EE 702 Computer Vision | Problem Setup

Meet Pragnesh Shah | 13D070003 | meetshah1995@ee.iitb.ac.in

April 11, 2016

1 Background & Abstract

A large number of videos contain one or more object of attention performing estimatable motion with a static natural background. This motion of the objects in most natural videos can be modelled using fairly simple trajectory estimation techniques. The motion plane of the object which is usually a part of the natural background remains fixed in these frames and moves only when the camera moves. The movement of the camera and the objects can be estimated from the relative movement of the background and trajectory estimates respectively. What I propose as a problem is the reconstruction of unkown intermediate frames from such videos given the initial and end frames taking guidance from the motion plane and object trajectory estimates.

2 Problem Description and Specifications

Given a video $V(x, y, t)$ with a few frames F_{missing} missing from a time interval T ($T < T_{\text{threshold}}$) ranging from $[t1, t2]$, estimate the following :

1. Object(s) in motion.
2. Motion plane P_{motion} of the object(s) in motion.
3. Motion estimate of the camera v_{camera}.
4. Trajectory estimate of the object(s) v_{object}.
5. Areas of the natural background that are occluded by each moving object(s) A_{occluded} in the missing frames.

Now from the information obtained in the (1), (2), (3) and (4) obtain the best estimate for the missing frames F_{estimate}. You may take the following assumptions :

- The interval T does not lie in the very begining or the very end of the video.
- The temporal gradient of the frames at the boundaries of F_{missing} are smooth i.e (There is no unsual motion in the missing frames).
- $T_{\text{threshold}}$ is small when compared to the video interval.

Now run this algorithm on a large number of **other** similar videos (with natural background and moving objects(s) and some frames missing) and create databases of object motion trajectories. Using this database try to reconstruct the missing frames in the above video using the best matching trajectory. Call this estimate F'_{estimate}. Now keep increasing the missing frames time interval T by some small constant on both sides and obtain F_{estimate} and F'_{estimate}. Compute the maximum missing frame interval $T_{\max}$ beyond which the video similarity measure between the reconstructed and original video falls below $alpha$ ($alpha$ being some video similarity [eg. ViSig] measure limit).

The following video can be taken as an example :

https://www.youtube.com/v/GAYYMgkTHFE?start=124&end=128&version=3 , where the objects in motion are the car and the helicopter, the plane of motion is the road.

Figure 1: Example input and expected output

Input with missing frames

Expected Output (ground truth)

3 Proposed Solution

An outline of the solution can be viewed in the following steps :

- Detection of the objects in motion and their trajecctories can be easily done by difference in object centroid across frames.
- Motion of the camera can be estimated using relative shifts (disparities) in the static background across frames.
- Once the objected is segmented from the frame and its trajectory estimated, the part of the image occluded by the object can be computed by shifting the object segment by its estimated velocity in the image.
- The difference in occluded regions across frames has to be reconstructed (which can be done using image completion algorithm as sugested by Huang et. al. [SIGGRAPH 2014])
- In order to curb unusual motion from being estimated a global objective funtion which is the sum of squares of temporal gradients across consecutive frames can be minimized.
- The database creation (training) part can be also done using deep neural networks where the network tries to minimize the above objective function given the video and trajectories as the input. The database would be nothing but a trained neural network which can then be used to reconstruct videos without computing the trajectory of the objects. This is just a rough idea to pose this a supervised learning problem.

4 Miscellaneous

A few other problems I wished to pose a problem but had highly cited papers already published on them :

- Velocity of moving body in single image from deblurring and depth estimate.
- Brain tumor segmentation and classification from MRI scans.
- Deforestation and soil erosion site detection using satellite imaging.
- 3D model reconstruction of iconic sites using large data (tourist photos)
- Video genre classification.
- Fast moving car number detection from multi-modal video
- Textbook scan from page-flipping video.