

小辉辉的博客

-----那些逝去的岁月，用文字记录我们的心路历程

[首页](#) [日志](#) [LOFTER](#) [相册](#) [音乐](#) [收藏](#) [博友](#) [关于我](#)

Making a Long Video Short: Dynamic Video Synopsis

2013-02-02 22:12:41 | 分类: [Computer Vision](#) | 标签: [运动检测](#) [费用函数](#) [能量最小化](#) [计算机视觉](#) [视频摘要](#)
[订阅](#) | [字号](#) | [举报](#)

此译文为本人原创，若要转载请注明。

作者: Alex Rav-Acha Yael Pritch Shmuel Peleg

联系方式: School of Computer Science and Engineering

The Hebrew University of Jerusalem

91904 Jerusalem, Israel

Email: {alexis,yaelpri,peleg}@cs.huji.ac.il

Abstract

在静止的图像上所建立的视频的作用主要在于呈现动态活动的能力。但由于视频本身固有的时空冗余性质的缘故，视频的浏览和检索并不方便，在视频中有一些时间段内没有物体在活动或者活动发生在一小块图像区域。video synopsis 目的在于呈现出一个紧凑的视频，同时也能够保留原始视频中的必要的活动。

我们提出了动态视频简介，在dynamic video synopsis中，通过同时显示多个活动，原始视频中大部分的物体运动被凝结起来，即使它们原来在不同的时间发生。举个例子来说，我们可以创造出一个“频闪电影”，其中多种移动物体的动态实例被同时播放。这是静止频闪图片的一种扩展。

过去video abstraction的方法大多数通过选择具有代表性的关键帧或者时间段来解决时间上的冗余。在dynamic video synopsis中，物体活动被转移到一显著缩短的片段中，并且片段中物体的活动更加紧凑。视频距离可以在网上搜索

到: <http://www.vision.huji.ac.il/synopsis>

1. Introduction

Video synopsis(or abstraction)是原始视频的一种时间紧凑的呈现，目的在于使得我们能够更容易地进行视频浏览和检索。因此我们提出了video synopsis的一种新的方法，它能最佳地减少视频中的时空冗余。举个例子，考虑图1中代表空时体积的概要视频剪辑。视频一开始有一个人行走在地面上，在一段时间静止之后，一只鸟飞行在天空中。不活动的帧在大部分的video abstraction方法中都被省略。通过同时播放人和鸟，video synopsis本质上更加紧凑。这个方法通过将时间从它们原始的时间段平移至同一空间位置没有其他活动发生的另一个时间段的方法，最优地利用了图像区域。该操作松散了事件的时间一致性，并且首次在[14]中被提出。

图1 输入视频显示了一个行人，在经过没有活动的一个阶段以后又出现了一个飞行的小鸟。

通过同时播放鸟和人，紧凑的video synopsis应运而生。

本文中提到的动态video synopsis不同于过去的video abstraction方法(在1.1部分中提到)，这主要体现在以下两个方面的性质：1 video synopsis本身是一个视频，表达出动态的场景。2 为了尽可能多地减少空时冗余，活动之间的相对时序可能会改变。后面的观点会解释本文中独特的贡献。

第二部分中，我们描述一个产生摘要视频的低级别的方法，这个方法使用马尔可夫随机场[9]的优化方法。

第三部分中，我们提出一个基于物体的方法，在这个方法中，物体被从输入视频中提取出来。相似的运动物体检测也在其他基于物体的video summary方法[7.5.16]中被研究。但是这些方法使用物体检测来鉴别有效的关键帧并且也没有把不同时间段的物体运动组合起来。运动物体的检测，即我们的实验中采用的方法将在1.2部分中描述。

在这项工作中提出的备选方案之一即显示同一个物体多种动态外观。这种效果是频闪图片一种概括，频闪图片曾经被使用在对运动物体的传统video synopsis[6,1]。

既然这项工作提出了一个视频到视频的转换，我们鼓励读者去查看位于<http://www.vision.huji.ac.il/synopsis> 的视频示例。

1.1. Related Work on Video Abstraction

对于video synopsis(or video abstraction)来说，主要有两种方法。在一种方法中，一系列的突出图片（关键帧）从原始视频序列中被选取出来。被选取的关键帧是最能够代表视频的帧[7,18]。在另一个方法中，短视频序列的一个集合被选取出来[15]。第二种方法相对不够紧凑，但是能够给出更好的序列动感印象。这些方法（以及其他的）被描述在关于video abstraction的详尽的研究中[10,11]。

在以上提及的两种方法中，整帧被使用作为基本的建立块。有一种不同的方法使用马赛克图片和一些元数据进行视频检索[6,13,12]。在这个方案中，静态的摘要图片包括不同时间的物体。

1.2. Activity Detection

这项工作假定每一个输入像素已经被标记了其活跃级别。对于活跃级别的评定已经超出了我们的工作范围，并且我们可以使用各种各样的方法来进行不规则性检测，运动物体检测以及物体追踪。

我们已经为我们的实验选取了一个简单而通用的方法作为活跃指示器，即当一个像素在坐标 (x,y) 处的颜色与该点的时间中值之差大于一个给定的阈值时，这个输入像素 $I(x,y,t)$ 被标记为“活跃”状态。

活跃的像素通过特性函数来定义：

$$\chi(p) = \begin{cases} 1 & \text{if } p \text{ is active} \\ 0 & \text{otherwise,} \end{cases}$$

为了将噪声从活跃指示器中清除，在继续任何摘要过程之前，首先对 x 进行中值滤波。

尽管也可以使用连续活跃度测量方法，但在本文中我们关注于二进制的情况。连续活跃度测量可以使用在本文中几乎所有的方程中，仅需要稍作修改。

2. Video Synopsis by Energy Minimization

用3D空时域 $I(x,y,t)$ 表示输入视频序列中的 N 帧，其中 (x,y) 是该像素的空间坐标，并

且 $1 \leq t \leq N$ 是帧号:

- 摘要后的视频 S 应该显著短于原始视频 I
- 摘要后的视频应该含有原始视频中最大化的“活动”
- 摘要后的视频中物体的运动应该与它们在原始视频中的运动相似
- 摘要后的视频看上去应该效果较好, 并且不应有看得见的接缝和零散的物体

具有以上性质的摘要后的视频 S 是通过一个映射 M 产生的。 M 将摘要后的视频 S 中的坐标分配到原始视频 I 中的坐标 (x, y, t) 。 本文中我们关注像素的时间平移并且保持空间位置固定。 因此, 任何摘要像素 $S(x, y, t)$ 可以来自一个输入像素 $I(x, y, M(x, y, t))$ 。 时间平移 M 通过求解能量最小化问题得到, 其中费用函数由下面给出:

$$E(M) = E_a(M) + \alpha E_d(M), \tag{1}$$

其中 $E_a(M)$ 表示在活跃程度上的丢失, $E_d(M)$ 表示接缝之间的不连续性。 活跃程度上的丢失将是原始视频 I 中含有的但是没有出现在摘要视频 S 中的活跃像素的数量。

$$E_a(M) = \sum_{(x,y,t) \in I} \chi(x, y, t) - \sum_{(x,y,t) \in S} \chi(x, y, M(x, y, t)). \tag{2}$$

间断费用函数 E_d 被定义为在摘要后视频和输入视频中相应的邻近视频序列之间空时领域间的接缝之间的颜色差异之和。(相似的公式可以在[1]中找到):

$$E_d(M) = \sum_{x,y} \sum_i \sum_{t=1}^K \| S((x, y, t) + e_i) - I((x, y, M(x, y) + t) + e_i) \|^2 \tag{3}$$

其中 e_i 是刘单元向量, 代表六个时空领域。 空时操作通过最小化费用函数(1)创建短 video synopsis 的示例如图2.a

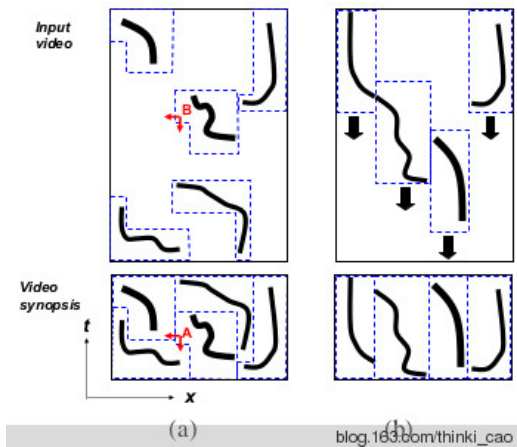


图2 在这个视频的空时表示法中, 运动物体创建了“活跃带”。上部分表示原始视频, 而下部分表示视频摘要。

(a) 更短的 video synopsis S 通过包括大部分活跃像素的方法产生于输入视频 I 。 为了保证平滑,

当视频 S 中的像素 A 对应于视频 I 中的像素 B , 它们的“交界”边缘应该是相似的。

(b) 摘要视频中的连续像素仅限于来自连续的输入像素

注意到费用函数 $E(M)$ (方程1) 对应于3D马尔可夫随机场(MRF)其中每一个节点对应于在输出视频的3D空间中的一个像素, 并且它能被分配对应于一帧输入的任何时间值。 节点上的权重由活跃度费用决定, 同时在节点之间的边缘由不连续费用决定。 费用函数因此可以通过类似于迭代图像分割[9]的算法被最小化。

2.1. Restricted Solution Using a 2D Graph

方程(1)的优化，从而能允许在video synopsis中的任何像素来自任何时间，是一个非常非常大型的问题。举个例子来说，有一个3分钟的输入视频，然后这个视频被总结为大约为5秒钟的视频摘要会产生一个大约有2的25次方个节点的图，每一个节点具有5400个标签。

在[2]中示出对于水平路径移动的动态纹理或者目标来说，通过将问题简化至1D问题可以使得3D马尔可夫随机场(MRFs)可以被有效地解决。在这项工作中，我们解决以更加一般方式运动的目标，因此我们使用了不同的约束条件。在摘要视频S中的连续像素被限制为仅能来自输入视频I中的连续像素。在这样的限制下，3D图表被简化为2D图表，2D图表中的每个节点对应于摘要后视频中的一个空间位置。每个节点M(x,y)的标签决定输入帧I中的帧数t，其中I是显示在S中的第一帧，如图2.b中所示。在S中如何M(x1,y1)不等于M(x2,y2)，那么在两个相邻的位置(x1,y1)和(x2,y2)之间存在一个接缝，而接缝上的不连续费用Ed(M)是所有S中的帧中在这点空间位置的颜色差异之和。

$$E_d(M) = \sum_{x,y} \sum_i \sum_{t=1}^K \| S((x,y,t) + e_i) - I((x,y, M(x,y) + t) + e_i) \|^2 \quad (4)$$

其中ei是4单元向量，描述4个空时领域。

每个节点的标签数量是N-K，其中N和K是输入视频和输出视频中各自的帧数。每个像素的活跃度丢失为：

$$E_a(M) = \sum_{x,y} \left(\sum_{t=1}^N \chi(x,y,t) - \sum_{t=1}^K \chi(x,y, M(x,y) + t) \right).$$

图3显示出一帧原始帧和使用这个限制求解得到的摘要后视频中的一帧。



图3 监控视频中的活动可以被浓缩到一个更短的摘要视频中。(a) 来自原始视频中的一帧典型帧。(b) 来自摘要视频中的一帧。

3. Object-Based Synopsis

早先描述的动态video synopsis的低级别方法受限于满足诸如避免可见接缝之类的局部性质。当物体能被检测到时，可以合并更高级别的基于目标的性质。举例来说，为了避免频闪效应，我们需要检测并且追踪录像中的每一个物体。本部分描述动态video synopsis的基于物体的实现方法。文献中含有几种基于物体的视频总结方法（举例来说[7,5,16]），并且他们都使用关键帧选择的方法进行物体检测。不同于这些方法，我们在时域平移物体并且创造出新的从未在输入视频序列中出现过的摘要帧，从而呢个个更好地利用时间和空间。

运动物体的检测如1.2部分描述的那样通过将每个像素与时域中值进行比较并且对两者的差值设定一个阈值。该部分后面紧跟着一个采用空间域中值滤波器实现的去噪器，并

且通过将时空连接的部件组合在一起，该过程产生了一系列物体，其中每一个物体**b**由它的特性函数来表示。

$$\chi_b(x, y, t) = \begin{cases} 1 & \text{if } (x, y, t) \in b \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

每个物体的切片通过选择该物体出现的帧中的子集而被创造。这样的切片能够代表不同的时间间隔，可能是以不同的采样频率采集到的。

Video synopsis **S**通过使用如下的步骤，将会从输入视频**I**中被构建：

1. 物体**b1.....br**被从输入视频**I**中分离出来
2. 一系列不重叠的切片**B**被从原始视频中删选出来。
3. 时域平移**M**被应用到每一个选择出来的切片中，创造出一个更短的video synopsis并且防止物体之间的碰撞实现无缝拼接。这部分可以在图1和图4中找到解释。图5中显示了一个例子。

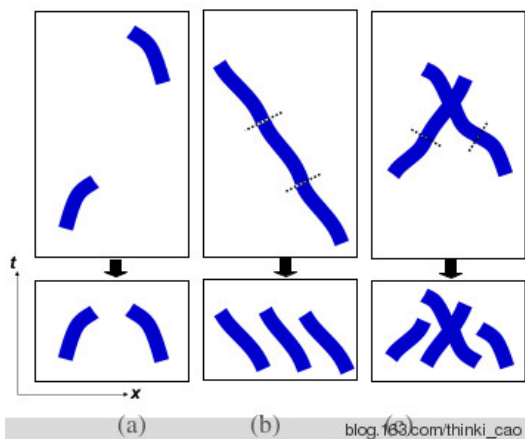


图4. 原理上对于物体在时域上倒换。移动或者激活物体创造了“活动带”。上部分代表了原始视频，下部分代表了摘要后的视频。

(a) 记录在不同时间上的两个物体在video synopsis中被平移进入同样的时间段中。(b) 在一长段时间内运动的一个单独的物体被分为若干个切片，每个切片有着更短的时间间隔，并且那些切片被同时播放出来创造一个动态的频闪效应。(c) 物体的交叉并不会干扰摘要过程，这时物体被分为若干个切片。



图5. 来自某个video synopsis并且具有动态频闪效应的一帧在4.b中例出。该视频位于 <http://www.vision.huji.ac.il/synopsis>

上面所述的步骤2和3是相互关联的，正如我们想要选出若干切片并且将它们时域上平移来得到一个短而无缝的video synopsis那样。

在基于物体方法的表示中，在产生的摘要中一个像素可能会有多种来源（来自不同的物体），因此当所有的物体被缝合在一起时，我们添加了一个后处理步骤。背景 图片是通

过提取像素在该点空间位置的所有帧序列中的时域中值产生的。然后选取出的物体可以被融入背景中，使用正比于距离的权值，这里的距离是指每帧图片中像素与中值图片中像素之间的距离。缝合的机制类似于在[6]中使用的机制。

我们定义被映射到单个摘要像素 (x,y,t) 的所有像素集，其中

$$(x, y, t) \in S \text{ as } src(x, y, t)$$

并且我们将一个物体（或者某个切片） b 中（活跃）像素的数量表示为

$$\#b = \sum_{x,y,t \in I} \chi_b(x, y, t).$$

然后我们定义一个能量函数用来测量切片 B 选取出的子集的费用和时域平移 M 的费用。这个费用包括活跃度丢失 E_a ，这是由于在物体 E_o 和条件 E_l 之间的闭塞而得到的长摘要视频的处罚：

$$E(M, B) = E_a + \alpha E_o + \beta E_l \quad (6)$$

其中：

$$E_a = \sum_b \#b - \sum_{b \in B} \#b$$

$$E_o = \sum_{(x,y,t) \in S} Var\{src(x, y, t)\}$$

$$E_l = length(S)$$

blog.163.com/thinki_cao

(7)

为了能够处理视频，即使运动物体的切片并不完美，我们容忍闭塞的代价还不是完全避免它们。这样，闭塞的处罚能够增加在对物体进行时域倒换时的灵活性，即使切片不完美，并且物体的像素可能包含一些背景。

我们还可以增加一些附加项，使得摘要视频的时间顺序能够偏向于输入视频的时间顺序。

最小化以上所有可能的切片选择 B 和时域平移 M 的能量是非常费力的，这时因为存在着大量的可能性情况。然后，问题可以通过限制求解方法而被显著地缩小。在接下来的部分将描述两个限制策略。

3.1. Video-Synopsis with a Pre-determined Length

在这段中，我们描述具有预定长度的短摘要视频从一段长视频中构建的情况。在这个策略中，每一个物体被分为具有长度 K 的重叠切片和连续切片。所有的切片被时域平移至从时间 $t=1$ 开始，然后我们只剩下决定哪一个切片要被包含在摘要视频当中。我们首先定义所有切片对之间的闭塞费用。把 b_i 和 b_j 作为具有出现时间分别为 t_i 和 t_j 的两个切片，让每个切片的支持由它的特性函数 x 表示（见方程5）。

在这两个切片之间的费用被定义为两个切片之间的颜色差之和，即当它们都被平移到时间 $t=1$ 之后。

$$v(b_i, b_j) = \sum_{x,y,t \in S} (I(x, y, t + t_i) - I(x, y, t + t_j))^2 \cdot \chi_{b_i}(x, t, t + t_i) \cdot \chi_{b_j}(x, t, t + t_j)$$

blog.163.com/thinki_cao

(8)

对于摘要视频，我们选取了一部分切片集 B ，即能最小化方程6中的费用，现在方程中的 E_l 是常数 K ，闭塞费用如下给出：

$$E_o(B) = \sum_{i,j \in B} v(b_i, b_j) \quad (9)$$

为了避免显示同一个空时像素两次（这时允许的但是比较浪费），我们对在原始视频中交叉的切片 b_i 和 b_j 设定

$$v(b_i, b_j) = \infty$$

另外，如果我们不希望有频闪效应，对于所有从同一个物体采样得到的 b_i 和 b_j ，我们可以通过设定

$$v(b_i, b_j) = \infty$$

而去除频闪效应。

模拟退火算法[8]被用于最小化能量函数。每一个状态描述被包含在摘要中的切片子集，并且相邻状态被采纳为一个集合，在这个集合中一个切片被移除，添加或者被另一个切片代替。

在切片选取之后，具有长度 K 的摘要视频通过把所有平移过的切片粘帖在一起而被构建。使用该方法的示例摘要如图5所示。

3.2. Lossless Video Synopsis

对于某些应用，比如视频监控，我们可能喜欢更长的摘要视频，但是在视频中，应该保证所有的活动都出现。在这种情况下，我们的目标并不是像在前面的部分那样选取一系列的物体切片集，而是找到物体切片的一种紧凑的时间重排方式。

再一次，我们使用模拟退火算法来最小化能量。在这种情况下，一个状态对应所有切片的时间平移的集合，如果两个状态的时间平移仅仅有一个切片不同，这两个状态被定义相邻状态。在这种情况下，我们应该注意两个问题：

- 出现在第一帧或者最后一帧的物体切片应该与摘要视频中保持一致。（否则他们可能会突然出现或者消失）。我们关心通过相应地固定所有这些物体时域平移，每一个状态将会满足这个约束条件。
- 输入视频的时域重排通常是一个能量函数的局部极小值，因此它不是初始化退火过程的最优选择。如果所有的物体都重叠的话，我们使用更短的视频来初始化我们的模拟退火算法。

使用该方法的摘要示例如图6所示。

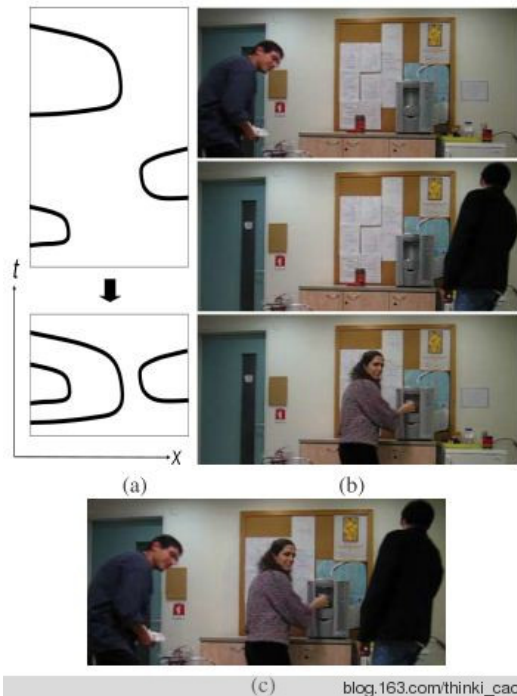


图6. 短摘要描述长视频序列的一个例子，在这个摘要中没有活动的丢失没有频闪效应。三个物体可以在时间上被平移从而能够被同时播放。(a) 原始视频（顶部）和摘要视频（底部）的原理上的空时关系图。(b) 原始视频中的三帧。(c) 来自摘要视频中的一帧

4. Panoramic Video Synopsis

当一个视频录像机正在扫描一个场景时，我们可以使用一幅全景图片来去除许多冗余。然而，目前的方法构建的单张全景图片中，场景动态被丢失。受限制的动态可以由频闪图片来表示，其中运动物体被显示在它们运动路径上的若干位置。

我们可以通过同时显示在场景的不同区域发生在不同时间的活动来创造全景video synopsis。由于每个物体的活动长度仅限于它被摄像机拍到的时间长度，我们得到了大幅度的视频浓缩。一种特别的情况是摄像机跟踪一个物体（比如图7中所示的奔跑的狮子）。在这种情况下，我们只能通过允许频闪效应来得到一个较短的video synopsis。构建全景的video synopsis在方式上与常规的video synopsis相似，但多了初步阶段，在这个阶段中，我们把所有的帧分配至一些参考帧。



图7. 当摄像机跟踪奔跑的狮子时，摘要视频是背景的全景拼接，前景包括奔跑的狮子的几份动态拷贝

5. Surveillance Examples

video synopsis一个有趣的应用可能就是触手可及的存储监控视频。当我们需要检查录像中的特定事件时，使用video synopsis可以帮助我们非常快速地完成。

图6示例出了video synopsis在将所有活动浓缩成一个短片段中的效果，并且没有丢失任何活动。该录像是由录像机监控一家咖啡厅收集到的录像。两个附加例子由真实的监控摄像机给出。图8使用由录像机监视城市街道捕捉到的录像，并且录像中时有行人偶尔穿过视野。他们中许多部分可以被浓缩成一个非常短的摘要。在图9中,video synopsis被应用到由摄像机监控安全围栏得到的视频。在围栏附近几乎没有活动，然后渐渐地我们看到一个士兵朝着围栏爬行。video synopsis同时显示所有爬行以及步行士兵的详情，或者也可以选择通过频闪地播放来使得摘要视频更加短。

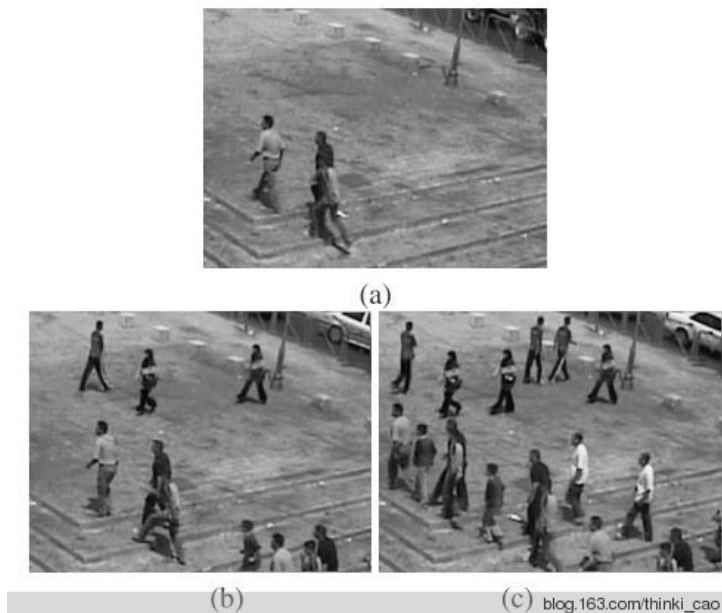


图8. 来自街道监控的video synopsis。(a) 来自原始视频的典型帧（22秒）(b) 来自摘要视频的一帧（2秒）显示了浓缩后的活动 (c) 来自更短的摘要视频中的一帧（0.7秒），显示了更加紧凑的活动

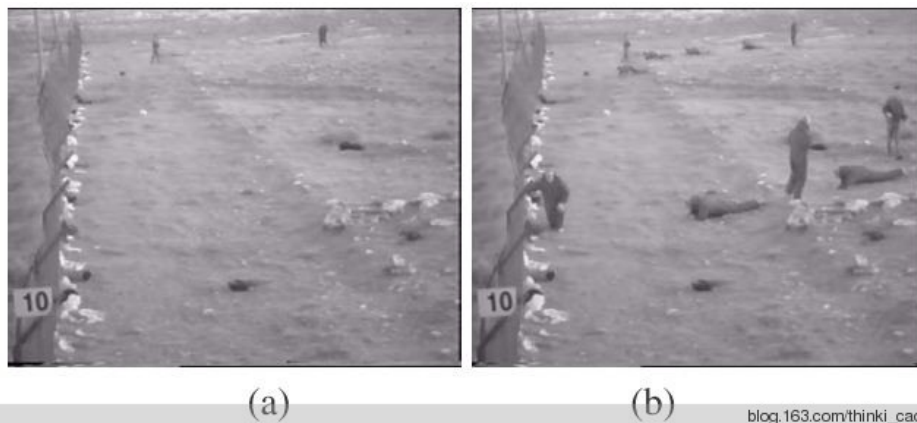


图9. 来自围栏监控的video synopsis。(a) 来自原始视频中的一帧 (b) 来自摘要视频中的一帧，同时显示几个不同爬行和步行士兵的动作

6. Video Indexing Through Video Synopsis

video synopsis可以用于视频检索，为用户提供视频中访问动作的有效直观的链接。这可以通过将每一个摘要像素联系起来，即指向原始视频中相应物体出现情况的指针。

在video synopsis中，视频的信息被投影到“活动的空间”中，其中只有活动的内容有效，不管他们的时间上下文（尽管我们仍然保留了空间上下文）。由于活动在一个短的片段中被关注，因此我们很容易在视频中找到特定的活动。

7. Discussion

video synopsis被提出为一种将视频中的活动浓缩为极其短的片段的一种方法。这种浓缩地表示能够使得我们高效地寻找到视频序列中的活动。

这里提出了两种方法：一种方法使用低级别图表优化，其中摘要视频中的每一个是该图表中的一个节点。该方法具有直接从输入视频中获得摘要视频的有点，然而求解方法可能非常复杂。一个替代方法是首先检测运动物体并且对检测的物体进行优化。虽然在第二种方法中需要一个运动切割的初步阶段，然而它快很多，并且基于物体的限制是可行的。

在产生的video synopsis中的活动比在任何普通视频中的更加凝练，观看这样的一个摘要对于没有经验的观看着来说似乎有些别扭。但是当目标是为了在短时间内观察很多信息时，video synopsis实现了这个目标。我们应该特别关注得到动态频闪的可能性。虽然我们允许在video synopsis的长度上进一步的减少，但是用户可能需要进一步地适应导致的动态频闪。我们确实需要进行一些训练来意识到单个物体在多个空间出现表明了一个更长的活动时间。

尽管我们已经详细说明了动态video synopsis的具体实现，许多扩展仍需可以很方便地实现。举个例子，除了使用一个二进制“活跃”指示器，活跃指示器也可以是连续的。一个连续的活动可以扩展一些可选项来创造摘要视频，举个例子来说，我们可以根据他们的活跃等级来控制显示的物体的速度。

video synopsis也可以被应用于由许多短片段构成的长电影。理论上，由于闭塞（或者不连续）的代价，我们的算法将不会把不同的场景拼在一起。在这种情况下，用于单个场景的简单的背景建模方法必须由可调节的背景估计器代替。另一个可以被应用于长电影的方法是使用目前用于场景边界检测的方法并且每个场景独立地创造出video synopsis。

Reference

- [1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. In SIGGRAPH, pages 294–302, 2004.

- [2] A. Agarwala, K. C. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin, and R. Szeliski. Panoramic video textures. In SIGGRAPH, pages 821–827, 2005.
- [3] J. Assa, Y. Caspi, and D. Cohen-Or. Action synopsis: Pose selection and illustration. In SIGGRAPH, pages 667–676, 2005.
- [4] O. Boiman and M. Irani. Detecting irregularities in images and in video. In ICCV, pages I: 462–469, Beijing, 2005.
- [5] A. M. Ferman and A. M. Tekalp. Multiscale content extraction and representation for video indexing. Proc. of SPIE, 3229:23–31, 1997.
- [6] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu. Efficient representations of video sequences and their applications. Signal Processing: Image Communication, 8(4):327–351, 1996.
- [7] C. Kim and J. Hwang. An integrated scheme for object-based video abstraction. In ACM Multimedia, pages 303–311, New York, 2000.
- [8] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. Science, 4598(13):671–680, 1983.
- [9] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? In ECCV, pages 65–81, 2002.
- [10] Y. Li, T. Zhang, and D. Tretter. An overview of video abstraction techniques. Technical Report HPL-2001-191, HP Laboratory, 2001.
- [11] J. Oh, Q. Wen, J. Lee, and S. Hwang. Video abstraction. In S. Deb, editor, Video Data Management and Information Retrieval, pages 321–346. Idea Group Inc. and IIR Press, 2004.
- [12] C. Pal and N. Jovic. Interactive montages of sprites for indexing and summarizing security video. In Video Proceedings of CVPR05, page II: 1192, 2005.
- [13] A. Pope, R. Kumar, H. Sawhney, and C. Wan. Video abstraction: Summarizing video content for retrieval and visualization. In Signals, Systems and Computers, pages 915–919, 1998.
- [14] A. Rav-Acha, Y. Pritch, D. Lischinski, and S. Peleg. Dynamosaics: Video mosaics with non-chronological time. In CVPR, pages 58–65, Washington, DC, 2005.
- [15] A. M. Smith and T. Kanade. Video skimming and characterization through the combination of image and language understanding. In CAIVD, pages 61–70, 1998.
- [16] A. Stefanidis, P. Partsinevelos, P. Agouris, and P. Doucette. Summarizing video datasets in the spatiotemporal domain. In DEXA Workshop, pages 906–912, 2000.
- [17] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In CVPR, pages 819–826, 2004.
- [18] X. Zhu, X. Wu, J. Fan, A. K. Elmagarmid, and W. G. Aref. Exploring video content structure for hierarchical summarization. Multimedia Syst., 10(2):98–115, 2004.