present 3DVFX, a pipeline shown in figure 2, to create VFX on videos with minimal user intervention. It uses IsoSfM to reconstruct 3D point clouds from the videos and enables the user to manually edit one of the reconstructed 3D point clouds and automatically transfers the edit to the rest of the 3D point clouds. The underlying process is extremely fast. The user can make multiple attempts at editing different images to pick the one that gives a desirable visuals on all the images. Figure 1 shows a use case of 3DVFX where we added two red blobs on the deformable face of Joker from the movie *The Dark Knight*.

2. Background

The use of 3D scene structure is common in VFX. The current methods may be categorized into the following three categories depending on how they obtain the 3D structure: *1)* Methods that assume rigidity and use SfM to build the 3D scene. [LHE*07], for example, uses a coarse 3D layout to add objects to the images. *2)* Methods that fit an existing rigid 3D model of the object by manually adjusting some global parameters. [HRC*18], for example, creates isometric deformations of the object by fitting a 3D model to the image. *3)* Lastly, methods that use an object-specific 3D Morphable Model (3DMM). [BSVS04], for example, exchanges the faces in two images by fitting them by a 3DMM, morphing the two models and rendering them to the images. However, 3DMMs are overly constrained and produce only a coarse geometry. They cannot be used for adding VFX on faces as it requires a highly precise 3D in order to augment finer details. Hence, manipulating generic deformable objects is still an open problem. For simple objects like faces, there are many mobile applications such as Snapchat which perform manipulations like adding makeup or accessories such as glasses. These applications fail due to tilting or turning of the head. They also cannot cope with the complex movements of the mouth and cheeks. Therefore, the industry mainly relies upon the manual editing of images in order to perform even simple VFX on deformable objects, which is extremely expensive. In contrast, the proposed 3DVFX pipeline uses IsoSfM to perform deformable object manipulation reliably with low manual and computational cost.

3. IsoSfM: Isometric (Non-Rigid) Structure-from-Motion

We give a short description of IsoSfM [PPB18] which reconstructs time-varying 3D structures from monocular images. It formulates deformation and reprojection constraints which force the surfaces to be related with each other by a 3D isometric transformation while fixing their projections in the images. It considers image point correspondences as input, which can be computed automatically using optical flow (OF) [SBK10] or keypoint matching methods such as SIFT [Low04] and outputs up-to-scale 3D point clouds. It evaluates a grid from the input correspondences which allows one to handle the missing data and also reduces the computation time.

IsoSfM uses the concepts of Metric Tensors (MT) and Christoffel Symbols (CS) from Riemannian geometry to draw local isometric constraints on surfaces by assuming them to be infinitesimally planar (IP). An IP surface can be assumed to be planar at a local level while it maintains its curvature at the global level. An MT is a physical quantity that describes the metric properties such as lengths, angles and areas on the surfaces. A CS is related to the

Figure 3: *2D vs 3D editing. 2D editing is unrealistic as it does not consider the underlying geometry of the object.*

curvature of the surfaces. Both MT and CS are preserved under an isometric deformation.

Given a surface $\mathbf{r}(u,v)$ where (u,v) are the normalized image coordinates, an MT is expressed in terms of the first-order derivatives, $\mathbf{r}_u$ and $\mathbf{r}_v$ of the surface. A CS is expressed in terms of the second-order derivatives, $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$ and $\mathbf{r}_{vv}$, in addition to the first-order derivatives of the surfaces. By expressing $\mathbf{r}(u,v) = \mathbf{z}^{-1}(u \quad v \quad 1)^\top$, where $\mathbf{z}$ is the inverse-depth function, both MT and CS can be expressed only in terms of $\mathbf{z}$ and its first-order derivatives, $\mathbf{z}_u$ and $\mathbf{z}_v$, as the second-order derivatives of $\mathbf{z}$ are neglected due to IP. $\mathbf{z}$ appears as a scaling factor in the expressions of MT and CS which are respectively quadratic and linear in terms of $\mathbf{z}_u$ and $\mathbf{z}_v$. Given two surfaces $\mathbf{r}^1$ and $\mathbf{r}^2$, the respective MT can be obtained in the following ways:

- Due to isometry, the MT at $\mathbf{r}^2$ can be expressed in terms of the one at $\mathbf{r}^1$.
- Due to isometry, the CS at $\mathbf{r}^2$ can be expressed in terms of the one at $\mathbf{r}^1$. Since it is a linear relationship, it allows one to express $(\mathbf{z}^2_u, \mathbf{z}^2_v)$ in terms of $(\mathbf{z}^1_u, \mathbf{z}^1_v)$. The $(\mathbf{z}^2_u, \mathbf{z}^2_v)$ thus obtained can be used to express the MT at $\mathbf{r}^2$.

By equating the MT at $\mathbf{r}^2$ obtained using these two ways, two cubic equations in terms of $(\mathbf{z}^1_u, \mathbf{z}^1_v)$ are obtained. IsoSfM accumulates these equations over all possible pairs by fixing one of the surfaces as $\mathbf{r}^1$ and minimizes their sum-of-squares to obtain the surface normals, which are eventually integrated to obtain an up-to-scale depth.

4. 3DVFX: VFX on Deformable Objects

The 3DVFX pipeline, shown in figure 2, consists of two blocks: *(a)* data preparation where image correspondences and 3D structure are computed and *(b)* interactive editing where the user edits the video.

4.1. Data preparation

1) Mark ROI. The user picks an image and marks the ROI.

2) OF tracking. Using the image from previous step as a reference, we use OF to obtain ROI point correspondences between the reference and next image in the video sequence. We then consider the recently treated image as reference and obtain ROI point correspondences on another image, thus sequentially establishing point correspondences between all images in the ROI.

3) Computing 3D ROIs. We obtain a 3D ROI for each image using IsoSfM. IsoSfM assumes that the camera is calibrated. If the calibration is unknown, we compute it either using SfM [Pho14] on the rigid part of the scene or we use [PBP18], which obtains camera calibration from deformable parts in the video.

4.2. Interactive editing

1) Manual isometric 3D editing. The user picks an image and edits its 2D or 3D ROI. The edit is simultaneously reflected to the 3D.

2) Automatic isometric 3D transfer. We automatically transfer the 3D edit to the remaining 3D ROIs using the isometric transformation computed in the previous block and reproject them back to the images.

At this stage, due to IsoSfM, all 3D ROIs are related to their corresponding 2D ROIs and other 3D ROIs as well. So, the edit made to one of them can immediately be reflected to the rest. Also, the edits on 3D ROIs are dependent on the 2D or 3D ROI the user manually edited. Hence, the user may execute this block several times by changing the image to be manually edited until a satisfying result on all 3D ROIs is obtained. An optional step may be to post-process the rendered images to add softness or sharpness to the edits.

5. Experimental Evaluation

We used 3DVFX to add VFX to some clips from the movie *The Dark Knight*, *Fantastic Beasts and Where to Find Them* and the TV show *The Big Bang Theory*. Figure 1 shows the result of adding red blobs on Joker's makeup. The added blobs adhere to the deformation of the face and look very realistic as they are added in 3D. Figure 3 shows the visual differences between the images edited without and with using the 3D data respectively. Unlike the 3D edit, the makeup added using only the 2D image data looks unrealistic as it does not follow the underlying deformation of the cheek.

Figure 4 shows the result of adding the Harry Potter mark on the Shaw's forehead printed on the flag which undergoes a large deformation. The added mark handles the cloth deformation and produces realistic results. Figure 5 shows the result of adding a moustache on Sheldon's face. The moustache follows the complex deformation of the mouth producing a visually realistic edit.

6. Conclusions

We presented 3DVFX, the first method to 3D edit the isometrically deformable objects in a video without using a shape prior. It is fast, precise and user intervention is limited to marking a ROI and manually editing only one of the frames. It uses IsoSfM, a recent and mature NRSfM method which produces reliable 3D reconstruction, which allows 3DVFX to augment finer details to the deformable objects. The experimental results are visually convincing even when there are complex deformations of the object. In future work, we would like to manipulate the light and shading in the deformable parts of the scene in order to deal with textureless surfaces. Also, we would like to extend this method to handle non-smooth and wrinkled surfaces.

a) Original images

b) Harry Potter mark added using 3DVFX

Figure 4: *Adding Harry Potter's mark to Shaw's forehead using 3DVFX.*

a) Original Images

b) Moustache added using 3DVFX

Figure 5: *Adding a moustache on Sheldon's face using 3DVFX.*

References

[ASKK09] AKHTER I., SHEIKH Y., KHAN S., KANADE T.: Nonrigid structure from motion in trajectory space. In *NIPS* (2009). 1

[BHB00] BREGLER C., HERTZMANN A., BIERMANN H.: Recovering non-rigid 3D shape from image streams. In *CVPR* (2000). 1

[BSVS04] BLANZ V., SCHERBAUM K., VETTER T., SEIDEL H.-P.: Exchanging faces in images. In *Computer Graphics Forum* (2004). 2

[GM11] GOTARDO P. F., MARTINEZ A. M.: Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 10 (2011), 2051–2065. 1

[HRC*18] HAOUCHINE N., ROY F., COURTECUISSE H., NIESSNER M., COTIN S.: Calipso: Physics-based Image and Video Editing through CAD Model Proxies. *Visual Computer* (2018). 2

[HZ00] HARTLEY R. I., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000. 1

[LHE*07] LALONDE J.-F., HOIEM D., EFROS A. A., ROTHER C., WINN J., CRIMINISI A.: Photo clip art. In *ACM SIGGRAPH 2007 Papers* (2007). 2

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110. 2

[PBP18] PARASHAR S., BARTOLI A., PIZARRO D.: Self-calibrating Isometric Non-Rigid Shape-from-Motion. 3

[Pho14] Agisoft Photoscan 1.0.4, 2014. URL: http://www.agisoft.ru/products/photoscan. 3

[PPB18] PARASHAR S., PIZARRO D., BARTOLI A.: Isometric Non-Rigid Shape-from-Motion with Riemannian Geometry Solved in Linear Time. *IEEE Transactions on Pattern Analysis and Machine Intelligence 40*, 10 (2018), 2442 – 2454. 2

[SBK10] SUNDARAM N., BROX T., KEUTZER K.: Dense point trajectories by GPU-accelerated large displacement optical flow. In *ECCV* (2010). 2