Transforming a Standard Video into Slow Motion using AI

A Seminar on Applications of AI by

Apurv Varshney, Piyush Doke & Rohan Teja

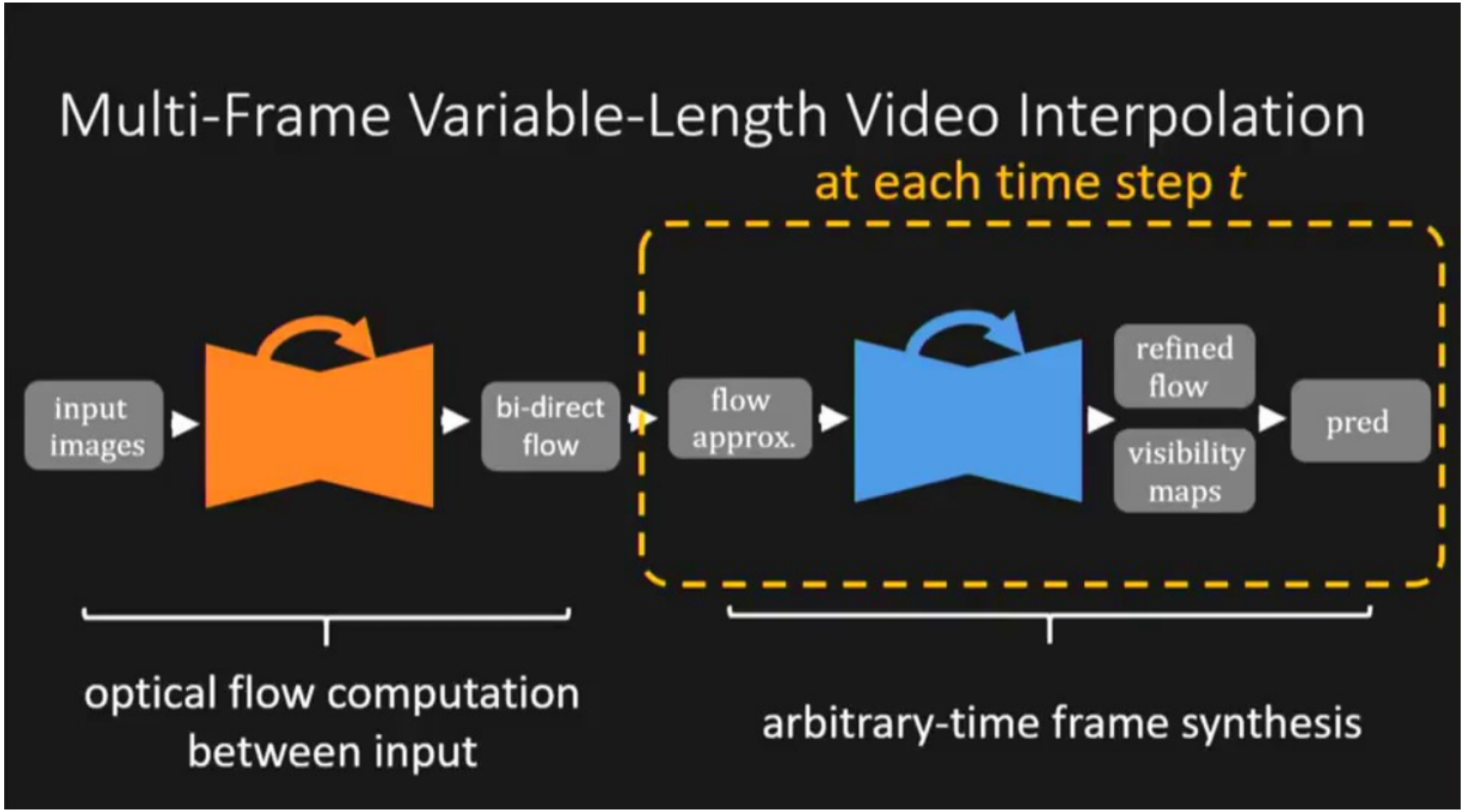
September 19, 2018

Motivation

There are many memorable moments in your life that you might want to record with a camera in slow-motion because they are hard to see clearly with your eyes: the first time a baby walks, a difficult skateboard trick, etc.

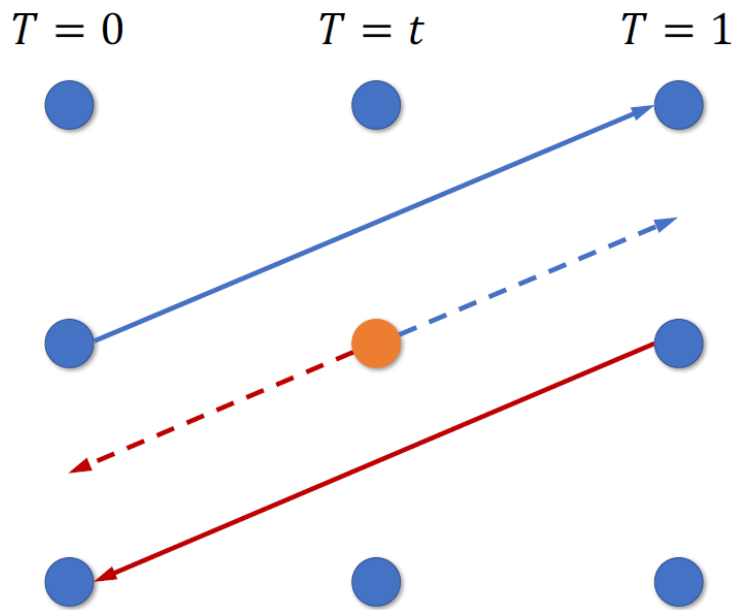
- Professional high-speed cameras are required for higher fps
- Moments we would like to slow down are unpredictable, and as a result, are recorded at standard frame rates.
- Recording everything at high frame rates is impractical–it requires large memories and is powerintensive for mobile devices.
- It also has intriguing new applications in self-supervised learning, serving as a supervisory signal to learn optical flow from unlabeled videos.

Multi-frame Variable-Length Video Interpolation



Optical Flow computation

Optical flow or optic flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and a scene.

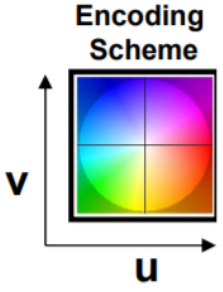


The orange pixel borrows optical flow from pixels at the same position in the first and second images.

Optical Flow - Example

Videos

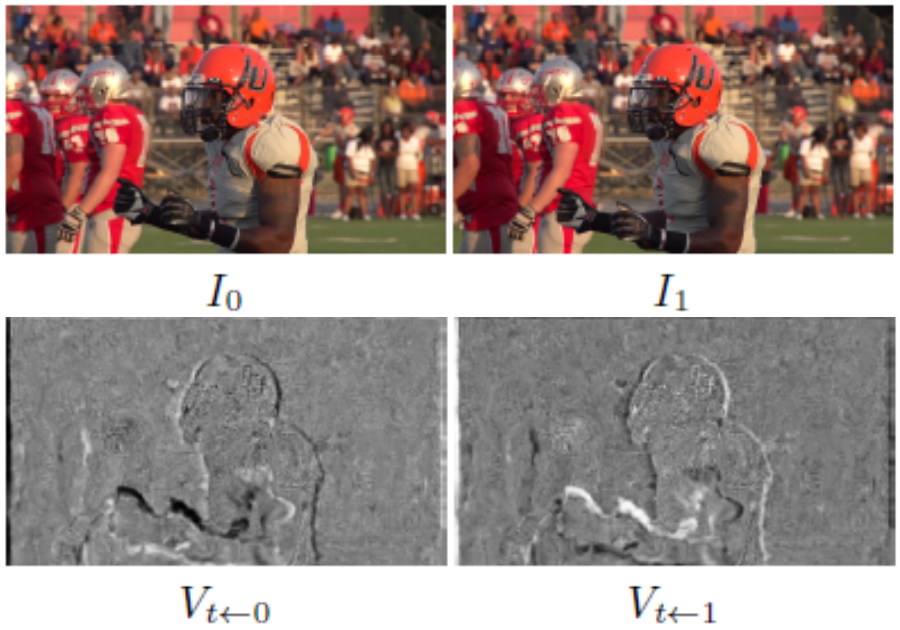
Color Coded Optical Flows



Refined Flow

- We take this Optical Flow and refine it further
- As approximation works well in smooth regions but poorly around motion boundaries, as the motion near motion boundaries is not locally smooth.
- To reduce artifacts around motion boundaries, which may cause poor image synthesis, we propose learning to refine the initial approximation.

Sample of predicted Visibility Maps



- The arms move downwards from T=0 to T=1. So the area right above the arm at T=0 is visible at t but the area right above the arm at T=1 is occluded
- The white area around arms in Vt→0 indicate such pixels in I0 contribute most to the synthesized It while the occluded pixels in I1 have little contribution.

Synthesizing the Intermediate Frame

- First use both the images (image at time t=0 and image at time t=1) and the corresponding refined bi-directional optical flows to backward warp these images.
- Then use the second neural network which takes these warped images, the visibility maps and fuses them together.

Why this method is the best?

- Produces as many intermediate frames as we want (time independent)
- Directly applicable to different scenarios without any modification
- Produces results that are less blurry and more fluidic than other approaches

Current Limitations

- It is not perfect and cannot match natively shot slow motion videos
- Not Opensource and not available to public
- Requires a basic amount of fps and quality of the original video for a legible output

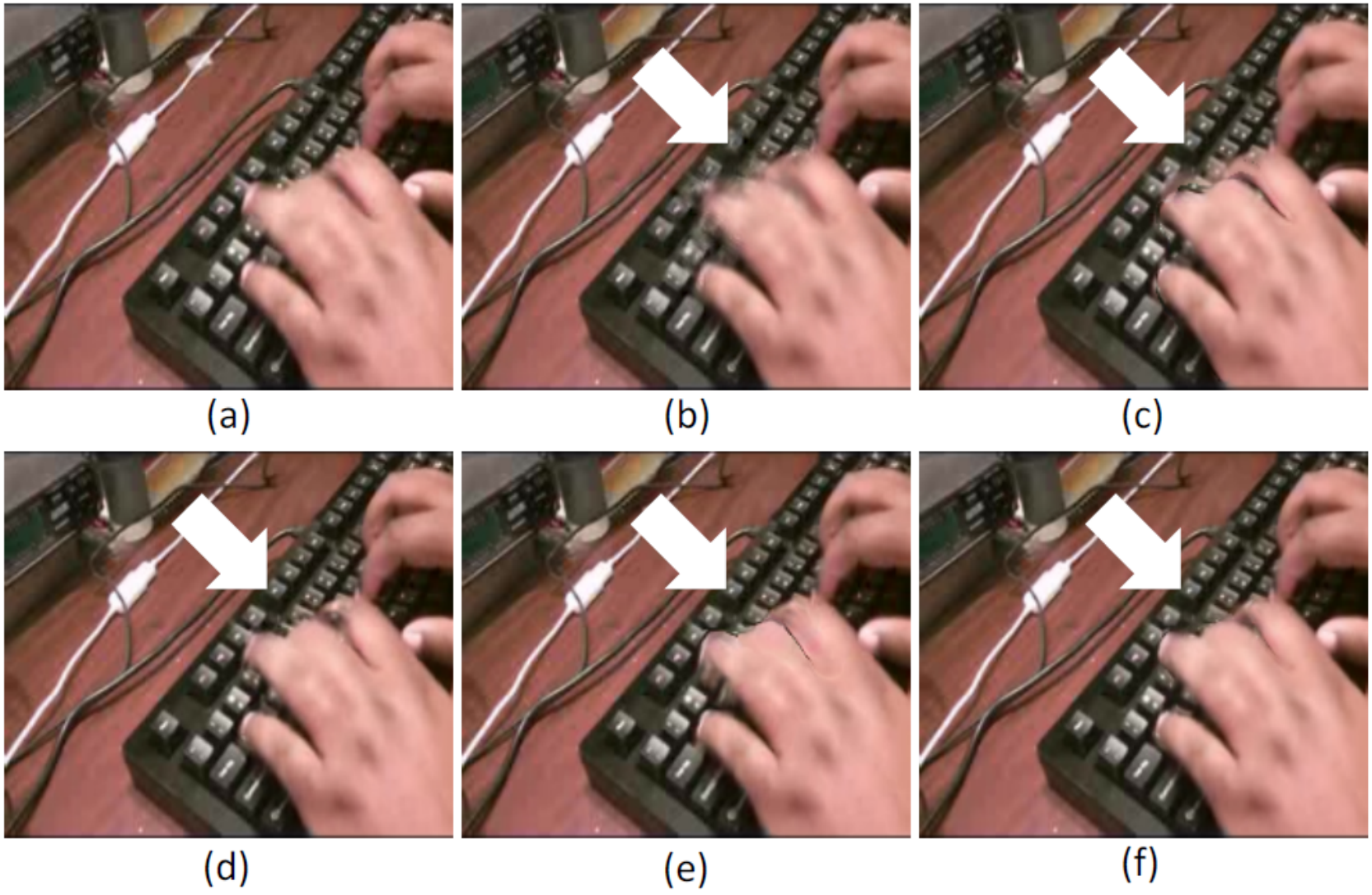


Figure: Visual Comparison with other Techniques

References

- https://www.youtube.com/watch?v=MjViy6kyiqs
- https://news.developer.nvidia.com/transforming-standard-video-into-slow-motion-with-ai/
- https://arxiv.org/abs/1712.00080
- https://www.youtube.com/watch?v=rkSzRbM4lIM
- Some Images taken from Google

Thank you!

apurv.varshney.16001@iitgoa.ac.in
piyush.doke.16001@iitgoa.ac.in
rohan.teja.16001@iitgoa.ac.in