Research Statement

Yinghao Cai

My research generally involves in object tracking, object recognition and incorporation of color information for object detection and description in the application of visual surveillance.

Visual surveillance in dynamic scenes aims to detect, recognize and track objects such as people and cars from image sequences at low level. Moreover, it analyses and understands objects behavior at a high level. Nowadays, more and more cameras are applied in surveillance to monitor activities over a wide area. The objective of my research is, specifically, given a network of cameras with overlapping or non-overlapping fields of view, to assign a unique ID to moving objects when they are traversing the fields of view of multiple cameras. That is to continuously track moving objects within and across cameras. There are mainly three issues in this problem: **object matching**, **topology estimation** and **data association**.

The most important part of my research is object matching. Object matching in my research means to establish correspondences between observations across cameras. Establishing correspondence of objects across different cameras is different from single camera tracking since no spatial continuity can be exploited. We rely on appearance information, more specifically, color cues and shape cues to identify moving objects across multiple non-overlapping cameras. However, it is a difficult problem. Since the fields of view of multiple cameras may not overlap, especially in cases of indoor and outdoor, the appearance based object matching must deal with several challenges such as variations of illumination conditions, poses and camera parameters. Therefore, the key problem of object matching is to build a representation which is robust to these variations. To deal with illuminations variations across cameras, one solution is through color quantization [1, 2]. The quantization is designed so that perceptually similar colors are mapped to the same quantized value. We also employ some color constancy methods to identify the color of the object in spite of illumination variations and receiver characteristics. Both color quantization and color constancy serve as a pre-processing step before object representation and matching. On the other hand, at the matching layer, we introduce diffusion distance [3, 4] into histogram matching to compensate for illumination changes and camera distortions. The diffusion distance models the difference between two histograms as a temperature field. The diffusion process on this temperature field diffuses the difference between two histograms by a Gaussian kernel. To effectively represent moving objects so as to be discriminable from other objects, we incorporate color-spatial information into object representation. Color-spatial information is important in discriminating one object from another since objects may have similar color components with different layout. We partition the human body according to major color regions of the body such as head, torso and legs in [3] and partition the blob of moving objects into regular patches for localization of color components in [1]. Besides these methods, inspired from part-based object recognition methods, pedestrians are represented by a set of region signatures centered at points sampled from edges in [2]. Then the frame-to-frame matching across cameras is formulated as finding corresponding points between a model image and a query image as a minimization of a cost function over the space of correspondence. The correspondence problem is solved under integer optimization framework where the cost function is determined by similarity of region signatures as well as geometric constraints between points. Recently, I am carrying out experiments on appearance modeling which incorporates both co-occurrence of features and relative location of features by conditional random field(CRF). Publishable results are expected soon.

The second part of my research is topology estimation. Topology estimation means to determine the connectivity or spatial adjacency of cameras so that activities viewed by this network can be inferred. The topology graph helps to predict the re-appearance of moving objects when they are leaving the field of view of one camera. My intuition is that if two cameras are related, some of the moving objects will appear in both cameras. In my method, if there are groups of people who are sequentially stepping out the field of view of one camera re-appear in another camera after a while with a small permutation of the order, then these two camera are connected. By this assumption, topology graph can be inferred automatically and on-line.

Finally, with object matching and topology graph indicating connectivity relationship between cameras, we can achieve our goal "continuously track moving objects within and across cameras". The process of assigning a unique ID to moving objects when they are traveling from one camera to another is called data association. Bayesian model is explored in [3] to associate observations from one camera to another. Appearance information, spatio-temporal information and topology relationship between cameras are combined in a maximum a posteriori (MAP) framework.

Tracking in overlapping or non-overlapping fields of view is a relatively new topic in visual surveillance. As I mentioned before, the most important part of my research is object matching. That is to build a representation which is robust to variations of illumination conditions, poses and camera parameters. Object representation and matching is also an essential part of many applications such as object recognition, content-based image retrieval and biomedical image analysis. I am drawn to multi-disciplinary fields. I believe that these areas often produce some of the greatest inspiration and potential for innovation.

References

1. Cai, Y., Huang, K., Tan, T.: Matching tracking sequences across widely separated cameras. accepted by IEEE International Conference on Image Processing (2008)
2. Cai, Y., Huang, K., Tan, T.: Human appearance matching across multiple non-overlapping cameras. accepted by International Conference on Pattern Recognition (2002)
3. Cai, Y., Chen, W., Huang, K., Tan, T.: Continuously tracking objects across multiple widely separated cameras. The 8th Asian Conference on Computer Vision **1** (2007) 843–852
4. Ling, H., Okada, K.: Diffusion distance for histogram comparison. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. (2006) 246–253