密级：无

编号：

国防科技大学

硕士研究生学位论文

开题报告

论文题目： 基于动态视觉传感器的高速图像帧重建研究

学　　号： 18023077 **姓名：** 张里蒙

一级学科： 集成电路工程

研究方向： 动态视觉传感器研究

指导教师： 陈吉华，王蕾 **职称：** 教授，副研究员

学　　院： 计算机学院

开题时间： 2020 **年** 3 **月** **日**

国防科技大学研究生院制

二〇一八年一月

说　明

一、　开题报告应按下述要求打印后装订成册：

1.使用 A4 白纸，双面打印；

2.封面中填写内容使用小 3 号仿宋字体；

3.表中填写内容使用 5 号楷体字体。

二、　封面中的编号由学院填写，采用八位数编码，前四位为审批日期，精确到年月即可，第五位为院别，后三位为审批流水号(按年)。如 15016100，为 6 院 15 年审批的第 100 位开题申请者，审批时间为 15 年 1 月份。院别代码与编制序列一致，海洋科学与工程研究院代码为 0。

三、　开题报告表中学员填写的内容包括学位论文选题的立论依据、文献综述、研究内容、研究条件、学位论文工作计划、主要参考文献等，指导教师认可学员开题报告内容后，对学员学位论文选题价值、对国内/外研究现状的了解情况、研究内容、研究方案等方面予以评价。

四、　开题报告评议小组一般由 3-5 名具有正高级专业技术职务的专家(包括导师)组成，其中一名为跨一级学科的专家。

五、　博士生开题报告会应面向全校公开举行，评议小组听取研究生的口头报告，并对报告内容进行评议审查。

六、　若开题报告获得通过，应根据评议小组意见对开题报告进行修改，并在开题报告会后两周内，将评议表和修订后开题报告纸质版原件交学院备案；若开题报告未获得通过，则填报延期开题申请，由原开题报告评议小组重新组织开题报告会。

1. 学位论文选题的立论依据

1.1 课题来源：

千万神经元规模的仿脑处理器体系结构研究。

1.2 选题依据：

近年来，计算机视觉领域出现了一种新的相机模型——事件相机，也称为动态视觉传感器(Dynamic Vision Sensor, DVS)。DVS 是受生物启发的传感器，其工作原理与传统相机截然不同。不是以固定的速率捕获图像，而是异步测量每个像素的亮度变化，输出编码了事件、位置和亮度变化极性的事件流。DVS128[8]是第一款商用的 DVS 相机。

传统的相机传感器是基于帧进行成像，在固定的间隔(曝光)时间内，像素阵列检测到的光电流被积分在电容器中，在每一帧时间内，每个像素的积累的电压电平以顺序方式传送出芯片。因此使用传统相机传感器时，每一帧都会将所有的信息传输出去，不管单个像素上是有变化进而判断是否需要进行信息传输。此外，由于光电流是在固定时间段(通常在20-30ms)完成积分，所带来的延迟对于高速场景会造成模糊。

DVS 相机相比于传统相机，具有杰出的属性：非常高的动态范围(140dB VS. 60dB)，非常高的时间分辨率(us 数量级)，更低能耗，没有动态模糊。因此，在高速、高动态范围等传统相机面临挑战的场景中，DVS 相机具有巨大的潜力。

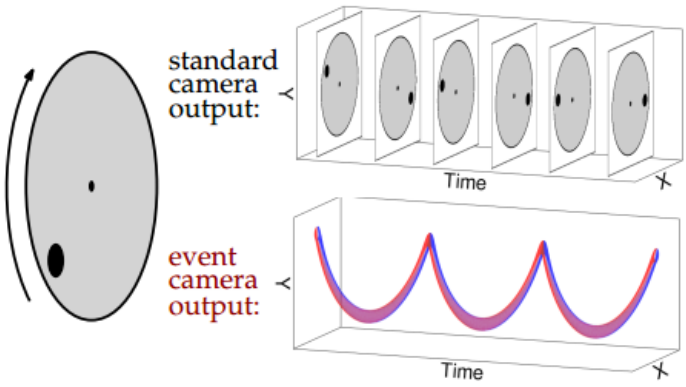

图 1-1 DVS 相机与传统相机输出[1]

由于事件相机的输出与传统相机有根本的不同，现有的计算机视觉技术不能直接应用于这些数据。因此，需要专门的定制算法来利用事件数据。这种专门的算法在从低层次视觉任务如特征检测[9-11]、光流估计到高层次任务如物体分类、姿态识别[12,13]等展示了令人印象深刻的表现。虽然一些工作已经通过将一组事件映射成类似于图像的 2D 表示如 Image plane[14]或者时间表面[12]上的事件积分去解决计算视觉领域的问题。然而，这些中间表示都不是传统意义上的自然图像，这意味着现有的很多计算机视觉工具不能有效地应用，很多传统的应用领域也无法应用 DVS 相机的数据。

图像重建是计算机视觉领域中一个重要的研究方向，传统的图像重建工作通常使用机器学习的方法如对抗生成网络、插值等方法去除运动所带来的运动模糊，场景光强引起的过曝、欠曝、逆光以及去马赛克。这些方法在高速、高动态范围等传统相机面临挑战的场景中能力有限。[15]首次提出使用 DVS 相机的数据进行图像重建，事件相机以异步事件流的形式报告亮度变化，而不是强度帧。与传统相机相比，它们具有明显的优势:高时间分辨率、高动态范围和无运动模糊。利用 DVS 相机事件，如果能够在高速、光照条件不理想的条件下更加高效地重建出清晰、连续的图像帧，计算机视觉各个领域的研究都可以采用更加清晰的数据，架起了 DVS 相机数据与传统计算机视觉领域研究的桥梁。

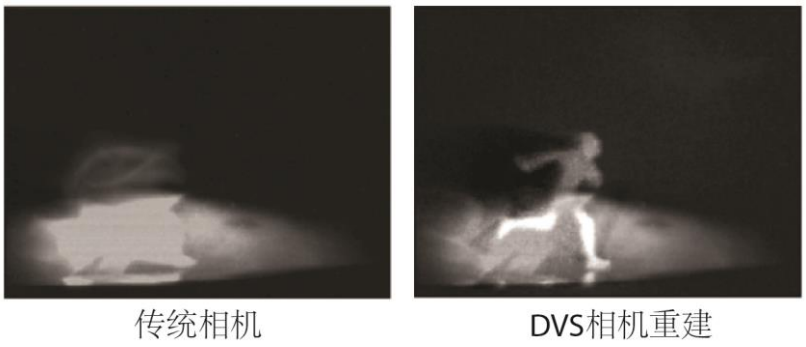

图 1-2 高速场景[5]

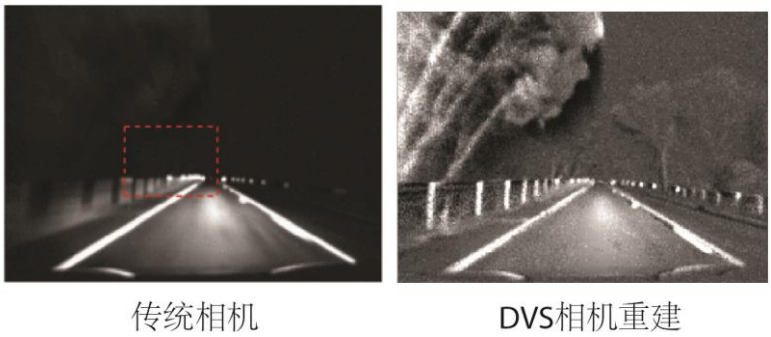

图 1-3 高动态范围场景[5]

所有的视觉传感器都有噪声，这是由于光子固有的噪声和晶体管电路噪声造成的。这种情况在事件相机上尤其明显，因为亮度变化信息的量化过程比较复杂，如何对噪声和非理想影响建模，从而更好地从事件中提取有意义的信息成为研究的重点。单纯的事件累计会由于噪声的累计而导致图像的质量快速下降，作为补救措施，一些工作提出用手工图像的先验知识来限制问题，但是这样限制会导致图像失真和伪影[2]。

目前，关于 DVS 重建帧的研究还处于开始阶段，相关的方法还比较少。主要应用方式分为两种：event-by-event，和 groups of events。其中每种又包含两类，model based 和 model free。基于模型的方法主要依赖于历史信息或者额外的传感器如一个传统的 RGB 相机，通过划窗累计一段时间的事件积累成帧，或者在 RGB 相机上叠加事件信息。Model free 的方法通常采用机器学习的方法，学习事件与梯度之间的关系然后通过泊松积分重建帧，或者通过生成对抗网络(Generative adversarial networks, GAN)。

循环神经网络(Recurrent neural networks, RNN)通常用来处理序列的信息，单纯的 RNN 因为无法处理随着递归、权重所带来的梯度爆炸和梯度消失，难以捕捉长期时间关联，长短期记忆神经网络(Long short time memory, LSTM)通过加入长期记忆可以很好的解决这个问题。视频中相邻的帧之间的差别通常是一个或者几个事件，如果将上一个时序状态的输出作为下一个时序状态的输入，再加上帧之间的事件变化，是一种解决图像重构的方法。

如果采用 RNN 对事件进行编码，通过序列建模的方法，利用事件恢复出连续图像，可以消除由于传统方法加入的先验的手工假设，让恢复出的图像更加自然。

1.3 研究意义：

1) 国内研究较少

DVS 相机是一种全新的视觉数据形式，是可能带来计算机视觉领域质变的一种可能性。目前国际上对于 DVS 数据的研究进步飞速，一大批将 DVS 数据应用到计算视觉各个领域的方法相继问世。但是在国内，DVS 数据的研究工作还处于起步阶段，一些工作研究着眼

于一些特定的领域，对于 DVS 数据的应用研究还有非常大的挖掘空间。

2) DVS 进行图像重建具有优势

DVS 数据具有的高时间分辨率，高动态范围，低数据量以及没有动态模糊的特点。进行高速图像帧重建具有很明显的优势。在帧重建领域，对于高速帧重建的研究还处于起步阶段。相当一部分工作虽然采用了基于模型的或者机器学习的方法，但这些工作的研究更多的集中在静态场景，即场景背景变化很小，并且提出了很多的先验手工假设，极少有研究在高速移动相机的状态下进行图像帧的重构。这为本课题的开展提供了可能，也赋予了本课题更多的现实意义。

3) 重建图像帧应用广泛

传统相机在具有挑战性的场景下表现不佳，计算机视觉领域急需解决在这种场景下所带来的问题，如高速无人机、自动驾驶、SLAM 以及极端光照条件下的各种应用。从 DVS 数据集进行帧重构从而将成熟的计算机视觉的工作直接应用到 DVS 数据集上，具有十分重要的意义。

2. 文献综述

（该领域在国内/外的研究现状及发展动态；阅读文献的范围以及查阅手段等。博士不得少于 2000 字，硕士不少于 1000 字。可附页）

动态视觉传感器是一种基于事件的生物启发的传感器，只在检测到光强变化时才产生事件，这些事件组成的数据集就是 DVS 数据集。相比于传统相机基于帧的数据，DVS 数据集有更高的动态范围和时间分辨率，更低的延迟和能耗。利用 DVS 数据集进行图像帧的重构可以解决这一领域在具有挑战性的场景下的不足，还可以将传统计算机视觉成熟的算法通过成帧技术应用到 DVS 数据集。有关 DVS 数据集的研究才刚刚起步，并且逐渐展现出优秀的成果，成为了计算机视觉领域研究的热点。文献综述将从以下几个方面介绍近些年有关 DVS 数据的研究和应用。

2.1 DVS 相机与 DVS 数据集

2.2 DVS 帧重建

2.3 彩色 DVS 帧重建

2.1 DVS 相机与 DVS 数据集

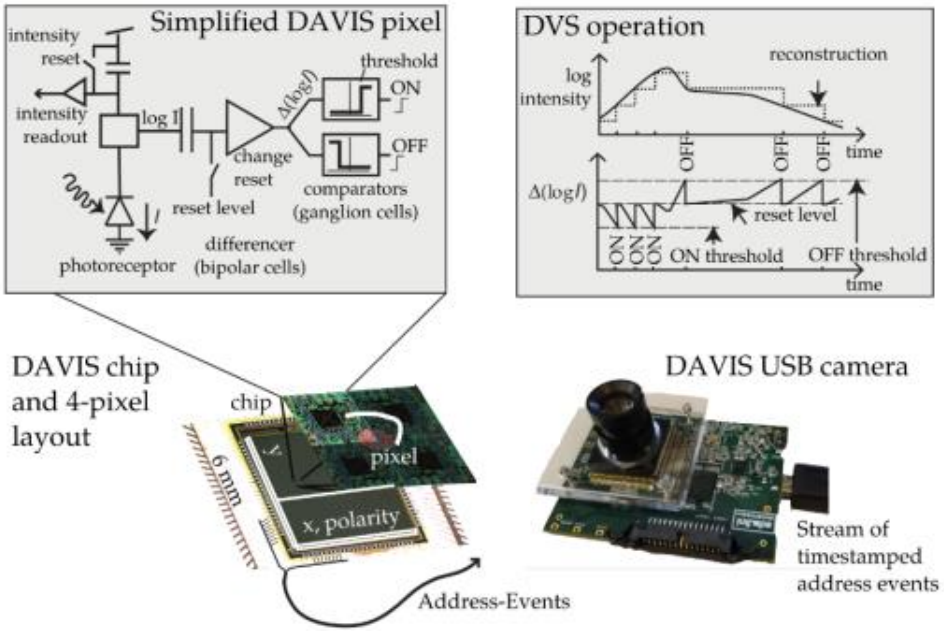

图 2-1 DAVIS 相机简易示意图[1]

如图 2-1 所示，DAVIS 相机包含了一个基于事件的动态视觉传感器 DVS 和一个基于帧的像素传感器 APS 在同一个像素阵列，每个像素共享相同的光电二极管。左上图是每一个 DVS 传感器像素的简化示例，每个像素记录着对数光强，当超过了一定的阈值之后就发射一个事件，包含了坐标、事件以及 1 位的极性，“ON”表示亮度增加，“OFF”表示亮度降低。事件通过共享的数字输出总线从像素阵列传输到外围，然后通过使用地址事件表示法(Address event representation, AER) [16]读出从摄像机传输出去。AER 总线可能会变得饱和，从而影响事件发送的时间。事件相机的读取速率从 2 MHz 到 300MHz，取决于芯片和硬件接口的类型。目前使用最广泛的事件相机包括 DAVIS128[8]、DAVIS240、DAVIS346、ATIS、DVS-Gen、Celex-IV，具体的参数如表 2-1 所示。

在模拟电路中，对亮度变化的监测是快速的，事件的读出是数字的，带有 1MHz 时钟，这意味着事件被检测到，并以微秒的分辨率记录时间。因此，事件相机可以捕捉到非常快

的动作，而不会遭受传统相机典型的运动模糊。每个像素独立工作，不需要等待帧的全局曝光时间：一旦检测到变化，它就会被传输。因此,事件相机有最小的延迟:大约 10μs(实验室)，和毫秒级(现实世界)。

由于事件相机只传输亮度变化，因此删除冗余数据，能量仅用于处理像素的变化。在芯片级别,大多数事件相机的功率为 10mW,还有原型实现不到 10μW。将传感器直接连接到处理器的嵌入式事件系统已经证明了系统级的功耗(感应加处理)为 100mW 以下[17]。

事件相机的动态范围(>120dB)明显超过传统高质量相机 60dB，使之能够在各种挑战的场景下工作。这是由于像素的感光器在 log 级别尺度下工作，并且每个像素独立工作，不需要全局快门。就像生物视网膜一样，DVS 像素可以适应非常暗的刺激，也可以适应非常亮的刺激。

	DVS128 [2]	DAVIS240 [4]	DAVIS346	ATIS [3]	DVS-Gen2 [5]	CeleX-IV [70]
Supplier	iniVation	iniVation	iniVation	Prophesee	Samsung	CelePixel
Year	2008	2014	2017	2011	2017	2017
Resolution (pixels)	128 × 128	240 × 180	346 × 260	304 × 240	640 × 480	768 × 640
Latency (μs)	12μs @ 1klux	12μs @ 1klux	20	3	65 - 410	-
Dynamic range (dB)	120	120	120	143	90	100
Min. contrast sensitivity (%)	17	11	14.3 - 22.5	13	9	-
Die power consumption (mW)	23	5 - 14	10 - 170	50 - 175	27 - 50	-
Camera Max. Bandwidth (Meps)	1	12	12	-	300	200
Chip size (mm^2)	6.3 × 6	5 × 5	8 × 6	9.9 × 8.2	8 × 5.8	-
Pixel size (μm^2)	40 × 40	18.5 × 18.5	18.5 × 18.5	30 × 30	9 × 9	18 × 18
Fill factor (%)	8.1	22	22	20	100	9
Supply voltage (V)	3.3	1.8 & 3.3	1.8 & 3.3	1.8 & 3.3	1.2 & 2.8	3.3
Stationary noise (ev /pix /s) at 25C	0.05	0.1	0.1	NA	0.03	-
CMOS technology (μm)	0.35 2P4M	0.18 1P6M MIM	0.18 1P6M MIM	0.18 1P6M	0.09 1P5M BSI	0.18 1P6M CIS
Grayscale output	no	yes	yes	yes	no	yes
Grayscale dynamic range (dB)	-	55	56.7	130	-	-
Max. framerate (fps)	-	35	40	NA	-	-
IMU output	no	1 kHz	1 kHz	no	no	no

表 2-1 商用 DVS 相机参数[1]

目前已经有越来越多的研究工作开始使用通过 DVS 生成的数据集，而不是基于传统帧数据集通过频率编码生成的神经形态数据集。前者更具有生物特征。然而，目前还没有大规模的事件序列与标签真实图像对应的数据集，这是由于在 DVS 相机擅长的具有挑战性的场景中，即高动态范围和高速场景中，传统相机获取的图形质量较差。没有大量标签数据集，就不能够采用强大的深度学习算法解决重建问题。

在传统相机领域，深度学习的快速发展对数据量的大量需求导致了模拟器的诞生，如 CARLA [18]，Microsoft Airsim [19]，UnrealCV [20]。[21]中提出了一个简单的事件摄像机模拟器，然而，这个模拟器没有实现事件相机的工作原理。相反，它只是比较两个连续帧之间的差异，以创建类似于事件相机输出的边缘图像。并没有模拟出 DVS 的异步和低延迟特性，因为来自一对图像的所有事件都分配了相同的时间戳。[22]使用定制的渲染引擎从三维高帧率的场景中渲染图像。虽然这种方法可以 DVS 相机的异步输出，但当亮度信号的变化比选择的固定渲染帧速率处理得更快时，它无法可靠地模拟事件数据。ESIM[6]采用了自适应的采样方案对事件进行精确模拟，与固定采样频率相比，通过光强的变化和连续两张渲染图之间像素的最大位移推断下一次渲染的时候，只在必要时渲染帧，效果更好，效率也更高准确。

2.2 利用 DVS 数据进行帧重构

事件到图像重建由于其应用范围广，是事件相机研究的热点。图像重建的方法可以分为三类，第一类是通过建立计算机视觉模型，加上人工的先验条件进行重建。第二类是直接事件集成的方法进行重建，不依赖于关于场景和运动的假设。第三类是通过机器学习的方法如卷积神经网络，生成对抗网络等进行重建。

第一类模型。Cook 等 [15]通过静态场景下转动的事件相机收集的大量事件中，重建单个图像，并利用通过亮度恒等[23]提供一个关于强度梯度和光流的方程，同时恢复强度图像，光流和角速度。Kim 等[7]开发了一种扩展的卡尔曼滤波器来重建旋转事件相机的 2D 全景梯度图像，然后通过 2D 泊松积分转换为强度帧。后来，他们将他们的方法扩展到静态 3D 场景和 6 自由度相机运动[24]。Bardow 等[25]提出通过能量最小化框架，从事件的滑动窗口同时估计光流和光强。他们提出了第一个适用于动态场景的事件帧重建框架。然而，他们的能量最小化框架采用了多个手工制作的正则化器，这可能会导致在重建中严重的细节损失。

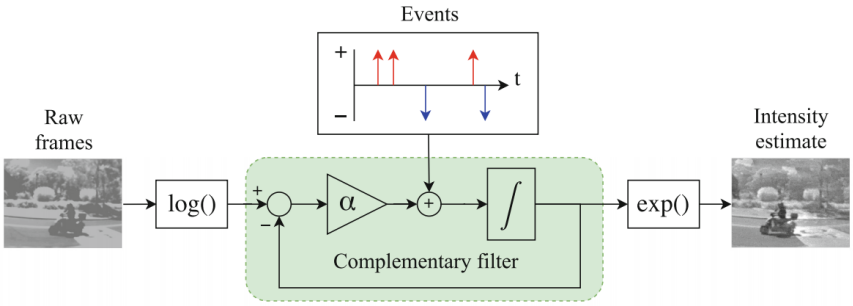

图 2-2 DVS 结合传统相机的重建[5]

第二类模型。Munda 等[26]将光强转换为定义在由事件时间戳引起的流形上的能量最小化问题。他们将事件积分与正则化相结合，在 GPU 上实现了实时性能。Scheerlinck 等[5]提出在集成之前使用高通滤波器对事件进行过滤，同时采用 RGB 相机的图像作为 DVS 相机的补充，因为传统 RGB 相机的帧包含了绝对信息。他们的方法如图 2-3 所示。这种方法在质量上可以与[26]相媲美，同时在计算上更加高效。虽然这些方法目前定义了最先进的技术，但是它们都受到直接事件集成事件固有的影响，也就是边缘模糊，这是由于对比度阈值(像素为触发事件而发生的最小亮度变化)在整个图像平面上既不是恒定的，也不是均匀的，所以重构的边缘会模糊。此外，事件的纯集成在原则上只能恢复强度到未知的初始图像强度，这将导致在重建序列中初始图像的踪迹仍然可见的幽灵效应如图 2-4 所示。

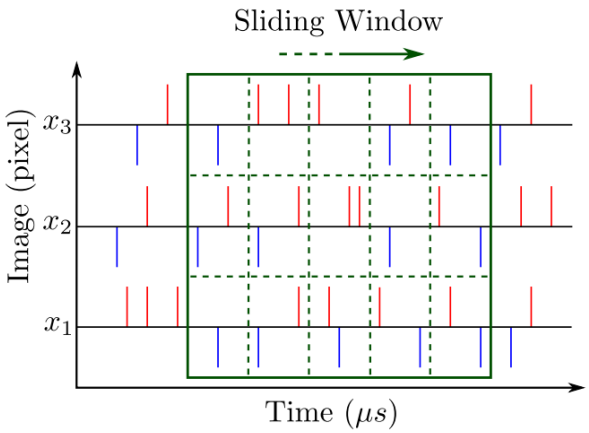

图 2-3 时间划窗[2]

第三类模型。Barua 等[29]首次提出了一种基于学习的方法来重建事件的强度图像。他们在模拟数据上使用 K-SVD[27]来学习将集成事件的块映射到图像梯度的字典，并使用泊松积分重建图像。[28]中使用生成式对抗网络从事件中生成真实的图像，提出了一种基于深度学习的框架综合方法，该方法由一个对抗结构和一个递归模块组成，利用事件相机的输出流来合成 RGB 帧的框架。Rebecq 等[3]表明，大量的模拟数据可以用来训练一个端到端的

全卷积循环神经网络 RCNN，从事件中重建高速、高动态范围的视频，如图 2-4 中 e 列所示。

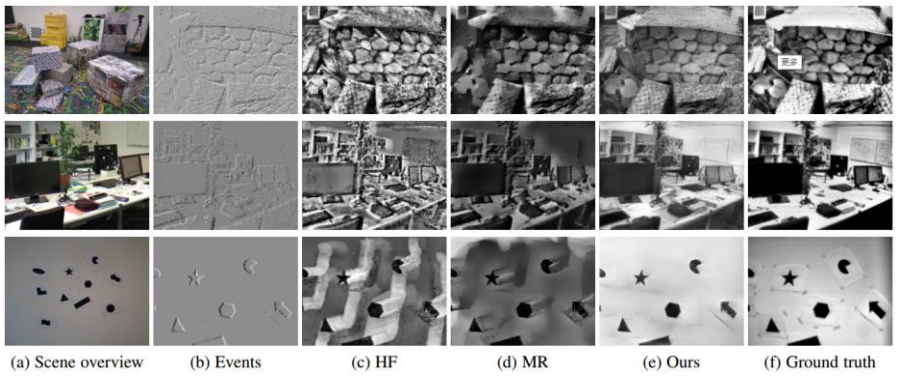

图 2-5 重建效果[3]

从图 2-4 中可以看出。采用机器学习的方法从事件流重建强度图像，比前两种类型依赖于任何手工制作的先验的效果更好。DVS 数据模拟器生成的大量标签数据集为机器学习训练提供了充足的数据。

2.3 彩色帧重构

事件相机大多是单色的，根据亮度的变化来生成事件，并丢弃颜色信息。随着一种可以感知颜色事件的传感器 Color-DAVIS346[30]的引入，这种情况发生了改变。彩色 davis346 由一个配备了彩色阵列滤波器(CFA)的 CMOS 芯片组成，形成一个 RGBG 滤波器图案。CFA 中的像素对其特定颜色过滤器的变化非常敏感，从而产生编码颜色信息的事件。

[31]分别使用了图像重建的方法 MR 或 GAN 重建了 RGBG 的四个单独的颜色通道，并通过四种通道重建出彩色图像，但是分辨率只有传感器的四分之一。

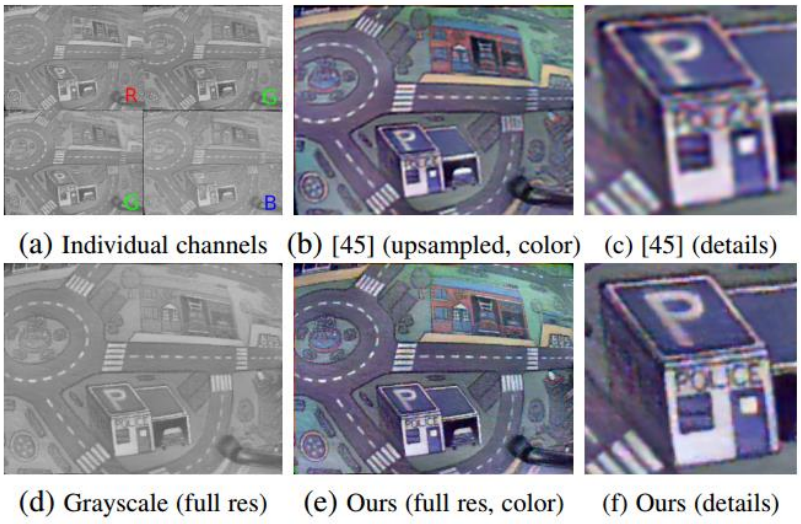

图 2-5 彩色帧重建[2]

[2]利用视觉系统对颜色差异的敏感度低于对亮度的敏感度提出了一种简单的方法增加分辨率。使用了同[31]的方法得到四个颜色通道，用双三次插值进行上采样，并将它们重新组合成一个低质量的彩色图像(2-5 (b))。然后，我们将这个彩色图像与没有彩色阵列滤波器得到的全分辨率灰度图结合，将的彩色图像投影到 LAB 色域中，用高质量的灰度重建代替亮度通道(2-5(d))。

3. 研究内容

3.1 研究目标

1) 提出基于动态视觉传感器的高速图像帧重建模型，能够有效地对静态和动态的高速运动进行图像重建。目标是所提出的模型能够清晰地，以高帧率地方式给出运动中的图像，解决传统相机在高速和高动态范围场景下表现不佳的问题。

2) 在高速图像帧重建的基础上使用最新的彩色动态视觉传感器重构出更加真实的彩色图像。

3) 优化高速图像帧重建模型，能够实时处理高速运动中的图像重建，并应用到场景重建 SLAM 中，提高重建的速度精度以及挑战环境下的场景重建。

3.2 主要研究内容及拟解决的关键科学问题和技术问题

3.2.1 主要研究内容：

在了解和掌握了国内外研究现状和发展动态，确定研究方向之后，本课题的主要工作包括以下方面：

1) DVS 数据集的制作。DVS 是无帧传感器，输出的是受光强变换产生的事件序列，传统相机在 DVS 相机性能优势的高速和高动态范围的环境下表现很差，很难制作出标签数据集用于训练。使用模拟器能够产生出干净准确的标签数据集。
2) DVS 数据的高速帧重建。本课题使用改进的循环神经网络能够保留之前的状态，通过大量标签数据训练神经网络能够提升 DVS 数据在高速条件下静态场景和动态场景的重建效果。
3) 在高速帧重建的基础上加入彩色 DVS 事件，重建彩色图像。
4) 优化重建模型。对于高速运动的控制场景，更加要求模型能够实时处理。

3.2.2 拟解决的关键科学问题：

1) 短时间内处理大量事件的网络模型训练。首先是训练数据集的构建，构建出的数据集能够有效训练出网络模型，训练出的模型能够有效地重建图像帧；
2) 彩色事件重建的低精度帧与全分辨率帧的融合。

3.2.3 拟解决的关键技术问题：

1) DVS 数据集使用模拟器构建；
2) 基于循环神经网络的模型使用 pytorch 训练；
3) 彩色事件帧与全分辨率时间帧的融合。

3.3 拟采取的研究方法、技术路线、实施方案及可行性分析

3.3.1 技术路线：

首先针对应用场景的不同，构建出用于训练的模拟器数据集以及真实的验证数据集。然后构建基于循环神经网路的高速帧重建模型，利用标签数据训练重建模型。然后结合彩色 DVS 事件，构建出彩色图像帧。最后，优化重建帧模型，能够实时处理。

四个研究内容之间的关系如图 3-1 所示。

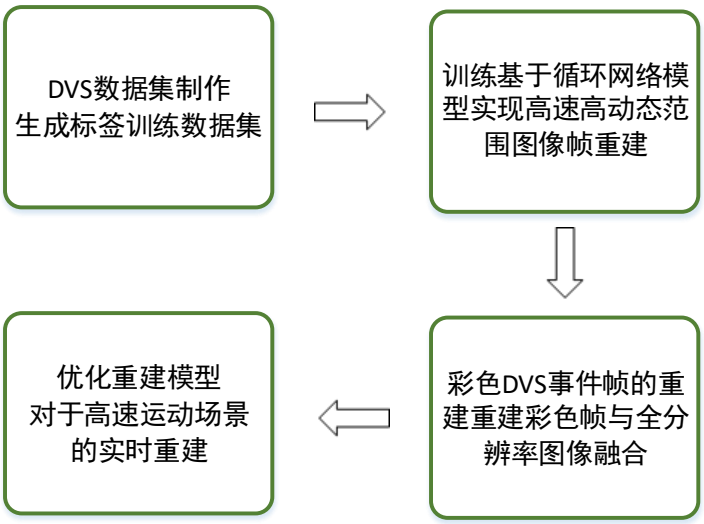

图 3-1 技术路线图

3.3.2 实施方案

1) DVS 数据集制作

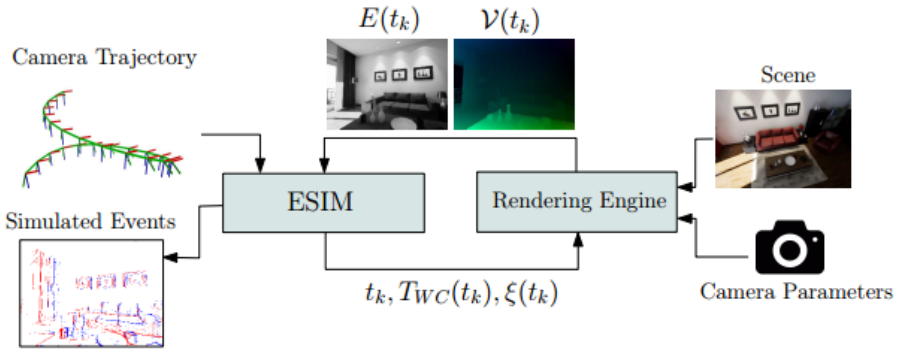

图 3-2 DVS 数据集制作示意图[6]

ESIM[6]是一个开源的事件相机模拟器，可以产生大量的有真实帧对应的事件数据。它采用了自适应的渲染方法，而不是传统的固定频率的渲染频率。制作数据集的关键是渲染引擎和事件模拟器之间的紧密耦合，即只在必要时渲染帧。

如图 3-2，在每一时刻 t_k，ESIM 会从用户设定的轨迹中获得一个新的相机姿态 $T_{wc}(t_k)$ 和场景的信息 sigma，并将它们传输给渲染引擎，渲染引擎会渲染出一个强度图 E(tk)和一个动态领域 $V(t_k)$，其中后者是用来计算亮度变化的，由于采用了动态调整的渲染策略，要模拟的数据量与场景中的运动量成正比。

动态调整策略包括两种，基于亮度变化和基于像素位移。前者的渲染帧率与图像的最大预期绝对亮度变化成比例，后者是一种更简单的策略，确保两个连续渲染帧之间像素的最大位移是有界的，这可以通过选择下一个采样时间 t_{k+1} 来实现。

2)神经网络模型训练

我们的目标是将一个连续的事件流转换成一个帧的序列。为了实现我们的目标，我们需要将事件流切割成时间和空间不重叠的时空窗口，每一个时间窗口都包含一定数量的。采用一种类似 Unet[32]的循环卷积神经网络网络，记录每一个中间状态 s_k，对于每一个新的时间序列，可以基于上一个状态 s_{k-1} 和新输入的时间序列生成一张新的帧。使用模拟器生成的表示 DVS 数据集进行训练。

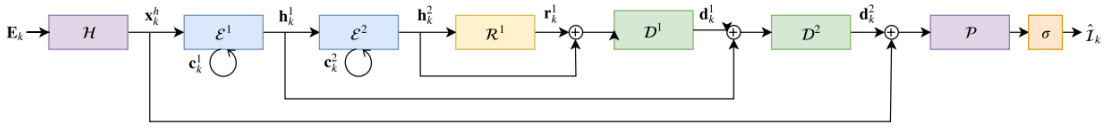

图 3-3 神经网络结构示意图[2]

为了处理使用循环卷积网络处理输入事件流，需要将包含事件的时空窗口转换成固定尺寸大小的张量表示 E_k 作为网络的输入。本课题使用体素网络[33]表示，每个时空序列被分成 B 个时间戳，张量的表示方法如公式(1)。t_i^*是归一化的时间戳。

$$\mathbf{E}(x_l, y_m, t_n) = \sum_{\substack{x_i=x_l \\ y_i=y_m}} p_i \max(0, 1 - |t_n - t_i^*|), \qquad (1)$$

图 3-3 是模型神经网络的结构示意图。从左至右分别是头层，紧接着是两个循环编码器，一个残差块，两个解码层和一个最终图像预测层。[34]中将对称的编码和解码层连接在一起。头层使用了卷积加 ReLU，编码层中包括一个步长为 2 的下采样以及一个 ConvLSTM[35]，它是 LSTM 的变体，主要是权重的计算变成了卷积运算，这样可以更好地提取出图像的特征。LSTM 计算单元的权值共享，每层 LSTM 都共享一份权值。中间加入残差快可以防止因为网络深度增加带来的梯度弥散和网络退化问题。解码块包括一个上采样的卷积块加 ReLU。

$$\mathcal{L} = \sum_{k=0}^{L} \mathcal{L}_k^R + \lambda_T C \sum_{k=L_0}^{L} \mathcal{L}_k^{TC}, \qquad (2)$$

为了更加能够反映重建的效果，使用了组合 loss 函数，一种描述重建的图像与真实图像的差，另一种描述连续重建图像之间的 loss。用 Perceptual Similarity(LPIPS)[36]来反映前者，真实的照片选择的是序列中最后一个事件时间戳时的。如公式(2)所示，loss 函数是连续 L 帧图像的重建损失和时间连续损失和总和。

3)彩色图像重建

color - davis346[30]引入了可以感知颜色事件的传感器。通过彩色阵列滤波器(CFA)，可以输出 RGBG 通道。CFA 中的像素对其特定颜色过滤器的变化非常敏感，从而产生编码颜色信息的事件。RGBG 也称为拜尔滤波器，滤镜中 50%为绿色，25%为红色，25%为蓝色，Bayer 首先发现人眼在红蓝绿三种中对绿色敏感性最高，第一个提出一红一蓝两绿的像素排列方式来模拟人眼对自然界的颜色感知。图 3-6 是拜尔滤镜的示意图。

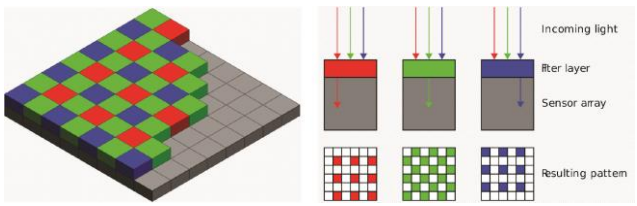

图 3-6 拜尔滤波器[30]

首先可以使用了同[31]的方法，将得到四个颜色通道用双三次插值进行上采样，并将它们重新组合成一个低质量的彩色图像。为了将这个彩色图像的分辨率提高，可以利用人眼对于亮度的敏感性比颜色更加高的特性，将低分辨率彩色帧投影到其他色域如 LAB 或者 YUV，LAB 是颜色-对立空间，L 表示亮度，a,b 表示颜色对立维度。使用同样的神经网络但是使用普通事件代替彩色事件可以得到一个全分辨率的灰度帧，再将灰度代替 LAB 色域中的亮度，保留颜色信息在 a,b 通道，从而可以得到视觉上全分辨率的彩色图像。

4)优化重建模型

虽然循环卷积神经网络可以较好的重建出高速和高动态范围运动中的图像。但是对于具有实时控制要求的场景，对于实时性的要求更高。由于高速运动的过程中会产生大量的事件，如果不能及时处理事件并对运动控制如高速无人机做出反馈，就会失去 DVS 数据在这种场景下的应用。

[2]的神经网络有 10M 参数，大型的神经网络虽然能够更好地重建出图像帧，但是在现代 GPU 上需要 30ms 才能实现 640*480 分辨率的重建。需要在能够保证精度的情况下优化网络模型，降低网络延迟。如针对之前提出的完全循环卷积神经网络，依赖于循环连接构建一个随着时间变化的状态，采用循环编码与解码的方式增大了参数量。如果使用单步长的卷积而不进行下采样，每一个卷积后接一个循环单元，其中循环单元可以使用 GRU 替代 LSTM，因为 GRU 相对于 LSTM 由于只有两个门而不是三个门，参数量更小，输出层可以采用一个 1*1 的单通道卷积，这样每一个事件都可以产生一个图像。

3.3.3 可行性分析

1)研究目标明确，思路清晰

基于事件的计算机视觉的最新研究成果和进展表明本课题的研究内容和关键技术是前沿关键问题。课题主要研究基于动态视觉传感器的高速图像帧重建技术，提出使用循环卷积神经网络解决高速动态场景中图像重建问题，并结合最新的动态视觉传感器重建出了彩色的图像。优化神经网络模型，降低神经网络的规模和延迟以满足重构模型的实时性。

研究目标明确、重点突出。DVS 视觉是一个新兴的研究领域，生态圈正在逐步扩大，新的研究成果不断在国际顶级会议收到认可，有很多可以学习借鉴的思路和方法。研究 DVS 数据进行帧重建的工作不仅仅计算机视觉中的一个新的方向，高质量的重建帧可以将传统的成熟的计算机视觉技术应用到新的数据集上。本课题以需求为牵引，以技术为推动，有目前最新的几款动态视觉传感器，通过实际软件模拟框架来验证科研成果，所采取的技术路线及方案是可行有效的。

2)国际国内学术交流环境良好

本人所在的课题组与多家国际国内科研机构保持有长期稳固的合作研究，包括瑞士苏黎世大学与苏黎世联邦理工大学神经信息研究所(简称 INI)、北京大学、清华大学、浙江大学、中科院、军事科学院的相关团队有良好的合作关系。指导老师王蕾曾在美国宾夕法尼亚州立大学访问一年，而且目前正有三名博士研究生在 INI 访问。INI 是 DVS 相机 DAVIS128 的提出者，目前正在积极拓展和维护动态视觉传感器的生态圈。这些都有利于在课题实施过程中同国际国内的先进研究机构不断交流，把握本领域国际学术前沿，提升研究水平。

3.3.4 预期创新点

1)已有的基于动态视觉传感器的帧重建方法只能在静态或者低速场景使用，本课题提出使用机器学习的方法在高速和高动态范围的场景下重建图像帧。

2)采用彩色动态视觉传感器在高速和高动态范围场景下重建出彩色图像帧。

3)通过压缩神经网络规模，降低神经网络规模和延迟，重建模型能够实时处理高速场景下的图像重建。

4. 研究条件

开展研究应具备的条件及已具备的条件，可能遇到的困难与问题和解决措施。

1) 已具备的条件：

- 高性能的 CPU 和 GPU 服务器；
- DAVIS346, Celex-IV 的动态视觉传感器；
- 相关领域的论文、书籍和其他资料；
- 指导老师有相关领域的研究、实践基础。

2) 可能遇到的困难与问题：

- DVS 数据研究是一个新的研究领域，刚刚接触，相关资料较少；
- 利用事件进行图像重建是一个跨领域的主题，包括计算机视觉、神经形态计算；
- 从数据采集到最终表现需要用到多种实验环境和代码，具有挑战；

3) 解决措施：

- 阅读 DVS 图像重建最新的研究成果，从传统的图像重建领域借鉴方法；
- 强化交流，多向老师请教学习，主动向国内外高水平研究团队交流学习；
- 强化对于数据处理，机器学习的学习，总结梳理，做好对比实验。

5. 学位论文工作计划

起讫日期	主要完成研究内容	预期成果
2020 年 3 月 1 日 —— 2020 年 4 月 1 日	DVS 数据集的制作，神经网络的训练以及高速动态图像帧重建。	掌握神经网络进行高速图像重建的方法，发表论文一篇
2020 年 4 月 1 日 —— 2020 年 5 月 1 日	彩色 DVS 数据重建彩色帧。	结合最新的彩色事件相机在原有基础上重建彩色图像
2020 年 5 月 1 日 —— 2020 年 6 月 1 日	神经网络优化压缩。	压缩网络模型大小，降低网络延迟，提高网络实时性，发表论文一篇
2020 年 6 月 1 日 —— 2020 年 9 月 1 日	探索高速动态图像重建应用。	将高速图像重建技术应用到高速无人机、SLAM 等应用场景，发表论文一篇
2020 年 9 月 1 日 —— 2020 年 10 月 1 日	撰写毕业论文。	毕业论文
2020 年 10 月 1 日 —— 2020 年 11 月 1 日	论文预审、论文修改。	准备答辩
2020 年 11 月 1 日 —— 2020 年 12 月 1 日	准备论文答辩。	完成答辩

注：每个子阶段不得超过 3 个月；预期成果中必须包含成果的形式、数量、质量等可考性指标该计划将作为论文研究进展检查的依据。

6. 主要参考文献(博士不少于 50 篇、外文不少于 25 篇，硕士不少于 30 篇、外文不少于 15 篇，可附页)

序号	文献目录(作者、题目、刊物名、出版时间、页次)
1	Gallego G, Delbruck T, Orchard G, et al. Event-based vision: A survey[J]. arXiv preprint arXiv:1904.08405, 2019.
2	Rebecq H, Ranftl R, Koltun V, et al. High speed and high dynamic range video with an event camera[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
3	Rebecq H, Ranftl R, Koltun V, et al. Events-to-video: Bringing modern computer vision to event cameras[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3857-3866.
4	Cedric Scheerlinck , Henri Rebecq , Davide Scaramuzza IEEE Winter Conf. Applications of Computer Vision (WACV), 2020.
5	Scheerlinck C, Barnes N, Mahony R. Continuous-time intensity estimation using event cameras[C]//Asian Conference on Computer Vision. Springer, Cham, 2018: 308-324.
6	Rebecq H, Gehrig D, Scaramuzza D. ESIM: an open event camera simulator[C]//Conference on Robot Learning. 2018: 969-982.
7	Kim H, Handa A, Benosman R, et al. Simultaneous mosaicing and tracking with an event camera[J]. J. Solid State Circ, 2008, 43: 566-576.
8	P. Lichtsteiner, C. Posch, and T. Delbruck, “A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor,” IEEE J. Solid-State Circuits, vol. 43, no. 2, pp. 566–576, 2008.
9	Hanme Kim, Stefan Leutenegger, and Andrew J. Davison. Real-time 3D reconstruction and 6-DoF tracking with an event camera. In Eur. Conf. Comput. Vis. (ECCV), 2016.
10	Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization. In British Machine Vis. Conf. (BMVC), Sept. 2017.
11	Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis. Event-based visual inertial odometry. In IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017.

序号	文献目录(作者、题目、刊物名、出版时间、页次)
12	Xavier Lagorce, Garrick Orchard, Francesco Gallupi, Bertram E. Shi, and Ryad Benosman. HOTS: A hierarchy of event-based time-surfaces for pattern recognition. IEEE Trans. Pattern Anal. Machine Intell., 39(7):1346－1359, 2017.
13	Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza, Jeff Kusnitz, Michael Debole, Steve Esser, Tobi Delbruck, Myron Flickner, and Dharmendra Modha. A low power, fully event-based gesture recognition system. In IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017.
14	Henri Rebecq, Timo Horstschafer, Guillermo Gallego, and Davide Scaramuzza. EVO: A geometric approach to eventbased 6-DOF parallel tracking and mapping in real-time. IEEE Robot. Autom. Lett., 2:593－600, 2017.
15	M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, “In-21teracting maps for fast visual interpretation,” in Int. Joint Conf. Neural Netw. (IJCNN), 2011, pp. 770－776.
16	K. A. Boahen, “A burst-mode word-serial address-event link-I: Transmitter design,” IEEE Trans. Circuits Syst. I, vol. 51, no. 7, pp. 1269－1280, Jul. 2004.
17	A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. D. Nolfo, T. Nayak, A. Andreopoulos, G. Garreau, M. Mendoza, J. Kusnitz, M. Debole, S. Esser, T. Delbruck, M. Flickner, and D. Modha,“A low power, fully event-based gesture recognition system,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017, pp. 7388－7397
18	A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. CARLA: An open urban driving simulator. In Proceedings of the 1st Annual Conference on Robot Learning, pages 1－16, 2017
19	S. Shah, D. Dey, C. Lovett, and A. Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In Field and Service Robotics, 2017. URL https://arxiv.org/abs/1705.05065.
20	Q. Weichao, Z. Fangwei, Z. Yi, Q. Siyuan, X. Zihao, S. K. Tae, W. Yizhou, and Y. Alan. Unrealcv: Virtual worlds for computer vision. ACM Multimedia Open Source Software Competition, 2017.

序号	文献目录(作者、题目、刊物名、出版时间、页次)
21	J. Kaiser, T. J. C. V., C. Hubschneider, P. Wolf, M. Weber, M. Hoff, A. Friedrich, K. Wojtasik, A. Roennau, R. Kohlhaas, R. Dillmann, and J. Zollner. “Towards a framework for end-to- end control of a simulated vehicle with spiking neural networks. In 2016 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR),pages 127 – 134, Dec. 2016
22	W. Li, S. Saeedi, J. McCormac, R. Clark, D. Tzoumanikas, Q. Ye, Y. Huang, R. Tang, and S. Leutenegger. Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset. In British Machine Vis. Conf. (BMVC), page 77, Sept. 2018.
23	D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, “Asynchronous, photometric feature tracking using events and frames,” in Eur. Conf. Comput. Vis. (ECCV), 2018.
24	H. Kim, S. Leutenegger, and A. J. Davison, “Real-time 3D reconstruction and 6-DoF tracking with an event camera,” in Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 349 – 364.
25	P. Bardow, A. J. Davison, and S. Leutenegger, “Simultaneous optical flow and intensity estimation from an event camera,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2016, pp. 884 – 892
26	G. Munda, C. Reinbacher, and T. Pock, “Real-time intensity-image reconstruction for event cameras using manifold regularisation,” Int. J. Comput. Vis., vol. 126, no. 12, pp. 1381 – 1393, Jul. 2018.
27	M. Aharon, M. Elad, and A. M. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311 – 4322, 2006
28	S. Pini, G. Borghi, R. Vezzani, R. C. U. of Modena, and R. Emilia. Learn to see by events: Color frame synthesis from event and RGB cameras. Int. Joint Conf. Comput. Vis.,Image and Comput. Graph. Theory and Appl., 2020.
29	S. Barua, Y. Miyatani, and A. Veeraraghavan, “Direct face detection and video reconstruction from event cameras,” in IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2016, pp. 1 – 9.
30	G. Taverni, D. P. Moeys, C. Li, C. Cavaco, V. Motsnyi, D. S. S. Bello, and T. Delbruck, “Front and back illuminated Dynamic and Active Pixel Vision Sensors comparison,” IEEE Trans. Circuits Syst. II, vol. 65, no. 5,pp. 677 – 681, 2018.

序号	文献目录(作者、题目、刊物名、出版时间、页次)
31	C. Scheerlinck, H. Rebecq, T. Stoffregen, N. Barnes, R. Mahony, and D. Scaramuzza, “CED: color event camera dataset,” in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2019.
32	O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015
33	A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, “Unsupervised eventbased optical flow using motion compensation,” in Eur. Conf. Comput. Vis. Workshops (ECCVW), 2018
34	A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, “EV-FlowNet: Selfsupervised optical flow estimation for event-based cameras,” in Robotics: Science and Systems (RSS), 2018.
35	X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” in Conf. Neural Inf. Process. Syst. (NIPS), 2015
36	R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2018.

7. 指导教师对开题报告的评语

（对 1-6 项逐项予以评价，并着重对国内/外研究现状的了解情况、研究内容的创新性等方面进行评价，最终给出是否满足博士/硕士层次学位论文研究要求的综合评价意见）

导师签名：

年　　月　　日