EgoSampling: Fast-Forward and Stereo for Egocentric Videos

Yair Poleg[1], Tavi Halperin[1], Chetan Arora[2], Shmuel Peleg[1]
[1]The Hebrew University, Jerusalem, Israel. [2]IIIT, Delhi, India.

In a Nutshell

Motivation While egocentric cameras like GoPro are gaining popularity, the videos they capture are long, boring, and difficult to watch due to constant motion of the camera because of motion of wearer's head. Fast forwarding (i.e. frame sampling) is a natural choice for faster video browsing but accentuates the shake present in the videos, making the fast forwarded video useless. Using video stabilization as a pre or post processing drastically crops the video to compensate large left right head motion.

Fast Forward For Egocentric Videos We propose EgoSampling, an adaptive frame sampling that prefers forward looking frames. Sampled frames optimize both video stability and the fast forward constraints, producing stable fast forwarded videos. Adaptive frame sampling is formulated as energy minimization, whose optimal solution can be found in polynomial time. See figure Fig. 1 for a schematic illustration. We compute viewing direction based on fast and robust $2D$ motion models, avoiding $3D$ estimations.

Bonus - Turning Egocentric Video to Stereo Egocentric video taken while walking suffers from the left-right movement of the head as the body weight shifts from one leg to another. We turn this drawback into a feature: Stereo video can be created by sampling the frames from the left most and right most head positions of each step, forming approximate stereo-pairs.

Fast Forward - Problem Formulation and Inference

We model the joint fast forward and stabilization of egocentric video as an energy minimization problem. We represent the input video as a graph with a node corresponding to every frame in the video. There are weighted edges between every pair of graph nodes, i and j, with weight proportional to our preference for including frame j right after i in the output video (see Fig. 2). There are three components in this weight:

1. Shakiness Cost ($S_{i,j}$): This term prefers forward looking frames. The cost is proportional to the distance of the computed motion direction (Epipole or FOE) from the center of the image.
2. Velocity Cost ($V_{i,j}$): This term controls the playback speed of the output video. The desired speed is given by the desired magnitude of the optical flow, K_{flow}, between two consecutive output frames. As a consequence, periods with fast camera motions are sampled more densely and stationary periods, such as waiting at red light, may be skipped.
3. Appearance Cost ($C_{i,j}$): This is the Earth Movers Distance (EMD) between the color histograms of frames i and j. The role of this term is to prevent large visual changes between frames. A quick rotation of the head or dominant moving objects in the scene can confuse the FOE or epipole computation. The terms acts as an anchor in such cases, preventing the algorithm from skipping a large number of frames.

The overall weight of the edge between nodes (frames) i and j is:

$$\mathcal{W}_{i,j} = \alpha \cdot \mathcal{S}_{i,j} + \beta \cdot V_{i,j} + \gamma \cdot C_{i,j}, \quad (1)$$

where α, β and γ represent the relative importance of various costs in the overall edge weight.

With the problem formulated as above, sampling frames for stable fast forward is done by finding a shortest path in the graph.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

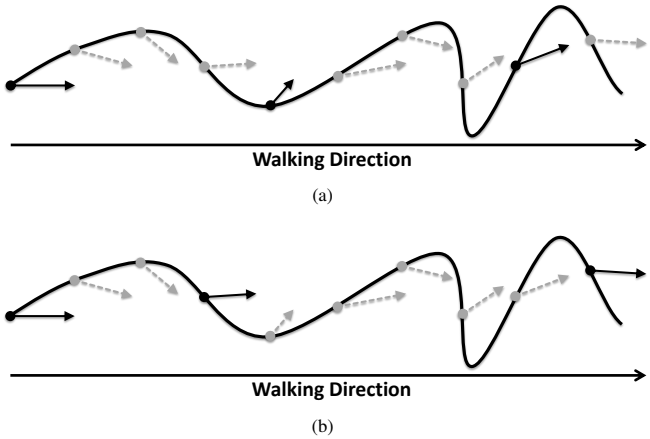

Figure 1: Frame sampling for Fast Forward. A view from above on the camera path (the line) and the viewing directions of the frames (the arrows) as the camera wearer walks forward during a couple of seconds. (a) Uniform $5\times$ frames sampling, shown with solid arrows, gives output with significant changes in viewing directions. (b) Our frame sampling, represented as solid arrows, prefers forward looking frames at the cost of somewhat non uniform sampling.

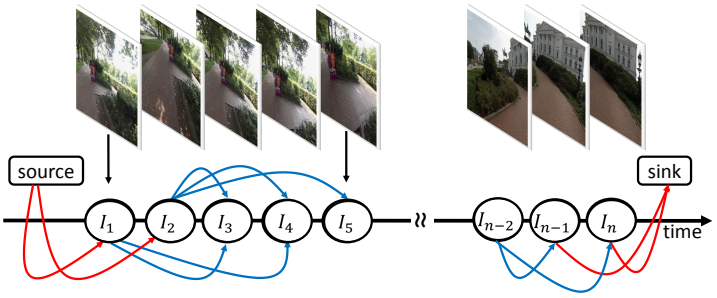

Figure 2: We formulate the joint fast forward and video stabilization problem as finding a shortest path in a graph constructed as shown. The edges between a pair of frames (i, j) indicate the penalty for including a frame j immediately after frame i in the output.

Turning Egocentric Video to Stereo

When walking, the head moves left and right as the body shifts its weight from the left leg to the right leg and back. Pictures taken during the shift of the head to the left and to the right can be used to generate stereo egocentric video.

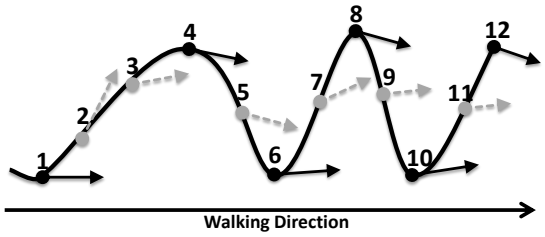

Figure 3: Frame sampling for Stereo: We pick the frames in which the wearer's head is in the right most position (frames 1,6,10) and left most position (frames 4,8,12) to form stereo pairs. Frame pairs (1,4), (6,8) and (10,12) form the output stereo video.