指导教师：杨涛　　提交时间：2016/3/17

CVPR2015 Paper Translation

No: 01
姓名：孟希
学号：2013302567
班号：10011305

动态场景分割相似度的研究

达米安·特尼特[1] 马修·布朗[2] 迪米特里[2] 彼得霍尔[2]

[1]卡内基·梅隆大学[2]巴斯大学

摘要

本文使用新颖的运动特性和监督学习方法，解决了包括动态纹理在内的任意运动的分割视频的问题。动态纹理是自然场景随处可见的，同时存在复杂的外观和运动模式（例如：水，烟，摇曳的树叶）。这些对于现有的分割方法来说都是很困难的，因为它经常违反光学流所需的亮度恒定的假设，并有超越统一的外观和运动的复杂段的特性。我们的解决方案使用自定义时空时空过滤器捕获的质感和运动线索，用一种新的度量学习框架，优化这种表示特定对象和场景。这是使用在一个层次同时基于分段设置产生先进的动态纹理分割的结果。我们也证明了我们的方法来总体目标和运动分割的适用性，同时显示了无监督分割和效果媲美最好的任务具体办法显著的改善。

1. 引言

我们研究的任意动态场景的视频分割，集中于动态纹理，例如水，火，或摇曳的树木[5]。这些现象是在自然场景的视频共同的地方，但在通用分割基准上[16，40，27]，主要涉及刚性的或光滑的非刚性运动时被表现的很差。动态纹理表现出复杂的外观和运动模式，即通常有超出一个简单的一致性指标的语义。一个描绘树木的序列，可能含有较小和较大的分支，一些静态和其他在风中摇曳的动态的，他们将一般无监督分割方

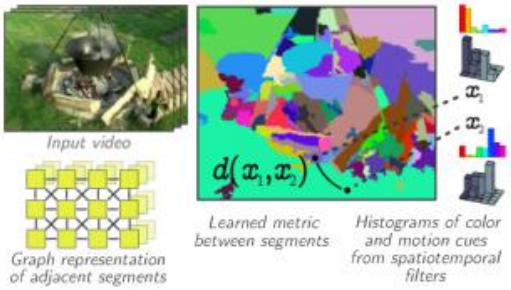


图 1. 我们延长了分层基于图的分割技术，并通过在运动和外观特征远程教育指标应用到动态场景分割。段是迭代的合并。空间和时间上。形成更大的和更大的片段。最“类似的”段之间发生合并，正如我们了解到指标决定

法来分配不同的段，而这些最好都应该为“树”的动态纹理的一部分。这样的中层的解释一个场景超出了大多数“supervoxels”和非监督的视频分割方法[43]。为了克服这些局限性,我们提出更合适的功能特性，其中有为了学习注释和地面实况分割形成的监督算法。

我们的贡献由两部分组成。首先，我们使用的答复时空颞滤波器组来捕捉动态纹理的外观和运动的特征。类

似于可以在静态图像捕捉结构的 2D 滤波器，这些 3D 滤波器可以在视频的音量模式商捕捉结构，如移动模式[11]。在银行的滤波器调谐到各种比例，方向和速度，并提供了一套丰富的表征图像的外观和动力学特征。不同的动态纹理往往会出现不同的运动数据。例如，在池塘的水波纹将表现出比大海的波浪更低频的动作。这个例子被用来作为单个的“水”的类别或者作为在实验的训练集中的一种被证明合法的现象。因此，其次我们展示了怎么样学习做彩色和基于过滤器运动线索直方图描述符之间的度量，这种度量允许在基于图形的框架上监督视频分割以分级[18](图 1)。我们使用地面实况分割生成描述符之间的成对约束，和学习-ric 预测部分是否应该合并分割。如果训练段设有语义标签，我们使用的（“相同”或“不同”级）片段之间的附加成对约束得知明确地分隔多个语义类的度量。这样允许各种各样的训练场景。视频分割总任务包括一些可以从中级模型中受益的应用。同时我们还在经典的动作分割项目和动作边界分割的适用性上评估我们方法[35]。

实际上，我们的工作延伸格伦德曼等人的分级基于图形视频分割技术，它构造减小粒度 supervoxels 的层次结构。该算法在本质上是适用于各种不同的任务，其中所希望的段的大小是不固定的或已知的先验。虽然已经在若干标准的证明是非常成功的算法但是其仍限于像素对外观和/或运动之前均匀的分组。另一个实际，其显著的缺点是它的大内存需求（用块处理如处理[45]），特别是我们提出了丰富的运动特征。我们通过学习解决这两个问题之间的度规的段描述符在监督环境,共同背景降维的描述符,以大大降低内存需求的原始算法。注意，这不等同于第一突出的例子数据 PCA，然后学习上低维的描述符的度量，它可能会失去判别的信息。我们宁可用一个目标函数同时优化任务来充分利用现有的监督。

综上所述，本文的技术贡献有三个方面。（1）一种新的方法来提高分层视频分割与监督学习。不像专为近邻分类现有的算法，我们优化段描述符之间的度量上标记的训练数据，使用适当的分层分割大利方法。（2）我们表征段内的外观和运动具有新颖一套基于时空时间滤波器的特征，它允许分割具有任意运动和复杂的动态纹理视频。（3）我们提供一种方法来优化降维度量标准,这大大降低了原始分割算法的内存需求。用大量的实验和数据集来评估这些贡献，同时证明其价值与现代技术相当。

2. 相关工作

监督视频分割 视频分割已经被广泛研究，尤其是在 supervoxels 的过度分割上。这些 supervoxels 通常在视频的高效表示中用于进一步更高级别的处理。监督式学习是提高分割为特定任务的性能的方式。Jiang 等. [26]

最近发现，监督学习的谱聚类包含可能对视频分割有显著改善的效果。

与我们的做法相关的阿尔珀特等人[1]学会了预测双段指导层次合并的图像分割。在相似性度量的形式下，Jain 等[24]最近还采用强化学习应用到三维医学图像的分割的方式研究学习对凝聚的聚集成的预测。

许多密切相关的中级任务可以配制成分割的问题。例如 Xu 等人[44]集中于分割框架跟踪汽车和人类，使用学习的模型，以改善这些特定类的实例的分割。对象[35,23]和运动[2,37,25]分割是在一般的分割框架内已用特定的方法[2,37,25]处理两个其它的任务。我们评估了中层的任务，显示出监督学习列入显著改善的结果。不同于以往的作品，我们的做法是能够处理更广泛的情况下，例如同时包含动态纹理和刚体运动混合的场景。

动态纹理分割 提取相关运动特征来表征动态纹理是具有挑战性的，因为它们常常违反亮度恒定和（例如，在水或烟的情况下）刚性运动的共同光流的假设。生成模型已经提出了与线性动态系统[12,7]，然后将其用于通过迭代[4,5]拟合分割直接模型图像。最近的作品提出了临时性的，手工制作的描述符。时空时间滤波器被早的方式来提取光流，通过响应给这样的导向滤波器组实际上提供比光流更加丰富的信息，建议。它们允许多个面向结构的捕获在任何时空位置，并同时处理运动（例如翻译对象）和非运动（例如闪烁效果）都具有相同的制剂。Derpanis 等人看着大量进入他们的认可动态纹理和场景的使用，同时 Tenley 和布朗最近使用它们作为视频分割的功能。摘要杆-年龄通过监督学习丰富的这些特性,特别是当使用语义(″类″)注释的训练例子。

度量学习 我们通过学习的相似性或距离对段之间的公制监督学习纳入分割框架。度量学习已经改善大多[41]的最近邻分类的目标研究。例如，Weinberger 等人[42]优化了广义马氏指标，这相当于输入功能的线性变换，进一步密切同等级的近邻，分开的不同类别。西蒙尼扬等人[34]通过优化非正方形线性投影矩阵执行判别维数的降低，而不是最近的邻居的性能。我们尝试这两种方法，发现西蒙尼扬等人在大幅制定执行在我们的案例（第 5 章），最终提高我们的段描述符来预测他们是否具有再分割合并的能力。

3. 分层视频分割

现在，我们将提供基于图形的层次的视频分割算法[18]的简要回顾，然后展示如何监督学习指导流程集成。

输入视频被连接的区域的曲线图表示，将其初始化为视频（图 1）的相邻像素的晶格。通常，图形的点用 N 表示，边用 ε 表示。每个点用 x_i 表示。边缘拓扑结构表示为 26 的相连的像素。

i 和 j 之间的边表示长度的权值定义为 $e_{ij}=d\left(x_i,x_j\right)$。这是距离函数是无监督的或者被学习提出。分估算法的反复进行，在每次的迭代中，合成节点 e_{ij} 对应节点 i 和 j。合并后的节点为 k，在分配一个新的描述符为 x_k 图结构 E 被更新为表示在视频体积空间和时间邻接，并且边缘加权 $e_{kl}\forall l$ 的相应的更新。节点代表正在增长的段数。他们在严格的增长，并且组成了段。大多数实现算法的分割在一些离散的“水平”分割，如当一个固定比例的边缘已被移除，但该算法有效地产生一个连续的树从两两合并区域[1]。

该段的外观由[18]的颜色直方图和基于过滤器运动线索（第 4 节）的直方图描述使用直方图作为描述符有两个主要优点。首先，它们能够有效地更新，这些原来的两个节点中的一个合并的节点是加权平均（用相对大小）的柱状图。第二，重要的是直方图可以有效的在不同规模的段内出现。分割算法和重新计算算法的相似性证明了分割算法的有效性，虽然其中掺杂了粗细的信息。然而，这样的结果是由于这些权重是使用更新描述符计算的，合并的边缘的权重不再单调递增。直观的来看，小的不同在大

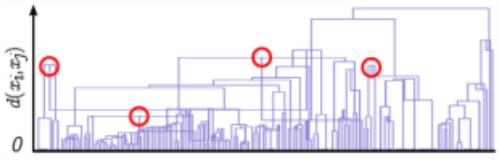


图 2. 一个典型的分段的树形图（图序列 7）。最低的像素级别没有被清楚地显示。纵轴表示合并段，其并不是单调递增的，通过在树形反转指示之间的距离（圈）。这源于段描述符在不同规模被重新计算，并激励采用不同的段大小不同的指标。

规模的图像的边沿的连接来说对其影响不大。这可以看作在合并处理（图 2）的树形图反转，这有可能在与重新计算的距离分级聚类（如例如在质心聚类[29]）。实际上，这一观察表明段之间的距离函数可以适用于段的实际大小受益被比较（第 5.2 节）。

4. 时空时间滤波器运动线索

我们使用 3D 时空时间滤波器[15，10]，这有助于在视频的音量显示的结构表征质感和运动视频。考虑类高斯的三个变量，$G(x,y,t)=e^{-(x^2+y^2+t^2)}$，我们使用他的三阶导数 $G3_{\hat{\theta}}(x,y,t)=\frac{\partial^3 G}{\partial\hat{\theta}^3}$ 和希尔伯特变换 $H3_{\hat{\theta}}(x,y$，导流到单位矢量 $\hat{\theta}$（G3 的过滤器的对称轴）的时空方位。我们用 v_σ 表示与弗吉尼亚州的空间尺度叠加帧的视频的音量 $\frac{1}{\sigma}$。在给定的规模

的σ对于给定的$\hat{\theta}$能量响应然后通过测定

$$E_{\hat{\theta},\sigma}=(G3_{\hat{\theta}}*v_{\sigma})^2+\left(H3_{\hat{\theta}}*v_{\sigma}\right)^2 \quad (1)$$

在这里，*表示卷积。 希尔伯特变换对应用于$\pi/2$ 的相移，正交对滤波器的$G3/H3$允许提取光谱强度独立的相位[15]。例如，一个过滤器转向到θ=（1，1，1）将响应斜型态 1 的 pX/帧移动，而过滤器转向到θ =（1,0,0）将捕捉垂直静态的结构。

滤波器响应（式 1）是敏感的图像中的对比。我们相对于全部过滤器的响应的总和与限定捕获缺乏在图像的非纹理区域结构的一个附加的信道，每个像素正规化减轻图[11]：

$$E'_{\hat{\theta},\sigma_j}=\frac{E_{\hat{\theta},\sigma_j}}{\sum_k E_{\hat{\theta},\sigma_j}+\in} \quad (2)$$

$$E'_{\in,\sigma_j}=\frac{\in}{\sum_k E_{\hat{\theta},\sigma_j}+\in} \quad (3)$$

在这里，$\in$是$E'_{\hat{\theta},\sigma_j}$的噪声阈值。测量量$E'_{\in,\sigma_j}$和$E'_{\hat{\theta},\sigma_j}$连在一起形成$L-1$规范化运动直方图。在这里，我们不总结多个过滤器的响应在现有的工作[11，39]，其中，作者寻求独立运动图像外观的测量。而我们认为这种外观的信息是有价值的，该过滤器分组随后可以在我们的指标中了解。

5. 学习段间的度量

在凝聚分割过程（第 3 节），段通过颜色和运动直方图描述，我们现在展示如何学习指标来比较这些段描述符。在细分的每次迭代中，两个连接段之间距离最小的发生合并，并且仅这个指标决定了整个过程的结果。我们

用带有$x_1,x_2\in\Re^d$比喻为广义马氏指标如下

$$d_L^2(x_1,x_2)=(x_1-x_2)^{\mathrm{T}}L^{\mathrm{T}}L(x_1-x_2) \quad (4)$$

其中，$L^{\mathrm{T}}L\in\Re^{d\times d}$是对称的半正定矩阵马氏定义的指标。

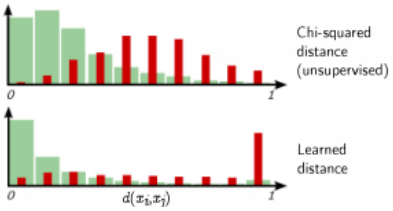


图 3. 我们寻求预测他们是否应合并(绿色）或不（红色）段的描述符之间的距离的功能。相比无监督度量诸如卡方距离（上部），度量与我们的大裕度优化好得多，其预测了解到和分离两个结果（低级）。

培训的形式提供数据与手工分割视频，从中我们产生一组对各种尺寸的段之间的约束的视频。为了获得现实的限制，我们模拟出一种无监督的距离度量，其中合并只允许在与地面实况相一致的层次分割过程。合并后，更新的边缘被添加到设置作为附加成对约束的训练。从形式上看，段j和i之间的一个这样的限制被描x_i和x_j和注释定义$y_{ij}=1$如果它们对应于同一个基本事实的标签，$y_{ij}=-1$。我们的目标是用 L 来优化方程 4，使得要合并

节点之间的距离是相对于所有其他对区隔不应该的小。最后，

$$d_L(\chi_i,\chi_j) \ll d_L(\chi_k,\chi_l) \quad (5)$$

$$\forall i,j,k,ls.ty_{ij}=+1,y_{kl}=-1 \quad (6)$$

注意，这与度量学习设计为最近邻分类（例如 LMNN[42]），其中的距离只应小至同一类的几个点. 这促使每个训练对 I，J 中使用的大裕度约束[34]：

$$y_{ij}(t-d_L^2(\chi_i,\chi_j))>1 \quad (7)$$

其中t是分离的阈值。此限制来预测y_{ij}与在t的任一侧的 1 缘的距离。约束分为使用 hinge-loss-耦合公式表示目标函数:

$$\arg\min_{L,t}\sum_{i,j}\max\left[1-y_{ij}\left(t-d_L^2\left(x_i,x_j\right)\right),0\right] \quad (8)$$

我们使用随机梯度下降优化这一目标如[34]。

5.1 联合度量学习和降维

检验公式 4，第一方面我们可以发现广义马氏距离等同于用在投影空间中的数据的欧氏距离用L,i,e变换的。

$$d_L^2(x_1,x_2)=\left\|\left(Lx_1-Lx_2\right)\right\|_2^2 \quad (9)$$

这使我们能够在预处理步骤中的段描述符与 L-变换，并让具有较低维数据分割算法工作。该降维易于集成到目标函数（8），因为我们优化在矩阵 L，相对于现在马氏矩阵矩阵 L'L。需要注意的是减少我们的段描述符的尺寸与线性变换与直方图通过加权平均的合并，这使所学指标作为一致段合并是一致的。

5.2 多尺度度量学习

该距离度量是用于比较增加大小作为分割算法进行的段。如在第 3 节所提到的，小的段的典型直方图可能跟那些较大的不同。因此，合并他们的最优准则可能也不同。换句话说，特征是在该图像（例如色彩一致性）一个小规模的关系可能不那么重要了具有大空间范围（其中，某些动作可以在更相关）的片段。因此，我们尝试用适应了解到距离不同的尺度。

对于每一对连接节段i,j我们表示由最小两个图覆盖的a_{ij}的图像区域。8 被定义为尺度a_{ij}的范围，不同的是L_s的范围是（$s=1....S$）。L_s用于比较一给定段i,j，他们的范围可以根据a_{ij}确定。校验值标与跨度相媲美是至关重要的。由于优化的目标（8）是不L中凸，我们通过局部重新优化的L_{s-1}获得以下比例的L_s（见算法列表）。这鼓励了跨尺度指标变化平稳,即使这些定义与固定的,硬阈值。这也有助于避免过度拟合大尺度的稀缺训练数据。请注意，使用比例依赖度量（8>1）不允许预伸出的数据转换成一维的下部空间。

6. 结论

我们评估上提供有监督的方式明确潜在的 3 分割任务的建议方法：动态纹理分割，刚体运动分割，并从运动边界的对象的分割。

我们把我们的结果与每个任务的不同的状态做对比。请参考补充材料有关评估协议提供额外的结果和细节。

完成 我们的分割算法的实现如下[17]。颜色直方图使用用 3×10 箱的实验室空间。运动直方图使用的滤波器组跨越 2 秤，8 方位，以及 0 和 2 像素/帧 9 之间的速度，以及用于捕获闪烁的过滤器。这导致存在尺寸 175 的段描述符（155 灰度的视频）。分割是在使用逐像素的颜色差异为边权重的最低水平自举，直到段有足够的规模（$a_{\min}$ =20 像素）考虑有意义的直方图[18]。

目标 我们比较对各自国家的最先进的方法，对每个任务进行考虑。我们也评估对三种备选方案所提出的贡献的价值：（1）使用直方图 2 之间的卡方距离的无监督分割（如[18]）（2）学习一个度量标准用一个简单的线性逻辑符号的回归量解释直方图。（3）学习国家的最先进的算法 ITML[9]，包含广义马氏距离，但没有降维的度量。

6.1 动态纹理分割

我们评估的动态纹理的合成器 DB [3]和 Dyntex[31]数据集的分割。合成器 DB 数据集[3]由火，水，烟，植被等的真实镜头 K 个补丁的拼贴画组成。该序列是非常具有挑战性的：影片灰度，一些纹理表现出非常相似的静态外形（图 4），同时相邻的纹理有时具有几乎相同的灰度值。捕捉运动是这个关键，实现动态纹理，只可以在其动态外观不同的地方分割（例如，在不同的方向上被风吹树枝在不同的频率，或吸烟的补丁移动）。我们使用具有 $K=2$ 和 $K=3$ 的序列分别用于训练和测试。地面实时的情况分割作为其训

算法 多尺度的监督学习，从成对约束段之间的大比建立指标。

输入：

$C=\{(x_i,x_i^{'},y_i,a_i)\}i$ 包含对段之间的约束

$x_i\in\Re^d$：段描述符

$y_i=+1/-1$：相同/不同的地面实况标签

a_i：每帧的最小的两个段的面积

$p<d$：段描述符的有效维数

$S\geq 1$：尺度的数目

λ：梯度下降的学习率

输出：

判别投影矩阵 $L_s(s=1...S)$

程序：

初始 $L\leftarrow p$ PCA 所有的最大分量 $x_i,x_i^{'}$ 在 C 里面

$t\leftarrow\frac{1}{2}[mean\left(d_L^2\left(x_i,x_i^{'}\right),y_i=+1\right)+$

$mean\left(d_L^2\left(x_i, x_i^{'}\right), y_i = -1\right)]$

从 C 中找到 S 的分布范围 a_i

对于每一个尺度 $s = 1...S$

选择子集约束 $C' \subset C$，范围 $a_i \in range_s$

从规模前初始化：$L_s \leftarrow L_{s-1}$，$t_s \leftarrow t_{s-1}$

循环

随机在 C' 中选择一个 i，$y_i = +1/-1$

IF $y_i(t - d_{L_s}^2(x_i, x_i^{'})) < 1$

$$L_s \leftarrow L_s - \lambda y_i L_s (x_i - x_i^{'})(x_i - x_i^{'})^{\mathrm{T}}$$

$$t_s \leftarrow t_s + \lambda y_i t_s$$

ELSE t_s，L_s 不变。

UNTIL 目标（8）不再减少抵抗验证集

训练和监督的对象，同时其具有 12 种分类。接下来，我们将在我们国家的先进结果基础上进行介绍。

无监督的 分割与强度直方图独自在这样的动态纹理表现不佳（图 4），这表明传统的外观为基础的分割方法的不足。即使当进行无监督分割，所提出的基于过滤器的运动特征显著改善了性能，这表明这些功能捕获相关信息。

手动分割学习 我们首先使用地面实况分割掩码，产生在训练例子我们相邻节段之间的约束（第 5 章）以此了解我们的距离度量。我们获得了无监督分割一致的 IM 改善（兰德指数从 72.7％提高到 89.7％）。我们在图 4 精确地展示了分割减少描述符的大小。这个数字表明，降维可以积极执行，合理降低性能和内存占用较原来的 10 倍。这表明原来的描述符中存在大的冗余。有趣的是，我们观察到的维度对应强度直方图在预测符号符时被给予很少的关注。此外，中等尺寸-实施减少甚至可能以通过丢弃无关维度的预测描述符的方式来避免过度拟合训练数据。这个“甜蜜点”为尺寸 p 的数量是通过对训练数据的交叉验证且容易辨认。

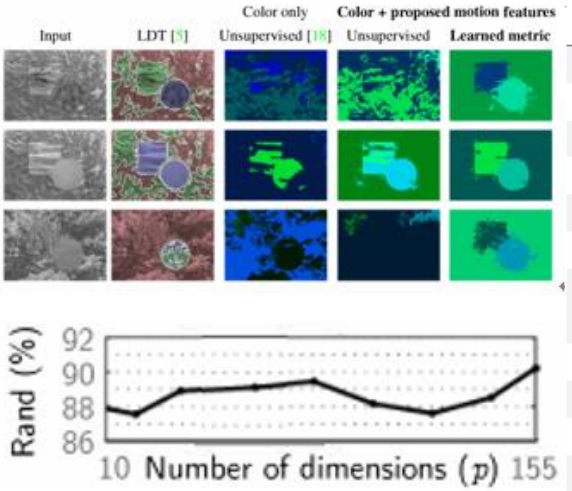
语义标签学习 然后，我们使用从训练数据为 12 类的语义标注（如草，河，蒸汽，池塘）来做额外的约束。约束标签实在训练段中随机选择的，其中 $y = +1/-1$ 对应相同或不同的标签(第五章)。学习距离是明确地训练来区分不同类别的纹理。这带来了在细分不同类型的纹理性能的附加改进（90.2％兰德指数）。我们对大多数 100 个测试序列进行了近乎完美的分割（图 4）。我们大范围报道此数据集的所有结果，特别是那些[39]，强迫静态段（而我们执行一个真正的视频分割），以及[6]使用手工，特别设计的描述。

现在，我们评估我们在 Dyntex 数据集[31]真正的视频能力段动态纹理。我们使用的合成器 DB 数据了解到的距离（从手动分割和语义标签，$p = d$）来分割由以前的作者使用场景

[13，6]。如图 5，这些序列是特别具有挑战性的。它们包括各种动态纹理（蒸汽，水喷溅和波纹，烟雾等），一些记录有移动摄像机。我们可以细分出这些具有非常精确的轮廓纹理。虽然没有基础事实是可用的，但我们的结果是定性的跟现有方法[13，6]一样好或更好。

6.2. 对象边界

接下来，我们考虑在 CMU 数据集[35]其运动的界限分割对象的任务。这个先前已处理，都用到无监督分割的方法[39]，并使用上训练图像中的候选边界（边缘）的分类器的方法。



方法		平均范围
特征	度量标准	
颜色	无监督	59.9
颜色+运动	无监督	72.7
颜色+运动	学习型，回归	77.1
颜色+运动	学习型，ITML	86.4
颜色+运动	从手动语义学习	89.7
颜色+运动	从手动语义学习，语义层 p=d	90.2
GPCA[33]		55.4
LDT[5]		89.4
DTM[3]		82.5
LBP/WLD[6]		88.4

图 4.我们获得有关动态纹理（合成 DB 数据集）分割的国家的先进成果，同时兰德指数计算超过了 100 个测试序列。我们所提出的降维只允许其性能轻微下降，同时还要提高超过 10 倍的内存占用。

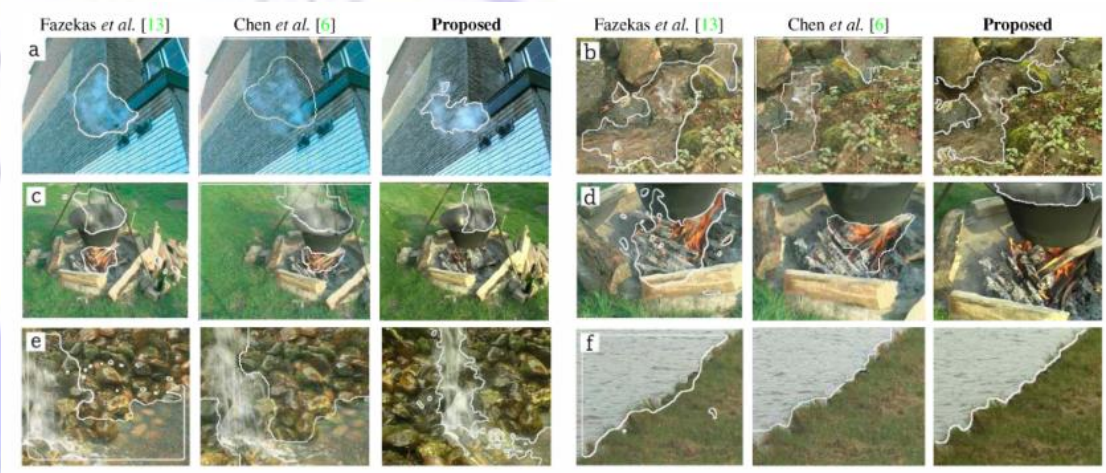


图 5. 在 Dyntex 数据集的复杂场景动态纹理分割。我们恢复了很多比现有方法更精确的界限。(a)蒸汽通风系统.（b）蜿蜒小溪间的鹅软石.（c）营火里面烹饪的蒸汽（d）同一场景的相似镜头(e)水在圆润的石头上翻滚然后聚集成池(f) 在阿姆斯特丹的一条运河和海岸线。序列（A，B，D，F）为场景与移动相机。

数据集是具有挑战性的，因为它包括刚性和非刚性运动，引起在不同深度的视差和由固有运动（动物和人）的混合。此数据集包含 30 个序列，这是我们用来训练用段注释的相似性指标，如第 5 章。我们展示（图 3）我们的优化算法是如何能够发现预测合并对相邻区段（使用用于可视化的地面实况绘出）的概率的距离度量，与两个之间的大裕度的。这转化为一个改善的无监督的分割（表）。我们能够恢复对象边界，性能接近这个的国家的最先进的方法的数据集。更优越方法是利用其它线索，如边界和段的几何特性。有一种可能性，探讨我们的框架扩展到使用比直方图之间的算术差的其他段的特性的一般功能上去。我们还评估维数降低的段描述符（图 6）的效果。我们再次观察较小段描述符性能所希望的平滑的下降，其允许可忽略的或仅在准确性适度损失内存在使用显著改善。较小的维数再次证明在避免了训练数据的过度拟合上是有益的。我们比较所提出的维数降低，以度量共同优化，针对一个微不足道的突起上最大 PCA 分量（图 6）。它显示了集成了一个小规模的描述符和高度歧视性指标的两个目标的联合方式具有明显的优势。

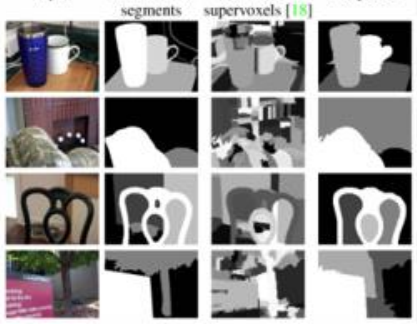
6.3 刚性运动

现在，我们考虑麻省理工学院人类标记的数据集[28]刚性运动的分割。它拥有具有内在的运动（例如汽车，狗），并在不同深度视差诱发运动的对象。由于数据集过小，我们采用交叉法评估。分割了距离函数为无监督分割对于大多数的数据序列提供了一个小的改进。我们相信，在这个数据集被监控方法的有限的优点是，用于分割使用无监督方法刚性运动变得相对容易。训练数据过小会导致学习方法出现过度拟合。对于“手”序列，如（图 7），其表现不佳，作为学习权重的外观特征（因为训练注释对应大多是刚性运动），即使他们在无人监管的环境下仍能完美分割。小训练集也是为什么采用度量多尺度最可能的原因（章节 5.2）因为其不会导致任何改善。总体来说，我们的性能还是比现有的一些方法[36，37]好，接近[39]国家的最先进的水平（见图 7 和补充材料详细说明）。

我们采用 Logistic 回归比较的学习算法与基线解法和国家的最先进的 ITML 算法。从我们的观察看到基于线性回归的距离不提供足够的自由度来捕捉段描述符的维度之间的关系。这支持广义马哈拉诺比斯的选择作为学习度量。相反地，ITML 算法似乎要么容易发生过度拟合要么无法限制相关的培训，最有可能的原因是由于刚性正则化（使用单位矩阵作为参考如[9]）。

7. 结论和未来的工作

本文讨论了使用一种新型的基于过滤器运动特征相结合，并有监督学习方法的动态场景的视频分割。我们使用时空时间滤波器的响应，包括光流失败[39]以捕捉广泛的现象。我们



方法		边界 F(%)	边界 AP(%)	段，范围（%）
特征	度量标准			
颜色直方图	无监督	51.1	43.2	60.3
颜色+运动	无监督	60.8	61.8	68.3
颜色+运动	S=1	64.1	65.3	72.1
颜色+运动	S=2	64.3	65.4	72.3
颜色+运动	S=3	64.4	65.5	72.4
斯坦和赫伯特[35]		66.7	63.7	
SARGIN 等[38]		57.0	58.0	
桑德博格等[32]		61.0	72.0	
HE 等[21]		47.0	43.0	
泰妮和布朗[39]		60.7	56.3	

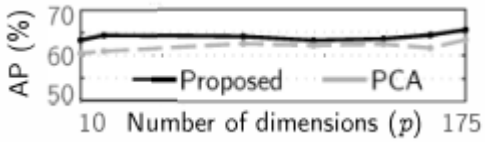
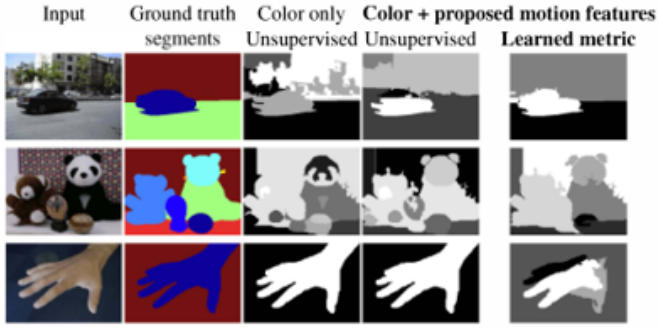


图 6. 左：从遮挡边界对象分割序列的结算数据集的相机翻译。右上：我们的结,F-measure 段边界(64.4%)相当与现有最好的特定任务的方法。右下:提出了降维、优化指标,执行在第一维 PCA 组件无监督分割投影。这突出显示降维和公制学习作为组合目标的重要性。



方法		平均范围
特征	度量标准	
颜色直方图	无监督	68.5
颜色+运动	无监督	78.8
颜色+运动	学习型，回归	76.3
颜色+运动	学习型，ITML	70.3
颜色+运动	S=1	83.0
颜色+运动	S=2	81.1
层++[36]		77.5
N 层[37]		82.3
泰妮和布朗[39]		83,2

图 7. 麻省理工学院人类标记的运动分割数据集。大多数序列受益于学习方法，但该数据集的小尺寸增加了过度拟合的风险。注释，其对应于刚性运动趋向于减轻权重的外观特征，即使他们是用于分割“手”序列（图片的最后一行）。

使用时空时间滤波器的响应，包括光流失败[39]以捕捉广泛的现象。我们使用的基于图的视频分割算法，通过一个学习度量扩展到测量段之间的相似性。客观度量学习减少了所述描述符的维数，减轻了原有的算法[18]的大存储器需求。我们的结果与最好的特定任务的方法做对比，获得了国家的最先进成果动态纹理的分割和论证适用性更广范一般的运动分割。这个广泛的评估带来了一个观点是一种通用的视频分割框架，可以通过将来自注解手动分段视频了解到具体的规则并成功地应用于各种任务。如前[16,28]所指出的存在于人的注释和在每个数据集的明显偏差，并且这将是有趣的未来处理的更多的一般规则和跨数据集性能的学习。这应包括注释或增加的视频分割基准。

引用

[1] S. Alpert, M. Galun, R. Basri, and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. In CVPR, pages 1-8, 2007. 2

[2] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In ECCV, pages 282-295. 2010. 2

[3] A. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. IEEE Trans. Pattern Anal. Mach. Intell., 30(5):909-926, 2008. 5,7

[4] A. Chan and N. Vasconcelos. Variational layered dynamic textures. In CVPR, pages 1062-1069, 2009. 2

[5] A. B. Chan and N. Vasconcelos. Layered dynamic textures.IEEE Trans. Pattern Anal. Mach. Intell., 31(10):1862-1879,2009. 1, 2, 7

[6] J. Chen, G. Zhao, M. Salo, E. Rahtu, and M. Pietikliinen. Automatic dynamic texture segmentation using local descriptors and optical flow. IEEE Trans. Image Processing,22(1):326-339, 2013. 2, 6, 7

[7] D. Chetverikov and R. Peteri. A brief survey of dynamic texture description and recognition. In Int. Conf. Computer Recognition Systems, pages 17-26. Springer, 2005. 2

[8] J. J. Corso. Cvpr tutorial on video segmentation. 2014. 3

[9] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In ICML, pages 209-216. ACM, 2007. 5, 7, 8

[10] K. G. Derpanis and J. M. Gryn. Three-dimensional nth derivative of gaussian separable steerable filters. In ICIP (3),pages 553-456, 2005. 3

[11] K. G. Derpanis and R. P. Wildes. Spacetime texture representation and recognition based on a spatiotemporal orientation analysis. IEEE Trans. Pattern Anal. Mach. Intell.,34(6):1193-1205, 2012. 1, 2, 4

[12] G. Doretto, P. Pundir, S.Soatto, and Y. N. Wu. Dynamic textures. In !1CV, pages 439-446, 2001. 2

[13] S. Fazekas, T. Amiaz, D. Chetverikov, and N. Kiryati. Dynamic texture detection based on motion analysis. !1CV,82(1):48-63, 2009. 6, 7

[14] c. Feichtenhofer, A. Pinz, and R. Wildes. Bags of spacetime energies for dynamic scene recognition. In CVPR, 2014. 2

[15] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. IEEE Trans. Pattern Anal. Mach. Intell.,13(9):891-906, 1991. 2, 3

[16] F. Galasso, N. S. Nagaraja, T. J. Cardenas, T. Brox, andB. Schiele. A unified video segmentation benchmark: Annotation, metrics and analysis. In ICCV, December 2013. 1,2, 8

[17] M. Grundmann, V. Kwatra, M. Han, D. Castro, and I. Essa.Graph-based hierarchical video segmentation. In CVPR Tutorial on Video Segmentation, 2014. 3, 5

[18] M. Grundmann, V. Kwatra, M. Han, and I. A. Essa. Efficient hierarchical graph-based video segmentation. In CVPR,pages 2141-2148, 2010. 1, 2, 3, 4, 5, 7, 8

[19] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In ICCV,pages 498-505, 2009. 5

[20] M. Haindl and S. Mike. Unsupervised dynamic textures segmentation. In Computer Analysis of Images and Patterns,volume 8047 of LNCS, pages 433-440. Springer Berlin Heidelberg, 2013. 2

[21] X. He and A. Yuille. Occlusion boundary detection using pseudo-depth. In ECCV, pages 539-552. 2010. 8

[22] D. J. Heeger. Model for the extraction of image flow. 1. Opt.Soc. Am. A, 1987. 2

[23] A. Humayun, O. M. Aodha, and G. J. Brostow. Learning to find occlusion regions. In CVPR, pages 2161-2168, june 2011. 2

[24] V. Jain, S. C. Turaga, K. L. Briggman, M. Helmstaedter,W. Denk, and H. S. Seung. Learning to agglomerate superpixel hierarchies. In NIPS, pages 648-656, 2011. 2

[25] H. Jung, J. Ju, and J. Kim. Rigid motion segmentation using randomized voting. In CVPR, June 2014. 2

[26] A. Khoreva, F. Galasso, M. Hein, and B. Schiele. Learning must-link constraints for video segmentation based on spectral clustering. In X. Jiang, J. Hornegger, and R. Koch,editors, Pattern Recognition, Lecture Notes in Computer Science, pages

701-712. Springer International Publishing,2014. 2

[27] F. Li, T. Kim, A. Humayun, D. Tsai, and J. M. Rehg. Video segmentation by tracking many figure-ground segments. In ICCV, 2013. 1

[28] C. Liu, w. T. Freeman, E. H. Adelson, and Y. Weiss. Humanassisted motion annotation. In CVPR, 2008. 2, 7, 8

[29] C. D. Manning, P. Raghavan, and H. Schtze. Hierarchical clustering. In Introduction to Information Retrieval, chapter 17. Cambridge University Press, Cambridge, UK, 2008.3

[30] A. Mumtaz, E. Coviello, G. R. G. Lanckriet, and A. B. Chan.Clustering dynamic textures with the hierarchical em algorithm for modeling video. IEEE Trans. Pattern Anal. Mach.Intell., 35(7):1606-1621, 2013. 6

[31] R. Peteri, S. Fazekas, and M. J. Huiskes. DynTex : a Comprehensive Database of Dynamic Textures. Pattern Recognition Letters. http://projects.cwi.nl/dyntex/. 5, 6

[32] M. Sargin, L. Bertelli, B. Manjunath, and K. Rose. Probabilistic occlusion boundary detection on spatio-temporal lattices. In ICCV, pages 560-567, 2009. 8

[33] J. Shi and J. Malik. Motion segmentation and tracking using normalized cuts. In ICCV, pages 1154-1160, 1998. 7

[34] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman.Fisher Vector Faces in the Wild. In BMVC, 2013. 2, 3, 4

[35] A. N. Stein and M. Hebert. Occlusion boundaries from motion: Low-level detection and mid-level reasoning. IlCV,pages 325-357, 2009. 2, 6, 8

[36] D. Sun, E. Sudderth, and M. J. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. Advances in Neural Information Processing Systems,23, 2010. 7, 8

[37] D. Sun, E. B. Sudderth, and M. J. Black. Layered segmentation and optical flow estimation over time. In CVPR, pages 1768-1775, 2012. 2, 7, 8

[38] p. Sundberg, T. Brox, M. Maire, P. Arbelaez, and J. Malik. 2092 Occlusion boundary detection and figure/ground assignment from optical flow. In CVPR, pages 2233-2240, 2011. 8

[39] D. Teney and M. Brown. Segmentation of Dynamic Scenes with Distributions of Spatiotemporally Oriented Energies. In BMVC, 9 2014. 2, 4, 6, 7, 8

[40] D. Tsai, M. Flagg, and J. M.Rehg. Motion coherent tracking with multi-label mrf optimization. BMVC, 2010. 1

[41] F. Wang and J. Sun. Survey on distance metric learning and dimensionality reduction in data mining. Data Mining and Knowledge Discovery, pages 1-31, 2014. 2

[42] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. The Journal of Machine Learning Research, 10:207-244, 2009. 3, 4

[43] C. Xu and J. 1. Corso. Evaluation of super-voxel methods for early video processing. In CVPR, 2012. 1, 2

[44] C. Xu, S. Whitt, and J. J. Corso. Flattening supervoxel hierarchies by the uniform entropy slice. In ICCV, 2013. 2

[45] C. Xu, C. Xiong, and 1. 1. Corso. Streaming hierarchical video segmentation. In ECCV, 2012. 2