摘要

随着社会的日益发展和视频技术的不断进步，视频已经涵盖安防、娱乐等诸多行业，并与人们的生活息息相关。因此，有效地进行视频的处理与分析也是一个亟待解决的问题。目前，在监控视频中的物体检测与跟踪，以及普通视频中的物体与场景的编辑，还是需要人们手动地进行处理，其处理效率十分低下。而视频帧的标注图可以标识出帧中的物体及其形状与位置信息，为自动地分析与处理视频提供了可能。对于视频中每一帧标注图的生成，如果采用完全手动的方法，需要大量人工标注，同时不能保证最终结果在时序上的平滑；而对于完全自动的标注，现有的技术不能完全适应各类场景，其应用存在局限性。

本文采用了半自动地视频标注信息的传递方法。由人工提供关键帧的标注信息，利用视频关键帧上的标注图以及非关键帧的纹理信息，生成所有非关键帧的标注图。在标注传递的过程中，首先利用光流的方法，获取视频中关键帧与非关键帧间的运动信息。依据关键帧的标注信息以及上一步获取的运动信息，将非关键帧的标注图进行初始化。之后，利用非关键帧图像的纹理信息进行优化。在本文中，标注信息的优化是通过最优化对应的能量方程。该方程包含数据项与平滑项两项。其中，数据项约束了最终的非关键帧标注图与其初始化标注图间的差异；平滑项保证了最终的非关键帧标注结果与相应的纹理图像的纹理一致性。最后，本文采用图割的方法，通过构造一个网络流图，并将能量方程映射到该网络流图上。通过最大流最小割算法求解该网络流图即可得到非关键帧的标注结果。本文在实验中采用了来源于纽约大学的数据集，其包含了丰富的室内场景的视频序列。实验结果显示本文提出的方法对于各类复杂场景具有较好的表现，同时在各类测试视频序列上都获得了较低的错误率。

另外，鉴于视频处理与分析领域对于时间效率的要求，本文提出了并行的视频标注信息传递框架，提高了标注信息传递的效率。通过分析视频标注信息的传递过程，针对该过程中的帧间传递模块利用 P-thread 技术进行线程级的并行优化；而针对该过程中符合单指令流多数据流的模块，利用 SSE 指令集进行指令级的并行优化。通过实验验证发现，本文提出的并行化视频标注信息传递框架能够有效地提高时间效率。最后，为了便于标注结果的展示、编辑与分析，本文开发了可视化的视频标注信息传递软件。

关键词：标注信息传递，图像纹理，图割，并行优化，Microsoft Foundation Classes

ABSTRACT

With the development of society and the improvement of video technology, video is related to people's life deeply in the area of security, entertainment and so on. Thus, effective analyzing and processing of video is an important issue. Nowadays, the object recognition and tracking task in the surveillance video and the object and scene edit in natural video depends on people's hand work. Thus it is inefficient. However, the label map of image can recognize objects and their shape and location. In this way, it is useful for the automatic video processing and analyzing. For each label map of the image in the video sequences, it can be generated by people's hand work. However, it needs lots of work and can't confirm the temporal correlation in the video sequences. Besides, the automatic labelling method has limitations in some complex scenes.

In this paper, we propose a half-automatic method to transmit the label maps by propagating from key frame to non-key frames with image guided. The label map of key frame is provided by people. The label map of a non-key frame is initialized by warping the label map of its corresponding key frame according to the motion estimation between them. Subsequently, the initialized label map is optimized with the guidance of its texture image. The optimization process minimizes an energy function which takes two constraints into consideration: (i) the data term measuring the similarity between an estimated label map and its initialized one, (ii) the regularization term enforcing the local smoothness in the label map and the consistency of region boundaries between the estimated label map and its corresponding texture image. A network-flows is constructed by the energy function. Graph cuts based computation process is finally performed to generate the optimized label map. We evaluate the proposed method on the NYU dataset. This dataset comprises of video sequences from a variety of indoor scenes and contains abundant and complex objects. Experimental results show that our method is robust to many complex scenes and obtain lower error rates than the baseline method.

Besides, time efficiency is also an important issue in the area of video processing and analyzing. In this paper, we propose the framework that parallel propagating the label map in video sequences. For the frame-level of propagating process, P-thread is used to parallelize the propagation. Subsequently, some modules of the propagating process are conform to Single instruction stream multiple data stream model and SSE

instruction can be used to optimize it. Experimental result shows that the time efficiency of propagating process is improved by using the proposed parallel optimization. Finally, the MFC based label map propagating software is developed for the visual display and edit of label map.

KEY WORDS: Label Map Propagation, Image Texture, Graph-cuts, Parallel Optimization, Microsoft Foundation Classes