

Event-based Moving Object Detection and Tracking

作者：Anton Mitrokhin, Cornelia Fermuller, Chethan Parameshwara, Yiannis Aloimonos

期刊：IROS

年份：2018

1、问题背景/研究动机

本文研究的主要问题：**解决具有挑战性的场景（低分辨率事件）中进行目标检测与跟踪的问题**

- **现有方法存在的问题：**

1、**挑战性场景下的目标检测**：现有方法（基于帧）在处理**快速运动**、**低光照条件**和**不断变化的照明**等具有挑战性的场景中面临困难。

2、**未知运动和场景的系统**：大多数问题需要对系统进行**实时分析**以处理快速变化的现实场景，现有方法对于**没有关于场景的先验知识**的系统存在问题，即**无法在未知场景的情况下准确检测**。

3、**低分辨率/高噪声**的影响：由于事件相机的工作原理，其只在感知到场景中发生变化时产生输出，且事件生成的阈值是动态的，这种工作方式会影响感知的分辨率；由于事件相机的高光敏度，在极端光照条件下，可能也会产生大量噪声。

2、解决方法

(1) 提出了一个新的基于事件的运动补偿流程。

(2) 提出了一种新的**事件流表示方法**：**时间-图像表示**，使其能够在每个时刻利用有关事件流的动态（时序）组成部分的信息，而不仅仅是空间组成部分。

(3) 用**参数模型近似事件流的3D几何结构**，以此进行运动补偿，无需任何形式的特征跟踪或明确的光流计算。并在迭代过程中检测不符合模型的运动物体。

(4) 提出了一个新的事件数据集——**极端事件数据集（Extreme Event Dataset, EED）**，记录了多种尺寸的物体在多种光照条件下以不同速度运动的特征。

(5) 在数据集中对所提出的方法进行全面的定量和定性评价。

3、具体实现

- 1) 本文算法受到了**3D点云处理技术**的启发，如Kinect Fusion，利用了变形场进行全局最小化处理。
- 2) 通过将**4个参数的运动模型**拟合到小时间间隔内的事件云中，实现相机的**全局运动补偿**。这四个参数分别是**x轴和y轴的平移、z轴的扩展和二维旋转**。
- 3) 利用**时间图像**和**事件计数图像**来计算**事件密度**和**全局运动补偿的误差**。
- 4) 通过最小化误差函数来寻找**不符合运动模型的事件簇**，并将其标记为**独立的移动区域**，同时对每个检测到的区域拟合运动模型。

- 数据表示

• 输入事件的表示

记 t 时刻的事件 (x, y) 到 t_0 时刻的位置 (x', y') 的二维位移为一个变形场

$\phi(x, y, t - t_0) : (x, y, t) \longrightarrow (x', y', t)$ 。该方法的目标是找到运动补偿偏差场

$\phi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ ，使得运动补偿事件投影到图像平面上时具有最大的密度。

这些**运动补偿事件**记为：

$$\begin{aligned} C' &= \Pi\{\phi(C)\} = \Pi\{\phi(x, y, t - t_0)\} \\ &= \{x', y', 0\} \quad \forall \{x, y, t\} \in C \end{aligned} \quad (1)$$

由于事件的几何性质，这样的变形场编码了每个事件的光流。其中， Π 是沿时间轴投影运动补偿事件的时间投影函数，用于对 $\mathbb{R}^3 \longrightarrow \mathbb{R}^2$ 的数据进行降维，简化最小化过程。

其中，使用**事件计数图像**和**时间图像**这两个用于表示事件数据的离散化数据结构。**事件计数图像**用于记录**每个像素上映射到的事件数量**；**时间图像**用于记录事件的**时序信息**

- 事件计数图像表示

将图像平面离散为特定大小的像素，将每个投影事件映射到某个离散的像素上，并将映射到该像素上的事件总数记为该像素的一个值。这种数据结构称为事件计数图像。

令 $\xi_{ij} = \{\{x', y', t\} : \{x', y', 0\} \in C', i = x', j = y'\}$ (2) 为事件轨迹：一组沿时间轴变形的事件，投影映射到像素 (i, j) 上。

事件计数图像为： $\mathcal{I}_{ij} = |\xi_{ij}|$

事件密度 \mathcal{D} 为：

$$\mathcal{D} = \frac{\sum_{i,j} \mathcal{I}_{ij}}{\#\mathcal{I}} = \frac{|C'|}{\#\mathcal{I}} \quad (4)$$

其中， $\#\mathcal{L}$ 表示至少有一个事件映射在它上面的像素的个数

- 时间图像表示

在高纹理场景的快速运动中，常遇到不同边缘产生的事件可以投影到同一个像素上的情况。为了缓解这个问题，利用事件的时序信息，使用一种新的表示方法，称为时间-图像 \mathcal{T} 。

- 时间图像 \mathcal{T} 是一种离散化的平面，其中每个像素包含了通过变形场映射到该像素上的事件的平均时间戳。
- 时间图像的生成：

$$\mathcal{T}_{ij} = \frac{1}{\mathcal{I}_{ij}} \sum t : t \in \xi_{ij} \quad (5)$$

- 时间图像 \mathcal{T} 遵循事件云的三维结构，其梯度提供了全局运动补偿误差的度量。

$(G_x[i, j], G_y[i, j])$ 表示 \mathcal{T} 沿 x 轴与 y 轴的局部空间梯度，全局误差可以通过计算时间图像的局部空间梯度得到：

$$Error = \sum \|G[i, j]\| = \sum (G_x^2[i, j] + G_y^2[i, j]) \quad (6)$$

- 全局误差可以分解为四个方向的误差，分别对应于模型参数的误差

$(h_x, h_y, h_z, h_\theta)$: 图像平面的平移参数 (dx, dy) , 扩展参数 (dz) , 和旋转参数 $(d\theta)$ 。

$$d_x = \frac{\sum G_x[i, j]}{\#\mathcal{I}}, \quad d_y = \frac{\sum G_y[i, j]}{\#\mathcal{I}} \quad (7)$$

$$d_z = \frac{\sum (G_x[i, j], G_y[i, j]) \cdot (i, j)}{\#\mathcal{I}} \quad (8)$$

$$d_\theta = \frac{\sum (G_x[i, j], G_y[i, j]) \times (i, j)}{\#\mathcal{I}} \quad (9)$$

- 时间图像的生成和利用可以**提高运动补偿结果的准确性**。
- 时间图像的像素值可以用来**量化局部运动补偿的误差**，并用于后续的**运动检测**。

• 最小化约束

时间图像 \mathcal{T} 的局部梯度和事件密度 \mathcal{D} 量化事件云运动补偿中的误差。用一个4参数的全局运动模型 $\mathcal{M}^G = h_x, h_y, h_z, \theta$ 来建模全局偏差场 $\phi^G(x, y, t)$ ，以描述相机运动在事件云中引起的变形，坐标变换为：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} - t * \begin{bmatrix} h_x \\ h_y \end{bmatrix} + (h_z + 1) * \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x \\ y \end{bmatrix} \quad (10)$$

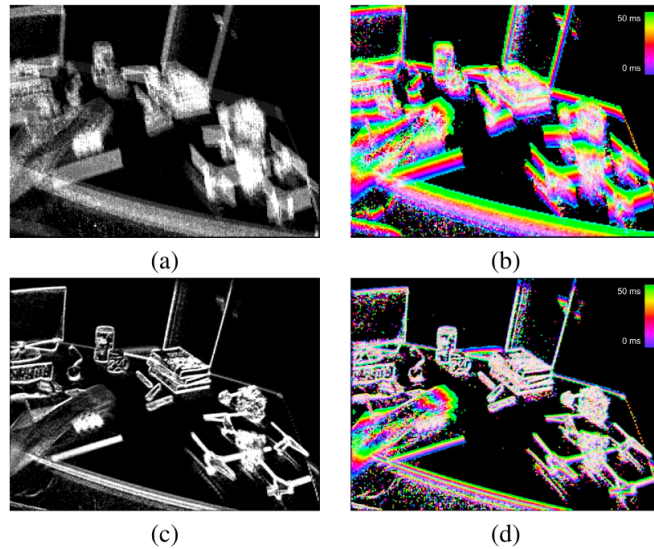


图1：参数化模型拟合示意图

(a)投影后映射到图像像素的事件计数；(b)平均时间戳；
(c)最小化阶段后得到的运动补偿事件计数；(d)时间图像

- 相机运动补偿

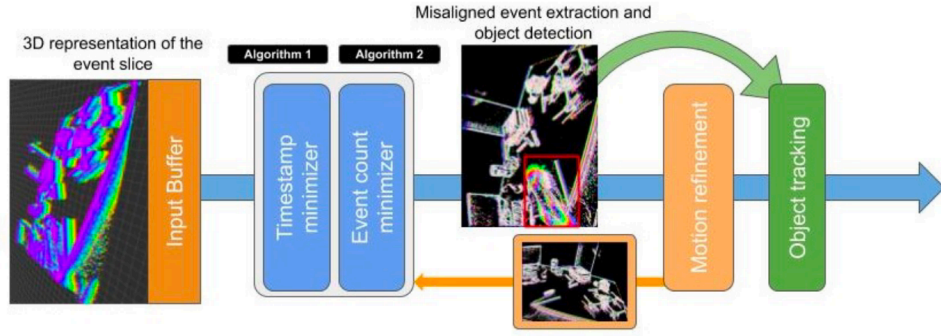


图2：运动补偿和检测独立运动物体流程

这一部分详细描述了方法中的相机运动补偿过程。该算法的流程包括：**相机全局运动补偿和后续的运动不一致性分析**，用于检测独立运动的物体。

为了补偿全局背景运动，使用前文提出的4参数模型 \mathcal{M}^G 。在**最小化全局运动**中，采用了**两个阶段**的策略：在 \mathcal{T} 上进行**粗略运动最小化**，在 \mathcal{L} 上进行**细致运动优化**。在此之后，剔除并跟踪发生失准概率较高的事件（图2中红框部分），对剩余的事件云再次进行运动补偿。

• 粗略运动最小化

Algorithm 1 Global motion compensation in event space using \mathcal{T} .

Data: $\mathcal{M}_{i-1}^G, C, d, \xi$

Result: $\mathcal{M}_i^G, C', \mathcal{T}$

```

 $C' \leftarrow \text{warpEventCloud}(C, \mathcal{M}_{i-1}^G)$ 
 $\mathcal{T} \leftarrow \text{getTimestampImage}(C', d)$ 
 $\mathcal{M}_i^G \leftarrow \text{updateModel}(\mathcal{M}_{i-1}^G, \mathcal{T})$ 
while  $\|\mathcal{M}_{i-1}^G - \mathcal{M}_i^G\|_2 > \xi$  do
     $C' \leftarrow \text{warpEventCloud}(C, \mathcal{M}_i^G)$ 
     $\mathcal{T} \leftarrow \text{getTimestampImage}(C', d)$ 
     $\mathcal{M}_{i-1}^G \leftarrow \mathcal{M}_i^G$ 
     $\mathcal{M}_i^G \leftarrow \text{updateModel}(\mathcal{M}_i^G, \mathcal{T})$ 
end

```

输入：前一帧模型、原始事件云 C 、离散化网格尺寸 d 和精度参数 ξ 。

- 使用前一帧模型对事件云进行变换，得到变换后的事件云 C' 。

- 根据 C' 生成时间图像 \mathcal{T} ，时间图像 \mathcal{T} 的颜色表示平均时间戳（蓝色为 t_0 ，绿色为 $t_0 + \delta t$ ）
- 根据梯度图像 G_x 和 G_y 计算运动模型参数 \mathcal{M}_i^G 的梯度，并使用梯度下降法更新 \mathcal{M}_i^G 的参数。

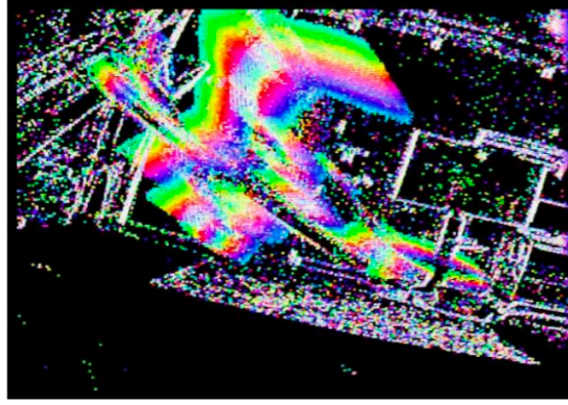


图3：运动补偿算法的一个实例输出

- 重复以上步骤，直到 \mathcal{M}_i^G 的参数收敛或达到最大迭代次数。

• 细致运动优化

通过最大化事件计数图像 \mathcal{I} 的密度 \mathcal{D} 来进行额外的精细运动细化。

Algorithm 2 Global motion compensation in event space with event count image

Data: $\mathcal{M}_{i-1}^G, C, d, \xi$

Result: $\mathcal{M}_i^G, C', \mathcal{I}$

```

 $\mathcal{M}_i^G \leftarrow \mathcal{M}_{i-1}^G$ 
 $C' \leftarrow \text{warpEventCloud}(C, \mathcal{M}_{i-1}^G)$ 
 $\mathcal{I} \leftarrow \text{getEventCountImage}(C', d)$ 
 $\mathcal{D} \leftarrow \text{getEventDensity}(\mathcal{I})$ 
 $\mathcal{D}' \leftarrow 0$ 
while  $\|\mathcal{D} - \mathcal{D}'\| > \xi$  do
     $\mathcal{D} \leftarrow \mathcal{D}'$ 
    for Parameter  $p$  in Model  $\mathcal{M}_i^G$  do
         $\mathcal{D}' \leftarrow \text{getEventDensityCloud}(C, d, \mathcal{M}_i^G + p)$ 
        if  $\mathcal{D}' > \mathcal{D}$  then
             $\mathcal{M}_i^G \leftarrow \mathcal{M}_i^G + p$ 
             $C' \leftarrow \text{warpEventCloud}(C, \mathcal{M}_i^G)$ 
             $\mathcal{I} \leftarrow \text{getEventCountImage}(C', d)$ 
        end
    end
end

```

输入：前一帧模型、原始事件云 C 、离散化网格尺寸 d 和精度参数 ξ

- 使用前一帧模型对事件云进行变换，得到变换后的事件云 C' 。
- 根据 C' 生成事件计数图像 \mathcal{L} ，事件计数图像 \mathcal{L} 的像素值表示映射到该像素的事件数量。
- 最大化事件计数图像 \mathcal{L} 的密度 \mathcal{D} ，通过对模型参数进行变化来获取导数并进行最小化。
- 重复以上步骤，直到密度 \mathcal{D} 的变化小于阈值或达到最大迭代次数。

输出：更新后的模型 \mathcal{M}_i^G 、变换后的事件云 C' 和时间图像 \mathcal{T}

- 多目标检测和跟踪

这一部分描述了多目标检测和跟踪的方法。该方法对时间图像 \mathcal{T} 的不一致性进行独立运动物体的检测，并使用传统的卡尔曼滤波器进行跟踪。

• 检测

- 使用前一帧模型对事件云进行变换，得到变换后的事件云 C' 。
- 使用阈值操作将像素检测为独立运动的像素，并将像素分组为对象。
- 每个像素 $i, j \in \mathcal{T}$ 与一个得分 $\rho(x_i, y_i) \in [-1, 1]$ 相关联，该得分量化了独立运动物体与背景的错位程度。

$$\rho(i, j) = \frac{\mathcal{T}(i, j) - \langle \mathcal{T}_{i, j} \rangle}{\Delta t} \quad (11)$$

- ρ 用作将一个像素分类为背景 \mathcal{B} 或独立运动物体 \mathcal{O}_k 的度量。

$$\mathcal{B} = \{(i, j) | \rho(i, j) \leq 0\} \quad \mathcal{O} = \{(i, j) | \rho(i, j) > \lambda\} \quad (12)$$

• 跟踪

为了解决漏检和误检的问题，特别是在存在遮挡的情况下，使用带有恒定加速度模型的**简单卡尔曼滤波器**对检测到的物体进行跟踪。

4、实验对比

- 数据集

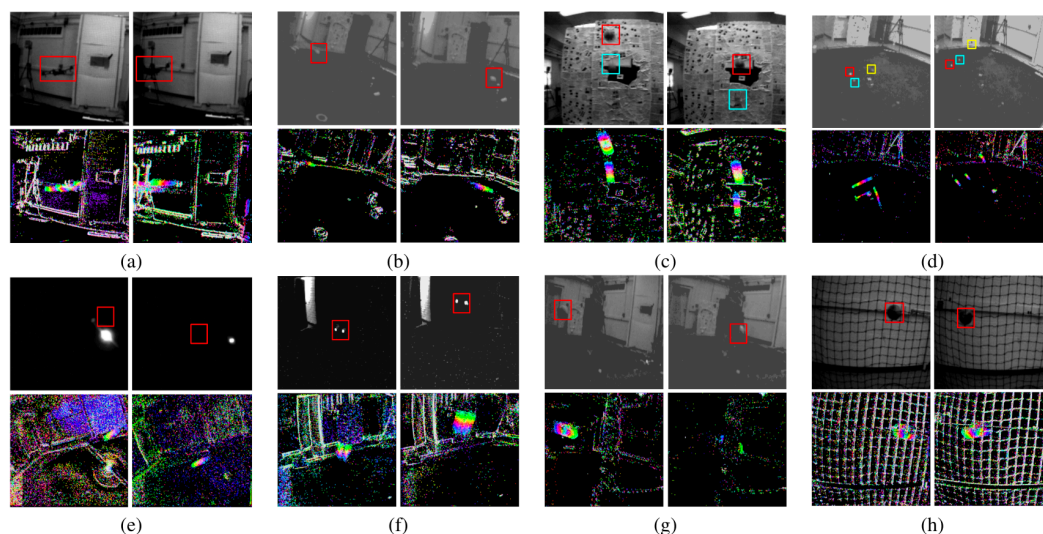


图3：运动补偿算法的一个实例输出：

上行显示被跟踪目标的叠加边界框的RGB帧；下行列显示相应的时间图像。

本文使用的**极端事件数据集 (Extreme Event Dataset, EED)** 是在两种场景下使用DAVIS240B生物启发传感器采集的。其中，大多数序列是在手持环境中创建的，对于四轴飞行器序列，修改了Qualcomm FlightTM平台，将DAVIS240B传感器连接到机载计算机上，在现实场景中采集数据。该数据记录了**多种尺寸的物体在多种光照条件下以不同速度运动**的特征。

根据场景的性质，将记录整理成若干序列。所有的记录都具有多种相机运动，包括旋转运动和平移运动：

- “Fast Moving Drone”——以单个小型遥控四旋翼飞行器为特征的序列。四旋翼具有丰富的纹理，在日光照明条件下，可以跟随多种轨迹在各种背景中运动。(a)
- “Multiple objects”——在正常光照条件下，包含1~3个运动目标的多个记录。物体都很简单，有些物体几乎没有纹理。物体以不同的速度运动，或者沿着直线轨迹运动，或者从表面碰撞。(b)/(c)/(d)：单个、两个和三个同时运动的物体

- “Lighting variation”——将闪烁频率为1 ~ 2 Hz的闪光灯置于暗室中，使事件传感器产生大量噪声。这是一个极具挑战性的序列，在其他方面类似于“Fast Moving Drone”。(e)/(f): 频闪光、低照度（无频闪）

- “What is a Background?”——在大多数跟踪算法评估中，目标在背景前移动。包括以下控制序列，表明即使当背景占据相机和物体之间的空间时，也有可能跟踪一个物体：一个简单的物体被放置在网格后面，运动只能通过网格看到。记录包含网格与相机之间的多种距离，物体以不同的速度抛出。(h)

- “Occluded Sequence”——该序列的唯一目的是测试在短时间内，在无法检测的场景中，进行跟踪的可靠性，有几段序列记录了遮挡场景中的物体运动。(g)

- 实验与评估

评价指标：通过手工标注记录的RGB帧来获得基本真值，对于每个时间切片对应的RGB帧，计算检测到的**与RGB帧中可见物体至少有50%重叠的物体**的百分比，作为成功率的评估指标。

TABLE I: Evaluation of the pipeline on the proposed dataset

Sequence	“Fast Moving Drone”	“Multiple objects ”	“Lighting variation”	“What is a Background?”	“Occluded Sequence”
Success Rate	92.78%	87.32%	84.52%	89.21%	90.83%

从表中可以看出，没有发现运动补偿表现不佳的序列。

“Lighting variation”序列相较于其他序列，性能较差，因为闪光灯的噪声完全覆盖了物体，并阻止了检测——这种场景上的噪声还会因为低光照条件而进一步放大。

在“What is a Background?”序列上，取得了较高的性能。因为物体被网格（位于相机与物体之间）部分遮挡，而网格的高纹理使得相机运动补偿更具鲁棒性。

5、总结

- 提出了第一个在**无约束条件**（全3D未知运动和未知场景）下**仅基于事件**的运动分割（检测与跟踪）方法。

- 使用新的基于事件时间戳的**时间图像表示**以及**事件计数图像**，进行误差计算和物体检测，提高运动补偿的鲁棒性。
- 借鉴了**3D点云处理技术的思想**，通过使用变形场对事件云进行全局最小化处理，利用全局运动模型和事件云数据来对相机运动进行补偿。
- 采集了一个**新的事件数据集（EED）**，包含多个在具有挑战性的条件中多个移动对象的序列。
- 在EED数据集中，使本文方法针对不同场景中的序列进行实验与评估，并证明在低光照和高噪声等挑战性场景中也具有较为鲁棒的性能。

6、启发

- **挑战性的场景中的检测**：本文所提出的运动补偿与物体检测与跟踪算法，在没有关于运动或场景的先验知识的情况下，也能够有效地检测移动物体。
- 在事件数据低分辨率的情况下，可以通过**相机运动估计与运动补偿**，使后续的处理算法更具鲁棒性。
- 未来的工作将扩展该方法以包含更精细的聚类 and 分割。目标是以完整的迭代方式实现三维运动估计和聚类，在检测所有运动物体的同时准确估计三维运动。