Bundled Camera Paths for Video Stabilization

Liu et. al., ACM Trans. Graph. 2013

2021.10.07 임은우

Index

Video Stabilization

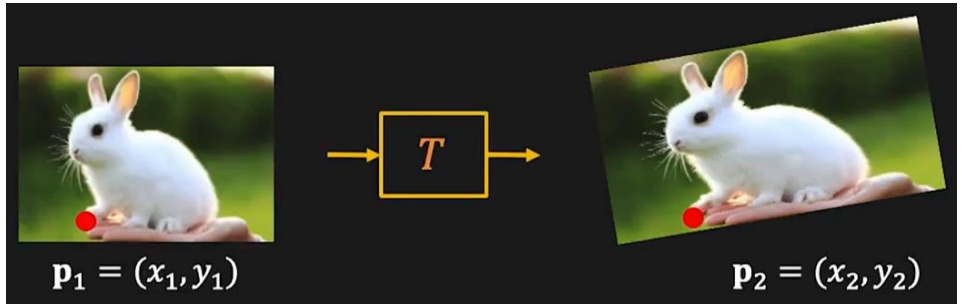
1. Improves the video quality by removing unwanted camera motion.

2. Goal: Stabilizing videos without additional hardware assisted equipment but with digital video editing software

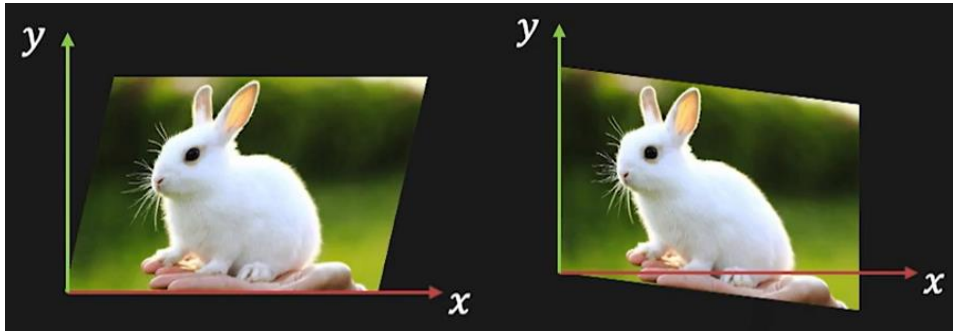
3. Understanding motion pattern and stable frame generation

4. How do we evaluate the performance of video stabilization model?

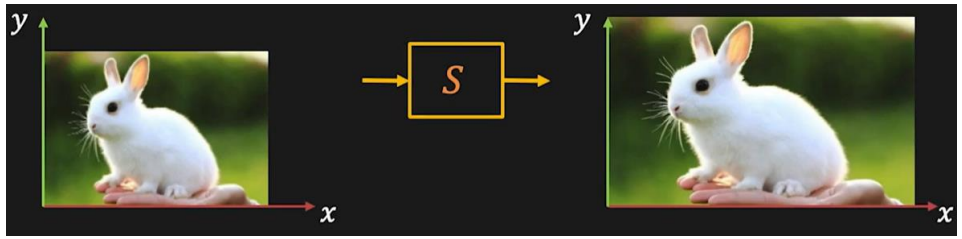
Linear Transformation



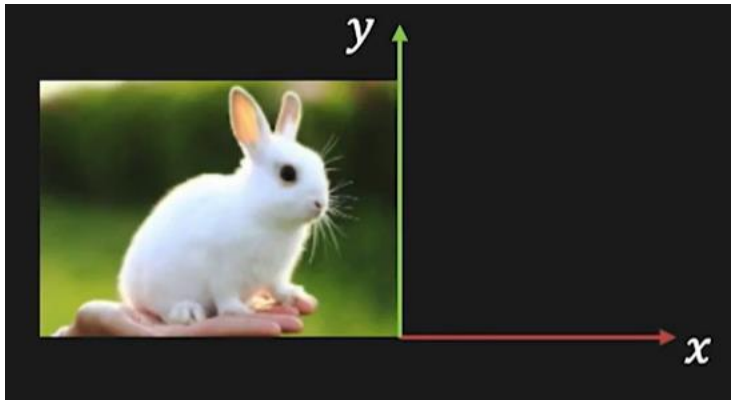
Rotation



Skew



Scaling

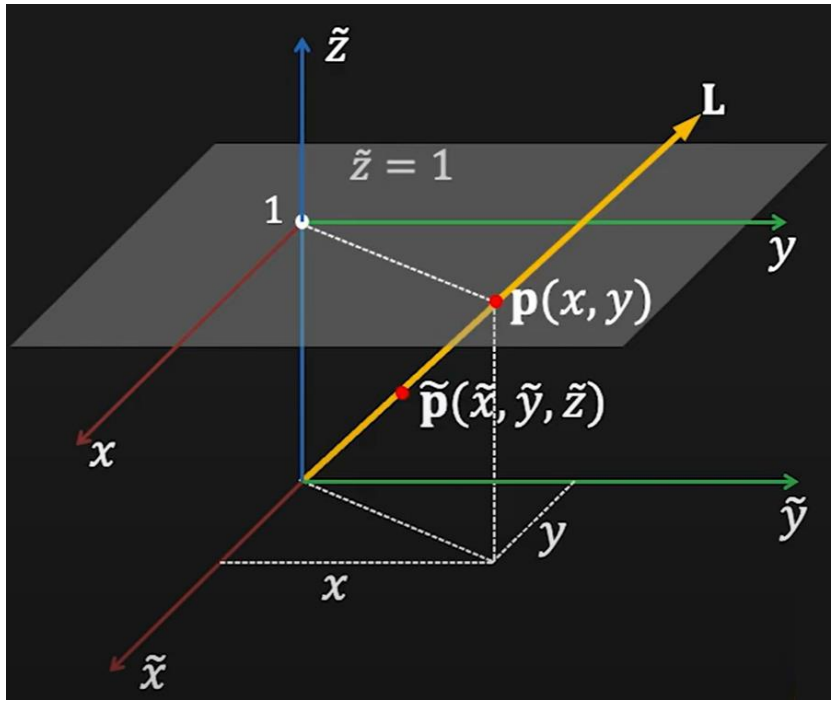


Mirroring

Homogeneous Coordinates

- The homogeneous representation of 2D point $p = (x, y)$ is a 3D point $\tilde{p} = (\tilde{x}, \tilde{y}, \tilde{z})$ where $\tilde{z} \neq 0$ such that $x = \frac{\tilde{x}}{\tilde{z}}, y = \frac{\tilde{y}}{\tilde{z}}$

$$p \equiv \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{z}x \\ \tilde{z}y \\ \tilde{z} \end{bmatrix} \equiv \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \tilde{p}$$



- Every point on line L (except origin) represent the homogeneous coordinate of $p(x, y)$

- Affine Transformation

Parallel remains parallel

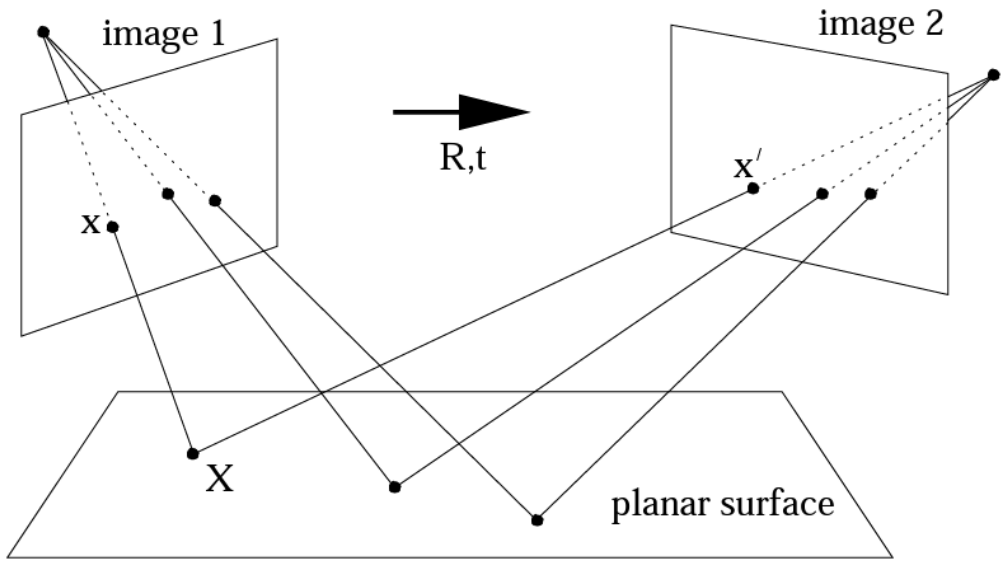
$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$$

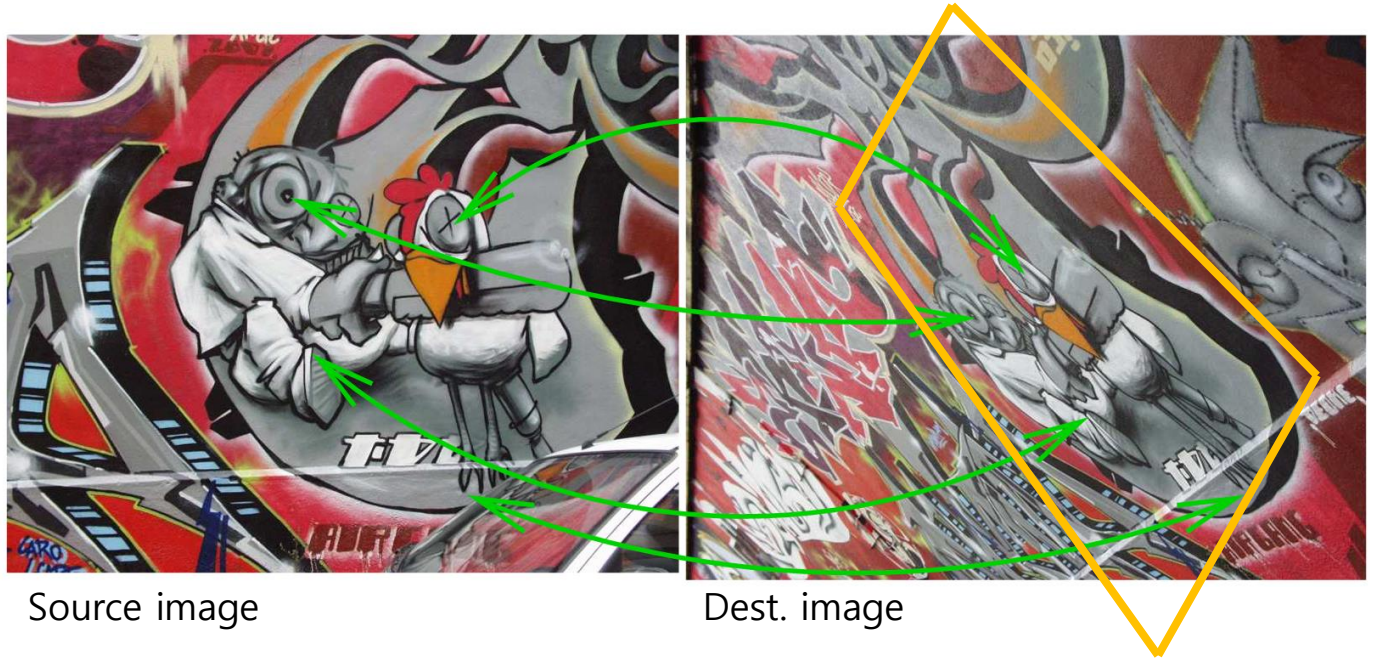
- Projective Transformation

Parallel does not remain parallel

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$$

Homography





Source image

Dest. image

Shaky video frames can be related with
Stable video frames as homography!

Local homography is called spatially-variant path

How can we compute homography?

$$\begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = H \begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix}$$

Metrics - Distortion

1. Compute warping transform $B(t) = C^{-1}(t)P(t)$ by SVD decomposition where $P(t)$ is optimal path, $C(t)$ is original path.

2. The anisotropic scaling measures distortion

3. Can be computed by the ratio of the two largest eigenvalues of the affine part of $B(t)$

4. Higher score signifies better preservation of the content.

   No shearing/skewing/wobble

5. Should be close to 1

Ground Truth

Distorted

Liu et al. Bundled Camera Paths for Video Stabilization. *ACM Trans. Graph.* 2013

Metrics - Cropping

1. Fit a global homography at each frame between input and output videos.
2. Compute the cropping ratio for each frame

   The cropping ratio can be directly computed from the scale component of the homography, and there is one global cropping for the whole sequence, and each frame provides an estimation.
3. Average these estimations at all frame as a final metric.
4. Measures preservation of visual information in generated frames.
5. Should be close to 1

Ground Truth

Much Cropped

Less Cropped

Liu et al. Bundled Camera Paths for Video Stabilization. *ACM Trans. Graph.* 2013

Metrics - Stability

1. Estimate paths to approximate the true motion (optical flow) in a video.

2. Extract translation and rotation components (1D temporal signal) from each path.

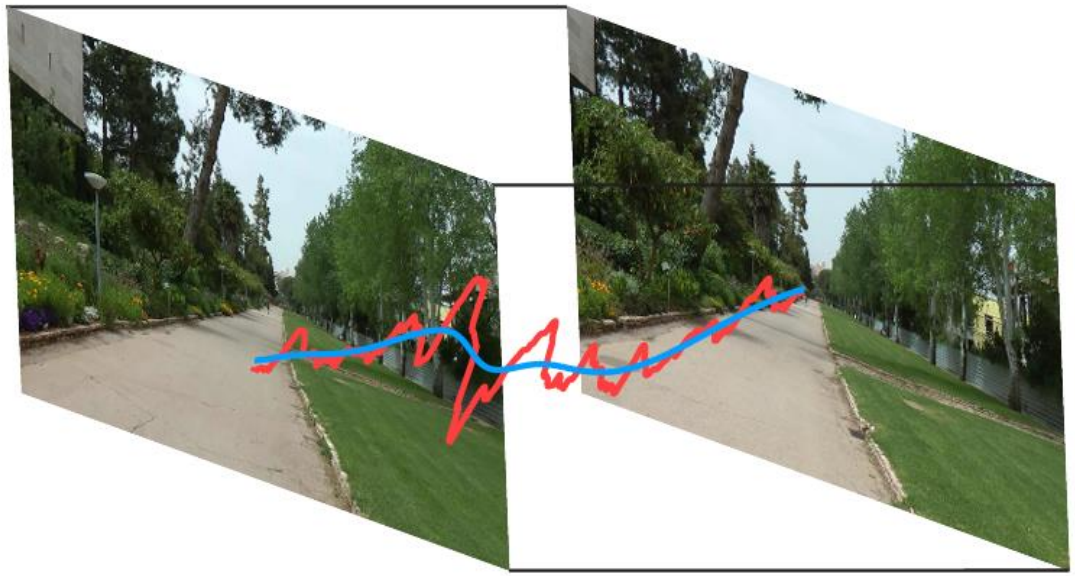
3. Apply 1D discrete Fourier Transformation.

4. Evaluate the energy percentage of the low frequency components (expect for DC components, 2~6 lowest frequencies over full frequencies).

5. Frequency analysis on estimated 2D motions from a video.

   More energy is contained in low frequency part of motion, the more stable a video is.

6. Should be close to 1



Liu et al. Bundled Camera Paths for Video Stabilization. *ACM Trans. Graph.* 2013

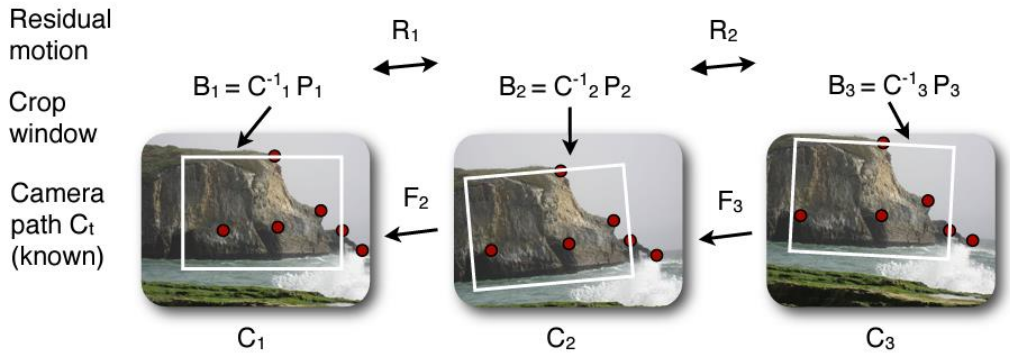
2D based Approaches

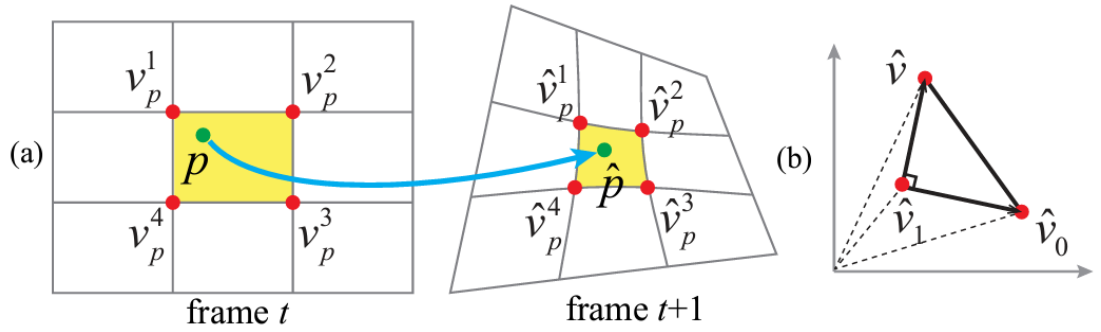
1. Track prominent features and stabilize their trajectories along the motion path.

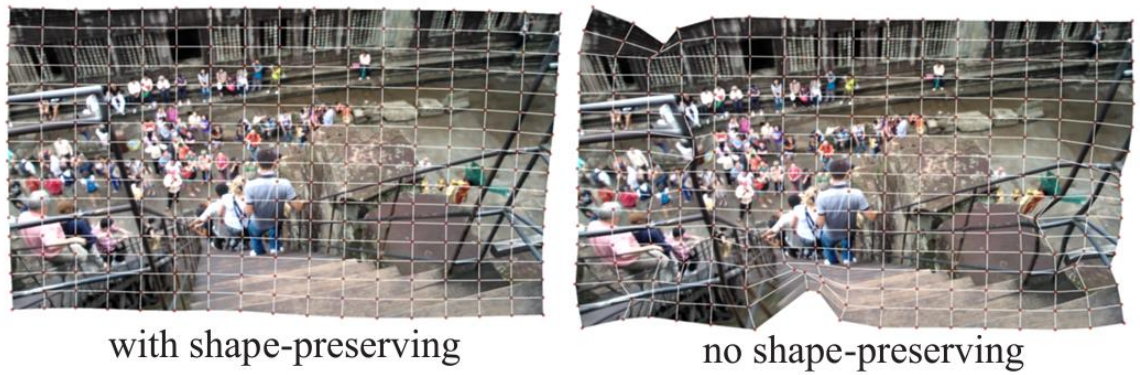
2. Smooth the 2D linear transformations (e.g. affinity, homography) estimated by consecutive frames.

3. Require low computational complexity in general.

4. Usually suffer from tackling parallax effects as 2D transformations are insufficient to model the entire 3D scene structures.

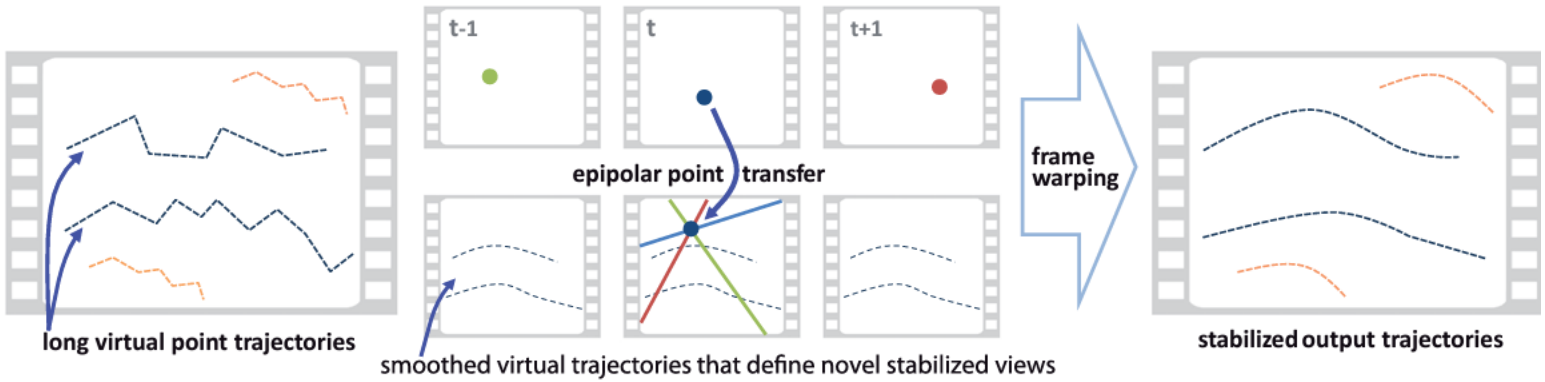






Grundmann et al. Auto-Directed Video Stabilization with Robust L1 Optical Camera Paths. In *CVPR*, IEEE, 2011
Liu et Al. Bundled Camera Paths for Video Stabilization. *ACM Trans. Graph.* 2013

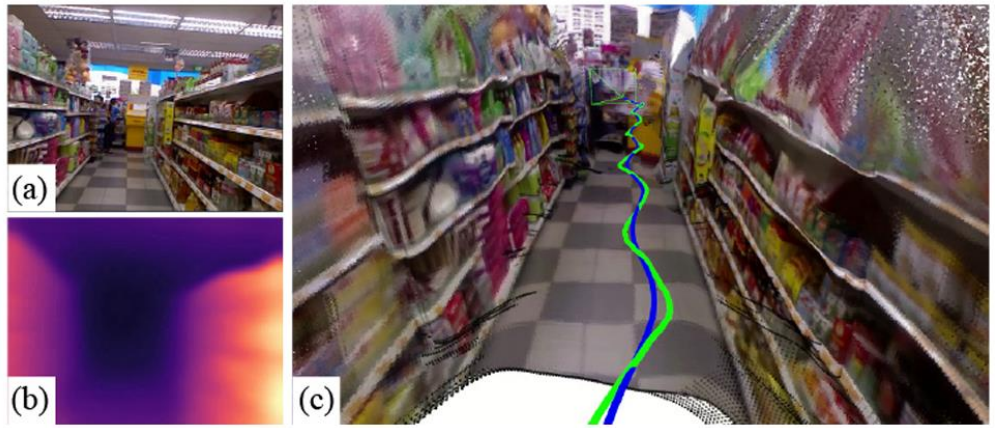
3D based Approaches



1. Camera trajectories and feature positions were projected along with feature tracks are reconstructed in a 3D space

2. Cannot handle dynamic scenes including moving objects

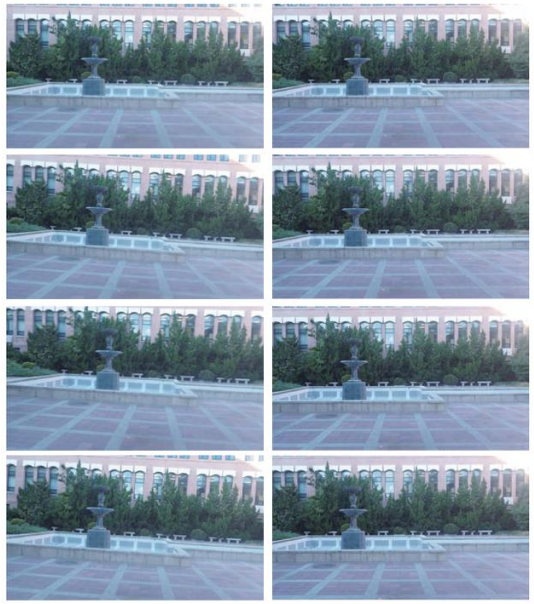
3. Require expensive costs for 3D reconstruction

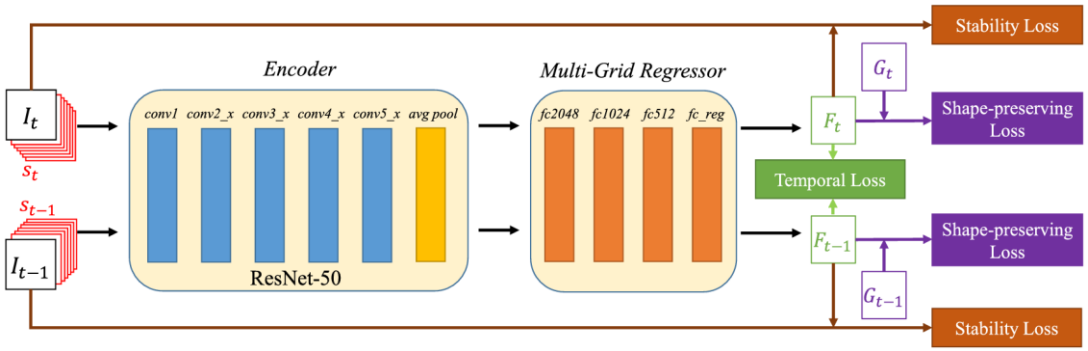


A. Goldstein and R. Fattal. Video stabilization using epipolar geometry. *ACM TOG*
Lee et al. 3D Video Stabilization with Depth Estimation by CNN-based Optimization. In *CVPR* 2021

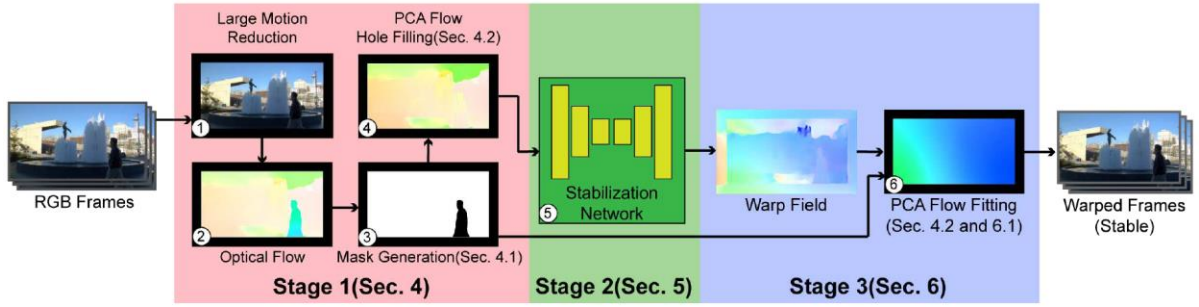
Deep Learning Approaches

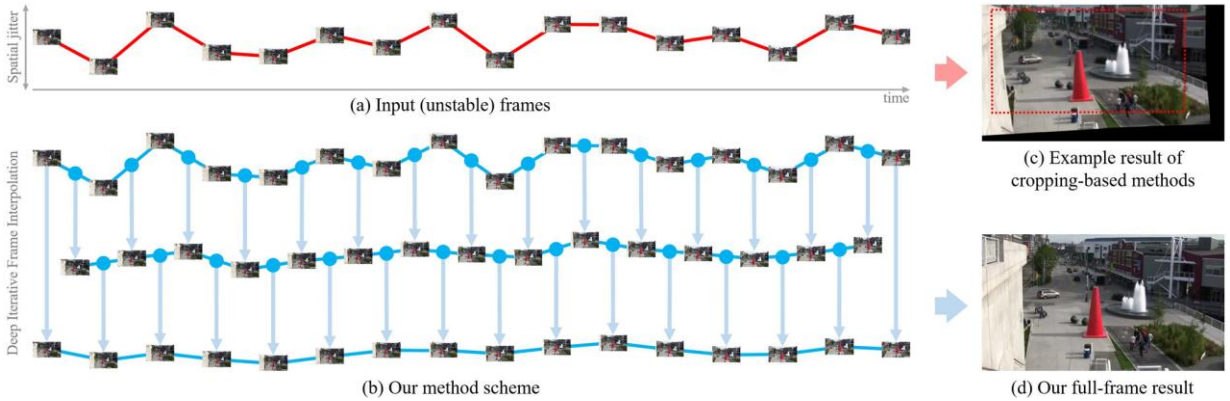
Supervised Learning





Self-Supervised Learning





Wang et. al. Deep online video stabilization with multi-grid warping transformation learning. *IEEE TIP*, 2019
Choi et al. Deep iterative frame interpolation for full-frame video stabilization. In *SIGGRAPH Asia*, 2019

Thank you ☺