

指导教师：杨涛

提交时间：2015/3/29

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名：马力

学号：2012302428

班号：10011202

高效图像和流媒体视频分割的光谱图还原

Fabio Galasso¹, Margret Keuper², Thomas Brox², Bernt Schiele¹

¹ Max Planck Institute for Informatics, Germany

² University of Freiburg, Germany

摘要

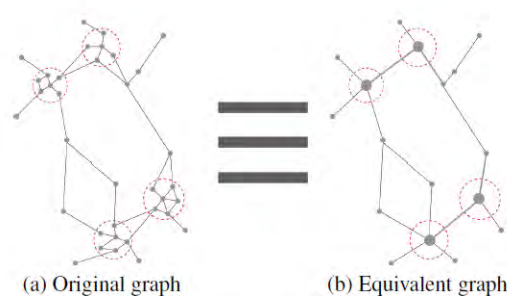
计算和存储成本把光谱技术的应用限制到了相当小的图像,这对视频分割是一个严重的限制。在本文中,我们提出了一个基于超像素的简化图的使用。对比以前的工作,简化图经过加权,因此在一定的假设下,相对于完整图,简化图所得到的分割是等价的。我们就正常化分割和其光谱聚类放宽来考虑等价。该方法减少了运行时间、内存消耗和图像视频分割的最终次数。此外,它为一种体现着最先进性能的新的流媒体视频分割方法提供了一种有效的数据表示和更新。

1.简介

光谱聚类方法为许多最先进的图像[2]和视频[12, 3, 7, 11, 24]分割技术建立了基础。光谱方法因为它们自然地包括远程亲和力的能力而被人们所接受。他们的全球化影响(参见[10]有助于鲁棒性而且是[2]中的一种重要的层次图像分割的一部分。而基于最小分割构想的分割技术适用于具有较强的一元条件的问题,光谱聚类技术的平衡性在焦点是成对的情况下是很有用的。

光谱聚类算法的资源需求对一定图像分辨率的图像分割仍是可达到的,但很容易对视频数据产生限制。[24]给出了对于整个视频[2]在GPU集群上运行的一种严重的像素级图像分割方法的并行扩展。由于这样的计算资源往往是不足的,视频分割通常是以一个两步走的方法实现。图是基于预先计算的超像素/超体素[27, 12, 16, 7],然后聚集,或聚类被应用到稀疏的轨迹,然后结果转换成在后处理步骤的一个密集分割[3, 20]。

在本文中,我们专注于第一组的方法,请注意,在与光谱聚类的结合中,它们在相同的预组合中是有限制的。考虑一个超像素完全覆盖住整个视觉对象的情况。理论上,



图一。在基于图形的分割中,像素间一定的关系可以假定为超像素形式。这些是必须连接限制。在光谱聚类算法中,为利用这些限制平衡起着关键的作用。我们提供两种图还原以保持归一化切割或其光谱聚类松弛。

这样的像素是可取的,由于它在模型误差最小限度增加的情况下提供了最大限度的计算还原。然而,在光谱方法[12, 11, 7]中平衡的解决方案是首选。因此,单节点集群,如单超像素覆盖住整个对象,几乎是不可能的。由[12, 11, 7]的超像素的标准图还原是不与光谱聚类的联系平衡兼容的。超像素的体积必须考虑。

为此,我们研究了能保证与原图在一定的假设下等价的两种图像还原方法(图1)。因此,假定一个超像素中的所有像素都受到必须要连接的限制。第一种方法是参照[21]绘图并减少原始图以便于保存归一化分割(NCut)。然而,发现NCut是NPhard并且所有的光谱技术借助于松弛解决问题。最常见的2类松弛[23, 19]被证明成功用于分割,但它可能会产生偏离NCut的解决方案。因此,(在理论上)等同于NCut实际上仍然会导致不同的解决方案。我们于是介绍一种在较强的假设下也能轻松解决的图还原方法。如果条件不满足,提出的还原会导致优化一种新的密度归一化分割,这在某些应用中可能是有利的。

所提出的还原模型(第3节)可以便利

地应用于图像和视频分割。运行时间和内存消耗在所有情况下都减少了。图像分割（第 4 节）的性能保持着最先进的水平。视频分割（第 5 节）则优于所有方法除过最近的一个方法。此外，还原技术允许了一种新的视频流分割方法（第 6 节），基于以上的图简化，它的性能超越了最先进的性能。

2.相关工作

图还原的两个主要目标是：1. 降低计算复杂度 2. 制约（偏置）解决方案。在一线的工作，这关系到减少图中数据点的个数的技术。这是非常理想的光谱聚类算法[23, 19]。在图像分割中，[9]通过采用 Nystrom 方法，即，随机取样本点和外推近似的对整个系统的样本解决方案而达到了目标。在其他地方，[30, 5]通过有代表性的点接近图（又名地标）。通过对比，我们用超像素近似图。

偏置一种方案也促使了大量的研究。偏置允许附加信息的包含，例如在解决方案中提供一个互动的方式。在图像分割中，[32, 8, 18]为了包括对光谱聚类方法的约束提出若干模型。对比我们的方法，这些方法处理整个图像，即昂贵的特征分解构成光谱聚类的基础是在整个关系矩阵上计算的。

最近，[21]已经解决了计算复杂度和约束，提出了一种保留（归一化）分割的还原方法。我们在图像和视频分割的范围内研究其还原方式。[26]为图像分割促使还原改变在精神上接近[21]。我们用考虑光谱松弛的构想补充这两种方法。

基于多重网格方法的特征值求解暂时减少节点的数量、加快计算并且保持原问题 [15]的解，但粗糙的网格不很好地适合图像结构。在[6]中一个近似的代数多重网格法已被提出，其重点在不同尺度下的不同特征的使用而不是一个单一网格问题的快速精确的解决方案。我们的视频分割部分是在精神上接近这方法。Maire 等人最近的工作。[17]结合传统的多重网格思想与在不同尺度上使用不同的关系的选择。

3.图形还原模型

本节为所提出的简化模型提供了数学的动机。我们简要地回顾归一化分割（NCut）及其与光谱聚类的联系并参照[28]得到更多详情。在根据 NCut 陈述[21]的等价模型后，我们为联系着松弛 NCut 目的的还原图引入一个新的加权方法。

3.1.NCut 与光谱聚类

考虑一个图 $G = (V; E)$ 的顶点集 V 和无向边 E 。我们把分割问题投射为图的划分。图中的顶点表示一个图像或视频片段的像素或超像素。边 e_{ij} 从对应的像素或超像素 i 和 j 之间的相互关系 w_{ij} 中取权重。NCut 提供了一个聚类目标其聚类的规格是按其大小平衡的。对于分割成的不相交的集合 A 和 B ， $A \cup B = V, A \cap B = \emptyset$ ，有

$$NCut(A, B) = \frac{cut(A, B)}{vol(A)} + \frac{cut(A, B)}{vol(B)}, \quad (1)$$

其中 $cut(A; B) = \sum_{i \in A, j \in B} w_{ij}$ ， $vol(A) = \sum_{i \in A, j \in V} w_{ij}$ 。确定最佳集合 A 和 B 需要解决下面的最小化问题：

$$\begin{aligned} \min f^T L f \quad \text{subject to} \quad & Df \perp \bar{1}, \quad f^T Df = vol(V), \\ \text{and } f_i = & \begin{cases} \sqrt{\frac{vol(B)}{vol(A)}} & \text{if } i \in A \\ -\sqrt{\frac{vol(A)}{vol(B)}} & \text{if } i \in B \end{cases}, \quad (2) \end{aligned}$$

其中 $L = D - W$ 是（非规范）图的拉普拉斯算子， W 是包含相互关系 w_{ij} 的矩阵， D 是对角度矩阵，其 $d_{ij} = \sum_{j \in V} w_{ij}$ 。用来求解 NCut 的是 NP-Hard。其松弛（允许 f 取任意实值）会得到光谱聚类：

$$\begin{aligned} \min g^T L_{sym} g = \min \sum_{i,j=1} w_{ij} \left(\frac{g_i}{\sqrt{d_{ii}}} - \frac{g_j}{\sqrt{d_{jj}}} \right)^2 \\ \text{subject to} \quad & g \perp D^{1/2} \bar{1}, \quad \|g\|^2 = vol(V), \quad (3) \end{aligned}$$

其中 g 即以坐标的变化代替 f ， $g := D^{1/2} f, L_{sym} = I - D^{-1/2} W D^{-1/2}$ 是归一化图的拉普拉斯算子。Rayleigh-Ritz 定理提供了（3）的解为对应于 L_{sym} 的次最小特征值 λ_2 的特征向量。接下来，我们考虑分为两个部分的情况，使数学公式更为简单、直观。参见[28]得到基于光谱聚类的多分割的一个归纳。

3.2 图的等价

给定一个把图像分成不相交的超像素并连接图中的像素通过必须连接的限制的点集 $\{I_1, \dots, I_m\}$, 我们要确定一个简化图 $G^Q = (V^Q; E^Q)$, 其中点集 $V^Q = \{I_1, \dots, I_m\}$ 是超像素集的简化, 边 E^Q 描述这些点之间的相互关系 w_{ij}^Q . 新图 G^Q 应该以一种有意义的方式反映原始量, 并允许一种平衡的分割. 这可以通过根据(2)中的 NCut 对原始 G 提供等价来实现. 其次的方法是通过根据 (3)

(光谱聚类的松弛) 来寻找等价来实现, 得到一个得出一个不同平衡条件的图的权的重调整.

归一化分割的等价 (G^Q NCut)

定理 1

决定于图 G^Q 的分割中的 NCut 和原始图 G 上对应的分割的 NCut 是相同的如果关系矩阵 W^Q 定义为

$$w_{IJ}^Q = \sum_{i \in I} \sum_{j \in J} w_{ij} \quad (4)$$

虽然从图形的视角在理论上很有吸引力, 先前的等价集中于切割的目的. 鉴于(2)的最优解通常是找不到的, 我们引入了一个图的权的重调整, 在一定的条件下, 为光谱聚类问题提供等价. 进一步的研究表明, 如果这些条件不满足, 所提出的权的重调整使一个图像的分割目的优化后与 NCut 略有不同. 稍后我们将经验地表明, 优化这一目标往往能带来良好的结果.

光谱聚类的等价 (G^Q SC)

定理 2

图 G^Q 等价于以相应的分割也就是关于光谱聚类的原图 G . G^Q 保存着放宽的光谱聚类优化问题 (3), 若

$$w_{IJ}^Q = \begin{cases} \sum_{i \in I} \sum_{j \in J} w_{ij} & \text{if } I \neq J \\ \frac{1}{|I|} \sum_{i \in I} \sum_{j \in J} w_{ij} - \frac{(|I|-1)}{|I|} \sum_{i \in I} \sum_{j \in V \setminus I} w_{ij} & \text{if } I = J \end{cases} \quad (5)$$

假如相等的 G 的元素的度数都限制在 G^Q 中:

当定理 2 的假设并不适用时, (5) 提

$$\sum_k w_{ik} = \sum_k w_{jk} \quad \forall i, j \in \mathcal{G} : (i, j) \in I, I \in \mathcal{V}^Q \quad (6)$$

出了一个新的平衡分割问题, 就是我们指的密度归一化分割 (DNCut):

定理 3

A^Q 和 B^Q 是图 G^Q 中 (3) 的解, 是根据 (5) 的 G 的还原, 当且仅当 A 和 B 是 G 关于 DNCut 问题的解:

$$DNCut(A, B) = \frac{cut(A, B)}{\sum_{I \in A^Q} \frac{vol(I)}{|I|}} + \frac{cut(A, B)}{\sum_{I \in B^Q} \frac{vol(I)}{|I|}} \quad (7)$$

其中 A 和 B 要满足 G 的必须连接的条件:

$A = \{i \in V \mid i \in I, I \in A^Q\}$, B 类似.

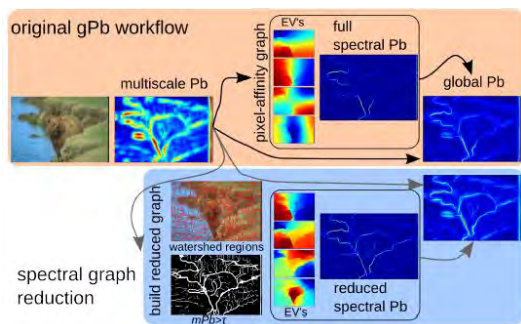
定理 1 和 2 的图还原得到了优于没有使用给出的还原方程的超像素的结果. 应注意的是, 在第 5 节我们在对视频分割的应用中, 定理 2 中的等度条件总是适用.

NCut 与 DNCut 在获得等大的超像素上是一致的, DNCut 提供了一个不同的平衡条件如果这不是问题的关键. 由于一个顶点集中的顶点数会趋向于增加它的基数, NCut 产生某个确定程度的聚类并且由它们的大小得到平衡. 在 DNCut 中, 每一个超像素的量由它们的大小归一化这样平衡就纯粹由点的联系而确定. 因此, 即使大的部分也要由它们的关系来决定. 在下面的章节中会看到, 这会在图像和视频分割中得到良好的结果.

4. 图像分割

(2) 中最先进的分层图像分割算法主要由三个部分组成 (图 2 中说明). 局部边缘信息由 mPb 获取, 这是颜色, 亮度, 和纹理梯度的一个多尺度、多方向的线性组合. 更多的全局信息是由 sPb 提供的, 这是一个介于 mPb 的在像素和特征向量变化率的计算之间的轮廓的相关矩阵的光谱分解. 这个光谱部分主要负责大的内存和计算资源的消耗. 最后一个封闭的层次边界图由 mPb 和 sPb (称为全球 Pb , 或 gPb) 的线性组合求得, 通过一个面向流域变换 (owt) 和超度量等值线图 (ucm).

我们提出的图还原确定昂贵的 sPb 步骤以将最确切的像素关系分组为基于 mPb 的超像素和还原 (2) 的单像素关系矩阵为一



图二。[2]的到 gPb 的工作流。像素图 G 上的 sPb 的计算以还原的超像素图 G^Q 中的计算取代。超像素是由流域提供的；在 $(mPb) > \tau$ 的地方分组被认为是 不确定的并且单像素保存为节点。（ $\tau = 0$ 使用像素，即原来的问题， $\tau = 1$ 意味着只使用超像素。）

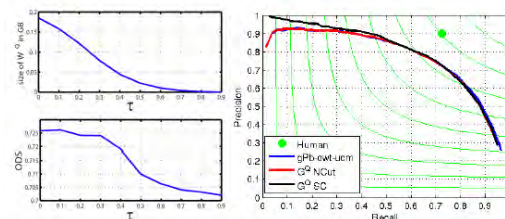
个较小的矩阵。原始矩阵对应于一个图 $G = (V; E)$ ，其中像素表示节点 V 。它们与它们的邻节点在一定的半径 r ($r = [2]$ 中 5) 内连接，通过由最大介于 mPb 沿线 ij 边界加权的边 $e_{ij} \in E$:

$$w_{ij} = \exp\left(-\max_{p \in ij} \max_{\theta} \frac{mPb(p, \theta)}{\rho}\right) \quad (8)$$

其中 ρ 是恒定的。为了减少光谱分析步骤的复杂性，我们通过用 mPb 值计算转折点分割将像素预分组为超像素，参见图 2。这个预分组对于边界检索是可靠的，即重要的边界不会被错过。然而， mPb 不支持一个精确的边界定位。因此，在 mPb 超过临界值的区域，我们为每一个还原图中的单一像素保持一个节点。最后我们从还原图中计算 sPb，其中关系根据定理 1 或定理 2 重新加权。我们分别用 G^QNCut 和 G^QSC 表示两张图。

4.1 实验评估

图还原的效果在 BSD 500 基准 [2] 上进行评估。在图 3(左)中，我们提出了一个对边界精度重算 (BPR) 和 sPb 关系矩阵相对 mPb 临界值迹的内存空间，这确定了流域内的超像素的集合。我们令 $\tau = 0.3$ ，我们选择 $= 0:3$ ，这不改变性能但是内存消耗降低了



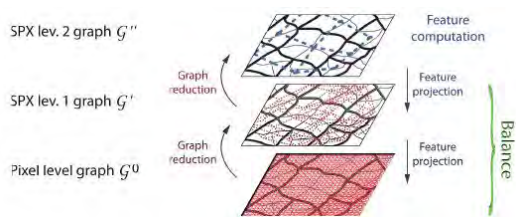
图三。（左）超像素/像素参数相对于边界精度重算 (BPR) 的性能 (顶部) 和所有 BSDS500 [2] 实验图像中 sPb (下) 中的关系矩阵的内存空间的分析。（右）[2] 的 BPR 的性能与提出的基于 BSDS500 图还原方法的比较。

	ODS	OIS	AP
[2] gPb-owt-ucm	0.73	0.76	0.73
G^QNCut	0.72	0.75	0.74
G^QSC	0.73	0.75	0.75

表一。[2] 的综合边界性能评估 (最优数据规模 [ODS]，最优图像尺度 [ODS]，平均精度 [AP]) 和我们所提出的在 BSDS500 上的方法。

50% 且 sPb 处理时间从 71.5 缩短到 33.5 秒。（这包括还原成本，在实验中可以忽略不计。）

图 3 (右) 所示还原图 G^Q 得到与 [2] 使用全像素图的原始算法相同的性能。还原图 G^QSC 甚至在高精度 BPR 的范围内得到更好的结果，并且因此，比原图达到一个更高的平均精度 (AP)，参见表 1。此外，由这种还原技术提取特征向量 (EV 的) 是比完整的图的一个更复杂的展现：在图 2 中，原始图的第一个 EV 的特征向量主要受全局图像坐标 (他们强调沿水平，垂直或对角方向分组) 的影响。通过对比，还原图中的第四特征向量的减少已经显现出了图像中的熊。注意，对 SC 也对 NCut 还原，已分组的像素的特征向量变得分段连续从而，就像我们的猜想，接近它们各自的不松弛的问题，即 DNCut 和 NCut。这可能进一步解释性能的略微改善。



图四。在视频分割中，判别关系是从 2 级超像素的粗糙图 $G'' = (V''; E'')$ 计算出的。更精细的 1 级超像素图 $G' = (V'; E')$ 或像素图 $G^0 = (V^0; E^0)$ 的设计和进一步还原对重平衡量是很必要的。我们提出了一种避免了昂贵的扩张的权重调整。

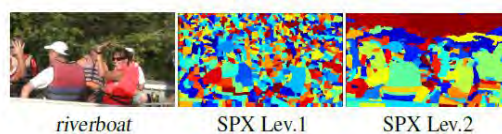
5.视频分割

我们用图还原和重新加权发展一种结合在一个粗糙的像素水平和像素或超像素水平的平衡图结构上计算的更多明显特征的视频分割方法。此外，这个方法得益于在粗糙水平计算解而要求较低的资源需求；参看图 4 所示整体概念示意。

该方法由在拉索等人[12]建立，他们为整个视频序列创建了一个图 $G' = (V'; E')$ ，其中节点 $V' = \{i_1, i_2 \dots i_n\}$ 是预先计算的超像素。边 E' 把每个超像素和其在空间和时间的邻节点连接起来。边的权重

w'_{ij} 是超像素 i 和 j 之间的成对关系，这是由动作，外观和像素的形状特征计算得到的。在[12]中，超像素由最低级（1 级）的[2]的层次图像分割组成，并且重加权使得超像素的量丢失。

我们用一个另外的，粗糙的图 $G'' = (V''; E'')$ ，其中节点 $V'' = \{I_1, I_2 \dots I_n\}$ 代表[2]中的粗糙超像素（2 级）；参看图 5 中的一个例子。2 级的超像素之间的关系 w''_{IJ} 是以[12]中的特征为基础而决定的，但它们被重新加权以获取 G' 的更精细的图形结构的更好的平衡。（参见简介，2 级超像素显示对平衡切割方案的计算有明显影响的尺寸的大变化。）在这种情况下，还原的等效 G^0 与 G'' 在于重加权关系 w'_{ij} 是相同的。



图五。从分层图像分割 [2] 中提取的 1 级（最低层）和 2 级（ $ucm = 0.12$ ）超像素样本。

重加权首先表明边 e'_{IJ} 对更精细的一级超像素图 $G' = (V'; E')$ 的投影。这是一个高阶的扩张问题，因为边 e''_{IJ} 通常连接着超过 2 个的 1 级超像素。整体扩展[1]可能产生一个数学公式：

$$\mathcal{E}' = \{e'_{ij} \mid \exists e''_{IJ} \in \mathcal{E}'', i \in I, j \in J\} \quad (9)$$

$$w'_{ij} = \sum_{e'' \in \mathcal{E}'': (I, J) \in e'', i \in I, j \in J} w''_{IJ}. \quad (10)$$

事实上，大代价的整体扩展图 G' 的计算是没有必要的，因为重加权可以直接做到粗糙图 G'' 。

根据定理 1，还原图 G^0 保存 G' 的归一化切割得到权重：

$$w^Q_{IJ} = \begin{cases} |I||J|w''_{IJ} & \text{if } I \neq J \\ |I|(|I| - 1) \sum_{K \neq I} w''_{IK} & \text{if } I = J \end{cases} \quad (11)$$

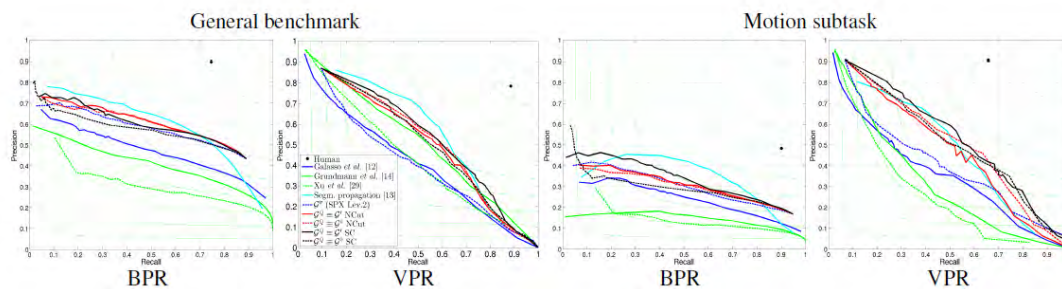
根据定理 2，还原图 G^0 保存 G' 的光谱聚类的解得到权重：

$$w^Q_{IJ} = \begin{cases} |I||J|w''_{IJ} & \text{if } I \neq J \\ (|I| - 1) \sum_{K \neq I} (1 - |K|)w''_{IK} & \text{if } I = J \end{cases} \quad (12)$$

不只是关于更精细的超像素分辨率的重平衡，根据原始像素等级图 G' 有效地重新计算权重也是可能的；参见图 4。

5.1 实验评估

[13]最近提出了一个具有挑战性的基于[25]中的高清视频的视频分割基准，[2]中的边界精度重算（BPR）度量和一种反映一个好的视频分割的性质，如时间上的一致性的量的精确重算度量（VPR）。该基准也考虑了各种子任务运动分割任务，其中只有移动对象才会被评估。我们考虑总的任务和运动的分任务。



	General benchmark						Motion subtask						Volume statistics	
	BPR			VPR			BPR			VPR			Length	NCL
Algorithm	ODS	OSS	AP	ODS	OSS	AP	ODS	OSS	AP	ODS	OSS	AP	$\mu(\delta)$	μ
Human	0.81	0.81	0.67	0.83	0.83	0.70	0.63	0.63	0.44	0.76	0.76	0.59	83.24(40.04)	11.90
A. State-of-the-art video segmentation algorithms														
*Corso et al. [6]	0.47	0.48	0.32	0.51	0.52	0.38	0.21	0.21	0.11	0.37	0.35	0.23	70.67(48.39)	25.83
*Galasso et al. [12]	0.51	0.56	0.45	0.45	0.51	0.42	0.34	0.43	0.23	0.42	0.46	0.36	80.17(37.56)	8.00
*Grundmann et al. [14]	0.47	0.54	0.41	0.52	0.55	0.52	0.25	0.34	0.15	0.37	0.41	0.32	87.69(34.02)	18.83
*Ochs and Brox [20]	0.17	0.17	0.06	0.25	0.25	0.12	0.26	0.26	0.08	0.41	0.41	0.23	87.85(38.83)	3.73
Xu et al. [29]	0.38	0.46	0.32	0.45	0.48	0.44	0.22	0.30	0.15	0.32	0.36	0.27	59.27(47.76)	26.58
Segm. propagation [13]	0.61	0.65	0.59	0.59	0.62	0.56	0.47	0.52	0.34	0.52	0.57	0.47	25.50(36.48)	258.05
B. Proposed graph reduction methods: G^Q equiv. to SPX Lev.1 (G') or pixel level (G^0) graphs, wrt NCut or spectral clustering (SC)														
* $G^Q \equiv G' \text{ NCut}$	0.62	0.66	0.54	0.54	0.57	0.52	0.39	0.49	0.28	0.51	0.54	0.48	69.08(40.91)	20.00
* $G^Q \equiv G^0 \text{ NCut}$	0.62	0.66	0.53	0.55	0.59	0.53	0.39	0.48	0.28	0.53	0.59	0.52	71.77(40.27)	20.00
* $G^Q \equiv G' \text{ SC}$	0.62	0.66	0.54	0.55	0.59	0.55	0.41	0.49	0.32	0.55	0.60	0.54	61.25(40.87)	80.00
* $G^Q \equiv G^0 \text{ SC}$	0.60	0.65	0.52	0.53	0.57	0.52	0.38	0.45	0.28	0.52	0.56	0.52	52.81(46.25)	100.00
* G'' (SPX Lev.2)	0.60	0.65	0.52	0.44	0.50	0.42	0.38	0.49	0.28	0.44	0.51	0.41	60.42(41.85)	18.00
C. Streaming algorithms: variants of methods ($G^Q \equiv G^0 \text{ NCut}$) and ($G^Q \equiv G' \text{ SC}$) above, wrt max number of nodes maintained														
*Stream ^{5k} $G^Q \equiv G^0 \text{ NCut}$	0.61	0.67	0.52	0.55	0.59	0.53	0.39	0.50	0.28	0.52	0.58	0.50	73.31(40.33)	15.63
*Stream ⁵⁰⁰ $G^Q \equiv G^0 \text{ NCut}$	0.61	0.66	0.52	0.55	0.59	0.52	0.40	0.48	0.27	0.51	0.56	0.48	68.61(41.21)	16.57
*Stream ^{5k} $G^Q \equiv G' \text{ SC}$	0.59	0.64	0.52	0.53	0.56	0.52	0.40	0.47	0.29	0.52	0.55	0.51	33.55(36.11)	116.78
*Stream ⁵⁰⁰ $G^Q \equiv G' \text{ SC}$	0.58	0.62	0.51	0.51	0.51	0.51	0.37	0.43	0.28	0.52	0.53	0.51	3.17(11.49)	586.47

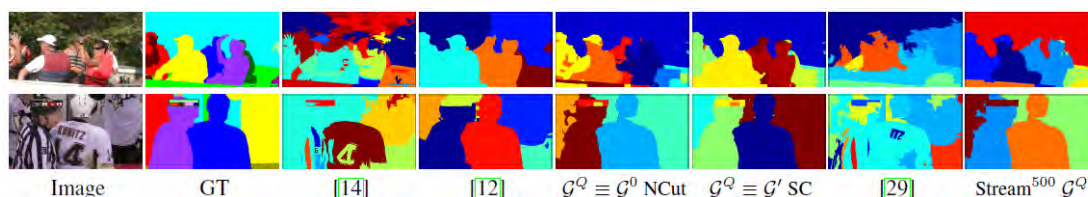
图六。最先进的视频分割算法[6, 14, 12, 29, 13]的比较（表中 A 块），（B 块）所提出的光谱图还原方法和（C 块）流变型，均在[13]的基准下。该图显示出边界精确度重算（BPR）和量的精确重算（VPR）曲线，为一般基准和运动分割子任务。该表显示了 BPR 和 VPR 的总的性能评估（最优数据规模[ODS]，最优分割尺度[OSS]，平均精度[AP]）并包含了长度统计（均值 μ 和标准差 δ ）和簇号（NCL）。（*）表示视频帧调整大小 0.5。每一块都突出了最好的性能。（图和块 A, B）所提出的基于图还原（ $G^Q \equiv G' \text{ NCut}, G^Q \equiv G^0 \text{ NCut}, G^Q \equiv G' \text{ SC}, G^Q \equiv G^0 \text{ SC}$ ）的方法超越了所有的最新方法除了分割传播[13]（见第 5 节）。（块 A, B, C）所提出的两种提出的还原方法的流变体 $Stream^X$ 稍微降低了它们的性能和并提供了最好的结果。（ $Stream^X$ 中的 X 表示保持在整个视频上的最大节点数。）（见第 6 节。）

在图 6（上）中，我们比较了不同的图还原的变型和最先进的视频分割。我们认为还原图 G^Q 等价于一级超像素图 G' 或在原像素之上的图 G' 。每一个版本都关于归一化分割（NCut）的等价，参见定理 1，和光谱聚类特征（SC）的等价，参见定理 2。

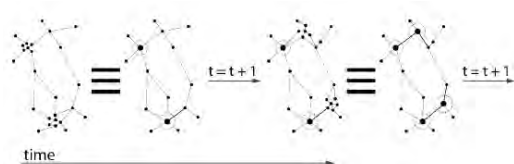
除了精密重算的图，我们也报告 BPR 和 VPR 的总体性能指标（最优数据集规模 [ODS]，最优分割尺度[OSS]，平均精度[AP]），对于一般的和运动的任务。块 A 中现有的方

法（除了[13]）优于块 B 中的图还原方法。在这些变型中， $G' \equiv G' \text{ SC}$ ，使用光谱聚类等价算法，是表现最好的。对最好的对整体任务的视频分割方法的平均改进程度是 7.3% 甚至对运动子任务是 11%。定性的例子如图 7 所示。

比较 2 级超像素图（不加权） G'' 和[12]的结果，其中 1 级超像素无重加权，显示出基于较大的超像素估计的关系是有利的。这可能是因为他们是在较大的均匀图像区域



图七。视频分割结果将[14, 12, 29]的算法和我们提出的光谱图还原方法与现有的地面情况[13]作了比较（我们报告每个算法中在 VPR 中性能最好的由粗到细的层次）。我们最好的两种执行方法（ $G^Q \equiv G^0 NCut, G^Q \equiv G^1 SC$ ）提高了算法[12]的质量，使之能更好的识别视觉对象。该流方法（ $Stream^{500} G^Q \equiv G^0 NCut$ ）提供了与它的批量版（ $G^Q \equiv G^0 NCut$ ）类似的结果。



图八。在流视频分割中，光谱图还原允许通过一种新的方法，在每个时间点 $t=i$ ，只有那些有很可能是可用的分组的节点被合并。根据假定的分组，一个等价的还原图 $G_{1:i}^Q$ 是确定的，这持续到了接下来的时间 $t=t+1$ ，迭代着过程。在时间 $t=i$ 最佳（超过 1: n 帧）视频分割是通过聚类等价图 $G_{1:i}^Q$ 得到。

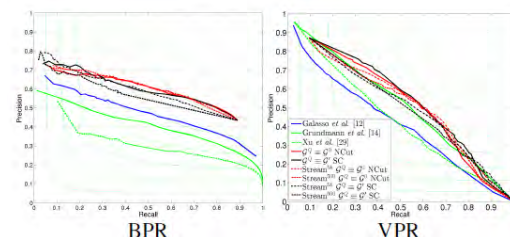
上计算，例如，运动可以计算得更可靠。此外，重平衡 G' 进一步提高性能，特别是关于 VPR。

算法[13]（分层图像分割[2]随着视线传播）仅部分超过所提出的方法。这可以解释为[2]的更复杂的图像特征，如纹理描述符。[12] 的低等级特征（平均颜色和运动）还会有很大的提高，我们就在这里使用。这进一步证实了我们对于运动任务的更好的结果，其中静态图像特征的作用是不突出的。

对于运行时间和内存使用情况， [12] 的算法的峰序列结果像考拉，熊猫和水牛，其中超像素(图 G' 中的节点)的数量是~200k ($K = 1000$)。对于相同的序列，2 级超像素 (G^Q 的节点) 的数量是~65K。平均来说，超像素的数量因为 3 的一个因素减少了，这意味着只有 10% 的运行时间和 20% 的内存是必要的。

6.流视频分割

在一个流的方案里，新的数据变得可用的只有顺序而且应该在处理后被抛弃。在视频分割的情况下，这就产生了两个限制：一、



图九。用 BPR 和 VPR 的流算法在一般[13]基准上的评估。所提出的我们的图还原算法的流的变型（虚线/红色点线/黑色曲线）保持着性能且优于视频分割[14, 12]和流算法[29]。

未来帧尚未得知；二、只有从先前帧中得到的有限的信息片段可以存在内存中。第一个限制出现在所有必须在第一时间作出决定应用程序里。第二个限制表现为大量的数据要在内存资源有限的情况下处理。

一些文献中的视频分割算法也可在流模式[14, 29, 4]（有时在线作为一个同义词）中得到，但他们都借助于一个严格的马尔可夫假设（参见[29]）。这意味着在当下帧的分割是从当下的图像 V_i 和先前帧的常数 n 确定的分割 S_{i-1}, \dots, S_{i-n} 中计算出的。通常多个图像 V_i, \dots, V_i 而不是一个单一的图像是在一个固定大小 m 的时间窗口中考虑的。

显然，这个弊端在于早期，次优的分割决定。尤其是当动作必须消除视觉对象的外观之间的歧义，必要的信息可能只在一个后来的帧中出现。而在流设置中是没有办法在第一帧避免这个问题的，在一开始缺乏的信息不应影响当前和未来的帧的图像分割的质量。

比起只考虑从固定的帧的数量中得到的信息，我们提出了一个更灵活的数据还原，

其中某些决定立即做出而且当必要的信息是可得到的不确定的决定会被推迟。这个概念称为在跟踪[22, 31]中的递延推理。

我们提出的图还原使这一概念在视频分割中的成就成为图 8 所示的那样。给定一个联系着所观察到的一个视频序列

$V_{1:i} = \{V_1, \dots, V_i\}$ 的部分的图 $G_{1:i}$, 我们建议只将这些节点合并这样一来一个分组是有较大的确定性是可用的, 以此确定一个还原图 $G_{1:i}^Q$ 持续下去。在每个时间点, 最优分割 $S_{1:i}^*$ 由聚类的还原图提供。从 $G_{1:i}^Q$ 中得出的最优解与 $G_{1:i}$ 在第三节中陈述的条件下是等价的, 如果 $G_{1:i}$ 得出的最优解是 $G_{1:i}^Q$ 中分组的超集。因此, 我们规划出从时间 $i-1$ 到 i 的流算法为两个基本步骤: 一、输入新的帧到当前图形, 即 $G_{1:i-1}^Q \cup G_i$; 二、将图聚类到假设的一个超集并因此还原它, 即 $G_{1:i}^Q$ 。

实施细节。如果找几十段, 一个假设的超集可通过成千上万的节点给出(我们用这个参数来试验)。聚类的步骤是一个光谱聚类的迭代: 用 k -均值计算特征向量和聚类。 k -均值需要两个扩展: 一、从现有的分组中产生解(聚类用随机分割的方法初始化, 在可用时使用从先前的迭代次数得到的标签); 二、在聚类中心的更新中考虑分组的多样性(估计的聚类方法由节点的基数加权)。

6.1 实验评价
我们用与在 5.1 节中描述的相同的计划来评估所提出的视频流分割。图 9 比较了现有的方法[14, 12, 29]和从上一节选出的还原图变型 ($G^Q \equiv G^0 NCut, G^Q \equiv G^0 SC$) 与提出的流变型 $Stream^X$ 。X 是在整个视频上的保持的节点的最大数目。我们在 5000 到 500 的范围内改变这个数目, 在 2 级全图中总共包含 ~33000 个节点。注意, 在 1.5 帧内 500 个节点与 2 级超像素的数目是一样的。图 6 中的表显示了相应的总体性能措施。一个好的流算法应尽可能接近它的批处理相应的性能。所有提出的流变型只降低几个百分点的性能, 但仍达到最好的结果。特别地, 基于 NCut 的流还原保持着最好的性能。它的性能对于 $Stream^{500}$ 1.9% 的下降与在格伦德曼等人[14]的流扩展 [29]中的 12.8%

的下降相比算是很小的了。基于光谱聚类的流方法确保着性能但提供着不太理想的分割(图 6 中的表的长度统计)。这是由于随着时间的推移改变了聚类之间的平衡, 例如对象进入镜头或变焦。

两种流变型进一步减少运行时间和内存消耗。在每个时间点, 加上新输入的 f 帧(根据应用 $1 \leq f \leq n$) $Stream^{500}$ 只在内存中保留 500 个节点 (~1.5 帧)。每一帧给出 ~330 个 2 级超像素, 还原阶段执行光谱聚类 $|V^Q| \approx 500 + 330f$ 。这分别降低了 [12] 的运行时间和内存需求为 5% 和 2% (我们令 $f=10$, 却为增强的鲁棒性增加了 G^Q 的 1 帧)。

7. 结论

我们已经解决了常用的视频分割与基于聚合超像素而不是单一像素的图的工作。我们已经表明, 相应的图应重新加权并分析了在一定的条件下保持全图的解的两构想。我们的实验鼓励结合了一种给出的重加权方案的超像素的使用。我们已经实现了最先进的图像和视频分割的性能。同时运行时间和内存消耗也明显减少了。此外, 我们也介绍了在视频流分割中递延的推理的概念, 这也得出了这个领域最好的结果。

致谢

我们承认部分资金由 ERC Starting Grant VideoLearn 提供。

参考文献

- [1] S. Agarwal, K. Branson, and S. Belongie. Higher order learning with graphs. In ICML, 2006.
- [2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. TPAMI, 33(5):898–916, 2011.
- [3] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In ECCV, 2010.
- [4] J. Chang, D. Wei, and J. W. Fisher. A video representation using temporal superpixels. In CVPR, 2013.
- [5] X. Chen and D. Cai. Large scale spectral clustering

- with landmarkbased representation. In AAAI, 2011.
- [6] J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. TMI, 2008.
- [7] X. Di, H. Chang, and X. Chen. Multi-layer spectral clustering for video segmentation. In ACCV, 2012.
- [8] A. Eriksson, C. Olsson, and F. Kahl. Normalized cuts revisited: A reformulation for segmentation with linear grouping constraints. In ICCV, 2007.
- [9] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the nystrom method. TPAMI, 26(2), 2004.
- [10] C. Fowlkes and J. Malik. How much does globalization help segmentation? Technical report, EECS – UC Berkeley, 2004.
- [11] K. Fragkiadaki and J. Shi. Video segmentation by tracing discontinuities in a trajectory embedding. In CVPR, 2012.
- [12] F. Galasso, R. Cipolla, and B. Schiele. Video segmentation with superpixels. In ACCV, 2012.
- [13] F. Galasso, N. S. Nagaraja, T. J. C'ardenas, T. Brox, and B. Schiele. A unified video segmentation benchmark: Annotation, metrics and analysis. In ICCV, 2013.
- [14] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph-based video segmentation. In CVPR, 2010.
- [15] D. Kushnir, M. Galun, and A. Brandt. Efficient multilevel eigensolvers with applications to data analysis tasks. TPAMI, 32(8), 2010.
- [16] A. Levinshstein, C. Sminchisescu, and S. Dickinson. Spatiotemporal closure. In ACCV, 2010.
- [17] M. Maire and S. X. Yu. Progressive multigrid eigensolvers for multiscale spectral segmentation. In ICCV, 2013.
- [18] S. Maji, N. Vishnoi, and J. Malik. Biased normalized cuts. In CVPR, 2011.
- [19] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In NIPS, 2001.
- [20] P. Ochs and T. Brox. Object segmentation in video: a hierarchical variational approach for turning point trajectories into dense regions. In ICCV, 2011.
- [21] S. Rangapuram and M. Hein. Constrained 1-spectral clustering. In AISTATS, pages 1143–1151, 2012.
- [22] D. B. Reid. An algorithm for tracking multiple targets. Automatic Control, 24(6), 1979.
- [23] J. Shi and J. Malik. Normalized cuts and image segmentation. TPAMI, 22(8):888–905, 2000.
- [24] N. Sundaram and K. Keutzer. Long term video segmentation through pixel level spectral clustering on gpus. In ICCV Workshops, 2011.
- [25] P. Sundberg, T. Brox, M. Maire, P. Arbelaez, and J. Malik. Occlusion boundary detection and figure/ground assignment from optical flow. In CVPR, 2011.
- [26] C. Taylor. Towards fast and accurate segmentation. In CVPR, 2013.
- [27] A. Vazquez-Reina, S. Avidan, H. Pfister, and E. Miller. Multiple hypothesis video segmentation from superpixel flows. In ECCV, 2010.
- [28] U. von Luxburg. A tutorial on spectral clustering. Stat. Comput., 17(4), 2007.
- [29] C. Xu, C. Xiong, and J. J. Corso. Streaming hierarchical video segmentation. In ECCV, 2012.
- [30] D. Yan, L. Huang, and M. Jordan. Fast approximate spectral clustering. In KDD, pages 907–916, 2009.
- [31] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. ACM Comput. Surv., 38(4), 2006.
- [32] S. X. Yu and J. Shi. Grouping with bias. In NIPS, 2001.