Team:
Research >> Intern

Members:

- Vaibhav Garg (20171005)
- Anchit Gupta (20171041)
- Neel Trivedi (20171015)

Topic: Seeing Arrow of Time

Basic Understanding

- The goal of this paper is to analyze the given video and determine whether the video is being played in forward or backward direction.
- This paper aims to learn the temporal relationship between subsequent frames of the given video.
- Spatial relations among frames is a very much studied area but there is not so much work done to understand the temporality among the frames.

Some Applications:

- Estimating Optical Flows in videos
- Video denoising
- Video decompression
- Filling the missing Frames
- Predicting what will happen next in incomplete video
  - A very much essential task in robotics
- And many more....

Dataset

- The authors have used a data set of YouTube videos containing **180 videos** which was obtained manually using more than **50 keywords**.
- There is also a train/test/validation split specified in the paper with **70 clips for training, 60 clips for testing and 50 clips for validation**
- All the videos are **6-10 seconds** long
- All videos are **HD** without any compression
- Among 180 clips, **155 are forward and 25 are backward** videos.
- There is also a **Tennis-ball Dataset** which contains **13 HD** videos of tennis balls being rolled along a floor and colliding with other rolling or static balls.

Approach

- The paper proposes **3** different methods and a baseline procedure to accomplish the task of finding the ‘Arrow of time’ in the given video.

1. SVM trained on SOE (Spatial - temporal Oriented Energy) *(baseline)*
2. Flow - words based method
3. Motion - Causation method
4. AR (Auto - Regression) method

Approach 1: Flow based words method

- **Flow words** are SIFT like features which capture local regions of motion in a video so that we can examine which types of motion exhibit temporal asymmetries that might be useful for detecting the Arrow of Time.
- A bag of flow-word descriptor representing the entire video sequence is computed.
- And once we have enough of such flow words for samples we train a SVM classifier to classify a video whether being played forward or backward.

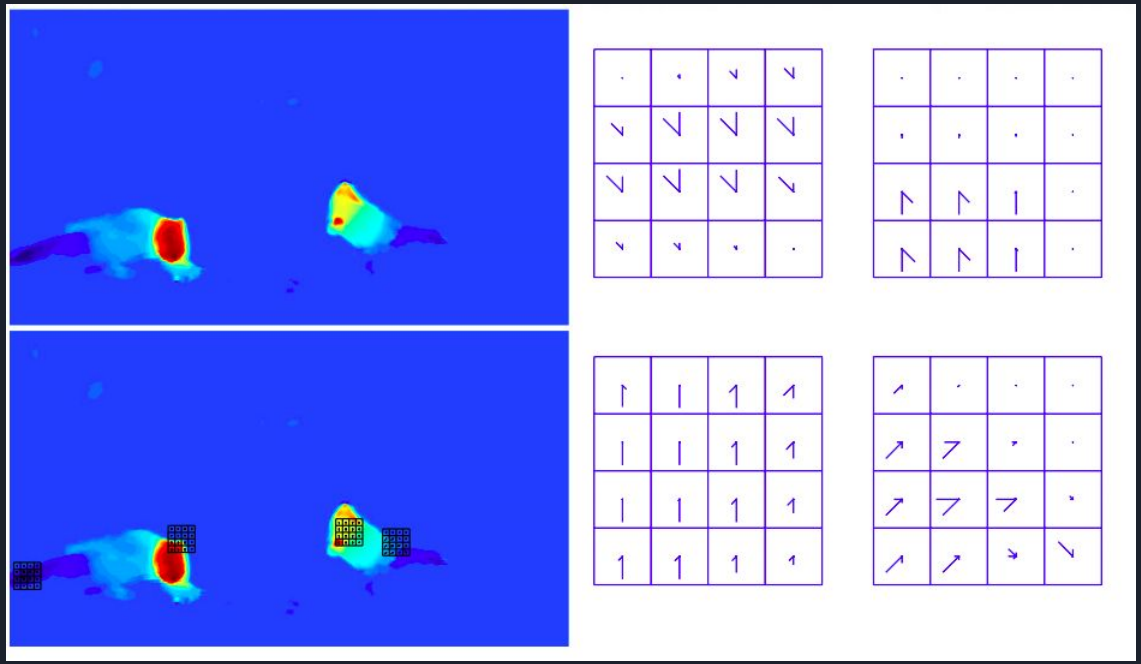
Steps:

Pre Processing

1. All the frames are first resized to width of 983 pixels.
2. For a frame at time stamp *t*, first frames at *t-1* and *t+1* are registered to frame at time *t*.
3. We then calculate **Optical Flow** between frames at *t-1* and *t+1*.
4. We divide X and Y components of this optical flow result into patches size 4x4 with stride of 3.
5. We threshold the patches based on optical flow value. (We discard the patches with optical flow less than **1200**)
6. And then we flatten and concatenate X and Y components of corresponding patches to get a ***#_patches x 32*** dimensional feature matrix.
7. We repeat this process for all the frames. (***T - 2*** times) and concatenate the results.
8. We repeat step from 1 - 7 for,
    a. the native direction of the video
    b. this video mirrored in the left-right direction
    c. the original video time-flipped
    d. the time flipped left-right-mirrored version
9. Now we concatenate the results from all the 4 above mentioned video directions to generate a feature vector of dimension $O(10^7)$.
10. For al the videos in the data set we calculate such feature vectors and then cluster them into **4000 clusters** using K-Means algorithm.

Training

1. As described in the pre processing step we calculate the optical flow and create ***#_patches x 32*** features for each frame and then using the clusters created in the pre processing steps we create a 4000 dimensional histogram.
2. Once we have such histogram for all the samples we train a SVM to classify a video whether being played forward or backward.



Optimization from our side

1. We parallelized the process of calculating optical flow using multi threading since videos and frames are independent of each other.
2. We have vectorized the pre processing step to achieve higher computational speed.

Results

- Paper claims to achieve 81% accuracy on test set.

What Next?

- We intend to achieve similar results for flow-words method as mentioned in the paper
- Once we are done with this method we will move to the next method which is **Motion - Causation method**.

Okay... That's it.
Thank You.