指导教师：杨涛　　提交时间：2016/ 3/13

CVPR2015 Paper Translation

No: 01

姓名：郭晓阳

学号：2013302521

班号：10011303

超越高斯锥：多功能堆叠的动作识别

摘要

最先进的动作特征提取器包括微分算子，作为高通滤波器和倾向衰减低频动作信息。这评价了产生的特点及产生—该病态特征矩阵。高斯金字塔—中已被用作一种功能增强技术，恩—码尺度不变特征的特征空间试图处理这种衰减。然而，在高斯锥的核是一个卷积平滑—操作，这使得它不能产生新的粗尺度特征。 为了解决这一问题—的问题，我们提出了一种新的特征增强技术多跳功能叠加（MIFS），这组特征用微分滤波器参数提取—化多时间跳跃和编码的平移不变性进入频率空间。补偿信息的方法—利用微分算子的夺回失去的过程粗尺度信息。这回信息让我们能够在不同的速度和范围匹配行动运动。 我们证明方法提高学习能力基于微分的特征指数。由此产生的特征矩阵方法具有更小的条件数和方差比传统的方法—ODS。实验结果表明，每—表现在具有挑战性的行为识别和事件—检测任务。具体而言，我们的方法超过了状态—在艺术上 hollywood2，ucf101 和 ucf5 数据集与对 hmdb51 艺术状态奥运体育数据集。方法也可以被用来作为一个最小或无特征提取的加速策略精度成本。

1. 简介

我们考虑的问题，提高视频表示对于动作识别，它变得越来越重要的是分析人类活动本身和作为更复杂事件分析的组件。

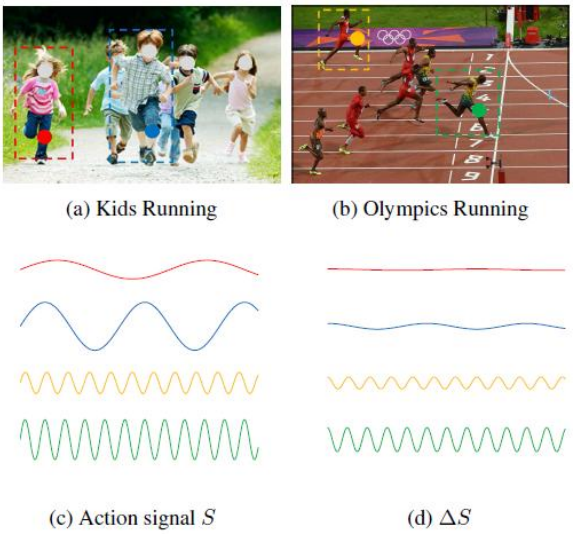


图 1：简化的动作信号（1c）从“跑”的行为（1A、1B）显示显著差异科目和场景。在行动中有着如此巨大的差异信号，一个差分算子与单尺度是不能覆盖全方位的动作频率和趋于失去低频信息（红色和青色的信号）。

事实上，大部分的定性对视觉分析的改进可以归因于介绍改进的陈述，从筛选[15]深度卷积神经网络[29]，规定[12]密弹道[38]。共同特征这几代人的视觉特征是他们所有的，在某种程度上，受益于多尺度表示的想法，一般认为，这是一个不分适用的工具，

可靠地产生性能的改善当应用到几乎所有的特征提取器。

在图像多尺度表示的核心是 要求没有新的细节信息应该被人为地发现在粗尺度的分辨率[10]。高斯基于此约束的塔式，一个独特的解决方案，生成一个家庭的图像精细尺度信息 高斯平滑抑制。 不过在动作识别中，我们往往渴望对方的要求。例如，在生成动作特性时使用微分过滤器，我们需要粗规模的特点：1）恢复已通过高通滤波器滤出的信息在细密的鳞片，如图 1c 的红色和青色的信号可能会过滤掉；2）产生功能在更高的频率在不同的速度和类似的动作匹配图中的橙色和绿色信号的运动范围 1C。这些要求都不能满足高斯锥表示。

在这项工作中，我们介绍了一个多跳功能堆叠（MIFS）表示，通过叠加特征一个家庭提取的微分滤波器参数化多时间跳过（尺度）。我们的算法依赖于这个想法，通过逐步降低的帧速率，功能差分滤波器提取信息提取关于行动的更微妙的动作。方法有几种有吸引力的特性：

- 这是一个不可分的适用的工具，可以可靠地容易通过的任何特征提取与微分滤波器，如高斯锥，
- 它生成的功能，在频率偏移不变性空间，因此更容易匹配类似的行动运动的不同速度和范围。
- 它栈功能在多个频率和趋于覆盖较常规的动作信号的更长的范围动作表示，
- 它生成的特征矩阵有较小的条件数和方差因此更强的可学习性相比传统的原始规模表示基于我们的理论分析。
- 它显着提高了性能的状态—基于实验结果的几种艺术方法现实世界的基准数据集。
- 它的易学性指数提高特征矩阵。因此所需的附加数量的尺度是对数的带宽的动作信号。实证研究表明，一个或 2 个额外的尺度是足够的恢复信息差分算子。因此，额外的的方法计算成本小。
- 它可以用来作为特征提取的加速策略以最小的或不准确的成本。如图所示我们的实验，结合特征提取在较低的帧速率视频（不同时间跳过）在原始的视频中进行比视频功能更好的性能帧速率在同一时间需要更少的时间过程。

在本文的其余部分，我们开始提供有关行动识别和多尺度演示。然后我们详细描述方法，其次从理论上证明了方法视频表示的可学习性指数。后这，我们的方法进行评估。进一步讨论包括潜在的改进结束。

2. 相关工作

有广泛的文学作品的行动这里我们只提到了一些相关的问题与国家的最先进的特征提取和特征编码方法。见[2]作深入调查。在传统的视频表示，特征和编码方法是相当大的进展

的主要原因在田野里。其中，轨迹为基础的方法[18，35，38，8，40]，特别是密集轨迹法王等人提出。[38，40]，连同费舍尔向量编码[26]产生艺术的当前状态几种基准动作识别的性能数据集。 彭等。[24，25]进一步提高性能通过增加码本密集轨迹大小，融合多个编码方法，并添加一个堆叠小鱼矢量。最近报道了一些成功的应用深卷积神经网络用于动作识别在视频。karpathy 等人。[9]培养了一种深层次的卷积使用 100 万个弱标记的神经网络 YouTube 视频和报道，用一种温和的成功它作为一个特征提取。Simonyan 和 Zisserman [32]报道一个有竞争力的结果，以提高密集的轨迹[40]通过训练卷积神经网络使用采样帧和光流。一种方法是所有可采用的工具上述特征提取器。

多尺度表示[1，14]已经非常流行对于大多数图像处理任务，如图像压缩，图像增强和物体识别。多尺度关键点检测 Lindeberg [13]了用洛[15]检测尺度不变的关键点使用拉普拉斯金字塔方法，其中的高斯平滑为每一个层次反复使用。Simonyan 和 Zisserman [33]报告的一个显着的性能改善利用多尺度的 ImageNet 挑战 2014 深卷积神经网络。在视频处理，时空兴趣点（STIP）[12]扩展 SIFT 通过寻找尺度不变特征的时域三维空间点。 邵等人。[30]也试图实现识别三维拉普拉斯金字塔行动尺度不变性三维 Gabor 滤波器。然而，没有意识图像与视频处理的根本区别，[30]与之相比并不是很成功的国家的最先进的方法。

在实验室 datasets 谨慎行动的人或模板可以 reliably 估计，动态时间翘曲（DTP）[5]，隐马尔可夫模型（HMM）与动态[41]贝叶斯网络（DBNs）[23]是好的 studied 方法为调整行为，有速度的变化。然而，在嘈杂的现实世界的行为，这些方法没有 shown 企业是非常强大的。

3。多功能叠加（MIFS）

我们现在正式我们的符号。目前讨论视频只是一个真实的三个变量的函数：

X = X(x, y, t). (1)

归一化坐标（x，y，t）∈R3 笛卡尔视频空间坐标。因为我们专注于时间域，我们省略（在）进一步讨论和表示一个视频（吨）。假定视频的长度将常态化，这是不∈[0，1]。在我们的模型中，内容视频中的一个线性混合的潜在的产生信号：

X= [¯x1, ¯x2, · · · , ¯xk]. (2)

每个潜在的行动信号¯西在时间的混合重量吨被记为我（吨）。因此，一个给定的视频生成作为

X(t) = ¯X α (t) + ϱ(t) (3)

α (t) =[_1(t), _2(t), · · · , _k(t)]T . (4)

在那ϱ(t)是次高斯噪声的噪声级

别_.

|_i(t)| ≤ 1 Et{_i(t)} = 0 (5)

Et{_i(t)2} ≤ 1 Et{_i(t) × _j(t)|i6=j} = 0 . (6)

假设的特征提取被建模为一个微分算子F[·，_]，时间跳跃参数化，给定一个固定的_，

F[X(t)，_] = [f (t1，_)，f (t2，_)，· · · ，f (tT ，_)] .

大多数行动的特征提取是不同版本的比如，规定[12]和[38]可以密集的轨迹源自f，1克：在那里，钾是一个帧的数目在视频。

MIFS堆多F[X(t)，_]和不同的_，通过叠加不同频率的多个特征，方法通过重采样在频域中寻找不变性时间域。图2显示了高斯差一个无约束的真实信号的金字塔和方法视频。很显然，由于平滑，高斯锥一旦被过滤掉，就无法恢复信号。随着水平越高，特征所产生的高斯金字塔只能变弱。而在方法，该生成的功能变得更加突出，可以恢复随着水平走高。

4。可学习的方法

在这一部分中，我们首先表明，在模型方程（2），标准特征提取方法不能产生特征矩阵条件好。然后，我们表明方法提高了特征提取的条件数矩阵指数。一个新奇的关键方法是，它也降低了特征的不确定性矩阵同时。这种减少是不可能的天真的方法。

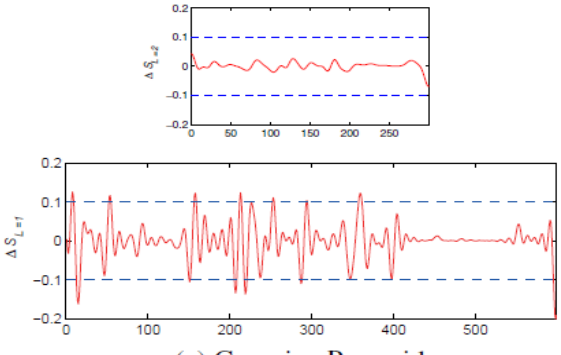
4.1。固定下的磷的状态数

在这一小节中，我们将证明，基于矩阵伯恩斯坦不等式[37]，该条件数不一定是一个小数目。在静态特征提取器如SIFT，权重系数在视频中，矩阵是独立的。流，动作信号是动态的，以衡量一个动作信号的动态，我们介绍我作为一个指数。

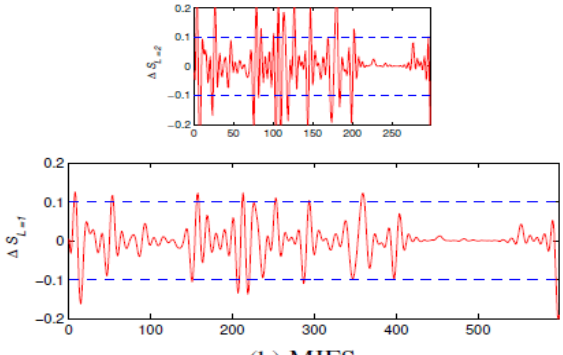
定义1。一个潜在的行动信号是如果给定一个动态非负常数 c ∈ [0, 1]，∀_ ∈ (0, 1]，。

1−(1+c) exp(−/_) ≤ Et|_(t)_(t+_)| ≤ 1−exp(−/_) ，

provided 1 − (1 + c) exp(−/_) ≥ 0.



(a) Gaussian Pyramid



(b) MIFS

图 2：一个真实的行动信号高斯金字塔和方法的比较。左图（一）显示为水平（升）上升（从 1 到 2），由此产生的特征（从微分算子）变得不那么显眼。因此，一旦特征已被过滤掉（假定为一个特征的阈值是 0.1），它不能被更高的回收高斯锥框架下的水平特征。右边的图（b）表明在方法，特征（S）成为更为突出的是，随着水平越来越高，可以表示在低水平已过滤的信号

因此，我们期待的相关性 $\alpha_i(t)$，$\alpha_i(t + \_)$，接近 0。或如果行动信是一个低频分量，相关指标应接近 1。为了简单起见，我们重新安排潜在的动作信号的频率有¯X。$_1 \leq _2 \leq \cdots \leq _k$.

在学习的问题，我们希望的特征矩阵良好的条件。给定特征矩阵我们可以恢复¯X 通过各种方法，如子空间聚类。任何恢复的采样复杂度算法依赖于体育的条件数。显然，当磷是病态的，我们需要大量的训练样本来估计¯X。的可学习性取决于它的条件数[3]，在返回取决于再次。在下面，我们将证明，对于一个固定的时间跳跃，不一定是很好的条件。因此，可学习性欠佳。F[X(t), _我们证明背后的直觉是当一个行动，信号有一个大, 然后一个小的时间跳过，将使该信号接近于零的系数。因此，磷是病态。形式上，我们有以下定理约束条件数，

$$\_(PPT) \leqslant (1 + c)\exp(-1/\_) + \_\_\exp(-k/\_) - \_\_ \quad (8)$$

$$\_(PPT) \geqslant (1 + c)\exp(-1/\_) - \_\_\exp(-k/\_) + \_\_ \quad . \quad (9)$$

那

$$\_\_ = 2rk1T(1 + c)\log(2k/\_) \quad (10)$$

$$T \geq 19(1 + c)k\log(2k/\_) . \quad (11)$$

定理1表明当特征数数量够大，条件数 _(PPT) 是一个随机数集中在它的期望 $(1 + c)\exp(-1/\_)\exp(-k/\_)$. Since $1 \ll k$,

分子是多大于的时候，是固定的。自从我们证明是基于伯恩斯坦的不平等，上限紧。这种力量 $\_(PP_T)$ 是一个比较大的价值。更具体地说，下面的推论表明，当，是指数大的期望。

推论 1。

When $k \geq (M + 1)1, E\{\_(PPT)\} \geq (1 + c)[\exp(1\_)]M \geq (1 + c)(1 + 1\_)M$.(12)

推论1显示视频中的动作在一个广阔的动态范围，特征提取单倾向于有病态的特征矩阵。一个幼稚的解决方案，这个问题是增加减少在预期的条件数。然而，这将增加方差__ of _(PPT) 因为一个较小的特征数。在实践中，一个大也增加了光流计算和跟踪困难。因此，作为也将在我们的实验中观察到，选择一个好的可以相当困难。直观地说，选择是特征偏差和方差之间的权衡。特征提取与一个大涵盖了一系列的行动信号但具有较少的特征点，因此产生的特点小偏差大变异。相似

的特征提取用一个小将产生具有大偏差的功能，但小变异

4.2。在多个条件下的多 t_

从结论一中得出，为了让 ppt 出现更好的状态,我们需要 t 越大越好，然而，当 t 特别大的时候。我们不能抽样足够高质量的特征点，为了解决这个问题,我们建议使用的方法 mips，逐步扩大 t，然后栈的所有功能在不同 t，形成特征矩阵。希望通过增加 t，我们提高条件数，_(PPT)，通过叠加，我们样本足够的功能，以减少方差。

假设我们有提取的功能，{_,2_，· · · m_}.提取次数特点是 Ti = [1/(i_)]. 下面定理边界条件数方法（见证据补充材料）。

推论 2

概率至少 1 − _,*ppt* 的条件数，该方法是有界的

_(PPT) ≤PiTiT 2(1 + c) exp(−1/_i) + __ P iTiT 2 exp(−k/_i) − __. (13)

*where*__ ≤ 2skP1i Ti(1 + c) log(2k/_) . (14)

5 实验

我们检验我们的假设和提出的方法表示任务：动作识别和事件检测。实验结果表明，方法表示优于传统的原始规模表示七个真实世界的具有挑战性的数据集。

改进稠密轨迹与矢量编码[40]是最真实的艺术的当前状态动作识别数据集。因此，我们用它来评估我们的方法。请注意，虽然我们使用改进密集的轨迹，我们的方法可以应用到任何地方功能，使用差分滤波器，例如，规定[12]。

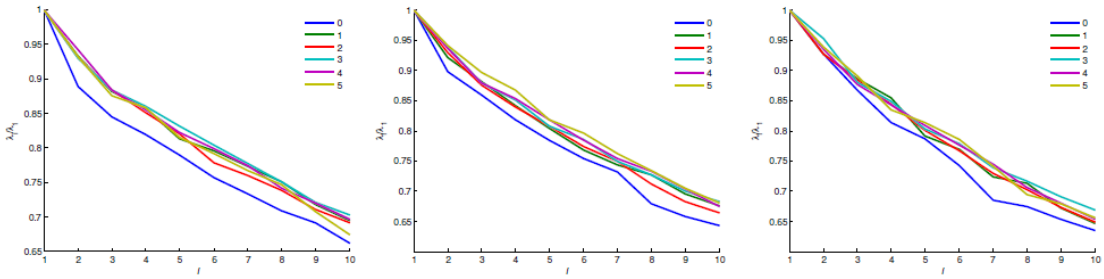
5.1 动作识别

问题制定本任务的目标是要认识视频短片中的人类行为。

数据集五个有代表性的数据集使用：hmdb51 集[11] 51 动作类和 6766 视频从数字化电影和 YouTube 提取剪辑。[11]提供原始视频和稳定的。我们只在本文中使用的原始视频和标准分割 MACC（平均精度）的性能进行评估。该数据集包含 12 hollywood2 [17] 行动类和 1707 个视频剪辑，收集从 69 个不同的好莱坞电影。我们使用标准分割[17]提供的训练和测试视频。平均平均精度（图）是用来评估这个数据集，因为多个标签可以被分配给一个视频剪辑。这个 ucf101 集[34] 101 动作类跨越 13320 YouTube 视频剪辑。我们使用标准分割由[34]和 MACC 提供培训和测试视频报道。的 ucf50 集[27] 50 动作类超过 6618 的 YouTube 视频剪辑，可以分分成 25 组。在同一组的视频剪辑一般很相似的背景。离开一个 groupout 使用和推荐的交叉验证平均精度（MACC）在所有的类和所有组的报道。奥运会体育数据集[20] 由 16 名运动员组成以 783 个视频为代表的练习运动剪辑。我们使用标准的分割与 649 个训练剪辑和 134 测试剪辑和报告图为在[20]中进行比较用途。

实验设置改进密集轨迹特征使用 15 帧跟踪，摄像机运动稳定、ROOTSIFT 归一化和描述轨迹， HOG, HOF, MBHx 和描述符。我们使用主成分分析，以减少这些描述符的维数由一个因素。原后，我们增强了描述符用三维归一化位置信息。方法和其他传统的唯一区别方法是利用特征点提取从一个时间尺度，我们提取和堆栈的所有原始不同尺度下的特征点编码。对于渔夫的矢量编码，我们映射的原始描述符 256 高斯为混合高斯模型从一组随机

抽取的 256000 个数据点的训练。电力和 L2 规范也用在连接不同类型的描述符的视频为基础的表示。另一个 L2 归一化后使用级联。这种重整化带来了我们在大多数的数据集的基础上的基线方法的改进除了奥林匹克运动。对于分类，我们使用一个线性有固定的 100 的支持向量机分类[40]和一对所有的方法是用于多类分类场景。



(a) HMDB51　　(b) Hollywood2　　(c) UCF101

图 3：衰减趋势的 hmdb51 特征矩阵的奇异值，好莱坞和 ucf101 数据集。0 至 5 表明方法水平和我指示与奇异值。从所有三个数据集，我们可以看到，MIFS 表示有一个较慢的奇异值衰减趋势相比，传统的陈述（蓝线）。

结论

我们首先检查是否正确的条件数 _(PPT) 改进的方法。然而，它不会有意义的计算，一种解决方法是检查腐烂特征矩阵奇异值的速度。奇异值除以最大奇异归一化价值_max，我们只绘制了前 10 个奇异值，因为由小奇异值所覆盖的子空间是噪声空间。显然，当方法提高了学习能力，我们应得到一个缓慢衰减的曲线的顶部价值观。图 3 所示为趋势_i_max，在三集：hmdb51，hollywood2 和 ucf101。在所有的三个数据集的方法，奇异值降低比传统的慢（0）。它也很有趣看有一个或两个额外的水平，我们有已经开发了大部分潜在的改进。

我们进一步研究如何性能变化与尊重在方法层面，如表 1 所示。首先，让我们比较 L = 0 的性能标准 locationinsentative 特征表示。我们的表现 hmdb51，hollywood2，ucf101 和 ucf50 数据集 62.1%，67%，87.3%和 93%分别与。这些数字高于王 & Schmid57.2%、64.3%、85.9%、91.2%、40 的结果，分别。这种改进很大程度上是因为我们的位置敏感特征表示与重整化。接下来，让我们检查的行为方法。对于完整性，我们列出单规模和堆叠性能。对于单一规模的性能，我们观察到 hmdb51，其性能从 62.1%增加到 63.1%然后迅速下降，类似的模式可以看出其他数据集，除了一些不增加在升= 1。这些结果与我们的观察，不同的行动需要不同规模的范围。他们也证实了我们的选择时间间隔的证明 t 是一个权衡之间特征偏差及其方差。如果计算成本是至关重要的，然后，我们可以选择只提取更高的单尺度特点，但痛苦最小或没有准确的损失和享受大的计算减少。现在让我们比较一下单尺度的表示方法。我们观察到方法表示，虽然仍有偏差和方差在不同的单尺度表示的权衡水平，他们都表现得比单一的表示和性能下降点比单尺度表示。我们也观察这个方法表示，大部分的性能改进来自 L= 1 和 L= 2，它支持我们在图 3 中所观察到的，在实践中，有一个或 2 个以上的规模足以恢复丢失的信息由于微分运算。更高尺度特征由于越来越困难，

变得不那么可靠了光流估计与跟踪。它也很有趣观察 hmdb51 享有更高的性能改进的方法比其他四个数据集。我们认为，主要的原因是，HMDB 数据集从两源混合视频：YouTube 和电影，在较大的动作速度范围内的结果比纯电影在其他数据集或 YouTube 视频。

与表 2 中的艺术状态相比，我们比较 L = 3 方法，执行所有的动作数据集，与国家的最先进的方法。从桌子 2，在大多数的数据集，我们观察到改善国家的艺术，除了 hmdb51 与奥林匹克运动，我们的 L = 3 方法给性能较差。请注意，尽管我们列出了一些最新的方法在这里进行比较的目的，其中大多数是不迪一直接与我们的结果由于使用不同一耳鼻耳的特点及表现。最具可比性王和 Schmid。[40]，我们建立我们的方法在。sapienz 等人。[28] 探讨的方式来进行并生成密集轨迹特征的词汇。Jain 等人。[7]的方法合并了一个新的动议描述符。oneata 等人。[21]专注于测试空间多个动作和事件任务的多个向量的向量。彭等。[24] 改善密集的性能通过增加码本大小和融合多个编码方法。karpathy 等人。[9] 培养了一个很深的卷积神经网络用 100 万个弱标记 YouTube 视频和报道 65.4%的平均精度在 ucf101 数据集。Simonyan 和 Zisserman [32]报道结果有竞争力的改进密集的轨迹用两种训练深度卷积神经网络采样帧和光学流动和获得 57.9%反贪委在 hmdb51 87.6% MACC 在 ucf101，哪比得上对王& Schmid 结果。彭等。[25]取得更好的结果比我们 hmdb51 和奥运会体育数据集相结合的层次的的的的的的小鱼矢量与原来的。

	HMDB51 (MAcc%)		Hollywood2 (MAP%)		UCF101 (MAcc%)		UCF50 (MAcc%)		Olympics Sports (MAP%)	
L	single-scale	MIFS	single-scale	MIFS	single-scale	MIFS	single-scale	MIFS	single-scale	MIFS
0	62.1		67.0		87.3		93.0		89.8	
1	63.1	63.8	66.4	67.5	87.3	88.1	93.3	94.0	89.4	**92.9**
2	54.3	64.4	62.5	67.9	85.5	88.8	92.2	94.1	88.1	91.7
3	43.8	65.1	60.5	**68.0**	81.3	**89.1**	89.7	**94.4**	85.3	91.4
4	24.1	**65.4**	58.1	67.4	74.6	**89.1**	84.3	**94.4**	85.0	90.3
5	15.9	**65.4**	54.4	67.1	66.7	89.0	76.7	94.3	82.3	91.3

Table 1: Comparison of different scale levels for MIFS.

HMDB51 (MAcc. %)		Hollywood2 (MAP %)		UCF101(MAcc. %)		UCF50 (MAcc. %)		Olympics Sports (MAP %)	
Oneata *et al.* [21]	54.8	Lin *et al.* [36]	48.1	Karpathy *et al.* [9]	65.4	Shi *et al.* [31]	83.3	Jain *et al.* [7]	83.2
Wang *et al.* [40]	57.2	Sapienz *et al.* [28]	59.6	Sapienz *et al.* [28]	82.3	Sanath *et al.* [19]	89.4	Adrien *et al.* [6]	85.5
Simonyan *et al.* [32]	57.9	Jain *et al.* [7]	62.5	Wang *et al.* [39]	85.9	Arridhana *et al.* [4]	90.0	Oneata *et al.* [21]	89.0
Peng *et al.* [24]	61.1	Oneata *et al.* [21]	63.3	Simonyan *et al.* [32]	87.6	Oneata *et al.* [21]	90.0	Wang & Schmid [40]	91.1
Peng *et al.* [25]	**66.8**	Wang *et al.* [40]	64.3	Peng *et al.* [24]	87.9	Wang & Schmid [40]	91.2	Peng *et al.* [25]	**93.8**
MIFS (L=3)	65.1	MIFS (L = 3)	**68.0**	MIFS (L = 3)	**89.1**	MIFS (L=3)	**94.4**	MIFS (L = 3)	91.4

Table 2: Comparison of our results to the state-of-the-arts.

5.2。事件检测

给定的问题制剂 A collection of 的视频，个事件的目标检测的任务是检测活动的兴趣如生日聚会和游行，只在基视频内容。冰的任务非常复杂 challenging 鸽行为和舞台。 通过对这项任务的评估，我们研究是否可以提高识别性能的方法非常复杂的行为。

TREC 视频检索评价（TRECVID 数据集）多媒体事件检测（MED）[22]是有组织的任务由 NIST （National Institute of Standards and Technology ） aimed at New Technologies for encouraging 检测具有复杂的事件，如生日派对。开始在 2010 年，NIST 已经逐渐建立数据库，包含了 8000 小时 40 of videos and events，which is by farthe largest collection 事件检测。 medtest13，14 数据集是标准的系统评估数据集发

布在 2013 和 2014 标准，分别。每一个他们包含约百分之 10 的整个地中海收藏并有 20 个事件。他们包括 2 个任务，即 ek100 和 ek10。ek100 任务有 100 个积极的训练而 ek10 样品 10。对于这两个任务，他们有约 5000 背景样本。一起，每一个数据集有 8000 个训练样本和 24000 个测试样本。

实验设置一个类似的设置在节讨论 5.1 应用除了我们使用五个文件夹交叉验证选择线性的罚参数支持向量机。对于每一个分类，选择中间 10−3, 10−2, 10−1, 1, 101, 102, 103. 我们只与测试方法在第 5.1 节中推荐的 L=3，因为提取密集轨迹特征，从这样的大数据集本身是非常耗时。我们花了 4 天时间来生成交涉为 medtest13，14 使用集群超过 500 的英特尔 e565 + 系列处理器。我们使用地图作为评价标准。

	HMDB51 (MAcc%)	Hollywood2 (MAP%)	UCF101 (MAcc%)	UCF50 (MAcc%)	Olympics Sports (MAP%)	Computational Cost (Relative)
L=0	62.1	67.0	87.3	93.0	89.8	1.0
L=1-0	63.1	66.4	87.3	93.3	89.4	0.5
L=2-0	63.9	67.6	88.5	93.8	91.9	0.75

Table 3: Performance versus relative computational cost for feature extraction

	MEDTEST13		MEDTEST14	
	EK100	EK10	EK100	EK10
Baseline	34.2	17.7	27.3	12.7
MIFS (L=3)	**36.3**	**19.3**	**29.0**	**14.9**

Table 4: Performance Comparison on the MED task.

结果表 4 列出了整体图（详细结果可以被发现在补充材料中。基线法是一个传统的单尺度表示与升= 0。从表 4 中，我们可以看到，无论 medtest13 和 medtest14 MIFS 表示，不断提高在原来的规模约 2%，在这两个 ek100 和 ek10。值得强调的是，医学是这样一个具有挑战性的任务，2%的绝对性能改善是相当重要的。

5.3。计算复杂度

0 级的方法表示具有相同的成本
其他单通的方法，例如，王和 Schmid。[40]。对于水平升，成本成为 1 /升的水平 0。所以一个方法高达 2 级，计算成本将不到两倍一个单一的通过视频的成本，但它可以显着提高单通方法。如果计算效率是至关重要的，该方法可以加快去除低尺度特征。例如，除去 L= 0（原创视频）将大大降低成本，但仍给予有用的改进如表 3 所示。我 -1 显示只使用功能从每第二帧的结果 L = 2-0 显示的特点相结合，从结果等级 1（每第二帧）和 2 级（每第三帧），但不升= 0。在大多数情况下，我们仍然可以得到更好的成本较低的结果。

6。结论

我们开发的多跳特性叠加（MIFS）为提高行动表示学习方法。方法采用

家庭差叠特征提取过滤器的参数化多时间跳跃和在频率空间中实现平移不变性。相比之下高斯金字塔，MIFS 生成特征在所有尺度并趋向于覆盖更长的动作信号范围。理论结果表明，方法提高动作表示的可学习性指数。大量实验七个现实世界的数据集显示，超过状态方法—艺术方法。未来的工作将决定不同的动作类型的适当水平。此外，我们想提高光学流量计算的质量粗鳞的跟踪。

7。确认

这项工作是由智力先进的部分支持的研究项目活动（IARPA）通过部门国家商务中心合同编号 d11pc20068。美国政府授权复制和分发政府目的转载尽管有任何版权注释。免责声明：此处包含的观点和结论是作者的，不应该被解释为一定代表官方政策或代言，任何明示或暗示的 IARPA，DOI / NBC，或美国政府。这工作也得到了支持部分由美国陆军研究办公室（w911nf—13-1-0277）。任何意见、发现、结论或在这材料中表达的建议是作者和不一定反映的意见 Aro。

工具书类

[1] E • H • 安德森，安德森 C H，J • R • 伯根，P。J。伯特，and
作者 J。M。金字塔方法在图像处理中的应用。RCA 工程师，29（6）：33 – 41，1984 年。2
[2] . Aggarwal 和 S 先生赖乌。分析了人类活动：
评论。ACM 计算调查（的心），43（3）：16，2011 年。2
[3] .楚和 T。林 Foundations and Advances in data。
矿业，体积 180。施普林格出版社，2005 年。4
[4]。ciptadi 古德温先生，美国，和 rehg J。M。运动模式
直方图行动识别和检索。在 eccv。2014 年。7
[5] .达雷尔 and A Pentland。空间-时间的手势。在 cvpr，1993 年。3
[6]。gaidon harchaoui，Z，和 C Schmid。活动描述 with Hierarchies 动议。国际杂志计算机视觉，107（3）：219 – 238，2014 年。
[7] M.杰恩，H J´ 恶狗，bouthemy 和 P.。更好地利用运动
为了更好的行动识别。在 CVPR，2013。七
[8]。江，问：，，，。基于运动轨迹的人体动作建模参考点。在 ECCV。2012。2
【9】A. karpathy，G. toderici，S. Shetty，T 梁，sukthankar R.，飞飞。大规模视频分类与卷积神经网络。在 CVPR，2014。2，7
[10] J. J. Koenderink。图像结构。生物控制论，50（5）：370——1984，363。二
[11] H 汉克，H.壮，如绞，T. Poggio，和 T. 塞尔。
HMDB：为人体运动识别大型视频数据库。在计算机视觉国际会议，2011。五
[12]一、拉普捷夫。关于时空兴趣点。ijcv，64（2-3）：107 – 123，2005。1，2，5，3
[13] T • 林德伯格。检测突出的斑点状的图像结构和他们的尺度空间原始草图：一个方法关注焦点。ijcv，11（3）：283 – 318，1993。二
[14] T. Lindeberg B.M. ter Haar Romeny。线性尺度空间
基本理论。施普林格出版社，1994。1，2
[15]D. G. Lowe。从尺度不变的图像特征关键点。ijcv，60（2）：91 – 110，2004。1，2
调查研究
人类视觉信息的表示与处理，
亨利霍尔特有限公司，纽约，NY，2 页 – 46，1982。一
[17] M.马尔沙莱克，I.拉普捷夫海，和 C. Schmid。语境中的行动。在 CVPR，2009。五
[18] P. matikainen，M. Hebert 和 R. sukthankar。trajectons：运动分析中的动作

识别特点。在计算机视觉国际会议研讨会，2009。二
[19] S. Narayan 和 K.R. Ramakrishnan。原因及影响运动轨迹的建模分析。在 CVPR，2014。七
[20] J. C. niebles，C. W. Chen，和我飞。建模时间活动分解运动节段的结构分类。在 ECCV。2010。五
【21】D. oneata，J.维贝克，C. Schmid 等人。行动和事件一个紧凑的特征集上的渔夫向量识别。在 ICCV，2013。七
[22]页，G. Awad J.国库，G.桑德斯。2013 – TRECVID 对目标、任务、数据、评估机制的介绍，和度量。2013。七
[23]美国公园和 J. K. Aggarwal。分层贝叶斯网络对于人类行为和相互作用的事件识别。多媒体系统，10（2）：164 - 2004，179。三
[24]。彭，属。王，X，和。袋视觉词汇渔船矢量。计算机视觉中的 – ECCV 2014，581 页 595 页。施普林格出版社，2014。2，7
【26】F perronnin，J 的´anchez Mensink，和 T.。改进大型图像分类的核函数。在 ECCV。2010。二
【27】K. K. Reddy 和 M. Shah。认识到 50 个人的行动网络视频类。机器视觉及应用，24（5）：981——2013，971。五
[28] M. F. cuzzolin，智慧，和 P. H.托。特征采样
大动作视觉词汇的生成与划分分类数据集。arXiv 预印本：1405.7545，2014。七
[29] P. sermanet，D.特征，张 X，M.马蒂厄，R.弗格斯，和 Y LeCun。overfeat：综合识别、定位
利用卷积网络检测。预印本论文：1312.6229、2013。一
[30] L，x，D，D，和 X。时空的拉普拉斯用于动作识别的锥编码。IEEE 交易论控制论，2168 页，2267 页，2013 页。2，3
【31】F 石，如北柳，laganiere 和 R.。抽样策略实时动作识别。在 CVPR，2013。七
【32】K. Simonyan A. Zisserman。双流卷积视频中的动作识别网络。预印本论文：1406.2199、2014。2，7
【33】K. Simonyan A. Zisserman。非常深的卷积大规模图像识别网络。预印本论文：1409.1556、2014。二
【34】K.姆罗，A. Zamir，M. Shah。ucf101：数据集在野生动物的 101 个人动作类。arXiv 预印本 arXiv：1212.0402，2012。五
[35] J.太阳，X. Wu，美燕，L • 畅，T. S. Chua 和李。行动的层次时空背景建模识别。在 CVPR，2009。二
[36]属太阳，K，T。陈，方，王，和。dlsfa：深入学习慢特征分析
[38]。klaser C. H. Wang，and C. L.施密德，刘。动作识别
通过密集的运动轨迹。在结果，2011 年。1，2，3
[39]王和 H. C.施密德。李尔公司在 INRIA thumos 工作坊。在 iccv 厂房，2013 年。7
[40] C. H. Wang，施密德等人。一种改进的动作识别的运动轨迹。在 iccv，2013 年。2、5、6、7、8
[41] J J 和 K 号，ohya，石井。recognizing 人行动在时间序列的图像使用隐马尔可夫模型。在结果，1992 年。3