

---

学号： 2015301500162

密级：

# 武汉大学本科毕业论文

## 基于深度学习的 视频阴影消除

院（系）名称： 计算机学院

专业名称： 计算机科学与技术

学生姓名： 李文洲

指导教师： 肖春霞 教授

二〇一九年五月

---

# 郑重声明

本人呈交的学位论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名：\_\_\_\_\_ 日期：\_\_\_\_\_

---

## 摘 要

在互联网日益蓬勃的今天，视频渐渐成为人们生活中不可或缺的一部分，对视频进行加工处理的重要性也日益显现。阴影作为自然界中常见的光学现象，十分容易被记录到图像、视频当中。而阴影区域亮度低的特点可能对图像和视频的进一步分析处理带来干扰，影响分析结果的准确性。视频阴影消除旨在准确检测视频中所包含的阴影区域，并且针对性地进行消除，将其恢复成正常光照条件下的区域，并保持视频帧间的连续性。视频阴影消除在计算机视觉和图形学的研究中有重要意义。

视频的阴影消除涵盖了许多亟待解决的问题。首先，视频的拍摄可能是由静态摄像机所拍摄或动态摄像机所拍摄，这导致了场景本身可能包含平移和旋转等复杂运动，给阴影消除后视频的时空一致性带来了巨大考验。其次，阴影的组成可能是多区域的，同一帧图像内可能有多个不同构成的阴影，按照传统处理方法，这需要区分场景，并分别检测出阴影区域，并用不同的光照还原参数分别进行还原。最后，因为阴影边界处拥有较大的光照变化，在边界处理上有严重丢失细节、光照还原出错等问题。因而如何处理阴影边界也是一个需要解决的问题。

针对上述问题，本文提出了基于深度学习和阴影关键帧的视频阴影消除方法。首先提取出视频的关键帧，利用深度学习将其进行图像阴影消除，并提取图像纹理特征进行阴影区域边界修复，然后运用原有视频的运动矢量信息将其他帧做相应处理，使其他帧依据阴影关键帧的信息进行对应的阴影消除，得到保留时间域信息的阴影消除后视频。

与其他方法相比，本文方法的贡献有：

1) 提出了一个新的图像阴影消除模型，该模型能有效恢复阴影区域的光照信息及纹理细节，并解决阴影消除后可能出现的阴影边界问题及阴影区域与非阴影区域之间的亮度不一致问题。

2) 提出了一个基于阴影关键帧的视频阴影消除算法。该算法可以处理动态摄像机拍摄的视频，并且解决了视频阴影消除后可能出现的帧间不连续及抖动问题，获得较为真实自然的无阴影结果。最为关键的，本文极大削减了所需计算资源，仅需较少的计算资源和时间即可获得和以往极长耗时方法得到近似的结果。

**关键词：**视频阴影消除；深度学习；阴影关键帧；时空一致性；边界修复

---

## ABSTRACT

As the Internet has been developing rapidly, video has gradually become an indispensable part of people's lives, and the importance of video processing has become increasingly apparent. Shadow, as a common optical phenomenon in nature, is very easy to be photographed and recorded in images and videos. The images and videos containing shadows are likely to hinder the further analysis and processing of shadow images. Video shadow elimination aims at accurately detecting the shadow areas contained in the video, and removing them pertinently, restoring them to the areas under normal lighting conditions, and keeping each frame containing continuous time information. Video shadow removal plays an important role in the later stage of vision and graphics research.

There are many related problems behind shadow removal in video. Firstly, the video may be shot by a static camera or a dynamic camera, which results in the complex motion of the scene itself, such as translation and rotation, which poses a great challenge to the temporal and spatial consistency of the video after shadow elimination. Secondly, the composition of shadows may be multi-region. There may be many different shadows in the same frame. According to the traditional processing method, it needs to be partitioned into scenes to recognize, then marked one by one, and restored separately with different illumination restoration parameters. Finally, because of the large illumination change at the shadow boundary, it may lead to the serious loss of details and illumination restoration errors in the traditional processing methods, so the smooth transition of the boundary is also a problem.

To solve these problems, this paper proposes a video shadow elimination method based on in-depth learning and shadow key frames. Firstly, we extract the key frames of the video, use the depth learning to remove the shadow of the image, and extract the texture features of the image to repair the boundary of the shadow area. Then we use the motion vector information of the original video to process the other frames accordingly, so that other frames can eliminate the shadow according to the information of the key frames of the shadow, and get the video after the shadow elimination which retains the information of the time domain.

---

Compared with other methods, the main advantages and contributions of our method are as follows:

1) A model is proposed to effectively restore the illumination and texture details of shaded areas, and to solve the problem of the possible shadow boundary and the inconsistency of brightness and color between non-shaded areas and shaded areas after shadow elimination.

2) A video shadow elimination algorithm based on shadow key frames is proposed. This algorithm can process video captured by dynamic camera, and solve the problem of discontinuity and jitter between frames after video shadow elimination, so as to obtain real and natural shadowless results.

**Key words:** Video Shadow Removal; Shadow Key Frame; Deep Learning; Spatio-Temporal Coherence; Boundary Rehabilitation

---

# 目 录

1	引言 .....	1
1.1	研究背景与意义 .....	1
1.2	研究现状 .....	2
1.2.1	图像阴影消除 .....	2
1.2.2	视频阴影检测 .....	3
1.2.1	视频阴影消除 .....	4
1.3	研究目的与贡献 .....	6
1.4	本文结构安排 .....	7
2	相关工作及技术介绍 .....	8
2.1	阴影相关背景 .....	8
2.1.1	阴影形成模型 .....	8
2.1.2	视频中阴影组成分析 .....	9
2.2	阴影消除相关技术 .....	10
2.2.1	基于关键帧的视频处理 .....	10
2.2.2	图像形态学处理 .....	13
2.2.3	使用 VGGNet 提取特征 .....	14
2.3	基于关键帧的视频阴影消除模型 .....	16
2.4	本章小结 .....	18
3	图像阴影消除 .....	19
3.1	深度学习图像阴影消除 .....	19
3.2	细节修正 .....	22
3.3	本章小结 .....	24
4	视频阴影消除 .....	25
4.1	视频阴影关键帧的提取 .....	25
4.2	视频的阴影消除 .....	27
4.3	进一步优化减弱可能的抖动 .....	29
4.4	实验算法、结果与分析 .....	29
4.5	本章小结 .....	34
5	总结 .....	35
5.1	全文总结 .....	35
5.2	算法局限 .....	35
5.3	未来工作 .....	35
	参考文献 .....	36
	致谢 .....	39

---

# 1 引言

## 1.1 研究背景与意义

在互联网技术飞速发展的当下，多媒体技术应用也日渐成熟，逐步走入千家万户。从随处可见的夺目视频广告到如今风靡全国的短视频；从我们熟知的传统电影、电视到现在日益崛起的自媒体 vlog（视频博客）；从遥远的太空卫星扫描到现在遍布大街小巷你我身边的监控摄像，视频拉近了人们之间的距离。视频作为可视化信息的重要承载媒体，也因其内容的多变、生活性渐渐成为人们生活中不可或缺的一部分。人类获取信息的方式早已从简单的静态图片与文字到现在能承载巨量信息的视频。同时，随着人们对视频的需求与日俱增，对视频进行相关处理分析的重要性也日益显现。

阴影是自然界中常见的光学现象，因而也大量存在于摄录的影视素材之中。阴影是由光源被物体部分或全部遮挡，然后被一个或多个物体反射投影到场景表面形成的。阴影在对图像视频的处理分析中有利也有弊。阴影的好处如：通过阴影，场景中的多个参数信息，如光照方向、光源个数、场景中的物体形状等等可以被较为准确地估算出来。除此之外，在非纯自然场景，即人工合成场景中，合理的阴影添加也能让场景渲染的画面看起来更加逼真。然而，阴影带来的消极影响也是非常棘手的。首先，不合理的光影往往直接给视觉带来不真实感，给人一种画面显假的感觉。其次，进行检测与识别目标时，阴影往往能和对象本体拥有较为一致的运动特征，这会对检测与识别造成一定程度的困扰。而在视频图像的本征分解中，阴影造成的纹理信息丢失也可能带来不利影响。综上所述，在对视频进行进一步处理之前进行阴影的检测和消除有着极为重要的作用。

视频的阴影检测指的是针对输入的视频，检测其中的阴影区域。视频的阴影消除指的是在保证视频阴影区域原有纹理结构的基础上，恢复阴影区域在正常光照场景下时的成像。传统的视频阴影消除要求视频的阴影检测作为前提和必要步骤，其检测结果的好坏将直接影响到消除算法的性能。视频中的阴影形态不一，并且可能存在运动的阴影，这为视频阴影的检测增加了难度。其次，摄像机可能存在平移和旋转等复杂运动，传统的基于背景建模的方法只能处理由静止摄像机所拍摄的视频，对于运动摄像机拍摄的视频往往无法进行有效的阴影消除。此外，

许多处理运动摄像机阴影视频的文献在面对阴影消除后视频时空一致性的保持问题时也不得不留下诸多未解决的困难。而因为人眼对时间维度上的不一致更为敏感，导致逐帧阴影消除对时间信息的丢失会加剧视觉上的不和谐，所以这一悬而未决的问题亟待解决。最后，如何解决由于阴影边界处光照变化引起的阴影消除后阴影边界失真情况仍然是一个十分具有挑战性的问题。

## 1.2 研究现状

目前，对视频阴影消除的研究较少，大多数的工作处理的都是由一个静止不动的摄像机拍摄的固定场景，其对视频的每一帧进行逐帧转换为图像，然后再进行图像阴影消除，最后将图片序列重新合成视频。在这样的处理中，时间信息的丢失导致了时空一致性无法保持，使得帧内容不连续，无法得到良好的视觉效果。最后，将一个视频所有帧转为图片序列，然后逐帧独立处理再拼接也是一件极为耗时耗资源的工作，这无疑加大了算法的时间复杂度。

### 1.2.1 图像阴影消除

在图像阴影消除上有众多前辈做了相关研究。一般而言，可以将前人的方法大致分成三类：基于梯度域的方法、基于光照迁移的方法和基于阴影抠图的方法。

基于梯度域的方法在梯度域上对图像进行阴影消除。Finlayson et al. (2002) 和 Finlayson et al. (2006) 提出了一系列利用梯度域处理的阴影消除方法<sup>[1][2]</sup>。其利用所提出的光照无关图和 Canny 边缘检测算子来计算初始阴影图像中的阴影区域。然后，基于如上阴影检测的结果，将阴影边界处像素的梯度值直接设为零值，以得到阴影消除后的图像。但是，该方法阴影消除的结果极易被边缘检测的精度所影响。除此之外，该方法前期准备过程中，需要对摄像机进行标定。因而，这种方法只能应用于较为简单的场景。

基于光照迁移的方法，计算图像中阴影区域和非阴影区域之间的线性映射关系。Reinhard et al. (2001) 采用近似颜色迁移的方法，将图像中非阴影区域的光照信息迁移至阴影区域，以达到阴影消除<sup>[3]</sup>。Shor et al. (2008) 首次提出了阴影消除的线性仿射模型<sup>[4]</sup>。其利用全局光照迁移算子，通过全局的迁移，得到了较好的结果。但是由于全局使用相同参数设置，所以该方法仅仅适用于纹理单一的阴影图像。Xiao et al. (2013a)、Xiao et al. (2013b) 对上述方法进一步改进，两篇工作分别提出了参数自适应的多尺度图像阴影消除方法和基于子区域纹理匹配的阴影消



除算法<sup>[5][6]</sup>。这些方法将图像中材质反射率的不同也作为参考条件，通过自适应参数的方法和多尺度的方法都得到了较好的结果。Zhang et al. (2015) 提出了一种应用局部光照的恢复算子<sup>[7]</sup>。该方法首先把图像拆分成许多重叠的图像块 (pitch)，然后，对每个图像块分别使用局部的光照恢复算子。最后，再对整幅图像进行全局参考的阴影消除，得到阴影消除后的图像。

基于阴影抠图的方法将原始图像当成是阴影图像和无阴影图像的线性组合。通过估计所给阴影图像和无阴影图像的线性组合模型的相关系数，实现阴影消除。Chuang et al. (2003) 提出了一种在自然场景中提取阴影，然后把提取到的阴影融合入另一个场景的方法<sup>[8]</sup>。但是该方法主要关注阴影的提取与合成。Wu et al. (2007) 把阴影的生成看成一个光源衰减的问题，并以此原理构建了全新的阴影形成模型，然后，通过求解泊松方程，得到无阴影图像<sup>[9]</sup>。

### 1.2.2 视频阴影检测

传统的视频阴影消除方法含有如下两个子过程：阴影的检测和消除。对传统方法而言，视频阴影检测是阴影消除的先行过程和必要步骤。Prati et al. (2003)、Sanin et al. (2012) 以及 Russel et al. (2016) 分别根据不同的分类标准对现有的视频阴影检测算法做了详尽的总结与分析比较<sup>[10][11][12]</sup>。

Jung et al. (2009) 提出了一种基于背景建模的方法，对单色视频进行阴影检测<sup>[13]</sup>。该方法先使用鲁棒的统计描述子，将背景图像进行建模。然后基于统计学习，提取视频中的阴影信息。但是，单色视频在面对前后帧光照变化较大的时候，由于无法建立有效的背景模型，往往无法准确地检测出视频中的阴影。而且，该方法仅能处理静止摄像机拍摄的单色视频。

Porikli et al. (2005) 提出了一种递归式的方法检测视频中的阴影<sup>[14]</sup>。该方法使用多变量高斯将视频中的阴影实现建模，然后在量化的颜色空间中，更细致地定义阴影流函数，以检测出视频中的阴影。但是，面对光照变化较大时，无法得到满意的结果。而且，该方法仅用于处理监控视频。Xu et al. (2005) 针对在视频分割时阴影往往被错误归类为运动目标的问题，提出了基于边缘检测算法的视频阴影检测算法<sup>[15]</sup>。但同样的，这种方法也面临着只能用于处理静止摄像机拍摄的室内场景的问题。

Zhang et al. (2007) 提出了一种基于比率边缘图的运动投射阴影检测方法<sup>[16]</sup>。该方法利用比率边缘这一光照独立的特征，区分前景物体中的运动目标和阴影。

但是该方法同样受限于边缘检测的性能，且同样只能处理静止摄像机拍摄的视频场景。

除此之外，还有很多学者提出了一系列基于背景建模以及统计学习的方法，以实现视频中运动投射阴影的检测。如 Zhou et al. (2007), Martel-Brisson et al. (2007), Martel-Brisson et al. (2008) 和 Huang et al. (2009) 等<sup>[17][18][19][20]</sup>。Zhou et al. (2007) 利用视频像素层、局部区域层和全局层的信息，提出了一种基于高斯混合模型的视频阴影检测方法，但是在面对光照条件变化较大时，该方法可能失效<sup>[17]</sup>。Martel-Brisson et al. (2007 & 2008) 提出了一种基于高斯混合模型及核函数学习的视频阴影检测方法<sup>[19]</sup>。但由于其较为复杂的建模过程，使得其方法较慢，并且鲁棒性不佳。Huang et al. (2009) 提出了利用视频中的物理特征信息，加上高斯混合模型的方法，检测视频中的阴影<sup>[20]</sup>。但是，该方法与上述同样基于高斯混合模型的方法也有着类似缺陷。Benedick et al. (2008) 及 Huerta et al. (2009) 针对健康视频场景分别提出了对应的视频阴影检测<sup>[21][22]</sup>，Amato et al. (2014) 对这些方法做了详尽的总结分析<sup>[23]</sup>。

从以上分析可以看出，现有的阴影检测方法几乎都专注于处理静止摄像机拍摄的固定场景内的视频。大体上，都是先利用简单学习、高斯混合模型 (Gaussian Mixture Model) 等建立视频的背景模型。然后通过背景减除法 (Background Subtract) 等手段提取视频中运动的前景物体。最后，根据运动目标与阴影具有相似运动特征的特点，对提取的前景物体作进一步统计分析进而识别出视频中的阴影区域信息。但是，上述这一类方法或多或少都拥有诸多不足。1、大多只能应用于静止摄像机拍摄的固定场景视频，对运动摄像机所拍摄的运动视频不适用。2、需要建模并训练学习模型参数，处理过多任务时可能导致耗时较大。3、应对较大光线、照明变化或非静止背景时，因背景模型建立的可靠性降低而可能导致失效。

### 1.2.1 视频阴影消除

与图像阴影消除的工作相比，关于视频阴影消除的研究就少了许多。而且，关于视频阴影消除的研究大多都专注于消除静态摄像机拍摄的固定场景的视频。

Zhou et al. (2007) 使用视频像素层、局部区域层和全局层的信息进行视频的阴影检测和消除<sup>[17]</sup>。该方法侧重于使用高斯混合模型的方式，将视频中转化为二值图像，再进行阴影的检测与消除。类似的，Martel-Brisson et al. (2007) 使用一种全

新的基于像素统计学习的方法对视频中运动的投射阴影建模<sup>[18]</sup>。可以消除视频中运动的投射阴影。Huerta et al. (2009) 提出了一种基于梯度和彩色模型的视频阴影消除<sup>[22]</sup>。其可以处理非白色光源照射产生的有色阴影，但是同样受到剧烈光照改变的影响。

Jung et al. (2009) 提出一种基于背景减除技术的灰度视频阴影检测消除算法<sup>[13]</sup>。其首先对输入的视频序列进行背景建模。再通过背景减除法，找出含有阴影的运动前景物体。然后，对颜色特征加以统计分析，进而识别阴影区域。最后，通过不包含前景物体的背景模型，从中恢复出视频阴影区域的光照信息。

Zhou et al. (2012) 提出了一种基于马尔科夫随机场的视频阴影消除方法<sup>[24]</sup>。该方法依旧采用混合高斯模型的建模方法。通过视频的全局统计特征信息，建立马尔科夫随机场，从而消除视频中固定阴影。该方法同样仅仅适用于监控视频，且文中给出的消除结果也仅包含检测的二值图。Amato et al. (2011) 提出了一种利用图像局部颜色恒常性假设进行视频阴影消除的方法<sup>[25]</sup>。其更关注于视频阴影的建模与检测，但没有给出消除结果。

Sunkavalli et al. (2007) 通过研究视频中每一个像素点的强度值随时间变化的曲线，以识别阴影<sup>[26]</sup>。其工作将视频看做一个三维立方体，然后使用矩阵分解的方式消除其中的阴影。该方法主要适用于静止摄像机拍摄的延时视频，而且时空复杂度都较高。Dammavalam et al. (2009) 提出了主要针对室外场景的视频阴影消除方法<sup>[27]</sup>。但在光变较大时无法得到有效消除结果，且只能处理静态摄像机拍摄的视频。

Kar et al. (2015) 提出了一种基于 HSV 三维立方体颜色空间的视频阴影检测及消除算法<sup>[28]</sup>。其不足在于依然只适用于静态摄像机拍摄的视频，且无法较好处理较大光变场景。Li et al. (2016) 提出了针对交通监控视频的车辆投射阴影的消除算法，该方法同样采用背景减除类似思想<sup>[29]</sup>。

上述的视频阴影消除方法大都利用复杂的建模方法进行阴影的消除研究，且大都只适用于静态摄像机拍摄的固定场景的视频，对运动摄像机拍摄的视频无法较好地进行阴影消除，且面对大光变场景也无较好结果，因而存在一定局限性。

### 1.3 研究目的与贡献

目前，对视频阴影消除的研究比较少，虽有一些学者提出了相关算法并取得不错的成果，但是其大都只能处理静态摄像机拍摄的固定场景的视频，且不能较好应对大光变的场景，也更多侧重于视频阴影的检测，或多或少也因为时空复杂度的原因导致耗费大量时间和资源。

目前的方法大都对每一帧使用图像处理的方法进行视频阴影的消除，然后再重新合成视频。然而，人眼对于时间维度上的不一致性往往更加敏感，因而时空一致性的保持十分重要，单帧消除重组的方式无疑将丢失时间信息，使得帧内容看起来不连续，且视频包含大量的帧信息也对时间复杂度带来了不利影响。

对在图像（单帧）阴影消除基于深度学习的方法而言，本文参考了多个公开数据集进行深度学习训练，通过多数据集训练出一个较为广泛适用的深度学习阴影消除模型，通过三层结构的神经网络从不同角度提取不同信息，注重纹理细节地实现阴影消除。而在过完一遍深度学习网络实现阴影消除后，又在该阴影消除模型后再经过一层额外的基于 VGG 的纹理修复深度卷积神经网络。旨在专门处理容易出现修复瑕疵的阴影边界，使得阴影消除后恢复光照的区域能够得到全面的细节修正，优化阴影区域的细节及颜色，使得阴影区域与周围环境能够保持较好的一致性。

而在对视频阴影消除的处理中，首先提取视频的阴影关键帧，再利用图像阴影消除模型完成关键帧的阴影消除，并根据视频前后帧内容的关联性实现帧间时空一致性处理。相比其他视频阴影消除方法，本文只需较少的计算内容，可以更为便捷地完成整个视频的阴影消除任务。该算法可处理动态相机拍摄的视频。

与其他方法相比，本文方法的主要贡献有：

1) 提出了一个能有效的恢复阴影区域的光照及纹理细节的模型，解决阴影消除后可能出现的阴影边界及非阴影区域与阴影区域亮度与颜色不一致的问题。

2) 提出了一个基于阴影关键帧的视频阴影消除算法，该算法可以处理动态摄像机拍摄的视频，并且解决了视频阴影消除后可能出现的帧间不连续及抖动问题，获得真实自然的无阴影结果。更为重要的是，该方法极大削减了原本需要的计算资源，仅需较少的计算资源和计算时间即可获得和以往极长耗时方法得到近似的结果。

---

## 1.4 本文结构安排

本文内容主要由五个章节组成，各章主要内容如下：

第一章为本文引言，大致介绍本文研究的背景和意义。从图形阴影消除、视频阴影检测和视频阴影消除分别探究研究现状。

第二章为阴影消除的相关工作及相关技术介绍。包含阴影本身的相关背景知识，诸如图形色彩、阴影形成模型等。然后介绍本文将会用到的阴影消除有关的技术，诸如关键帧提取与处理、形态学处理、深度学习等。最后，介绍总技术路线。

第三章为本文图像阴影消除算法的介绍。本文提出一种基于深度学习的目标阴影消除及基于深度卷积神经网络 VGG-19 纹理不变层的阴影边界细节优化模型。

第四章为本文视频阴影消除流程介绍，通过第三章得到的深度学习图像阴影消除算法对视频中提取出的阴影关键帧进行图像阴影消除，再利用视频编码运动补偿的预测及还原得到包含时间信息的连续阴影消除图像序列，如有必要再进行抖动优化，最终合成出保持时空一致性的消除了阴影的视频。

第五章为全文的总结，总结本文提出的视频阴影消除算法，提出不足及可发展方向。

## 2 相关工作及技术介绍

本章介绍本文的研究内容及涉及到的相关工作及技术原理，介绍了阴影的相关背景知识，包括形成模型、组成分析，及视频阴影的检测和消除相关技术和算法，最后进行技术路线总结及本章小结。

### 2.1 阴影相关背景

#### 2.1.1 阴影形成模型

在自然界中，阴影是一种非常常见的光学现象，根据 Barrow et al. (1978) 提出的成像理论<sup>[30]</sup>，图像的形成过程可由如下公式 2-1 描述：

$$I_x(\lambda) = R_x(\lambda) \cdot L_x(\lambda) \quad (2-1)$$

其中， $I_x(\lambda)$  表示波长为  $\lambda$  的可见光照射于场景画面中  $x$  处反射光的强度，即图像在点  $x$  处的颜色值大小。 $R_x(\lambda)$  表示波长为  $\lambda$  的可见光在  $x$  处反射率的大小。 $L_x(\lambda)$  表示波长为  $\lambda$  的可见光在  $x$  处入射光的强度值。

若场景内只有单光源，且由于物体遮挡形成阴影区域，由可见光源发射出的可见光会在场景内多次反射，场景中的各个物体之间也会发生表面反射，所以场景中某点  $x$  处的入射光由两部分组成：(1) 入射光线直接从光源发出照射到  $x$  处，称之为直接光源  $L_x^d$ ；(2) 另一部分入射光线是由场景中其他表面经过多次反射照射到  $x$  处，称这部分入射光为环境光  $L_x^a$ 。

根据场景中物体表面的光照信息可将物体表面分为非阴影区域 (unshaded) 和阴影区域 (shaded)，其中非阴影区域的入射来源包括直接光源  $L_x^d$  和环境光  $L_x^a$ ，阴影区域则可进一步划分为本影区 (umbra) 和半影区 (pen-umbra)。场景物体阴影分区如图 2-1 物体所在阴影分区所示。

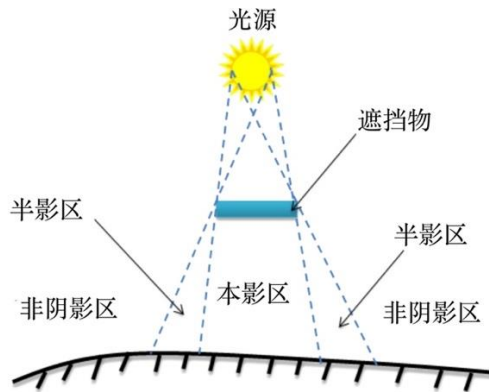


图 2-1 物体所在阴影分区

所以，本影区的入射光只有环境光  $L_x^a$ ，半影区的入射光同时拥有直接光源  $L_x^d$  和环境光  $L_x^a$ 。因而，场景中物体表面某点  $x$  处的入射光总强度值可由公式 2-2 描述：

$$L_x = \alpha L_x^d + L_x^a, \alpha \in [0, 1] \quad (2-2)$$

其中，当  $\alpha = 0$  时， $x$  位于本影区；当  $\alpha = 1$  时， $x$  位于非阴影区域；当  $0 < \alpha < 1$  时， $x$  位于半影区。

由公式 (2-1) 及公式 (2-2) 可得，场景中任一点  $x$  处的反射光强度值，也即图像中该点颜色值大小可以由如下公式描述：

$$I_x = R_x \cdot (\alpha L_x^d + L_x^a), \alpha \in [0, 1] \quad (2-3)$$

其中，当  $\alpha = 0$  时， $x$  位于本影区，只有环境光  $L_x^a$  照射；当  $\alpha = 1$  时， $x$  位于非阴影区域，入射光包括全部的直接光源  $L_x^d$  和环境光  $L_x^a$ ；当  $0 < \alpha < 1$  时， $x$  位于半影区，入射光包括部分直接光源即  $\alpha L_x^d$  和环境光  $L_x^a$ 。 $R_x$  表示可见光在  $x$  处反射率的大小。

可见，对图像中单个像素点而言，因为其环境光是不会变化的，所以单个点的亮度几乎完全由  $\alpha$  决定，也就是其相对阴影投射的位置决定，阴影内的点受不到直接光源照射因而较暗。成像时，即便此点与其受直接光源照射的阴影区域外的对应点的反射率是一致的，但是仍可能因为成像设备等原因导致直接进行亮度值改变时产生色偏，因而光照还原绝不仅仅是改变亮度达到和阴影区域外的点亮度一致即可。

### 2.1.2 视频中阴影组成分析

视频中的阴影相较图像（单帧）而言更加复杂，在于其每一帧都包含时间信息。按照运动特征的不同，视频中的阴影可以分为由固定摄像机拍摄的静态场景中的阴影和由运动摄像机拍摄的动态场景中的阴影。按照阴影区域光照强度不同，视频中的阴影可以分为硬阴影（浓阴影）和软阴影（淡阴影）。按照遮挡光源对象的不同，视频中的阴影可以分为投射阴影和自阴影。除了上述分类方法外，视频中的阴影还可能分布在多个不同区域，每个区域的阴影形状、大小、运动特征、纹理特征都可以不尽相同。

简单地利用图像阴影消除算法进行单帧阴影消除后合成视频往往得不到较好的阴影消除效果。因为时间域信息的丢失，视频将会在时间上不连续，产生抖动闪烁等瑕疵。此外，简单的帧之间内容的互相替换无法解决视频阴影消除的问题。

首先，视频中可能存在静止的阴影，这意味着该阴影区域无法在视频的其他帧序列中找到对应的无阴影区域。其次，拍摄视频的摄像机可能存在平移、旋转等构成的复合运动，因此视频每一帧的视角不尽相同，这也意味着视频不同帧之间相同平面位置的内容并不完全相同。因此不论是对固定的摄像机拍摄的静态场景的视频还是运动摄像机拍摄的运动视频，都需要一个更为鲁棒的算法进行视频的阴影消除。

## 2.2 阴影消除相关技术

### 2.2.1 基于关键帧的视频处理

关键帧的提取涉及到视频压缩编码，阴影关键帧可充当视频序列的关键帧。这里主要关注 FFMPEG 包含的 MPEG-2 视频编码格式。MPEG-2 的编码包含视频流层结构、帧间编码结构、类与极、运动估计等<sup>[31]</sup>。在帧间编码及运动估计处进行切入实现关键帧的阴影消除，再使用编码特性实现关联帧的阴影消除，可以有效地保证关键帧前后的时空一致性。

MPEG-2 的视频流包含图像序列层 VSL (Video Sequence Layer)、图像组层 GOPL (Group of Pictures Layer)、图像层 PL (Picture Layer)、宏块条层 SL (Slice Layer)、宏块层 ML (Macroblock Layer) 和块层 BL (Block Layer) 总六个部分。在图像组层 GOPL 中，规定单个关键帧及其前后包含的相关帧为一个图像组 GOP (Group of Pictures)。

MPEG-2 为了追求更加效率的压缩编码和可随机存取的高压缩比图像，其定义了三种类别：I 帧、P 帧以及 B 帧。全称：关键帧 I 帧 (Intra Picture)、前向预测帧 P 帧 (Predicted Picture) 和双向预测帧 B 帧 (Bidirectional Picture)。其中，I 帧是包含完整信息的一幅图像的帧，而 P 帧和 B 帧都是根据其他帧（并不必须只是 I 帧）进行运动补偿帧间预测算法解码得到的。利用这点，可以进行阴影消除后的阴影宏块区域迁移。

在 IBP 帧压缩编码中，帧内编码即 I 帧的编码本身就作为基准图像，其编码得到的图像称之为帧内编码图 ICP (Intra Coded Pictures)。其数据量最大，因为包含完整图像所以只能进行帧内压缩，如 DCP 等压缩算法，通过相邻像素、行亮度、色度等进行空间内压缩，所以压缩度较低。预测编码即 P 帧的编码则需要前一个 I 帧或 P 帧作为基准图像，在这只是进行运动补偿来解码推算得到完整图像。其



称之为预测编码图 PCP (Predictive Coded Pictures)。其由于只能进行前向预测 FP (Forward Prediction)，若连续编码多个 PCP，同时前一个 PCP 出现错误编码，则有可能引起错误扩散，一路传递下去直到解码到下一个 ICP。而双向预测编码，即 B 帧的编码不仅需要前一帧的 ICP 或 PCP，还需要后一帧的 ICP 或 PCP，以两帧为基准图像进行运动补偿解码计算，才能推算到完整图像，因而称为双向预测编码。

因为序列相邻两帧的运动部分具有连续的时间相关性，所以可以看做当前帧的内容是前一帧图像的位移，这也是基于关键帧进行阴影消除的核心思路，只需要消除关键帧的阴影再对其进行位移即可。虽然位移的方向和位移的幅度在各个部分可能不一，但是可以通过光流等运动信息推算便可预测出当前帧，这称为前向预测 FP。对应的，根据位移的运动矢量预测出之前的帧中可能出现的图像信息则称为后向预测 BP (Backward Prediction)。BCP 是前向预测 FP 和后向预测 BP 同时作用取均值运算后得到的图像，因而也称为平均值预测图。

根据 IBP 帧压缩编码的原理，P 帧和 B 帧都只包含其相对于别的帧的预测值和运动矢量，在解码时根据运动矢量以参考帧为基准，将预测值与差值相加以得到完整的图像。因而对于包含阴影的任何区域，其只包含区域的运动信息，在将关键帧的阴影进行图像阴影消除后，其按照原有阴影区域图像加上原有阴影的运动信息还原则可以得到去除阴影的新图像，且此时前后帧是必然含有运动连续性的，这样就保证了时空一致性。

而根据其他帧重建图像这一步需要运动补偿的预测及差值。这就需要在编码时估算好运动矢量及位移以便后期运动补偿。将缓存中前一帧图像重建后得到结果中包含块 Block，将得到的块用以求算运动矢量位移就可以得到所需的运动补偿 MC (Motion Compensation)。通过运动补偿，可以压缩时间冗余度 TR (Temporal Redundancy)。

运动补偿，就是假设画面中的部分区域可能通过不同的方式做平移等运动，这样就可以根据区域推测画面。之所以是根据区域，是因为画面中的不同区域可能有不同的运动轨迹或方式。运动补偿之前首先要进行运动估计，为的是尽量减少预测误差。运动估计是需要完整的抽取运动的信息的技术，其结果的好坏、参数特点等等可以直接或间接地决定运动补偿的最终性能。运动估计由多个步骤组成，其大致的预测实质是根据前一帧（第  $n - 1$  帧）推断出当前（第  $n$  帧）像素

的估测值。再通过运动矢量编码，传输第  $n$  帧的还原后完整像素值和估测值的差值。运动补偿示意图如下所示。

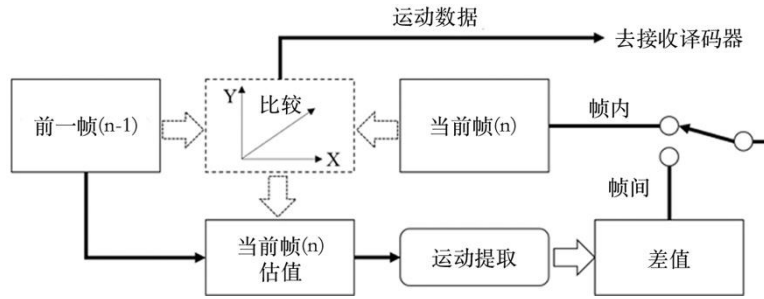


图 2-2 运动补偿过程示意图

假设宏块为  $M \times N$  的矩阵块，运动估计需要比较第  $n-1$  帧和第  $n$  帧的宏块。即将第  $n$  帧中的宏块和第  $n-1$  帧中的限定搜索区 SR (Search Realm) 内全部对应大小的宏块比较。假设第  $n-1$  帧的亮度信号为  $f_{n-1}(i, j)$ ，而第  $n$  帧的亮度信号为  $f_n(i, j)$ 。其中， $(i, j)$  为第  $n$  帧的  $M \times N$  宏块的任意位置。假设第  $n$  帧中的一个  $M \times N$  的宏块可以被认为是第  $n-1$  帧内的某个宏块 MB (Macro Block) 平移而来。同时，要求这一个满足要求的宏块内部的所有像素必须拥有完全一致的位移值  $(k, l)$ 。根据如上提出的要求后，在第  $n-1$  帧的 SR 内执行范围搜索，必然能够找到符合要求的某一宏块。这样，第  $n-1$  帧中的满足要求的宏块与第  $n$  帧中要匹配的对应该宏块的像素差值的绝对值达到极小值。根据这两个差值最小的宏块计算它们的差值，就可以算到这一组关联对应宏块的运动矢量数据。在第  $n-1$  帧对应宏块的原始图像数据加上差值运动矢量的数据，就可以得到第  $n$  帧的相应的估算值。循环往复如上过程，直到第  $n$  帧内的  $M \times N$  宏块的所有位置  $(i, j)$  的像素完全可以通过第  $n-1$  帧的像素预测出来。即第  $n$  帧和第  $n-1$  帧的相关性可由函数  $F(k, l)$  的绝对值表示，如下公式描述：

$$F(k, l) = \min \sum_{i=1}^M \sum_{j=1}^N |f_n(i, j) - f_{n-1}(i + k, j + l)|, (k, l) \in SR \quad (2-4)$$

其过程可见示意图 2-3 运动补偿的过程：

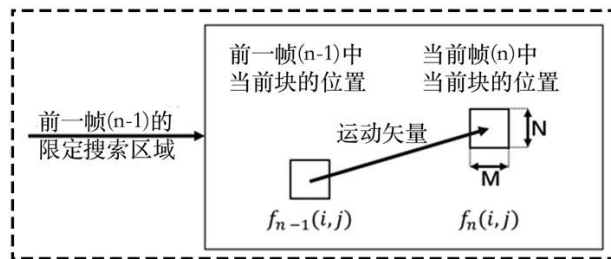


图 2-3 运动补偿的过程

当待匹配的宏块完成匹配并且获得宏块间矢量位移值  $(k, l)$  后，将帧之间传输的数据差值尽可能降低到极小值，且能够补偿宏块位移，就是运动补偿。在较为静止的画面中连续的视频序列画面中有较多的相同像素，图像序列差值非常小，导致时间冗余度较高。此时，使用运动补偿方式，可以省去非常多不需要的画面信息。只需要在解码器中将运动差值估计值配合关键帧数据进行补偿还原即可得到和压缩前差别不大的画面。当然，进行运动补偿参数估计时也需要关联帧同时输入，用运动矢量数据和完整帧的复原数据合并计算才能将帧完整预测还原。实际操作上，根据处理视频摄像机的静止运动情况，视频内物体的多少、大小、运动快慢等因素，可能包含完整图像的关键帧出现的次数不尽相同。拍摄视频的摄像机可能存在平移、旋转等构成的复合运动，因此视频每一帧的视角不尽相同，这也意味着视频不同帧之间相同平面位置的内容并不完全相同。因此，可能有连续的不完整、只包含运动估算和差值的帧，且可能出现前向预测和双向预测的帧。比如，出现 IBPBBPBI 这样的长不完整帧组，两个关键帧相距离长达 7 个非关键帧，这些非关键帧在解码还原时可能存在互相依赖，需要按照严格的非线性顺序进行解码。当然实际情况可能比这还要长得得多。这时候就需要以组 GOP (Group of Pictures) 为单位进行打包，单次解码也以组为单位，一整组可能存在互相依赖关联，需要按序依次解码。

在帧间编码和运动估计并还原补偿的算法下，通过对关键帧 I 帧进行图像阴影消除，再使用 B 帧 P 帧原始的运动补偿信息进行前后图像还原就可以得到包含时间信息的其他阴影消除图像，这样就可以在保证时空一致性的情况下完成图像序列也即视频的阴影消除。

### 2.2.2 图像形态学处理

形态学图像处理指的是对于输入的图像，通过数学形态学的方法从中提取出对于表达和描绘图像区域形状有用的分量。形态学图像处理可以对图像中的内容从形态学的角度进行有效的分析，获取许多重要信息。形态学处理一般包括腐蚀和膨胀两个基本操作。膨胀操作即使目标物体轮廓向外扩张，腐蚀操作即使目标物体轮廓向内收缩。膨胀操作可以消除图像中的多余空间，也可以突出目标。腐蚀操作可以去掉图像中多余噪声。腐蚀和膨胀的操作一般是通过改变结构元素，即图像中具有特定性质的像素位置的集合。操作由给定原点开始，遍历图像中的

每一个像素位置，在结构元素覆盖的像素区域内通过所要求的操作对覆盖区域内的像素值进行修改以达到形态学处理目的。

设  $f(x,y)$  为一幅灰度图像， $b(x,y)$  为定义的结构元素，则膨胀操作为加强亮部区域，减少暗部区域，其定义为：

$$[f \oplus b] \cdot (x,y) = \max_{(s,t) \in b} \{f(x+s, y+t)\} \quad (2-5)$$

其中， $[f \oplus b] \cdot (x,y)$  是经过膨胀操作后图像在  $(x,y)$  处的灰度值， $(s,t)$  是结构元素中相对于原点的坐标位置。膨胀操作示意图如下。



图 2-4 膨胀操作。左图为原图，右图为膨胀后

类似的，腐蚀操作为加强暗部区域，减少亮部区域，其的定义为：

$$[f \ominus b] \cdot (x,y) = \min_{(s,t) \in b} \{f(x+s, y+t)\} \quad (2-6)$$

其中， $[f \ominus b] \cdot (x,y)$  是经过膨胀操作后图像在  $(x,y)$  处的灰度值， $(s,t)$  是结构元素中相对于原点的坐标位置。腐蚀操作示意图如下。



图 2-5 腐蚀操作。左图为原图，右图为腐蚀后

通过上述定义可以得到，执行膨胀后图像中的暗特征降低，亮特征增强，因而灰度图像中黑色的间隙部分将消失。执行腐蚀后的图像暗特征增强，亮特征降低，因而灰度图像中的噪声部分将消失。腐蚀膨胀是两个重要、具有广泛引用的图像形态学操作，常被用于二值图像处理、图像分割、目标检测的前期准备工作。

### 2.2.3 使用 VGGNet 提取特征

本文在进行深度学习阴影消除、后序边缘细节修正和关键帧提取时都将使用到深度卷积神经网络 VGG (Visual Geometry Group)<sup>[32]</sup>。VGG Net 主要研究了卷积

神经网络 CNN (Convolution Neural Network) 的深度如何影响性能，因其原本用于图像语义识别，因此其中间层拥有较强的图像特征不变性。本文用了 VGG-19 用于语义提取时其中间层特征不变的特点，来进行提取图像特征，分别进行阴影初步消除、后序纹理修复和关键帧提取。

VGG 是 Alex Net 的进化版，同样由卷积层、全连接层构成，如下图所示。

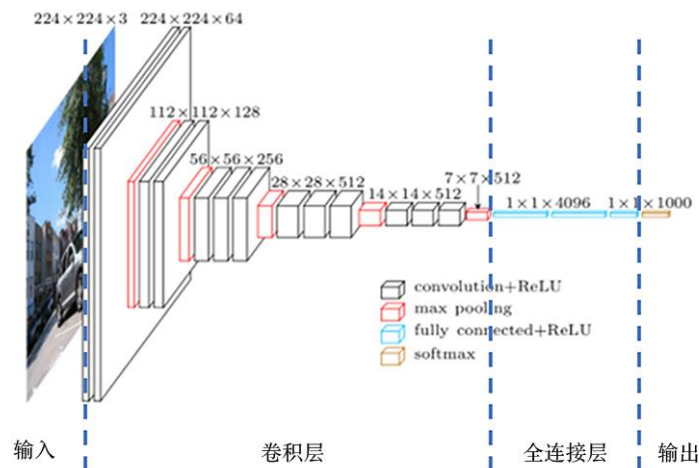


图 2-6 VGG 示意图<sup>[32]</sup>

VGG 具有如下特点：

- (1) 结构简洁。VGG 仅由 5 层卷积层、3 层全连接层、softmax 输出层构成。层与层之间使用 max-pooling（最大化池）分开。使用 ReLU 函数作为隐层激活函数。
- (2) 小卷积核和多卷积子层。大的卷积层并没有在此被使用，取而代之的是多个 3x3 的卷积核。参数数量降低的同时等同进行更多的非线性映射。最终实现的是加强网络的拟合/表达能力。VGG 并没有使用类似 AlexNet 的 7x7 大卷积核尺寸，而是使用了 3x3 大小的取而代之。这样一来，增加卷积子层数以取得同样的性能。但是 VGG 有从 1 到 4 卷积子层，AlexNet 只有 1 子层。通过双 3x3 卷积堆叠，获得相当单个 5x5 卷积的感受野。而三个 3x3 卷积的堆叠时，则可以得到约为一个 7x7 的卷积的感受野。这样，通过增加非线性映射，可以有效减少参数。
- (3) 小池化核。VGG 全部采用 2x2 的池化核。
- (4) 通道数多。VGG 网络从首层 64，层层翻倍至最多 512 个，以得更多信息。
- (5) 具有更深的层和更宽特征图。由于卷积核致力于增大通道数，而池化致力于减小宽和高，使得模型架构上更深更宽，且控制了计算量的增加规模。

(6) 全连接转卷积。将全连接替换全卷积，使得到的全卷积网络可以接收任何数值的宽和高的输入，不受全连接的限制。在测试阶段，这有重大意义。

根据如上介绍的 VGG 层次特点，VGG-19 一共可以提取 36 个特征图。其 conv-1\_1 层及 relu-1\_1 层的特征提取如图 2-7 所示。

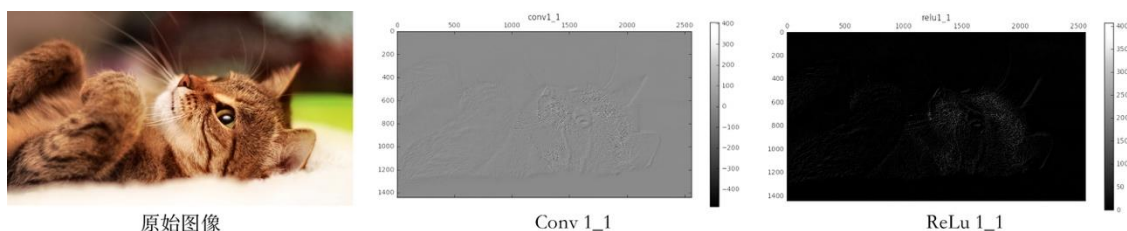


图 2-7 VGG 特征提取

根据 Yang et al. (2016) 的研究<sup>[33]</sup>，其 ReLu 3-1 层及 ReLu 4-1 层通过单独输出可以得到具有较强的图像纹理特征。其对应权重维度信息为：ReLu 3-1 核大小 3×3，共 128×256 个；ReLu 4-1 核大小 3×3，共 256×256 个。通过输入图像并提取这两层，可以得到较好的特征。

实际操作中，本文并没有根据 VGG 的原理进行从无到有开始训练构造，而是直接用了大数据集 ImageNet 上已经训练好的 VGG 模型，固定前面层的参数，微调 (FineTuning) 最后两层的参数，用后文提到的数据集进行权重更新。除上述 ReLu 3-1 和 ReLu 4-1 层之外，pool 5 的 fc6 和 fc7 的特征效果也可以用于提取特征，可以直接输出 pool 5 的最后一层。具体的，使用 TensorFlow 调用时，提取特征可以直接用 TensorFlow 规定的语句（Python 下）`sess.run(vgg.层数, feed_dict={images: batch})` 提取，不赘述。

然后，将提取到的特征应用于阴影消除、边界区域的纹理修复和关键帧检测，以实现视频阴影消除。

## 2.3 基于关键帧的视频阴影消除模型

阴影消除后阴影区域像素点的颜色值会改变，在视频阴影消除过程中，如果每帧单独进行阴影消除，不同帧间阴影区域像素点颜色值的改变可能会使得阴影消除后的视频结果出现帧间抖动的情况，特别是在阴影边界处。

因此，本文提出了一个基于阴影关键帧的视频阴影消除算法，该算法将视频的阴影消除问题转换为阴影关键帧的阴影消除问题，然后利用视频前后帧之间内

容的关联性及时空一致性优化来对整个视频进行阴影处理，完成视频的阴影消除工作。本文完整思路如下图所示。

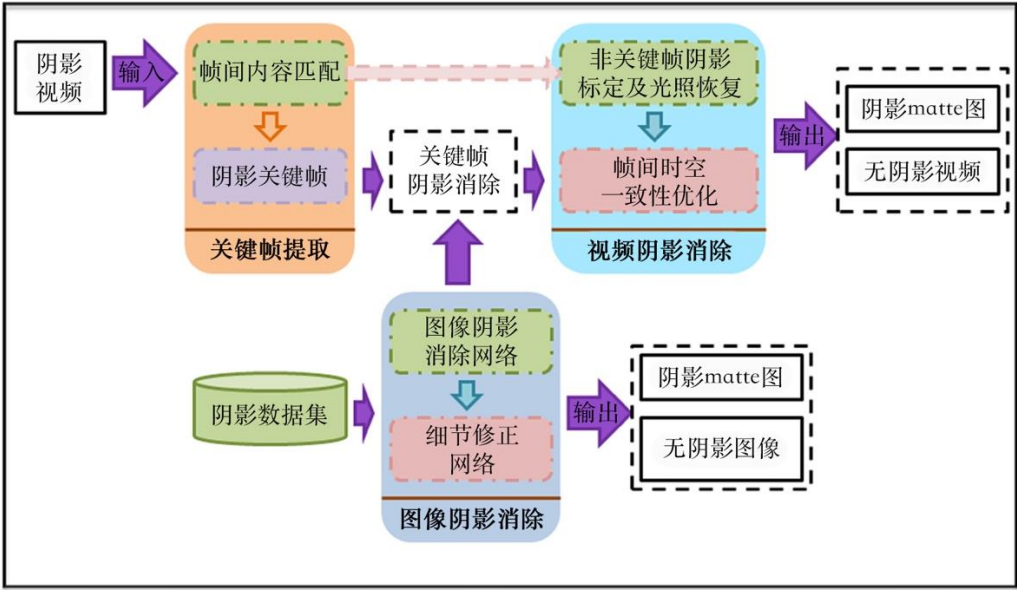


图 2-8 基于阴影关键帧的视频阴影消除技术路线

**基于数据集的图像阴影消除：**首先需要构建一个适用于阴影消除的深度学习网络，专注于学习数据集中原始的包含阴影的图像与将阴影抠出后得到的阴影matte图之间的映射关系。得到了这样一个抠图网络之后，就可以利用其对输入的包含阴影的图像进行阴影消除。这个网络结构可以同时实现阴影检测和阴影消除。将阴影检测和阴影消除作为同一个任务来处理时可以拥有更好地连续性，减少分开工作后产生的偏差，在完成阴影检测同时就可以得到阴影消除的结果。得到初步阴影消除结果后，常常会遇到初步阴影消除后，阴影区域出现的细节丢失、特征结构模糊及阴影边界失真等问题。所以，阴影消除后，还需要对恢复光照后的图像进行纹理细节方面修正及色彩上的修正，使得消除阴影后，图像的阴影区域部分的光照及内容与图像阴影区域周围的环境保持一致。

**基于阴影关键帧的视频阴影消除：**首先通过视频压缩时前后帧内容的关联性提取，根据压缩流程确定起决定性参考作用的视频关键帧，也即阴影关键帧。得到阴影关键帧后，利用上述图像阴影消除模型对关键帧进行阴影消除。得到关键帧阴影消除结果后，利用关键帧与其前后帧内容的关联性处理非阴影关键帧中的阴影，这一步操作能确保帧序列保持时空一致性，从而优化完成整个视频的阴影消除任务。这种利用阴影关键帧处理整个视频中阴影的方法解决了视频阴影消除后可能出现的抖动问题，保证了视频阴影消除结果的时空一致性及视觉真实性。

---

## 2.4 本章小结

本章介绍了自然界中阴影形成模型，阴影区域包含本影区和半影区，这直接关系到组成阴影的浓阴影和淡阴影。同时也介绍了阴影的多种分类方法。对于本文将要用到的几个主要技术，本章介绍了 FFMPEG 规定的 MPEG-2 编码下的视频处理技术，包括视频编解码、基于帧的压缩、分块、运动估计、补偿等等，为后文关键帧提取奠定技术基础。然后介绍了基本的两种形态学处理。以及著名的深度卷积神经网络 VGGNet，这将被全文多个地方的算法用到。最后，介绍了全文的总技术路线，完成本章基础工作铺垫。



### 3 图像阴影消除

本文首先构建一个全自动地端到端的深度学习网络来对图像进行阴影消除，然后利用一个细节修正网络对阴影消除结果进行调整，增强阴影区域的纹理细节并修正可能产生的阴影边界问题，保持恢复光照后的阴影区域与周围邻域亮度及细节信息保持一致。

#### 3.1 深度学习图像阴影消除

如第 2.1 节中所提到的，图像是由场景中的光照与物体相互作用产生的结果，任意一点的颜色值模型最终可定义为  $I_x = R_x \cdot (\alpha L_x^d + L_x^a)$ ,  $\alpha \in [0, 1]$ ，其中  $I_x$  表示观察到的图像  $I$  在像素点  $x$  处的颜色值， $R_x$  表示像素点  $x$  处的物体的反射率， $L_x^d$  表示场景中的直接光源在假设无遮挡的像素点  $x$  处的强度， $L_x^a$  表示场景中的环境光照在像素点  $x$  处的强度， $\alpha$  表示直接光源的衰减程度。光照强度  $L$  还可以化归为像素点上光照的衰减度和场景光照（无阴影遮挡）的乘积，设  $\beta$  为场景中每个像素对应的光照衰减因子， $L'$  为在无阴影遮挡下所对应的光照强度，则可以化归结果如下：

$$L = \beta \cdot L' \quad (3-1)$$

假设非阴影区的图像表示如下：

$$F = R \cdot L' \quad (3-2)$$

则阴影图像  $I$  可表示为：

$$I = \beta F \quad (3-3)$$

其中  $\beta$  对应的图像可理解为阴影 matte 图。阴影 matte 图可以做为阴影消除中间层额外对比确认的信息，可以和包含阴影 matte 图的数据集做比较，调整中间层以得到更好结果。目前深度学习存在六种公开的阴影图像数据集，如下表。

表 3-1 目前的阴影图像数据集

数据集	数量	内容	目的
UIUC	76	阴影图像、无阴影图像	阴影消除
LRSS	37	阴影图像、无阴影图像	阴影消除
SRD	3088	阴影图像、无阴影图像	阴影消除
SBU	4727	阴影图像、阴影区域标记图	阴影检测
UCF	245	阴影图像、阴影区域标记图	阴影检测
ISTD	1870	阴影图像、无阴影图像、阴影区域标记图	阴影检测和消除

对上述训练集放入深度卷积神经网络 CNN (Convolutud Neural Network) 进行训练，得到训练模型。神经网络构建部分参考了 Qu et al. (2017) 发表的论文<sup>[34]</sup>，通过多层结构分别提取信息，分别有：高层语义信息、中层表面信息和局部图像细节，分别对应三个子网络。该网络不仅从全局的角度理解整张图像，还关注局部提取细节纹理。

对全局信息，使用局部化网络 G-Net 提取全局阴影特征，代表场景中的全局结构和高层语义纹理。使用第 2.2.3 节提到的 VGG 提取特征的特性，将 VGG 的最后两个最大池化层的步幅设置为 1 来获得更密集的预测，并将全连接层替换成一个  $1 \times 1$  的卷积层和一个  $1 \times 1$  的反卷积层来获得完全卷积。然后，使用表层模型网络 A-Net 从 G-Net 的浅层提取表面信息，对 G-Net 的结果进行矩阵转置，将反卷积的结果 Decv2-1 和对原图的卷积 conv2-1 再多次卷积得到 A-Net 最终输出。使用 S-Net 从 G-Net 的深层提取语义信息，类似 A-Net，将 G-Net 的结果进行矩阵转置，进行反卷积的结果 Decv3-1 和原图卷积 conv3-1 再多次卷积得到 S-Net 最终输出。

因参考论文已经做过反卷积后可视化论证，转置后得到图像可以看出在 A-Net 大致能提取出阴影区域还原后的色彩信息，在 S-Net 大致能提取出物体边界信息，证明其各项特征提取合理性，加以组合即可得到阴影消除后图像，故本文不加以验证赘述。神经网络构建图如图 3-1 所示。

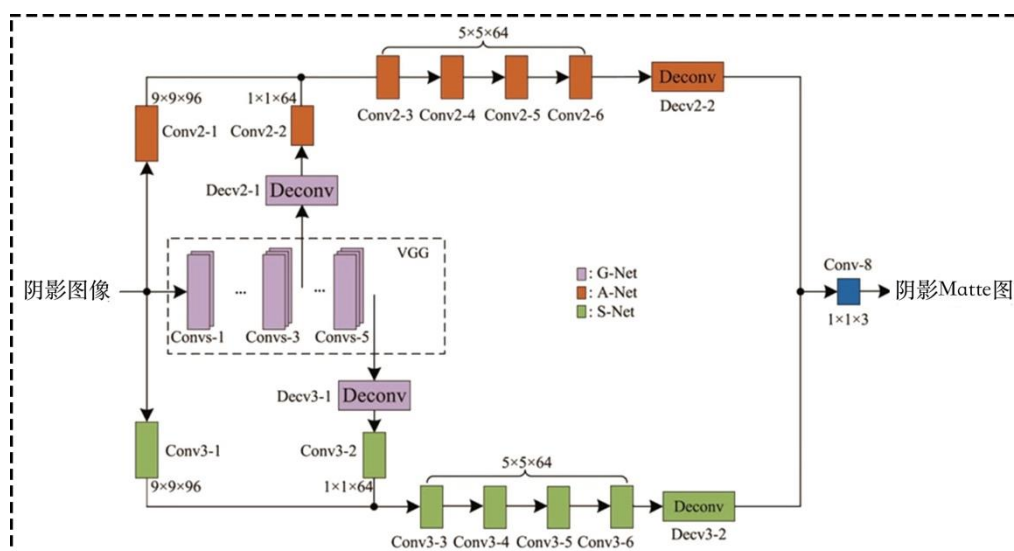


图 3-1 神经网络结构示意图<sup>[34]</sup>

训练时，使用包含  $224 \times 224$  的数据集作为阴影图像和阴影 matte 图，并分别转换为 log 空间再输入。为了防止过拟合，采用如下策略：首先多阶段训练，分

开训练 G-Net + A-Net 和 G-Net + S-Net，然后再整体训练；然后多尺度训练，从粗糙的  $64 \times 64$  到  $128 \times 128$ ，最后再用  $224 \times 224$ 。使用 MSE (Mean Squared Error) 在 log 空间作为损失函数，卷积层、池化层、全连接层相关参数如下表 3-2 所示（缩进需要， $N \times N$  用  $N^2$  表示）。

表 3-2 神经网络参数<sup>[34]</sup>

Net	Layer	convs					Decv	
		1	2	3	4	5	2-1	3-1
G-Net	#of convs	2	2	3	3	3	1	1
	#of channels	64	128	256	512	512	256	256
	Filter Size	$3^2$	$3^2$	$3^2$	$3^2$	$3^2$	$8^2$	$8^2$
	Conv. Stride	1	1	1	1	1	4	4
	Zero-Padding	1	1	1	1	1	2	2
	Pool-Size	$2 \times 2$	$2 \times 2$	$2 \times 2$	$2 \times 2$	$2 \times 2$	-	-
	Pool-Stride	2	2	2	1	1	-	-
	Output Size	$112^2$	$56^2$	$28^2$	$28^2$	$28^2$	$112^2$	$112^2$
A-Net (S-Net)	#of channels	96	64	64	64	64	64	3
	Filter Size	$9^2$	$1^2$	$5^2$	$5^2$	$5^2$	$5^2$	$4^2$
	Conv. Stride	1	1	1	1	1	1	2
	Zero-Padding	4	-	2	2	2	2	1
	Pool-Size	$3 \times 3$	-	-	-	-	-	-
	Pool-Stride	2	-	-	-	-	-	-
	Output Size	$112^2$	$112^2$	$112^2$	$112^2$	$112^2$	$112^2$	$224^2$

设置 G-Net 的学习率为  $10^{-5}$ ，A-Net 和 S-Net 为  $10^{-4}$ 。使用 Caffe 进行训练大概需要 5 周左右时间。因配置需要，将 Caffe 模型转换为 TensorFlow 模型，之后就可以直接将待处理的图片输入。以上未提到的可以使用 TF 默认参数，得到阴影消除后的图像。下面使用 8 个常用阴影消除视频样例，经过第 4.1 节将会使用的关键帧提取处理后，得到各个视频的首个关键帧，也就是视频的第一关键帧，转换为图像后放入上述深度学习网络进行阴影消除，得到结果如下图 3-2 所示。



图 3-2 通过深度学习网络得到的阴影消除后的图像，第一行为关键帧原始图像，第二行为阴影消除后的图像

### 3.2 细节修正

在进行上述深度学习阴影消除后，得到的阴影消除图片的阴影区域的边界处，由于亮度的变化可能会造成阴影边界处纹理的丢失及颜色失真，这种纹理的丢失及颜色失真造成阴影消除后的结果在阴影边界处出现颜色及亮度失真，如上图 3-2 中部分结果图所示。

因此，本节再次利用 VGG-19 (Visual Geometry Group - 19) 模型的纹理特征层设计了一个细节修正步骤，对阴影消除后的结果进行纹理细节及颜色的修正，以解决阴影消除后阴影区域可能出现的细节丢失、特征结构模糊及阴影边界失真等问题，并保证消影后阴影区域的光照及内容与周围环境一致，得到忠实于实际场景的阴影消除结果。这部分的流程图如下图所示。

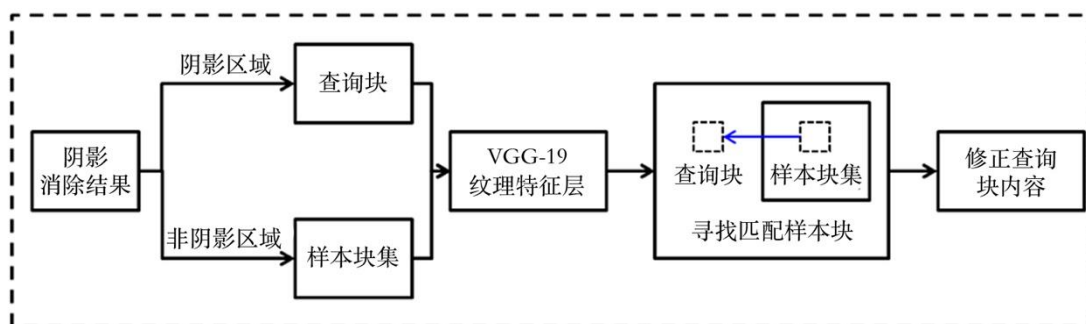


图 3-3 细节修正流程图

在此之前，首先对原图进行 2.2.2 中提到的形态学处理，对阴影 matte 图在消除后的图上进行腐蚀操作，再将除 matte 图外的地方进行腐蚀操作，取两者交集处用平均像素值填补。使用 Yang et al. (2016)<sup>[33]</sup>提到的内容填充网络 (Content Network) 进行填充。将图 3-4 的 8 张阴影消除后图像经过内容感知填充网络后得到各个视频的首个关键帧阴影边界修复内容填充如下图 3-4 所示的 8 个结果。

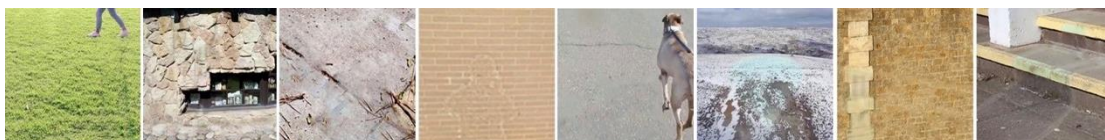


图 3-4 内容填充后的阴影消除图像

随后，根据 2.2.3 节介绍，用从 VGG-19 提取到的特征进行纹理优化。这里，取 VGG-19 的 ReLU 3-1 层及 ReLU 4-1 层提取纹理属性。

参数  $R$  为阴影区域， $D$  为阴影消除后的结果图， $t$  为细节修正优化的次数， $D(t)$  为当前待优化的结果， $D(t-1)$  为上一次优化的结果， $E_{global}$  为全局数据项，



$E_{local}$  为局部纹理约束项,  $E_{smooth}$  为平滑项,  $\lambda_1$  和  $\lambda_2$  为平衡权重。为了完成阴影区域的细节修正, 最小化下面这个能量方程来优化阴影消除的结果:

$$E = E_{global} + \lambda_1 \cdot E_{local} + \lambda_2 \cdot E_{smooth} \quad (3-4)$$

**数据项  $E_{global}$  :** 是从全局上约束阴影区域的内容, 设  $i$  为阴影区域  $R$  中的一点, 其内容可表示为:

$$E_{global} = \sum_{i \in R} (D_i(t) - D_i(t-1))^2 \quad (3-5)$$

**局部纹理约束项  $E_{local}$  :** 用来约束阴影区域的纹理外观与周围环境一致。VGG-19 网络是用来进行语义分割, 取其中间层 ReLU 3-1 和 ReLU 4-1, 其对于纹理特征具有强的不变性。将结果图中的阴影区域分为很多个固定大小的块作为查询块, 将非阴影区域分为同样固定大小的块, 作为样本块集。比较块之间纹理信息的损失值, 通过寻找最小损失值为查询块在样本块集中找到最相似的样本块。 $P$  为查询块,  $R^P$  为所有查询块集,  $f$  为块的纹理特征向量, 则纹理约束项表示为:

$$E_{local} = \frac{1}{|R^P|} \sum_{P \in R^P} \sum_{i \in P} (f_p(t) - f_p^{nn})^2 \quad (3-6)$$

其中,  $|R^P|$  为需要处理的查询块数量,  $f_p(t)$  为当前查询块对应的纹理特征向量,  $f_p^{nn}$  为查询块对应的最相似的样本块。

**平滑项  $E_{smooth}$  :** 通常, 局部窗口中的像素具有相似的外观。平滑项的目的就是用来约束这种局部相似性, 并且定义为:

$$E_{smooth} = \sum_{i \in R} \sum_{j \in N_i} (D_i(t) + D_j(t))^2 \quad (3-7)$$

其中,  $N_i$  为点  $i$  的 4 邻域。t 为迭代次数, 且  $D_i(0) = D$ 。通过多次使用纹理优化网络 (VGG-19 的纹理特征层) 进行边界纹理的迭代优化, 修复边界瑕疵, 最终可以得到比较好的阴影消除结果, 如下图 3-5 所示。

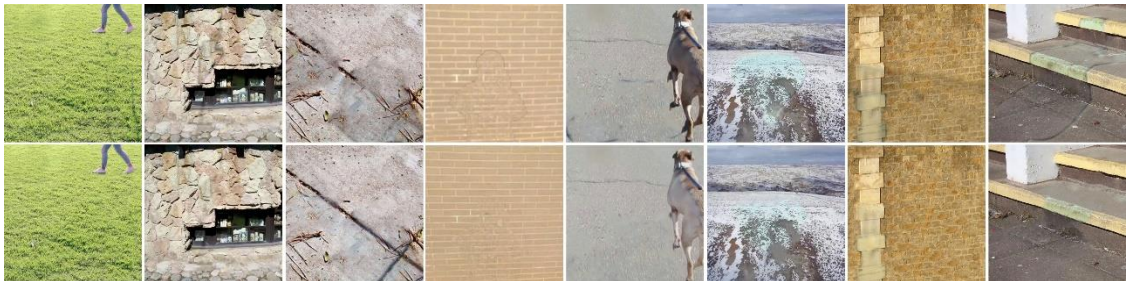


图 3-5 阴影边界修复结果。第一行为初始阴影消除图像, 第二行为边界修复结果

---

### 3.3 本章小结

本章主要介绍了基于深度学习的视频阴影消除方法。通过卷积神经网络 CNN，得到输入一个阴影图，输出一个阴影消除后图像的深度学习网络，并利用这个网络进行阴影消除。在阴影初步消除后，再接着使用 VGG-19 的中间层特征不变特性对特征提取，采集纹理特征，对阴影消除后的阴影区域边界瑕疵进行修复，得到整体纹理、光照更为统一的无阴影图。

## 4 视频阴影消除

### 4.1 视频阴影关键帧的提取

阴影关键帧是指视频中阴影区域出现新内容或新目标的视频帧。在视频处理过程中，如果在一个视频体结构的阴影区域中出现了新的内容，则将当前帧定义为阴影关键帧。同时，因为视频的第一帧按照定义必然直接影响后向序列，其包含序列起点、组长、属性等相关信息，故可以直接设置为阴影关键帧。

在 FFMPEG 定义的 MPEG-2 视频规范中，根据帧间压缩算法将帧划为了三种帧类型：I 帧、B 帧、P 帧。如第 2.2.1 节所介绍，其压缩编码定义中即定义 I 帧为参考帧，也是关键帧，B 帧及 P 帧都在其之上通过运动补偿建立并还原出完整帧信息。而在 FFMPEG 原有对 I 帧的查找定义内容之上，本文进一步加强了关键帧及关联帧的判定。

对于每一帧，我们首先将当前帧分为很多固定大小的块，块的大小为  $w \times w$ 。对于每一个块构建一个体结构，体结构的纵向大小为固定块大小的两倍，横向大小（时间轴范围）为当前帧和前一帧，为每一个块在体结构中寻找一个纹理最相似的块，并且这个最相似的块位于前一帧。这里我们同样使用已经训练好的 VGG-19 网络中的纹理特征层来获取块的纹理信息，通过比较块间的纹理差异来获得最相似块。如果这个相似块在上一帧中标记为了阴影块且两个块的相似度小于设置的阈值，就将当前帧标记阴影关键帧。寻找关键帧的流程如图 4-1 阴影关键帧提取流程所示。

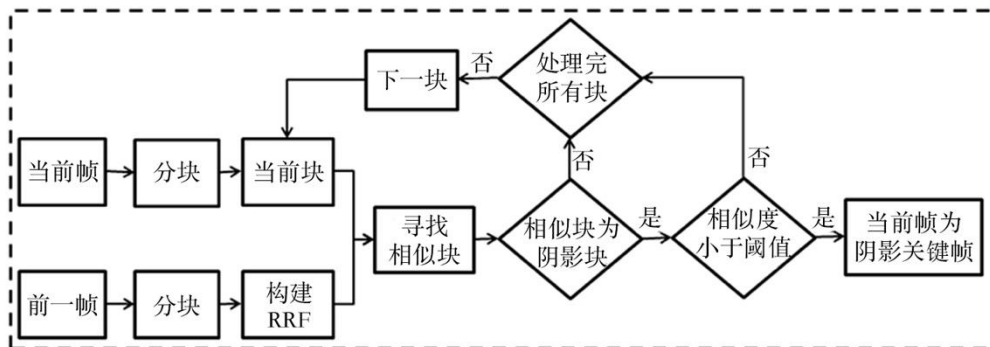


图 4-1 阴影关键帧提取流程

根据上述提取关键帧的方法，对之前使用的 8 个阴影消除样例视频分别通过上述流程，提取到阴影关键帧，按列排布，各列之上而下代表各个视频时间顺序

出现的各个关键帧，因各个视频长度、运动内容不尽相同，故各个视频提取到的关键帧数量也不同，所以各列长度不同，各个视频的关键帧如图 4-2 所示。



图 4-2 各个视频分别提取出的关键帧



## 4.2 视频的阴影消除

根据上一节内容，在得到了视频的阴影关键帧后，利用第 3 章已构建好的图像阴影消除模型对阴影关键帧进行阴影消除，并标定出关键帧的阴影区域。对原始视频，根据 2.2.1 节的内容，标定得到的阴影关键帧为新的 I 帧，其他帧都以这些 I 帧作为参考帧进行运动估计和差值补偿，对原始视频进行重编码。然后利用第 3 章的图像阴影消除方法，对关键帧转换为的图像进行阴影消除，如图 4-3 所示，为各个视频的首个关键帧图像阴影消除后结果和原始关键帧图像对比。



图 4-3 首个关键帧阴影消除前后对比，第一行为关键帧原始图像，  
第二行为阴影消除后图像

接下来，将图像阴影消除的结果替换原有视频中的阴影关键帧，并保持前后帧的运动估计和差值补偿不变，以保证之间内容的关联性，得到其它视频帧对应的阴影区域及阴影消除结果。算法流程如图 4-4 视频阴影消除算法流程图所示。

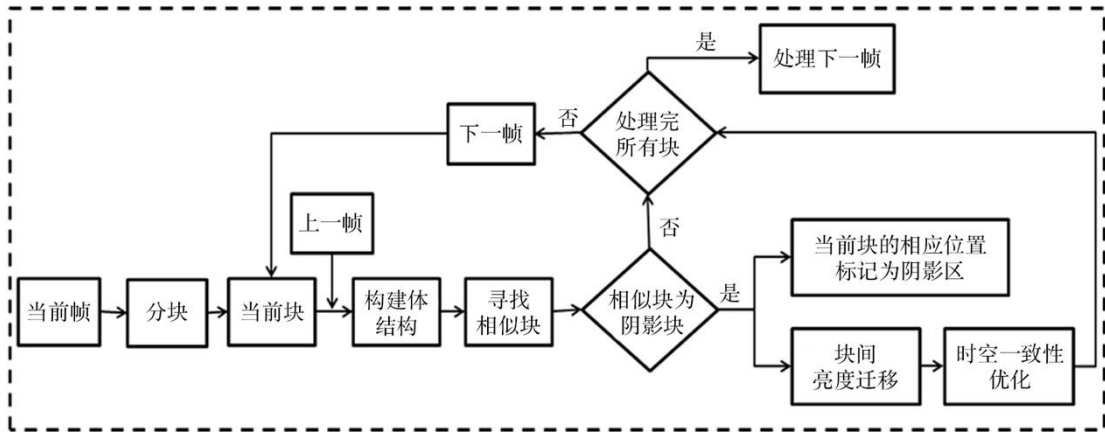


图 4-4 视频阴影消除算法流程图

将当前帧分为很多固定大小的块，在寻找阴影关键帧是已经为这些块在前一帧寻找好了相似块，接着我们就可以来处理当前帧中的阴影区域了，具体步骤如下：

A.如果当前帧为阴影关键帧，直接利用已构建好的图像阴影消除模型来获得当前帧的阴影消除结果；

B.如果当前帧为非阴影关键帧，并且相似块为非阴影块，则当前块的亮度不变，沿用原始视频编码该块信息，继续处理当前帧中下一个块；

C.如果当前帧为非阴影关键帧，并且当前块对应的相似块为阴影块，则当前块对应区域也标记为阴影区域，并将相似块对应的阴影消除后的亮度迁移到当前块的阴影区域，消除当前块中的阴影。具体的，若有运动矢量信息，则根据前一帧该块的替换后内容加上运动补偿得到该块，即得到阴影消除后的本帧块；若本帧此块运动补偿信息丢失，则根据前一块近似位置的帧内压缩相似块替换信息应用于本帧找块替换恢复当前块亮度。还可以通过 matte 图和消除后对应信息的关系完成恢复，如 3.1 节所述。

在第 3 章消除阴影的同时，我们可以获得当前帧的阴影 matte 图。这里使用局部亮度恢复算子来恢复当前块阴影区域的亮度：

$$T_i = \frac{q+1}{\alpha_i q + 1} I_i \quad (4-1)$$

其中  $I_i$  为当前帧点  $i$  处的颜色值， $T_i$  为点  $i$  阴影消除后的颜色值。

$$q = \frac{I_i - I_i^{recover}}{\alpha_i I_i^{recover} - I_i} \quad (4-2)$$

$$I_i^{recover} = \frac{\sigma(L)}{\sigma(S)} (I_i - \mu(S)) + \mu(L) \quad (4-3)$$

其中  $\alpha_i$  为当前帧阴影 matte 图在点  $i$  处的值， $\sigma(L)$  为当前块对应的相似块颜色值的方差， $\sigma(S)$  为当前块对应阴影区域颜色值的方差， $\mu(L)$  为相似块颜色值的平均值， $\mu(S)$  为当前块对应阴影区域颜色值的平均值。

处理完当前帧的阴影区域后，执行当前帧对应的视频的编码。首先初始化，选择 DC 和量化器。然后对图像，也就是帧，进行霍夫曼编码，将一个图像切分成多个宏块 MB (Macro Block)，并写入相应的图片头信息。编码 I 帧，也就是上述确定的阴影关键帧的 MB，因为 YUV 比例不一致，可能是 422 / 420 / 444 等，所以每次编码的块数量要根据原视频信息决定，这里在读取视频时需要直接提取原视频信息用于后面参数定义。编码块，任何一个块其中包含三个编码对象，分别为：DC 系数、AC 系数和游程长度。其中，DC 系数的编码使用差分编码，AC 系数和联合长度同时编码。帧内宏块编码结束后，进行帧间宏块编码，根据第

2.2.1 节内容，进行运动估计和补偿信息编码其他关联帧。最终，进行熵编码并控制码率后，打包封装完成整个阴影消除后的视频编码。

### 4.3 进一步优化减弱可能的抖动

若使用了局部光照恢复算子，在视频中所有的阴影区域都处理完成后，可能发生前后帧阴影区域颜色值分别作出改变，对应视频帧间可能会出现抖动的情况。因此，为防止或减弱这种抖动的情况，我们利用前后帧内容的关联性对每一帧的阴影区域进行时空一致性优化处理。时空一致性优化通过优化如下方程实现：

$$\sum (T_i^{free} - T_i)^2 + \lambda_3 \sum w_i (T_i^{free} - T_i^{before})^2 \quad (4-4)$$

其中， $i$  为当前帧中阴影区域内一点， $T_i^{free}$  为点  $i$  处待优化的结果值， $T_i$  为点  $i$  处阴影消除的结果， $\lambda_3$  为平衡权重， $T_i^{before}$  为上一帧在对应点  $i$  位置的已优化的颜色值， $w_i$  为当前帧点  $i$  与上一帧对应位置在输入视频帧中对应的相似度。

此处操作只需要将视频解码分解为图像序列，分别对前后一组图像序列进行优化方程操作，最后再重新拼接成视频即可。

### 4.4 实验算法、结果与分析

本章通过对各种类型的视频场景进行试验，证明本文提出的算法的有效性，视频阴影消除结果如 4.3 节的图 4-5 所示。本文使用了 C++、Lua、Python 分别对不同步骤进行编程，实验平台为装有 AMD Ryzen 7 1700X 8C16T 3.7Ghz 处理器、32G RAM、NVIDIA GTX 1070 显卡的计算机，算法运行在 Ubuntu 18 LTS 64 位及 Windows 10 64 位操作系统下。

本章视频阴影消除的算法的伪代码如**算法 4-1** 所示。

---

#### 算法 4-1 视频阴影消除算法

---

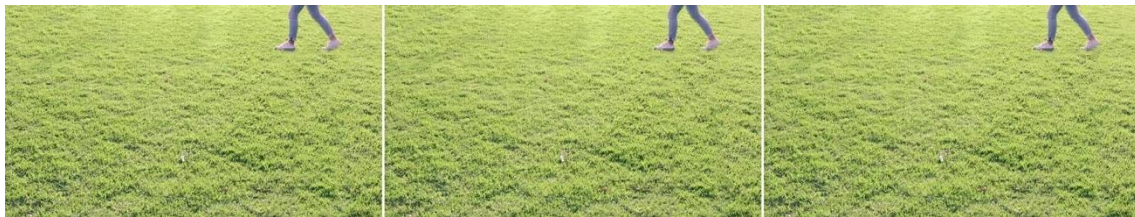
输入：包含阴影的原始视频  $V$ ，视频帧数为  $n_f$

输出：阴影消除后的视频  $V'$

- 1: **for**  $f = 1:n_f$
  - 2:     检测当前帧是否符合关键帧标准
  - 3:     若符合，则标记并提取关键帧  $P_i$
  - 4: **end for**
-

- 
- 5: **for**  $i = 1 : n_i$
  - 6:     将关键帧  $P_i$  转换为图像依次输入深度学习网络得到阴影消除后的图像
  - 7:     对阴影消除后的图像再次进行阴影边界细节修复得到  $R_i$
  - 8: **end for**
  - 9: 用原阴影关键帧  $P_i$  作为 I 帧，按格式重编码原视频序列，得到对应的前后运动估计及差值补偿
  - 10: 在重编码视频序列上用处理后的关键帧  $R_i$  替换  $P_i$ ，保持原有前后运动估计及补偿不变，得到阴影消除后的完整视频序列
  - 11: 对可能的抖动进行进一步优化
  - 10: 得到阴影消除后的视频
- 

完整合成阴影消除后的 8 个样例视频前 3 帧如下图 4-5 所示。可以看出经过一系列优化，视频的阴影区域得到很好消除的同时维持了时间连续性。



(a) 静止手持摄像



(b) 静止手持摄像



(c) 静止手持摄像





(d) 横向运动摄像



(e) 纵向运动摄像



(f) 纵向运动摄像



(g) 原地运动摄像



(h) 原地运动摄像

图 4-5 阴影消除视频前 3 帧，第一列为首帧，第二列为第二帧，第三列为第三帧

本文提出的算法优势在于，在保持和其他视频阴影消除模型相近消除结果的前提下，通过提取关键帧，仅仅对关键帧进行图像阴影消除处理，其他帧通过与关键帧的关联性，使用运动补偿等通过帧间信息推测的方式，将非关键帧进行处理进而恢复得到完整帧。这样，就只需要优化少部分区域就可以极为便捷地解决

一个连续帧组的大部分连续区域的问题。对比传统方法将视频逐帧转换为图像，并对所有帧依次进行阴影消除，然后再将所有的帧依次顺序拼接，使用全局光流法等方法进行全局或局部运动检测，再通过检测结果进行时空连续性优化，本文使用的方法因为阴影消除要计算的帧仅仅只有关键帧的少数帧，所以大大减少了所需计算量。同时，因为直接使用了视频压缩编码时的运动检测及补偿方式，在时空连续性优化上具有天然优势，大大减少了进一步优化的需要。相较于以往的方法，在许多地方通过关键帧的处理方法减少了大量工作。降低了进一步优化的复杂度和耗时、占用资源等。

本文阴影消除算法的主要运行时间开销约由五个部分组成，上述 8 个样例耗时平均开销大致如下表所示。

表 4-1 大致时间开销

组别	提取关键帧	图像阴影消除	边界优化	视频编码	抖动优化
1	2 分钟	2 分钟	20 分钟	60 分钟	-
2	3 分钟	3 分钟	30 分钟	80 分钟	-
3	3 分钟	3 分钟	30 分钟	80 分钟	-
4	3 分钟	3 分钟	30 分钟	80 分钟	-
5	2 分钟	2 分钟	20 分钟	60 分钟	-
6	6 分钟	6 分钟	60 分钟	160 分钟	-
7	2 分钟	2 分钟	20 分钟	60 分钟	-
8	4 分钟	4 分钟	40 分钟	100 分钟	-

#### 4.5 算法结果比较

本节将对比 Hu et al. 基于 DSC 模型的结果<sup>[35]</sup>。对比不同关键帧图像，结果如图 4-6，可见本文方式取得了不错的效果。



(a) 静止手持摄像





(b) 静止手持摄像



(c) 静止手持摄像



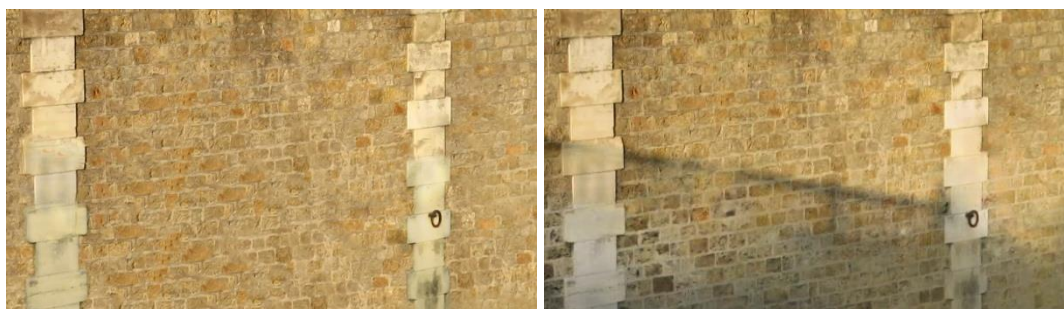
(d) 横向运动摄像



(e) 纵向运动摄像



(f) 纵向运动摄像



(g) 原地运动摄像



(h) 原地运动摄像

图 4-6 阴影消除结果对比，左侧为本文结果，右侧为 GAN 结果。

如上所列示例图，在这几类测试集下，本文方法在阴影区域标定识别、光照还原区域色彩准确性、纹理细节保留程度、阴影区域及非阴影区域的边界过渡衔接等方面都取得了较好结果。

## 4.6 本章小结

本章内容在前面所有工作的基础上完成了图像阴影消除到视频阴影消除。引用了运动估计、差值补偿的想法，及 VGG-19 提取特征找出关键帧。通过将原有带阴影的关键帧替换为阴影消除后的关键帧，并保留原有其他帧序列运动估计、差值补偿的信息，在解码时直接将其他帧序列还原成阴影消除后的帧序列，实现了帧序列时间信息的保留，得到的阴影消除视频能较好地解决时空一致性的问题。对于还可能出现的抖动问题，也提出了进一步优化的可能方法。



---

## 5 总结

### 5.1 全文总结

视频阴影的检测和消除在图形学和计算机视觉研究上具有重要作用。其对视频分割、目标检测等任务会带来不少阻碍。针对视频阴影的消除将会提高计算机视觉和图形学上许多已有研究的算法性能。对于现有的许多视频阴影消除的工作都专注于解决静态摄像机拍摄的固定场景的视频，而全局处理动态视频的方式又不太理想，本文提供了一种较为广泛适用的视频阴影消除方法，可以处理由静态及动态摄像机拍摄的阴影视频。本文通过深度学习，利用深度学习网络模型对关键帧进行阴影消除，然后利用消除完的图像替换视频关键帧，用原有运动矢量信息构建其余非关键帧，合成一组完整保持时空一致性的图像序列，构成视频，极大节约了时间和资源消耗。同时，本文使用的边界瑕疵消除方式也能较好的修正边界纹理信息损失及光照恢复错误。

### 5.2 算法局限

本文采用的基于关键帧的阴影消除模型，在面对拍摄物体高速移动、场景频繁剧烈变化及强烈光照变化时将出现十分密集的关键帧，不得不以场来运动估计，此时因为只有极少的预测帧，密集的关键帧可能导致再次出现时空不一致的问题。

其次，面对场景纹理种类过多、过于复杂的场景时，因本文使用的是 VGG-19 的中间层特征不变原理，对于全局可能提取出不正确的特征，导致阴影边界修复失败。

最后，本文视频关键帧的阴影消除算法是基于深度学习的原理，数据集和待处理的视频关键帧间的关联性会影响阴影消除的结果，如果待处理的视频和数据集包含的视频相差较远，可能导致模型不适用，得不到较好的阴影消除结果。

### 5.3 未来工作

未来可以努力的方向有：对视频处理做进一步提升，使得面对更为复杂的场景，不得不使用密集关键帧时也能更好地应对；构建更为完整的深度学习数据集，以应对更多不同的场景。

---

## 参考文献

- [1] Finlayson G D, Hordley S D, Drew M S. Removing Shadows from Images[C]// European Conference on Computer Vision. 2002.
- [2] Finlayson G D, Hordley S D, Lu C, et al. On the removal of shadows from images[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(1):59-68.
- [3] Reinhard E, Ashikhmin M, Gooch B, et al. Color Transfer between Images[J]. IEEE Computer Graphics & Applications, 2002, 21(5):34-41
- [4] Shor Y, Lischinski D. The Shadow Meets the Mask: Pyramid-Based Shadow Removal[J]. Computer Graphics Forum, 2010, 27(2):577-586.
- [5] Xiao C, She R, Xiao D, et al. Fast Shadow Removal Using Adaptive Multi-Scale Illumination Transfer[J]. Computer Graphics Forum, 2013, 32(8):207-218.
- [6] Xiao C, Xiao D, Zhang L, et al. Efficient Shadow Removal Using Subregion Matching Illumination Transfer[J]. Computer Graphics Forum, 2013, 32(7):421-430.
- [7] Zhang L, Zhang Q, Xiao C. Shadow Remover: Image Shadow Removal Based on Illumination Recovering Optimization[J]. IEEE Transactions on Image Processing, 2015, 24(11):4623-36.
- [8] Chuang Y Y, Dan B G, Curless B, et al. Shadow matting and compositing[J]. Acm Transactions on Graphics, 2003, 22(3):494-500.
- [9] Wu T P, Tang C K, Brown M S, et al. Natural shadow matting[J]. Acm Transactions on Graphics, 2007, 26(2):8.
- [10] Prati A, Mikic I, Trivedi M M, et al. Detecting moving shadows: algorithms and evaluation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2003, 25(7):918-923.
- [11] Sanin A, Sanderson C, Lovell B C. Shadow detection: A survey and comparative evaluation of recent methods[J]. Pattern Recognition, 2012, 45(4):1684-1695.
- [12] Russell M, Zou J J, Fang G. An evaluation of moving shadow detection techniques[J]. Computational Visual Media, 2016, 2(3):195-217.

- 
- [13] Jung C R. Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences[J]. IEEE Transactions on Multimedia, 2009, 11(3):571-577.
- [14] Porikli F, Thornton J. Shadow Flow: A Recursive Method to Learn Moving Cast Shadows[C]// Tenth IEEE International Conference on Computer Vision. 2005.
- [15] Xu D , Li X , Liu Z , et al. Cast shadow detection in video segmentation[J]. Pattern Recognition Letters, 2005, 26(1):91-99.
- [16] Zhang W , Fang X Z , Yang X K , et al. Moving Cast Shadows Detection Using Ratio Edge[J]. IEEE Transactions on Multimedia, 2007, 9(6):1202-1214.
- [17] Zhou L, Huang K, Tan T, et al. Cast Shadow Removal Combining Local and Global Features[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2007.
- [18] Martel-Brisson N, Zaccarin A. Learning and removing cast shadows through a multidistribution approach.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(7):1133-1146.
- [19] Martel-Brisson N, Zaccarin A. Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2008.
- [20] Huang J B , Chen C S . Moving cast shadow detection using physics-based features[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2009.
- [21] Benedek C, Sziranyi T. Bayesian foreground and shadow detection in uncertain frame rate surveillance videos.[J]. IEEE Transactions on Image Processing, 2008, 17(4):608-621.
- [22] Huerta I, Holte M, Moeslund T, et al. Detection and removal of chromatic moving shadows in surveillance scenarios[C]// IEEE International Conference on Computer Vision. 2009.
- [23] Amato A, Huerta I, Mozerov M G, et al. Moving Cast Shadows Detection Methods for Video Surveillance Applications[M]// Wide Area Surveillance. 2014.

- 
- [24] Zhou L, Huang K, Tan T. Cast Shadow Removal in a Hierarchical Manner Using MRF[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2012, 22(1):56-66.
- [25] Amato A, Mozerov M G, Bagdanov A D, et al. Accurate Moving Cast Shadow Suppression Based on Local Color Constancy Detection[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2011, 20(10):2954-66.
- [26] Sunkavalli K, Matusik W, Pfister H, et al. Factored time-lapse video[C]// Acm Siggraph. 2007.
- [27] Dammavalam S R, Piciarelli C, Micheloni C, et al. Shadow Removal in Outdoor Video Sequences by Automatic Thresholding of Division Images[C]// International Conference on Image Analysis & Processing. 2009.
- [28] Kar A , Deb K . Moving cast shadow detection and removal from Video based on HSV color space[C]// International Conference on Electrical Engineering & Information Communication Technology. IEEE, 2015.
- [29] Li Q, Bernal E A, Shreve M, et al. Scene-independent feature- and classifier-based vehicle headlight and shadow removal in video sequences[C]// IEEE Winter Applications of Computer Vision Workshops. 2016.
- [30] Barrow H G. Interactive aids for cartography and photo interpretation[J]. Stanford Univ Report, 1978.
- [31] Haskell B G, Puri A, Netravali A N. Digital Video: An introduction to MPEG-2[J]. Journal of Electronic Imaging, 1998, 7(1):539-545.
- [32] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [33] Yang C, Lu X, Lin Z, et al. High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis[J]. 2016.
- [34] Qu L , Tian J , He S , et al. DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.
- [35] Hu X , Zhu L , Fu C W , et al. Direction-aware Spatial Context Features for Shadow Detection[J]. 2017.

---

## 致谢

本研究及学位论文是在我的导师肖春霞教授的亲切关怀和悉心指导下完成的。他严肃的科学态度、严谨的治学精神、精益求精的工作作风深深感染和激励着我。肖老师不仅在学业上给予我精心指导，同时还在思想上引领我前行，在此谨向肖老师致以诚挚的谢意和崇高的敬意。同时，也要感谢不断给予我帮助的张玲学姐，本文的顺利完成同样离开学姐的耐心帮助，还有卓工班导师团队一路走来对我从学习到精神上的指路。

其次，也感谢这篇论文所涉及到的各位学者，本文引用了数位学者的研究文献，若没有各位学者的研究成果的帮助和启发，我也将难以完成本篇论文的写作。再则，也要感谢学院为我提供良好的毕业设计写作环境。

最后，再次对关心、帮助我的老师和同学们表示衷心地感谢！