学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（情報科学）　氏名　梁　棟

学 位 論 文 題 名

Dynamic Scenes and Appearance Modeling for Robust Object Detection and Matching Based on Co-occurrence Probability

（共起確率に基づいて動的シーンと外観のモデリングを用いたロバストな物体検出とマッチング）

Detecting moving objects plays a very important role in an intelligent surveillance system. It is often integrated with various tasks, such as tracking objects, recognizing their behaviors and alerting when abnormal events occur. However, object detection suffers from non-stationary scenes in surveillance videos, especially in two potentially serious cases: (1) sudden illumination variation, such as outdoor sunlight changes and indoor lights turning on/off; (2) burst physical motion, such as the motion of indoor artificial objects, which include fans, escalators and auto-doors. If the actual background includes a combination of any of these factors, it becomes even more difficult to perform detection. State-of-the-art algorithms, i.e. Gaussian Mixture Model (GMM) and Kernel Density Estimation (KDE) handle gradual illumination changes by building the statistical background models progressively using long-term leaning frames. In practice, however, this kind of independent pixel-wise model often fail to avoid mistakenly integrating foreground elements into the background, and it is difficult to adapt to sudden illumination change and burst motion. On the other hand, spatial-dependence model, i.e. Grayscale Arranging Pairs (GAP) and Statistical Reach Feature (SRF), shows promising performance under illumination change and dynamic background.

This study proposes a novel framework to build a background model for object detection, which is evolved from GAP method and SRF method. It is brightness-invariant and able to tolerate burst motion. We name it Co-occurrence Probability-based Pixel Pairs (CP3). In order to model the dynamic background, spatial pixel pairs with high temporal co-occurrence probability are employed to represent each other by using the stable intensity differential increment between a pixel pair which is much more reliable than the intensity of a single pixel, especially when the intensity of a single pixel changes dramatically over time. The model performs robust detection under outdoor and indoor extreme environments. Compared with the independent pixel-wise background modeling methods, CP3 determines stable co-occurrence pixel pairs, instead of building the parameterized/non-parametrized model for a single pixel. These pixel pairs maintain a reliable background model, which can be used to capture structural background motion and cope with local and global illumination changes. As a spatial-dependence method, CP3 does not predefine any local operator, subspace or block, and it provides an accurate detection criterion even though the gray-scale dynamic range is compressed under weak illumination.

The proposed method can be used for modeling the appearance of an image to realize image template matching. Theoretically speaking, both the object detection and image matching can be seen as a model matching problem. The differences is that the object detection is to seek the regions of interest

(ROI) which violate/mismatch the background model, while the image matching is to seek the ROI which can match the image model optimally. Therefore, in this study, we further extend the use of CP3 to the robust image matching task.

This thesis is organized in to the following chapters:

Chapter 1 introduces related works in object detection. Some general problems are discussed and considered. Furthermore, the motivations and contributions of this research are described.

Chapter 2 presents the details of CP3 based on co-occurrence pixel pairs. We tested it on video datasets for both qualitative and quantitative analysis. Experiments using several challenging datasets (PETS-2001, AIST-INDOOR, Wallflower and a real surveillance application) prove the robust and competitive performance of object detection in various indoor and outdoor environments. For quantitative analysis, Precision (also known as positive predictive value), Recall (also known as sensitivity) and F-measure (a weighted harmonic mean of the Precision and Recall.) were utilized. The three evaluation metrics measure the exactness, fidelity and the completeness of foreground. We compare our algorithm with three methods: (1) GMM method, which is a standardized method among independent pixel-wise models; (2) Sheikh's KDE method as a representative method among spatially dependent models; (3) our previous method GAP. In addition, the accelerated version of CP3 can effectively reduce the time cost of the background modeling stage.

Chapter 3 proposes the framework of CP3 for modeling the appearance of an image to realize template matching. We detail the learning phase and present the similarity measure procedure and present the experimental results. Although an additional learning stage is necessary, the experiment results show that the proposed method is robust under several imaging cases and it also outperforms SRF.

Chapter 4 presents the discussions of the proposed methods, concludes the main contributions of our work, and shows the future works of this study.