

doi: 10.3969/j.issn.1006-1576.2010.10.012

一种基于关键帧的战场视频目标挖掘方法

郭忠伟, 张少兵, 钱晓进, 黄秋波

(解放军炮兵学院 2 系, 安徽 合肥 230031)

摘要: 一种基于关键帧的战场视频目标挖掘方法。首先对战场视频进行基于内容的视频检索, 提取出关键帧, 然后对关键帧进行目标识别, 从而快速提取出目标信息。实例证明该方法有效可行。

关键词: 目标挖掘; 战场视频; 关键帧; 神经网络; 目标识别

中图分类号: TN911.73 **文献标识码:** A

A Mining Method of Battlefield Video Target Based Key Frame

Guo Zhongwei, Zhang Shaobing, Qian Xiaojin, Huang Qiubo

(No. 2 Department, Artillery Academy of PLA, Hefei 230031, China)

Abstract: A mining method of battlefield video target based on key frames. Firstly, adopts the CBVR model and extracts the key frames of battlefield video, then recognize the battlefield target from key frames, extract the target information quickly. Program result proves the method is effective through a battlefield target mining prototype.

Keywords: target mining; battlefield video; key frame; neural network; target recognition

0 引言

随着信息技术的发展与广泛应用, 在战场上具有视频侦察能力的装备种类和数量也越来越多, 提供的战场视频情报也在大规模增长。这些视频情报中包含有许多战场态势信息, 不仅信息数据量大, 而且有较大的冗余, 如何从这些视频情报中实时、快速地挖掘出目标, 获取有效的战场信息, 是当前迫切需要解决的一个问题。

关键帧是反映所在镜头中主要信息内容的一帧或若干帧图像。提取关键帧能减少场景中的冗余信息。关键帧的提取方法很多, 主要是利用图论、曲线分裂、聚类 and 奇异值分解的方法, 把一帧看成是多维特征空间中的一点, 选择的关键帧是点的子集, 这些点集或者能覆盖特征距离内的其他点, 或者能表明场景内容的变化。有些方法比较简单, 容易实现且速度较快, 但精确度不高; 有些方法虽然精度较高, 但计算量也明显增加。对于战场视频, 关键帧的选取必须能够反映镜头中的主要事件, 即描述应尽可能准确完整; 同时, 为便于管理, 数据量应尽量地小, 且计算不宜太复杂, 耗时要短。故针对战场视频给出一种简单有效的关键帧提取方法。

1 战场视频关键帧提取

1.1 战场视频图像的互信息量

战场视频某一帧图像可以看作是一个二维的随

机变量, 其熵值表示该帧图像所包含的平均信息量的大小, 熵值越大, 该帧图像包含的信息量越大。对于一帧图像, 可认为其各自像素的灰度值为独立样本, 则这帧图像的灰度分布为 $p = \{p_1, p_2, p_3, \dots, p_{L-1}\}$, p_i 为灰度值为 i 的像素点的个数与该帧图像总的像素点之比。设帧图像的大小为 $M \times N$, 帧图像 A、B 具有相同的灰度级别 (如都是 256 级灰度图像), 设帧图像的灰度级为 L , 令 $p_A(a)$ 和 $p_B(b)$ 分别表示帧图像 A、B 的概率密度函数。概率密度函数可以方便地由帧图像的直方图除以帧图像总的像素个数得到。令 $p_{AB}(a, b)$ 表示帧图像 A、B 的联合概率密度, 它是在求出的帧图像 A、B 的联合灰度直方图的基础上, 除以图像总的像素个数得到。

对于相邻的连续两帧, 分别独立地计算其 3 个 RGB 分量的互信息量。由互信息量公式可知, 相邻视频帧 t 、 $t+1$ 在 R 分量上的互信息量可以表示为:

$$I^R(t, t+1) = \sum_{a=0}^{a=L-1} \sum_{b=0}^{b=L-1} P_{AB}(a, b) \log \frac{P_{AB}(a, b)}{P_A(a)P_B(b)}$$

帧 t 、 $t+1$ 之间总的互信息量可表示为:

$$I(t, t+1) = I^R(t, t+1) + I^G(t, t+1) + I^B(t, t+1)$$

1.2 基于互信息量的战场视频关键帧提取

采用相邻帧间的互信息量 $I(t, t+1)$ 作为特征参

收稿日期: 2010-04-03; 修回日期: 2010-05-17

基金项目: 国家自然科学基金项目线性鉴别特征提取方法的深化研究 (60975006)

作者简介: 郭忠伟 (1974-), 男, 河南人, 博士, 从事作战指挥、指挥自动化、目标识别的教学和科研工作。

数参与帧间变化的计算, 选取的特征参数能有效地描述帧间变化。用相邻帧 f_t 和 f_{t+1} 间互信息量来描述战场视频变化的可能性, 互信息量 $I(t, t+1)$ 的值越小, 则发生视频镜头变化的可能性越大, 反之则越小。计算所有 f_t 和 f_{t+1} 间的互信息量 $I(t, t+1)$, 然后使用基于滑动窗口的自适应阈值的方法, 取滑动窗口 W 的长度为 25, 以当前帧 f_t 为窗口中心, 计算 f_t 和 f_{t+1} 间的互信息量 $I(t, t+1)$ 以及整个窗口内相邻帧间的互信息量的均值 \bar{I} , 若 $\bar{I}/I(t, t+1)$ 的比值大于某阈值 θ_1 时, 就认为视频镜头发生了显著变化。

利用帧间互信息量作为提取关键帧的依据, 可得到镜头序列为 $S = \{S_1, S_2, S_3, \dots, S_N\}$, 镜头 S_i 帧序列为 $S_i = \{f_1, f_2, f_3, \dots, f_N\}$, 镜头 S 相应的相邻帧间的互信息量为 $I_S = \{I_{1,2}, I_{2,3}, \dots, I_{N-1,N}\}$ 。则镜头 S 互信息量的均值 \bar{I}_S 为:

$$\bar{I}_S = \frac{\sum_{t=1}^{N-1} I_{t,t+1}}{N}$$

镜头 S_i 、 S_{i+1} 互信息量的平均值的差异 $\bar{I}_{S_i, S_{i+1}}$ 为:

$$\bar{I}_{S_i, S_{i+1}} = \bar{I}_{S_i} - \bar{I}_{S_{i+1}}$$

选取的互信息量特征参数能有效地描述镜头的变化, 而且对镜头突变和渐变有较强的鲁棒性, 分类后的镜头序列则使用聚类的方法进行关键帧提取。算法中使用参数 θ_2 来控制聚类的个数。首先, 将第一个镜头 S_1 序列的相邻帧间互信息量的值赋值给第一个聚类 C_1 , 然后将其后的镜头 S_2 序列的相邻帧间互信息量的值也加入聚类 C_1 , 根据镜头 S_i 、 S_{i+1} 互信息量平均值的差异公式, 计算该聚类中的互信息量差异的平均值, 若 $\bar{I}_S > \theta_2$, 则表示视频内容已经发生变化, 此时将 C_1 作为一个单独的聚类, 并将刚刚加入的镜头 S_2 相邻帧间序列的互信息量形成一个新的聚类 C_2 , 并重复前述过程。通过聚类, 镜头 S 被重新分成镜头组, 即 $C = \{C_1, C_2, C_3, \dots, C_k\}$ (k 是形成的聚类的个数), 每个聚类 C_i 均由相似

的帧序列组成。然后对每个聚类取互信息最大帧、最小帧、接近 C_k 类中平均互信息量 \bar{I} 的帧做为该镜头组的候选关键帧。最后, 如果一个聚类 C_i 中的互信息最大帧和互信息最小帧间差大于每秒帧率的一半, 就选取最大和最小 2 个作为关键帧; 如果一个聚类 C_i 中的互信息最大帧和互信息最小帧间差小于每秒帧率的一半, 就选取最接近类平均互信息的那帧作为关键帧, 从而得到该视频的关键帧序列。

以一登陆视频为例, 运用上述提取关键帧算法, 取阈值为 1.18, 编程实现提取结果如图 1。这里阈值的取值与提取结果关系很大, 如果阈值过小, 就会出现过检测, 提取出多个相似的关键帧, 而如果阈值过大, 就可能出现错漏实际的关键帧。经过多次实验比较, 阈值取 1.18 比较合理。在提取的关键帧图示中, 无论从查全率还是查准率来进行分析与评估, 这些提取的关键帧基本上都较好地表示了该登陆视频的主要内容和特点, 为以后的目标识别处理提供了方便, 大大减少了数据量。不过图 1 中也出现了多个相似的关键帧, 这是由于算法中只考虑了颜色直方图特征的原因。其实视频帧中目标的运动和形状也是关键帧提取的重要特征, 需要综合考虑, 从而提取出准确的关键帧。

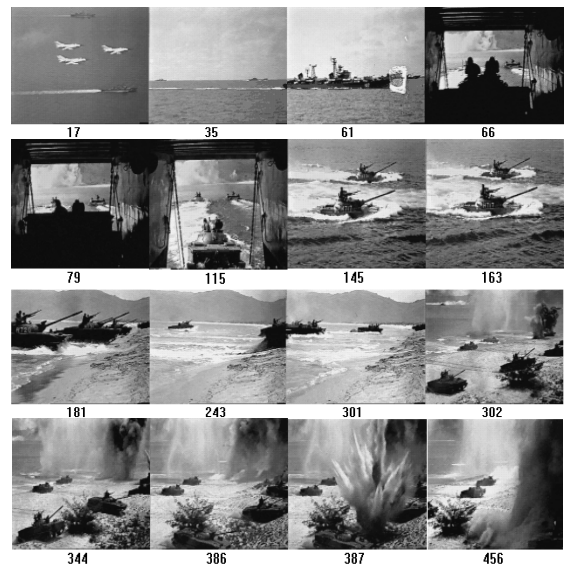


图 1 提取的关键帧图像

2 关键帧图像目标识别

2.1 关键帧的不变矩特征

由于在战场侦察中的拍摄角度不同, 视频中的目标形状会发生改变, 而目标和摄像机的相对运动

则会导致目标相对距离的改变。因此，在进行关键帧中的目标识别时，需要找出一种对目标在视场中的不同位置、大小和方向变化不敏感的特征，而不变矩特征就具有旋转、比例及平移变换不变性，故在对关键帧图像中的目标进行识别时，提取的特征是不变矩。

矩是一种完备的数学表示，其优点是能直接用于包含感兴趣目标的区域而不需要先把目标分离出来。区域的矩是用所有属于区域内的点计算出来的，不太受噪声等的影响。矩是基于区域的物体形状表示方法。假设 R 是用二值图像表示的物体，则 R 形状的第 $p+q$ 阶中心矩为：

$$\mu_{p,q} = \sum_{(x,y) \in R} (x - x_c)^p (y - y_c)^q$$

其中， (x_c, y_c) 是物体的中心。为获得缩放无关的性质，可以对该中心矩进行标准化操作：

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma}, \quad \gamma = \frac{p+q+2}{2}$$

基于这些矩，Hu 提出了利用二阶和三阶中心矩构成的矩组，称为 Hu 不变矩，一共有 7 个矩：

$$\begin{aligned} \phi_1 &= \mu_{2,0} + \mu_{0,2} \\ \phi_2 &= (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2 \\ \phi_3 &= (\mu_{3,0} - 3\mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2 \\ \phi_4 &= (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} + \mu_{2,1})^2 \\ \phi_5 &= (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] + \\ &\quad (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{2,1})^2 - 3(\mu_{0,3} + \mu_{1,2})^2] \\ \phi_6 &= (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1})^2] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1}) \\ \phi_7 &= (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] + \\ &\quad (\mu_{3,0} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1})[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2] \end{aligned}$$

2.2 目标样本库的构建

为获取较好的识别效果，需要构建目标样本库。样本库中的样本要尽可能多，才能提取目标更多本质特征，使分类器得到充分训练，才能得到最佳的权值进行具有高识别率的目标识别。另外，在训练样本与测试样本数量保持一定的比例下，足够多的训练样本可以对更多的测试样本进行识别。故建立了坦克、轮船、车辆等 3 类目标的图像数据库，并对每个样本从不同角度、不同姿态提取 7 个不变矩特征，为分类器训练与识别做好准备。

2.3 目标识别分类器设计

分类器采用 BP 神经网络结构进行设计。BP 网

络属于多层形状的人工神经网络，其本质功能是通过简单非线性函数的数次复合来实现输入到输出的高度非线性映射，隐含表达现实物理世界存在的极其复杂的非线性函数关系。设计的 BP 神经网络分类器包括输入层、隐含层、输出层，输入层神经元为 7 个，隐含层神经元为 15 个，对于输出层，仅考虑车辆、轮船、坦克 3 种类型，对应着 3 种结果，所以，输出层神经元为 3。神经网络各层神经元的输入输出映射是通过激活函数来实现的，激活函数采用的是 sigmod 型函数 $g(x) = 1/(1 + \exp(-x))$ ， $g(x)$ 具有单调递增的特性，当 x 趋于正负无穷时，函数 $g(x)$ 分别达到它的最大值和最小值 1 和 0，即该函数输出区域为 [0,1]，与期望输出数值范围一致。采用 VisualC++ 6.0 编程工具开发了分类器。

2.4 关键帧目标的识别

利用所构建的目标样本库，对设计的分类器进行训练测试。满足误差要求后，对从战场视频提取的某一关键帧（图 2），经过预处理（图 3）以及特征提取后，利用分类器进行识别，得到如下结果：

0.999 999 895 418 452 8 0.000 001 876 260 734 0
0.000 043 430 682 521 1

对计算结果进行四舍五入（精确到千分位），则以上的输出结果整理为：1 0 0

即对应的目标为坦克，与实际情况相符。



图 2 识别提取的关键帧

图 3 预处理后的关键帧

3 结论

该方法用关键帧代替战场视频，大大减少了数据量，起到了消冗的效果。实例表明，该方法为从战场视频中快速挖掘目标信息提供了一个有效可行的选择，也为战场图像情报处理提供了一个先进的技术手段，可作为支撑技术广泛应用到战场视频情报处理系统之中。

参考文献：

[1] Jiawei Rong, Wanjun Jin, Lide Wu. Key Frame Extraction using Inter-Shot Information[C]. Taiwan: IEEE International Conference on Multimedia and Expo for

Video Technology, 2004.

[2] U Gargi, R Kasturi, S H Strayer. Performance Characterization of Video-Shot-Change Detection methods[J]. IEEE Trans. on Circuits and Systems for Video Technology, 2000, 10(1): 1-13.

[3] 章毓晋. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003.

[4] 薛模根. 数字图像处理及其军事应用[M]. 北京: 解放军出版社, 2005.

[5] 周明全. 基于内容的图像检索技术[M]. 北京: 清华大学出版社, 2007.

[6] 艾海舟, 武勃. 图像处理分析与机器视觉[M]. 北京: 人民邮电出版社, 2003.

[7] 郭忠伟, 张中廷, 等. 一种战场图像目标角点提取方法[J]. 兵工自动化, 2009, 28(8).
