

文章编号:1671-6833(2013)03-0102-04

# 一种适合于监控视频内容检索的关键帧提取新方法

周 兵<sup>1</sup>, 郝伟伟<sup>1</sup>, 袁社锋<sup>2</sup>

(1. 郑州大学 信息工程学院, 河南 郑州 450001; 2. 河南农业职业学院 电子信息工程系, 河南 郑州 451450)

**摘要:**为减少监控视频的存储容量,便于用户迅速查看视频关键信息,提出一种适于监控视频内容检索的关键帧提取方法.首先,使用背景差分法检测含有运动物体的关键视频段,记录其起始帧和结束帧,然后,在关键视频段中基于联合直方图的方法进行关键帧的粗提取,最后,根据监控视频帧序列的连续性特征,通过图像的信息熵进一步精确提取关键帧.实验结果表明,该方法能有效地提取出监控视频中的关键帧,大大减少了视频数据的存储量,且便于用户浏览关键信息.

**关键词:**监控视频;关键帧提取;信息熵

**中图分类号:** TP391

**文献标志码:** A

doi:10.3969/j.issn.1671-6833.2013.03.025

## 0 引言

监控产生的视频具有数据量大而有用信息少的特点,因此如何在海量的监控视频信息中快速有效地存储和浏览信息,对现今监控视频的进一步分析处理有着重要的意义.而关键帧技术的提出,为其提供了良好的方法.

关键帧被定义为一个有限的视频帧子集,其代表了视频序列的主要内容.早几年,针对不同的应用目的,关键帧提取已有了一定的研究. Mona Omidyeganeh 等人<sup>[1]</sup>使用广义高斯密度特征向量之间的相对熵(KLD)进行镜头聚类边界的选取,进而基于相似性和相异性标准提取关键帧. Jiang Peng 等人<sup>[2]</sup>介绍了一种基于视觉注意模型的视觉注意力(AVI)描述来提取关键帧. Liu Gentao 等人<sup>[3]</sup>提出了镜头边界检测与在镜头内提取关键帧的统一架构. Gwo-Cheng Chao 等人<sup>[4]</sup>提出“增强三维关键帧”来浓缩监控视频片断有意义的内容信息,但算法实现复杂、计算量大.以上文献都是针对视频中所有帧序列进行计算来提取关键帧,现实监控视频中多数含有大量的纯背景帧,而仅包含运动目标的视频片断才是人们想要查看的关键序列.

根据以往方法的总结以及自己工作的实验结果,本文结合具体应用,针对视频监控提出一种基

于运动目标检测的关键帧提取方法.首先,设计一种稳健而且快速的背景差分算法检测监控视频中的运动物体,记录进入和离开视频画面的起始帧和结束帧;然后,基于联合直方图的对称性来判定图像的相似程度,对含有运动对象的关键视频段进行初步的关键帧提取;最后,为防止数据的冗余,对关键帧集合基于图像的信息熵进一步进行提炼.

## 1 关键帧提取方法

### 1.1 运动目标检测

运动目标的检测过程,实际是对运动目标在视频序列中定位以及跟踪的过程.目前运动目标的检测主要有光流法、帧间差分法和背景差分法<sup>[5-7]</sup>.其中背景差分法又被称为背景减,通过背景模型检测当前帧与背景差异来检测运动目标,该方法在速度和稳健性上都较有优势.经典的高斯混合背景模型<sup>[8]</sup>,当背景更新时需要排序运算,计算量很大,不利于高清视频处理.本文简化高斯混合背景模型,背景模型数固定为两个,在变化像素检测时增加邻域像素模型符合性检测,兼顾速度和稳健性两个方面.

若某一像素当前强度值为  $x_t$ ,则该像素属于两个背景模型的概率计算公式如下式:

收稿日期:2012-12-21;修订日期:2013-03-01

基金项目:国家发改委产业化项目(CNGI090213);郑州市科技攻关项目(ZZ20120101).

作者简介:周兵(1964-),男,河南郑州人,郑州大学教授,博士,主要从事关键帧的提取研究, E-mail:iebzhou@qq.com.

$$H_k(x_t, \nu_{k,t}, \gamma_{k,t}) = \frac{1}{(2\pi)^{d/2} |\gamma_{k,t}|^{1/2}} e^{-\frac{1}{2}(x_t - \nu_{k,t})^T \gamma_{k,t}^{-1} (x_t - \nu_{k,t})} \quad (1)$$

式中:  $k \in (1, 2)$  分别表示两个高斯模型公式;  $H$  服从  $d$  维高斯分布;  $T$  表示确定高斯分量个数的阈值;  $\gamma_{k,t}$  和  $\nu_{k,t}$  分别是在  $t$  时刻高斯模型的协方差和均值。

使用背景差分法将每帧像素强度值与其对应的背景像素强度值做差处理, 其处理结果大于设定阈值的像素即判定为变化像素, 模型  $m_k$  ( $k \in (1, 2)$ ) 中当前像素是否是变化像素的计算公式如下:

$$m(x, y) = \begin{cases} 0, & \text{If } |x_t - \nu_{1,t}| < C \cdot \gamma_{1,t} \text{ and } |x_t - \nu_{2,t}| < C \cdot \gamma_{2,t}; \\ 1, & \text{otherwise.} \end{cases} \quad (2)$$

其中,  $C \in [2.5, 3]$  为系数常量。

为了消除背景中杂乱的虚假变化像素, 在判定当前像素是否为变化像素时, 除了检查背景中像素  $(x, y)$  之外, 还同时检查其邻域像素, 即使用背景差分法将每帧像素强度值与其对应的背景像素强度值和相邻的背景像素强度值分别做差处理, 只有当前像素与背景中的这些像素差值结果均大于阈值时, 才认为是运动像素。

使用  $\diamond(x, y)$  表示点  $(x, y)$  的邻域像素坐标, 则变化像素检测公式修改为:

$$m(x, y) = \begin{cases} 0, & \text{If } |x_t - \nu_{k,t}| < C \cdot \gamma_{k,t} \text{ and } |x_t - \nu_{k,t}(\diamond(x, y))| < C \cdot \gamma_{k,t}; \\ 1, & \text{otherwise.} \end{cases} \quad (3)$$

另外, 因背景变化产生的前景大多数表现为小尺寸的像素块, 本文通过尺寸滤波加以消除, 以降低监控系统背景的错误。

## 1.2 关键帧提取

根据背景差分法确定含有运动物体视频片段的定位信息, 即起始帧及结束帧, 在该关键视频段内进行关键帧提取。笔者基于联合直方图的关键帧提取方法, 使用其对称性来判定图像的相似程度。

联合直方图表示两幅大小相同的图像  $I_i$  和  $I_j$  间其对应位置上像素对的灰度组合出现的频率。对于同为  $M \times N$  的图像  $I_i(x, y)$ 、 $I_j(x, y)$ , 对应的像素值对  $(p, q)$  的联合概率表示为:

$$F(p, q) = \frac{1}{MN} \sum_{y=1}^M \sum_{x=1}^N \eta(I_i(x, y), p) \cdot \eta(I_j(x, y), q). \quad (4)$$

$$\text{其中, } \eta(I, p) = \begin{cases} 1, & \text{If } (I = p); \\ 0, & \text{If } (I \neq p). \end{cases}$$

根据上式可知, 对所有可能的像素值对  $(p, q)$  求  $F(p, q)$  值, 可得到图像  $I_i(x, y)$  和  $I_j(x, y)$  的联合直方图。联合直方图对称性定义为

$$\varepsilon = \frac{\alpha \sum_p F^2(p, p) + \sum_{p, q} \beta F(p, q) F(q, p)}{\alpha \sum_p F^2(p, p) + \sum_{p, q} \beta F^2(p, q)}. \quad (5)$$

式中:  $\alpha$  是联合直方图对角线上的权值, 在此为小于 1 的正常量; 而  $\beta = (p - q)^n$  则表示远离对角线元素的权重; 公式中  $n$  为整数;  $\varepsilon$  更直观地表示出了两帧之间的相似性, 当  $\varepsilon$  越趋近于 1, 则表示联合直方图越对称, 即说明两图像相似越大。当目标快速出现、亮度等视频内容发生较为明显变化时, 帧间相似性也发生相应变化, 一般相邻帧间的相似性  $\varepsilon \in (0, 1]$ 。为避免关键帧的遗漏, 文中设阈值  $T' = 0.9$ 。

根据监控视频连续的特性, 在一段连续变化的视频序列中, 连续的前后视频帧存在的特征值是渐变的, 即相邻帧的图像信息值变化不大。图像的熵信息体现了图像包含的信息量, 而关键帧则是一段视频主要内容的代表, 因此关键帧所含的信息量也相对较大。为减少数据的冗余, 本文提出在间隔较近的视频序列中选取图像的信息熵值最大的一帧作为关键帧。图像信息熵的计算公式为

$$Z = - \sum_{i=1}^N p(x_i) \times \log p(x_i). \quad (6)$$

式中:  $N$  指图像的灰度级数;  $x_i$  表示像素  $(x, y)$  的灰度值;  $p(x_i)$  为各灰度级出现的概率。为防止提取的关键帧中由于光照变化等原因造成关键帧存在冗余, 笔者在相邻较近的候选帧中选取信息熵最大的一帧作为关键帧。实验表明, 在一段连续的监控视频段中, 相邻帧间的差异较小, 帧间熵差很难区分出来, 而间隔 20 帧的非相邻帧间熵差则可明显地区分出来。因此, 文中关键候选帧序列间隔小于 20 时, 选取信息熵最大的一帧作为关键帧。这样得到的关键帧即可有效地反映连续视频序列的主要内容, 又减少了数据的冗余。

关键帧的提取具体步骤如下:

(1) 对视频  $I$ , 基于背景差分法监测视频中含有运动物体的关键视频段, 记录关键视频段的起始帧  $I_1$  与结束帧  $I_n$ ;

(2) 解析视频得到关键视频序列  $(I_1, I_2, I_3, \dots, I_{(n-1)}, I_n)$ ,  $I_1$  作为当前关键帧  $K_{\text{new}}$ , 即  $K_{\text{new}} = I_1$ , 加入到关键帧的集合  $K$  中,  $i = 1$ ;

(3)  $i = i + 1$ , 计算  $I_i$  与  $K_{\text{new}}$  间的相似性  $\varepsilon$ , 若大于阈值  $T'$  ( $T' = 0.9$ ), 则  $I_i$  为新的关键帧, 将该

关键帧记录进关键帧集合  $K$  中,并标记为当前关键帧,即  $K_{\text{new}} = I_i$ ,否则不处理;

(4)检测是否是帧序列最后一帧,若是则转(5),否则转(3)继续处理;

(5)检测初步提取的关键帧集合  $K$ ,若相邻关键帧之间  $I_i, I_j, |i-j| < 20$ ,则选取信息熵最大的值作为关键帧;

(6)检测是否是集合  $K$  中最后一帧,若不是,

则转(5),否则结束.

## 2 实验结果

本实验使用一个普通像素的摄像头检测室外一小时中含有运动目标的关键视频段.每个关键视频段可提取若干个关键帧,实验结果中只显示各个关键视频段中中间一帧,结果如图 1 所示.

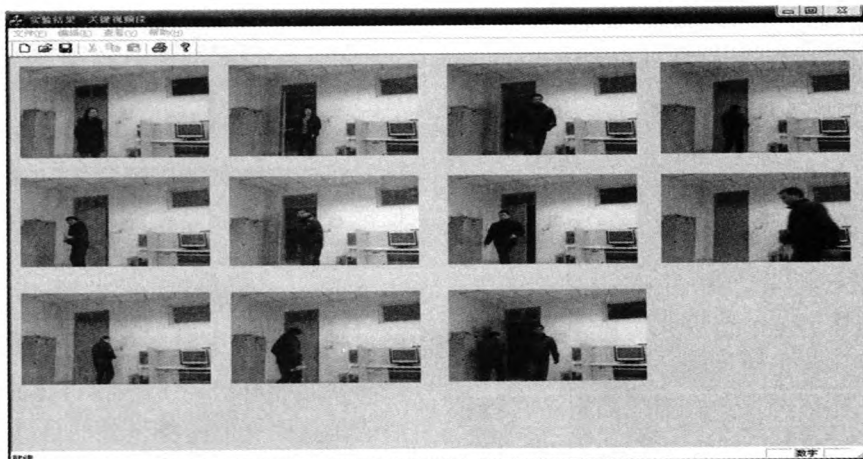


图 1 各个关键视频段中代表性关键帧

Fig.1 Representative key frame of each key video segment

例如第一个关键视频段有进入视频画面和离开画面组成,该视频段共 180 帧,程序解析该过程的部分连续视频帧序列如图 2 所示.

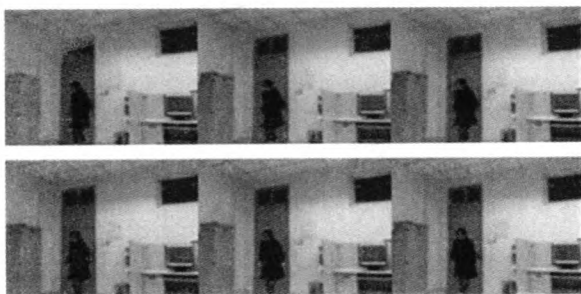


图 2 第一个关键视频段中的部分帧序列

Fig.2 Part of the frame sequence of the first key video segment

该关键视频段通过关键帧的提取,可得关键帧序列如下:在系统用户查看的页面中显示关键帧集合中的中间一帧,既第二关键帧.如图 3 所示.监控得到的视频信息为:帧宽 640、帧高 480、数据速率为 302 kbps,总比特率为 350 kbps,帧速率为 30 帧/s.1 h 录像大小为 150 MB,经过视频数据的处理,提取含有运动目标的关键视频段,对相应视频进行关键帧提取.第一个视频段提取关键帧数据存储大小为 80 kB,该 1 h 录像提取关键

帧后,共保存数据大小为 756 kB,其数据的存储量,远小于原监控视频的数据量,且涵盖了主要的有效信息,便于用户迅速查看.



图 3 第一个关键视频段的关键帧

Fig.3 The key frames of the first key video segment

## 3 结论

提出一种在监控视频中提取关键帧的方法,该方法提取的关键帧很少存在冗余数据,大大减少了视频数据的存储量,且系统显示出每段关键视频段的主要内容,便于用户浏览关键信息.上述提取的关键帧可能不是人的正面图像,而人的正面信息才更有效直观的显示运动目标,因此下一步将引入人脸检测算法,解决这一问题.

## 参考文献:

- [1] OMIDYEGANEH M, GHAEMMAGHAMI S, SHIRMOHAMMADI S. Video keyframe analysis using a seg-

- ment-based statistical metric in a visually sensitive parametric space[J]. IEEE Transactions on Image Processing, 2011, 20(10): 2730–2737.
- [2] JIANG Peng, QIN Xiao-lin. Keyframe-based video summary using visual attention clues[J]. IEEE MultiMedia, 2010, 17(2): 64–73.
- [3] LIU Gen-tao, WEN Xiang-ming, ZHENG Wei, et al. Shot boundary detection and keyframe extraction based on scale invariant feature transform[C]//2009 Eighth IEEE/ACIS International Conference on Computer and Information Science, Shanghai: IEEE Computer Society, 2009: 1126–1130.
- [4] CHAO G C, TSAI Y P, JENG S K. Augmented 3-D Keyframe extraction for surveillance videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(11): 1395–1408.
- [5] 谭鑫, 肖南峰. 运动人体的检测跟踪方法的研究与实现[J]. 计算机应用与软件, 2010, 27(4): 30–31.
- [6] BARNICH O, VAN DROOGENBROECK M. ViBe: A Universal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing, 2011, 20(6): 1709–1724.
- [7] 李毅, 孙正兴, 远博, 等. 一种改进的帧差和背景减相结合的运动检测方法[J]. 中国图象图形学报, 2009, 14(6): 1163–1168.
- [8] LI Ying-hong, TIAN Hong-fang, ZHANG Yan. An improved Gaussian mixture background model with real-time adjustment of learning rate[C]//2010 International Conference on Information Networking and Automation (ICINA), Kunming: IEEE, 2010. 5636758, 2010(1): 512–515.

## A Novel Method of Key Frame Extraction for the Surveillance Video Content Retrieval

ZHOU Bing<sup>1</sup>, HAO Wei-wei<sup>1</sup>, YUAN She-feng<sup>2</sup>

(1. School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. School of Electronic Information Engineering, Henan Vocational College of Agriculture, Zhengzhou 451450, China)

**Abstract:** This paper proposes a new method of the key frame extraction aiming at reducing storage and fast retrieval of the surveillance video. First, an improved background subtraction is used to detect the key video segment which contains the moving objects, and record the first frame and the last frame of the segment. Then the key frames are extracted coarsely in the video segment based on the joint histogram. Finally, according to the continuity of the key video frames, we further refine the key frames by image information entropy. The experimental results show that the method can extract the key frames of the surveillance video effectively. It greatly reduces the storage of the video data, and helps the users to browse key information easily.

**Key words:** surveillance video; key frame extraction; information entropy