Bridging the Gaps between Cameras
Dimitrios Makris et al.
(assignment 2)

Yi Tian Xu
260520039

November 15, 2014

1   Motivation and Past Works

Research has been done on tracking multiple targets within the viewfield of a single camera and tracking a single target that is simultaneously visible on multiple overlapping cameras. Yet, little is done on how to connect the tracks of targets that travel through blind regions between two spatially adjacent cameras. Solving this problem may be beneficial in applications such as building a complete record of a target's activity from the moment it enters to the moment it leaves.

Examples of proposed methods include using supervised learning to label trajectories with features that may not be reliably used to match targets across multiple cameras; or constructing a network-based activity model by finding the correspondence between regions at the border of the camera views, which is not a total correspondence.

To compute the connectivity of the cameras, one can consider the topology, which describes the spatial adjacency of the cameras views. One approach is to use camera calibration, which determine the transformation of pixel coordinates to 3D coordinates. In this paper, the authors propose an algorithm that can automatically calibrate the camera network, creating a topographical mode with temporal and spatial information characterized by probabilities. They name this process "network calibration".

2   Approach

Correlation between the signals received from two zones of different camera views may suggest an association between the cameras corresponding to those camera views. The author focus on the correlation for the entry and exit zones in the camera views, which are learnt prior to performing network calibration using Expectation-Maximization. The signal received at each entry/exit zone is modelled as an Gaussian distribution. Each new track is then broken down into three components: the appearance point, the disappearance point and the transition time (time between appearing and disappearing). Network calibration takes each track, uses Maximum-a-Posteriori to classify it to an entry-exit link and accumulate that information to the cross-correlation of the signals associated to the two zones in that link for a specific discretized transition time. Then, the covariance of the signals at each possible entry-exit link and each discretized transition time can be estimated using the cross-correlation. If two zones are adjacent or overlapping, then the covariance must have a peak. A threshold for the peaks is chosen to filter in the valid peaks; and a transition probability is estimated for those links.

The transition times of the valid links define the topology of the camera. Suppose the covariance for a pair of entry-exit zones that are each captured by a distinct camera shows a significant peak for some transition time τ. If $\tau \approx 0$, then the zones in that link overlap. If $\tau > 0$, then the targets go through an area between the two zones that is invisible to the two cameras. If $\tau < 0$, then the target enters a zone before it exit the other zone; that is the two zones are either partially or entirely visible to both cameras.

3 Result

An experiment to test network calibration is performed on a set of 6 cameras observing street traffic flows. The starting point and ending point of trajectories in a dataset collected earlier are derived in 13 hours. Coincident entries and exits are merged for simplicity.

The method automatically detects the links and gives their most popular transition time and probability. It successfully infers the topology of the camera using the covariance and the transition time. That is, for each pair of cameras, any link between the cameras with a transition time that is not too large (e.g.: 20 seconds) indicates that the cameras are adjacent or overlapping. Whether two cameras are disjoint, adjacent or overlapping is determined with the sign of the transition times between their zones.

The result shows that the covariance between the signals from the same camera view is almost always negative as on the same view, the entrance is observed before the exit. Most detected links are unidirectional as they correspond to vehicle traffic flow. Some links may be redundant since they can be expressed as a sequence of many other links; and to eliminate those redundant links, as the authors state, is a problem for further research.

4 Review

The method depends on some assumptions about the real world. Namely, the model assumes that the signals received at the entry/exit zones are Gaussian and stationary. The Gaussian assumption is mainly presented in network calibration to be used when applying the conditional probability of the signal given a zone to classifying a new track using Maximum-a-Posteriori. If the signals are distributed otherwise in some setting, one can easily replace the formula for the conditional probability with the one that matches the correct distribution.

On the other hand, the stationary assumption may be a less trivial one as the authors themselves have stated that it is wrong. Although they dropped that assumption when estimating the covariance, once network calibration finishes running, the transition probabilities for the valid links are fixed. It may be possible that pedestrians trajectory behave according to season or climate. There is no guarantee that the data gathered for one day is a good predictor of tomorrow. Furthermore, prior to network calibration, the links are only learnt for a finite period of time. Thus, if there is an outlier target entering and exiting from zones never detected before (perhaps a thief that jumps out from a bus and crashes into a window), this system may fail to reconstruct an accurate record of this activity.

Although automatic calibration is a step higher than manual calibration, due the 13 hours of processing trajectories in the authors' experiment, it may be unlikely that this process can be efficiently used in surveillance applications that may need to readjust the system's "believes" regularly.

Thus, as an idea for future investigations, a system that can not only automatically calibrate itself, but also update itself when observing new information may offer more power.