

基于对象的视频摘要技术^①

刘彩云, 曹建荣, 李洪艳

(山东建筑大学 信息与电气工程学院, 济南 250101)

(山东省智能建筑技术实验室, 济南 250101)

摘要: 视频摘要技术是当前多媒体领域研究的热点之一。视频摘要生成方法归结为两类: 基于关键帧的视频摘要和基于对象的视频摘要; 对基于关键帧的视频摘要方法做了简要的介绍, 并重点总结了历年来出现的基于对象的视频摘要的生成方法。最后对视频摘要技术的发展做出了总结和展望。

关键词: 视频监控; 视频摘要; 关键帧; 对象

Object-Based Video Synopsis Technology

LIU Cai-Yun, CAO Jian-Rong, LI Hong-Yan

(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

(Shandong Provincial Laboratory of Intelligent Buildings Technology, Jinan 250101, China)

Abstract: The video summarization technology is one of the hot spots in the field of multimedia research. This paper presents the generation method of video summarization and it can be divided into two classes: video summarization based on video key frame and video synopsis based on video object. Video summarization is briefly introduced and the method of the video synopsis is principally described. At last, the development trend and the application perspective of video summarization and synopsis are presented.

Key words: video monitoring; video synopsis; key frame; object

当今社会, 随着人们对安全性能要求的不断提高, 视频监控越来越广泛的应用于军事、商业、教育等领域。由此便产生了大量的监控视频, 在对它们进行视频检索和分析时, 如大海捞针一般费时费力, 且大量视频的存储也是一个重要问题。如果能够把视频内容进行浓缩, 提取出其中最主要的部分, 用一个简短的片断的视频摘要来表示出主要内容, 可有效提高检索和分析的效率。所以视频摘要的研究近年来已经引起国内外的广泛关注。例如美国 MIT 实验室、明尼苏达大学、德国 Mannheim 大学, 国内的亚洲微软研究院、清华大学、国防科技大学等都进行了此领域的研究^[1]。

视频摘要技术发展到今天主要有两种模式: 基于关键帧的视频摘要和基于对象的视频摘要。二者的相同之处就是都可以大大缩短监控视频的长度, 大大方便用户对监控视频的观看、分析以及检索等。而二者

又有明显的区别: 前者只重点考虑视频的关键帧; 而后者在对视频中对象的检测、分割的基础上, 选取合适的算法, 最大限度的减少时间-空间冗余, 从而方便了对视频的分析、检索以及索引等, 同时也节省了大量的存储空间。表现出了对象随时间的动态变化过程, 因此其表现的内容比基于关键帧的视频摘要要丰富的多。

1 基于关键帧的视频摘要

基于关键帧的视频摘要多为静态视频摘要模式, 也就是通过一系列的关键帧组成相应的语义单元, 概括表示镜头中的内容^[2], 并提供快速检索。在早期关键帧提取的基础上, 现今基于关键帧的视频摘要主要向着根据高层语义特征对视频提取视频摘要的方向发展。

① 收稿时间:2011-05-04;收到修改稿时间:2011-06-18

经过多年的发展,关键帧的提取方法也在不断的改进中,文献[3-10]中介绍了近年来出现的较为成熟的关键帧提取方法。文献[11]提出了一种基于关键帧聚类表示的视频摘要方法,文献[12]提出了一种草图形式的视频摘要。文献[13]还提出了基于关键帧的一种新的动态视频摘要的生成法。目前国内外视频摘要算法的研究主要是基于关键帧进行的。

虽然关键帧提取算法很多并得到了很大的改进,但仍存在明显的缺陷:帧选取依赖于阈值的选择;计算量太大,不可能做到实时处理;仅有颜色特征不能很好地表达视频的语义信息^[14]。近来,基于对象的视频摘要技术的兴起和迅速发展使得视频摘要技术有了新的突破和发展。

2 基于对象的视频摘要

基于对象的视频摘要是近年来提出的一种新的视频摘要技术,基于对象的视频摘要为动态视频摘要模式,即保持了视频内容随时间动态变化的特征。一般是由能够表达原视频内容的小片段加以编辑而成。这样形成的视频摘要更大程度的减少了时间-空间冗余。随着视频摘要技术的发展,各种基于该思想的新的方法算法也被广泛的应用到该研究中,其中比较经典的有如下的几种方法:

2.1 动态视频摘要

动态视频摘要^[15]的主要思想是将原始视频中不同时发生的活动在视频摘要中无遮挡的情况下同步播放。产生一个在时间和空间上相对紧凑并且包含原始视频中必要活动的视频摘要。

如果在相对较长的一段视频中只有单个或较少对象出现,可以生成一个“频闪电影”形式的摘要,既在同一小的帧序列中显示出同一对象在各个小的时间段里的位置及运动情况。其具体的实现如图 1 所示:

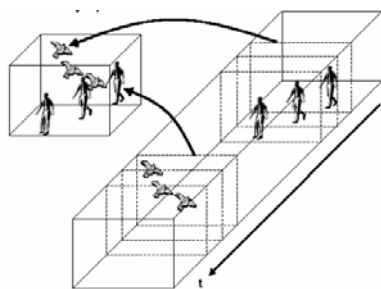


图 1 动态视频摘要形成过程

如图中所示,输入图像中首先出现一个行人行走的片段,经过一段没有运动对象的空闲时间后,出现一只飞行的鸟,行人和飞鸟在视频中出现在不同的时间,但通过将行人和鸟在同一段视频中同时播放的方法可以形成一段比较紧凑的视频摘要。

该方法得到的视频摘要有两个特点:(1) 该方法得到的摘要本身仍为一段视频,表现镜头中的动态变化。(2) 最大限度的减少了时间-空间冗余,但活动间的时序关系可能会发生变化。

2.2 基于相似活动聚类的视频摘要

基于相似活动聚类的视频摘要^[16]是一种新的产生简短连贯视频摘要的方法,在视频摘要中,如果同一画面中播放的活动是类似的,这样将能更有条理的表现出视频内容,所以该方法的中心思想是先将类似的活动聚类,然后以 2.1 中的方法为基础来形成视频摘要。

该方法用将类似活动聚类且只同时播放类似活动的方法来达到提高浏览效率的目的。该方法有三大特点:(1) 类似的活动比较有效地集结在一起以形成简短的视频摘要;(2) 所得的视频摘要条理非常清晰,大大提高浏览效率,还使得多重的类似活动便于查看到;(3) 非正常的活动更容易被检测到。

2.3 基于 3D 空间-时间描述中的“管带”排列视频摘要

3D 空间-时间描述中的“管带”排列视频摘要^[17]主要用于视频中感兴趣对象是运动对象的情况。该方法的中心思想是将视频转化到一个 3D 的空间-时间描述中,那些运动的对象是我们感兴趣的,每一个对象都可以看作一段“管带”,将对象分割提取并建好背景模型后,将我们提取的“管带”以最优的排列整合到由背景图像组成的 3D 空间-时间描述中,将“管带”的重要度计算出来,按其重要度由高到低选取,改变选取的“管带”数量就可以按照用户的要求来形成不同长度的视频摘要。

2.4 基于用户关注空间与注意力分析的视频摘要

基于用户关注空间与注意力分析的视频摘要^[18]的主要思想是利用视觉定位的方法来选取操作员感兴趣的区域,从而不需要形成语义规则,只需注意操作员是如何观看视频和跟踪运动目标的。人类如果把注意力放在多个运动对象的中心部位,可以同时追踪 5 到 8 个运动对象,那些离注意力中心部位较远的对象就可以看做是被操作员忽略的对象,既

可以将这些被忽略的对象包含到视频摘要中,形成最终的视频摘要。也可以使最终的视频摘要中只包含操作员注意到的对象。

2.5 多视频摘要

多视频摘要^[19,20]是一种全新的视频摘要方法。它提出了将多个视频在同一画面中同时播放的思想。该方法的主要思想是先选取一段较长的视频作为主视频,一些比主视频短的视频可作为增补视频,一个主视频可以插入几个增补视频要看主视频和增补视频的长短和内容而定。先将视频转换到一个3D的时间-空间描述空间中,找出主视频中由没有对象或感兴趣对象较少的部分所构成的一个3D的“孔洞”,再将我们的增补视频插入到这个3D的“孔洞”中,从而形成我们的视频摘要。其具体的实现方法如下图所示:

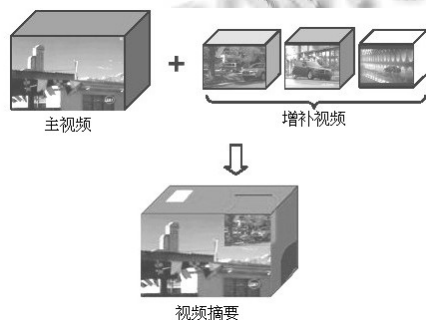


图2 多视频摘要形成过程

如上图所示,首先选定一段主视频和几段增补视频,将增补视频经过合理的排列组合整合到主视频帧中,这样就大大缩短了视频播放的总时间,形成了较为省时的视频摘要。

2.6 基于带状雕刻剪裁的视频摘要

利用基于带状雕刻剪裁的视频摘要^[21]方法形成最终的视频摘要与前文中的动态视频摘要技术形成的视频摘要在最终的观感上相似,都是将视频序列中处于不同帧的物体在无遮挡的前提下,放到同一帧或帧序列中,从而减小时间空间冗余。而本方法的实现却与动态视频摘要中提及的方法有所不同。该方法比较适用于像在高速公路上对车辆等定向运动的物体的监控。这样就可以将视频的时间-空间3D描述空间转化为一个类似于2D的空间,因为在x-y平面上,在物体定向运动时,图像中的像素对应的坐标的其中一个分量是不变的。这样对视频进行运动目标检测后,检测

出来的运动目标都是定向运动的,这样各个被检测出来的目标在该2D空间中就是基本平行的。每两个运动目标中间都会有一些空白,我们可以选取适当的算法,将中间的空白最大程度的去掉,这样就相当于减少了视频的帧数,也就压缩了视频。

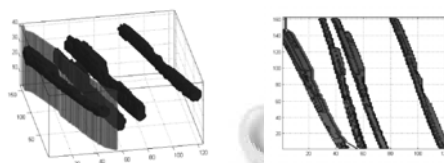


图3 类2D空间

图3表示了将视频的时间-空间3D描述转化成了一个类似于2D的空间描述中

2.7 基于对象轨迹表示的视频摘要

随着研究的深入和技术的发展,出现了许多新的生成视频概要的方法,像Gwo-Cheng Chao等人^[22]提出了一种对象轨迹表示的方法。该方法分为两个步骤:第一步内容提取,就是将视频进行分析后,提取视频中的运动对象,然后将运动对象的运动轨迹分解出来,并用细线连接着的小圆点来表示,线的末尾的箭头表示运动对象体的运动方向;第二步是内容的综合,就是将第一步中得到的运动对象的轨迹综合到一张或多张图像中。可以按照实际情况来决定轨迹的布局安排。这样在最终得到的图像中,每个条轨迹就成为了该对象在原始视频中的一个索引,我们选择某一条轨迹,就可以链接到原始视频中,这样就可以观看到该运动对象在原始视频中的运动情况。

该方法得到的视频概要内容简捷明了,只有对象的运动轨迹,且每个轨迹都是原始视频的一个索引,可以方便的查找到对象在原始视频中的运动情况。

3 总结与展望

虽然视频摘要技术已经得到了长足的发展与进步,但仍有很多理论和实际问题需要进一步改进,现归纳如下:

(1) 如何准确提取出视频的客观内容是成功提取视频摘要的关键之一;如何能准确表达人的主观感知视频摘要也是一个重要课题;最后因为视频内容理解是复杂、模糊而且歧义的,这些方面需要更为深入的

研究。

(2) 视频摘要技术必将向着多元化、智能化的方向发展,对语义及情感语义的监控视频分析和摘要是目前发展的主要方向,在对视频中最精彩、最重要、最激动等情节的提取中,结合严格的数理模型定义,将是当前研究的难点及热点之一。

参考文献

- 1 余卫宇,曹燕,谢胜利.视频摘要的现状和研究进展.计算机工程,2008,25(7):1948-1952.
- 2 欧阳建权,李锦涛,张勇东.视频摘要技术综述.计算机工程,2005,31(10):7-9.
- 3 Tonomura Y, Akutsu A, Otsu GK, et al. VideoMAP and video spaceicon: tools for automatizing video content. Proc. of ACM Interchi'93 Conference 1993:131-141.
- 4 Ueda H, Miyatake T, Izawa YS. M PACT:an interactive natural-motion-picture dedicated multimedia authoring system. Proc. of ACM Chi'91 Conference. 1991: 343-350.
- 5 Ruiy, Huang TS, Mehro TS. Exploring video structure beyond the shots. Proc. of IEEE Intemational Conference on Multimedia Computing and Systems Texas:1998:237-240.
- 6 Sum ZH, Fu P. Combination of color and object outline based method in video segmentation. Proc. of SPIE Stroage RetrMethods ApplMultimed 2004:61-69.
- 7 Pentland A, Picard R, Davenport G, et al. Video and image semantics: ladvanced tools for telecommunications. IEEE Multimedia, 1994,1(2):73-75.
- 8 Girgensohn A, Boreczky J. Time-constrained key frame selection technique. Proc. of ICMCS. 1999:756-761.
- 9 Narasmha R, Savakis A, Rao RM, et al. A neural network approach to key frame extraction. Proc. of SPIE IS&T Electronic Imaging Stroage and Retrieval Methods and Applications for Multimedia 2004:439-447.
- 10 Zhao L, QI W. Key-frame extraction and shot retrieval using nearest feature line. Proc. of ACM Multimedia Workshops on Multimedia 2000: 217-220.
- 11 Damjanovic U, Fernandez V, Izquierdo E, Martinez JM. Event Detection and Clustering for Surveillance Video Summarization Ninth International Workshop on Image Analysis for Multimedia Interactive Services.
- 12 陈佳,滕东兴,杨海燕,等.一种草图形式的视频摘要生成方法.中国图像图形学报,2010,15(8):1139-1143.
- 13 程文刚,须德,蒋轶苇,等.一种新的动态视频摘要生成方法.电子学报,2005,33(8):1461-1466.
- 14 Yu WY, Yu YL, Xie SL. Representtation on image semantics structure information. Proc. of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics. Orland: 2004:253-256.
- 15 Alex RA, Pritch Y, Peleg S. Making a Long Video Short: Dynamic Video Synopsis. Proc. of the IEEE Conf. Computer Vision and Pattern Recognition, 2006: 435-441.
- 16 Pritch Y, et al. Clustered Synopsis of Surveillance Video. IEEE Advanced Video and Signal Based Surveillance, 2009, 195-200.
- 17 Pritch Y, Acha AR, Gutman A, Peleg S. Webcam Synopsis: Peeking Around the World. Proc. IEEE, Conf. Computer Vision, Oct. 2007.
- 18 Vural U, Akgul YS. Eye-gaze based real-time surveillance video synopsis. Pattern Recognition Letters, 2009,30:1151-1159.
- 19 Li T, Mei T, et al. Video M: Multi-Video Synopsis. IEEE International Conference on Data Mining Workshops. 2008, 854-861.
- 20 Li T, Mei T, Kweon I, et al. Multi-video synopsis for video representation. Signal Processing, 2009,89:2354-2366.
- 21 Li Z, Ishwar P, Konrad J. Video Conden sation by Ribbon Carving. IEEE Trans. on Image Processing, 2009,18(11): 2572-2583.
- 22 Chao GC, Tsai YP, Jeng SK. Augmented keyframe J. Vis. Commun. Image R.21, 2010. 682-692.