摘　要

随着“平安城市”等计划的实施，视频监控系统将逐渐覆盖各种重要场合，生成的监控数据也会成千上万倍的增加。以人眼观看为主的监控视频分析已经不能满足这种海量数据的需求。因此，智能视频监控的必要性越发明显。在智能监控系统中，最重要的功能之一就是场景中目标对象和背景的分离，这通常是通过背景减除技术实现的。

虽然背景减除技术已经被研究了十几年，它仍然面临着很多的挑战，对象密集场景就是其中之一。在这类场景中，大部分对象运动缓慢甚至静止不动，传统的背景减除算法(GMM、KDE 等)会失效，因为它们只能检测运动前景对象。一种有效地检测静止前景的方法是特征背景减除方法，它利用多个特征背景重构当前帧的背景。然而，在对象密度较高时，前景信息会融入到特征背景中，导致较高的漏检和误检。为了解决这个问题，本文提出了选择式的特征背景减除方法。本文的主要创新点包括：

第一，提出了基于最佳特征背景选择的块级背景减除算法。该算法对视频帧进行分块，对每一图像块单独处理，大大降低了训练样本中前景对象的比例和计算的时空复杂度；此外，为了在背景重构时将尽量少的前景信息引入到背景中，该算法在检测阶段为每个图像块选取最佳的特征背景，而不像传统方法利用所有的特征背景进行背景重构。实验结果表明，在对象密集场景中，该算法的性能高于传统的背景减除算法。

第二，提出了基于虚拟帧的像素级选择式特征背景减除算法，对特征背景减除算法作了进一步改进。一方面，该算法从视频中选取纯净的像素点构建不包含前景对象的“虚拟帧”。以虚拟帧作为训练和更新样本，特征背景的稳定性和纯净性大大提高。另一方面，在前景检测阶段，该算法为每个像素选取最佳的特征背景进行背景重构，从而使每个像素均获得最佳的背景重构结果。实验结果表明，与基于最佳特征背景选择的块级背景减除算法相比，该算法的性能有了更大的提升。

第三，将背景减除技术应用在了视频分析与编码中，开发了两个基于背景减除的智能监控系统。第一个是事件检测系统 eSur，包括针对人的监控的 eSur_Person 子系统和针对交通监控的 eSur_Vehicle 子系统。利用前者我们在 2009 和 2010 年的 TRECVID-SED 国际评测中均获得了优异的成绩。第二个是基于背景建模的视频编解码系统 SmartCam，包括软件实现与部分硬件实现。该系统将背景建模算法应用于 AVS-S 视频编码中，在保证视频质量的情况下大大降低了编码码率。

综上所述，本文针对对象密集场景的背景减除进行了研究，同时对其在智能监控系统中的应用作出了初步的探索。大量的实验结果展示了本文工作的可行性与有效性。

关键词：背景减除；背景建模；前景对象检测；特征背景；智能监控；视频分析；视频编码；密集场景

Research and Application of Selective Eigenbackground Method for Background Subtraction in Crowded Scenes

Hu Zhipeng (Computer Application Technology)
Directed By Huang Tiejun

With the “safe city” plan，video surveillance systems are gradually covering all crucial places and the surveillance data generated would increase hundreds of times. The traditional video analysis based on eye monitoring can not satisfy the need brought by such rapid increasing data. Therefore, intelligent video surveillance becomes more and more necessary. One of the most important functions in intelligent surveillance system is separating target objects from background, which is usually utilized through background subtraction.

Although background subtraction technology has been researched more than a decade, it remains confronted with many severe challenges, one of which is the foreground detection in crowded scenes. In such scenes, most foreground objects keep still, which leads to inefficiency of the traditional background subtraction algorithms, such as GMM, KDE, etc, because they can only detect moving foreground objects. Eigenbackground method can detect motionless objects, where several eigenbackgrounds are used to reconstruct the background. However, when the scene becomes crowded, some foregrounds may be absorbed into the eigenbackgrounds, leading to severe miss detections and false alarms. In order to solve this problem, selective eigenbackground methods are proposed in this thesis. The main contributions can be summarized as follows:

Firstly, this thesis presents a block-level eigenbackground algorithm based on best eigenbackground selection, where the original video frame is divided into blocks and each block is processed independently. Through this blocking strategy, the foreground proportion in the training samples and the spatio-temporal complexity of the algorithm are significantly reduced. Moreover, in order to introduce the foreground information into the background as less as possible, the algorithm selects the best eigenbackground for each block to reconstruct its background, rather than using all the eigenbackgrounds in the traditional eigenbackground method. Experimental results show the performance of the proposed algorithm exceeds those of some states-of-arts on crowded scene dataset.

Secondly, to improve the eigenbackground method furthermore, this thesis proposes a pixel-level selective eigenbackground algorithm based on virtual frame. On one hand, clean pixels are selected from the video to construct “virtual frames”, which contain no foreground objects. With virtual frames as training and update samples, the stability and the purity of the eigenbackground model are raised significantly. On the other hand, in the detection stage, the best eigenbackground is selected for each pixel to reconstruct its background value, thus each pixel can get the best background reconstruction result. Experimental results show that the background subtraction performance is further raised with the proposed algorithm.

Thirdly, background subtraction technology is applied in video analysis and video

encoding, where two intelligent surveillance systems based on background subtraction are developed and evaluated. The first one is the event detection system “eSur”, which includes the subsystem “eSur_Person” oriented to person surveillance and “eSur_Vehicle” oriented to traffic surveillance. Particularly, “eSur_Person” is used in the TRECVID-SED evaluation, where we got outstanding results. The second system, “SmartCam”, is a video encoding and decoding system based on background modeling, which includes the software implementation and partial hardware implementation. In this system, background modeling is applied in video encoding. Experimental results show that the encoding bit rate is greatly reduced under the same video quality.

In summary, this thesis studies the background subtraction in crowded scenes and makes a preliminary exploration of its application in intelligent surveillance system. Tentative experimental results demonstrate the feasibility and effectiveness of the work in this thesis.

Keywords: background subtraction, background modeling, foreground object detection, eigenbackground, intelligent surveillance, video analysis, video encoding, crowded scene