

结合视觉舒适度的无参考立体视频稳像效果评价

吴剑荣, 黄华*

(北京师范大学人工智能学院 北京 100875)
(huahuang@bnu.edu.cn)

摘要: 立体视频稳像效果评价是评价立体视频稳像算法性能的有效途径。针对当前缺乏立体视频稳像效果客观评价方法的问题, 提出一种结合视觉舒适度的无参考立体视频稳像效果评价方法。将立体视频稳像前后的运动平滑度和视觉舒适度变化作为视频稳像前后的变化特征, 结合主观评价训练得到立体视频稳像效果评价的支持向量回归模型; 回归模型通过学习立体视频稳像前后的变化特征与主观评价结果之间的关系, 最终获得直接评价任意立体视频稳像效果的能力。使用收集的 55 条仿真视频训练模型, 并在 10 条真实视频上进行实验的结果表明, 所提方法的稳定性较好, 在视频量达到 180 条时, 模型的评价结果趋于稳定, 且模型评价结果与主观评价结果相关性达到 93%, 可用于立体视频稳像效果的客观评价。

关键词: 立体视觉; 视频稳像; 视频质量评价; 视觉舒适度

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024.19967

Non-Reference Assessment of Stereoscopic Video Stabilization Joint Visual Comfort

Wu Jianrong and Huang Hua*

(School of Artificial Intelligence, Beijing Normal University, Beijing 100875)

Abstract: Assessment of stereoscopic video stabilization is an effective way to evaluate the performance of stereoscopic video stabilization algorithm. For the deficiency of stereoscopic video stabilization assessment method, a novel non-reference assessment scheme is proposed joint visual comfort. The variation of motion stability and visual comfort are employed to represent variation characteristics between stereoscopic video and the stabilized video, and a support vector regression model is trained for stereoscopic video stabilization assessment, combining with subjective assessment. The regression model get ability that can correctly assess stereo video stabilization algorithm's product by learning the map between variation characteristics and subjective assessment. 55 simulation videos were collected to train the model. The experiments on 10 real videos demonstrate that the proposed schema has good stability, the model is convergent with the train video's number up to 180. The result's correlation between model and human subjective assessment is 93%, suggests the model can be used for object assessment of stereoscopic video stabilization.

Key words: stereo vision; video stabilization; video quality assessment; visual comfort

视频稳定(简称稳像)指利用相关的算法减少原始视频中的抖动, 使得视频中的运动更平滑。视频稳像一方面是为了让人眼观感舒适, 另一方面

也有助于后续的压缩、分析和理解等任务。立体视频稳像是用于立体视频的稳像技术。有效的立体视频稳像效果评价方法可以让研究者高效地判断

立体视频稳像算法优劣,促进立体视频稳像算法的研究。

依据评价形式,立体视频稳像效果评价方法分为主观评价方法和客观评价方法。主观评价方法需要设置合适的问题来统计人的主观感受;客观评价方法则是通过统计视频图像的相关信息,计算符合人主观感受的指标。与客观评价方法相比,主观评价方法虽然能更准确且直接地反映算法效果,但收集人的主观感受耗费的成本较高。因此,需要高效且准确地反映人的主观感受的客观评价方法,才便于对大量算法进行多次测试。

单目视频稳像评价和立体视频质量评价是 2 个与客观立体视频稳像效果评价密切相关的方法。单目视频稳像评价方法使用二维模型描述视频的运动,很难被扩展到三维;且因为单目视觉机制不同于双目视觉机制,所以单目视频稳像评价方法并不适用于立体视频稳像效果评价。立体视频质量评价方法从图像质量、深度感知和视觉舒适度 3 个方面进行评价,仅有少数立体视频质量评价方法^[1-2]借助运动信息评价视频的质量;但这些运动信息只考虑视频运动的速度和方向,没有考虑运动变化的剧烈程度对观感的影响。视频稳像解决的首要问题是平滑视频的运动,所以立体视频质量评价方法也难以直接应用于立体视频稳像效果评价。

在观看视频的过程中,观看者通过运动感知判断视频是否抖动,通过立体视觉感知获得三维的观看体验。视频稳像需要通过图像变形平滑视频的运动,在这个过程中,立体视频的运动和视差都发生了改变,从而引起视觉舒适度的变化。所以,立体视频稳像效果评价考虑的主要因素应该包含稳像前后视频运动平滑度和立体视觉舒适度的变化。

根据有无准确的参考真值,质量评价可以分为有参评价和无参评价 2 种。由于视频稳像是一个反问题,很难获得准确的参考视频,因此本文基于稳像前后立体视频的运动平滑度和视觉舒适度的变化,利用主观评价数据建立一个无参考立体视频稳像效果评价的回归模型。鉴于高精度的深度图可以提高模型评价的准确性,本文针对立体视频稳像评价问题,收集一个包含精确深度信息的抖动立体视频数据集,用于回归模型训练。该数据集通过模拟仿真获得,将运动传感器接入物理引擎,使虚拟相机的运动和传感器的运动保持一致,并借助物理引擎获得场景的高精度深度图。

1 相关工作

1.1 立体视频质量评价

立体视频质量评价方法通常被分为 2 类。

(1) 由单目视频质量评价方法扩展而来。首先用单目视频质量评价方法对左右视图的视频分别进行评价,然后结合左右视图的单目评价结果评价立体视频的质量。这类方法使用峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity, SSIM) 等针对单目视频设计的评价指标,但因为这些指标无法有效地表示视频场景的深度信息,所以不适用于立体视频质量评价。

(2) 基于人眼视觉机制的立体视频质量评价方法。一些方法依据人眼对运动物体更敏感的特性设计评价方法, Galkandage 等^[1]利用人眼对各个方向运动敏感度不一致的特性,赋予各个像素不同权重; López 等^[2]通过视频的运动速度直方图评价立体视频质量。双目的融合与竞争机制被广泛应用在立体视频质量评价中, Yang 等^[3]将左右视图叠加,对叠加图进行稀疏编码,可以有效地表征立体视频的视差信息; Chen 等^[4]发现,把左右视图相减获得的残差图也可以表示图像的视差信息; Yang 等^[5]在时域上应用双目竞争机制,发现交替提取左右视频的时空信息比单侧视频的时空信息能更贴合人的主观感受。受到视觉神经系统设计网络结构的启发, Feng 等^[6]使用三维卷积提取时空特征,并引入多阶段注意力增长机制模拟人眼视觉的多阶段处理。

以上方法大多侧重于视频图像的空域信息而忽略了视频的运动信息。考虑了运动信息的立体视频质量评价方法,要么只考虑像素速度和方向对观感的影响^[1],忽略了速度连续变化造成的影响;要么使用速度直方图特征评价立体视频质量^[2],丢失了运动的时序关系,也无法反映视频的平滑程度。综上所述,现有的立体视频质量评价方法都无法准确地反映稳像前后运动的变化。

1.2 单目视频稳像评价

单目视频稳像评价方法提取视频的时空信息,根据时空信息变化的剧烈程度评判视频稳像的质量。依据是否有参考真值,单目视频稳像评价方法可以分为有参视频稳像评价方法和无参视频稳像评价方法。

有参视频稳像评价方法常以专业云台拍摄的平滑视频作为真值,在平稳视频上添加额外的运动,合成抖动视频。这类方法借助 SSIM、帧间距离^[7]和频域变化比例^[8]等视频质量评价方法,计算稳像视

频与真值之间的差异,进而评价视频稳像质量。

无参视频稳像评价方法中,使用能反映观看者主观感受的客观指标评价视频稳像质量。对于运动幅度较小的视频,可以使用图像的 PSNR^[9]和 SSIM 计算相邻 2 帧的相对偏移量,以此评价视频稳像效果。但当运动幅度较大时,帧间图像差异过大,难以准确地评价稳像质量,使得这些方法可靠性差。一些方法使用运动模型表述视频的运动,借助运动模型的属性估计视频的平滑度。例如,在频域上计算运动的平滑度,将视频的二维运动轨迹作傅里叶变换,使用低频部分代表运动的平滑度^[10];或使用单应矩阵描述视频的运动,在流形空间中计算单应矩阵序列的几何曲率,以几何曲率作为运动平滑度^[11]。

单目视频稳像评价方法大多基于二维的运动模型,没有涉及视差,而视差是产生深度感知的重要因素,也是影响立体视频视觉舒适度的重要因素。所以,单目视频稳像评价方法无法准确地反映立体视频稳像算法对立体视频造成的观感影响。

1.3 立体视频数据集

根据数据来源的不同,目前具有深度真值的立体视频数据集可以分为真实数据与合成数据 2 类。

(1) 真实数据集。如 KITTI^[12]和 DrivingStereo^[13]等,通常来自于自动驾驶以及 SLAM 等领域,这些数据集对空间分辨率要求更高,导致采样频率低。如 KITTI 的采样频率为 10 Hz,而消费级相机的采样频率通常高于 30 Hz。且这些视频是在驾驶过程中拍摄的,运动的模式比较单一,无法全面地反映立体视频稳像算法的效果。一些立体视频稳像算法提供的数据集^[14-15]较小且没有精确的视差信息且无法准确地评判立体视频稳像算法对视差的影响。

(2) 合成数据集。如 SceneFlow^[16]和 UrbanScene3D^[17]等,它们借助计算机图形学技术合成仿真视频,并通过模型的参数信息计算视频的视差信息。但是,虚拟相机都依照预定的平滑轨迹运动,合成的视频已经足够平滑。

现有的立体视频数据集在运动模式、采样频率和数据量大小等方面都存在一定的不足,无法用于立体视频稳像效果评价。因此,需要构建一个运动模式多样、采样频率适当且数据量充足的立体视频数据集。

2 立体视频稳像客观评价的特征描述

本文依据稳像前后立体视频运动平滑度和视

觉舒适度的变化,评价立体视频稳像效果。但是,这 2 方面都存在各自的难点:(1) 立体视频的三维运动难以被精确表示,也难以定义其平滑度;(2) 立体视觉舒适度感知是人的主观感受,很难被量化。

运动场是一种被广泛应用的运动表述方法^[18]。在单目视频稳像的研究中,Liu 等^[19-20]提出基于运动场的稳像算法,指出运动场的平滑与视频运动的平滑有着密切的联系。所以可以使用三维运动场的平滑度表达立体视频的运动平滑度。

视频稳像需要通过图像变形改变视频的运动,所以视频稳像在改变视频运动的同时必然会引入图像形变。Du 等^[21]通过实验探究了立体视频的运动速度对视觉舒适度影响;Kooi 等^[22]发现,在图像的形变超过一定阈值时,会明显降低视觉舒适度。受到以上 2 项工作的启发,本文分别将运动视觉舒适度变化和形变视觉舒适度变化作为评价稳像视觉舒适度变化的特征。其中,运动视觉舒适度变化反映稳像后运动速度变化对视觉舒适度的影响;形变视觉舒适度变化反映稳像后图像的形变对视觉舒适度的影响。

下面分别介绍立体视频运动平滑度变化、运动视觉舒适度变化和形变视觉舒适度变化 3 个特征。

2.1 立体视频运动平滑度

首先使用场景流表示立体视频三维运动场,然后将二维运动场的平滑度计算方法扩展到三维,最终结合人眼对局部差异敏感的特性计算立体视频运动平滑度变化。

2.1.1 三维运动场

运动场是一个关于坐标的函数,描述了运动向量在空域上的分布。在单目视频中,常使用光流表示二维运动场,而对于立体视频,则使用场景流表示三维运动场。场景流是光流的扩展,在光流的基础上,场景流增加了对图像的视差估计和视差变化量的计算,即场景流包含了 (x, y) 处像素在 t 时刻的视差 $d := d(x, y, t)$,以及对应的三维运动向量 $M_t(x, y) = [u_t, v_t, q_t]^T$;其中, u_t , v_t 和 q_t 分别表示 t 时刻像素对应的三维点在水平、垂直和深度方向上的运动偏移量。

最直接的场景流算法是分别求解 2 帧立体图像的视差图,然后利用光流估计得到相邻 2 帧图像像素的对应关系,进而求解出视差变化量。但因为光流估计与视差估计本身都存在一定的误差,使得直接组合求解得到的场景流不可靠,所以需要合适的场景流算法计算立体视频的三维运动场。

2.1.2 基于三维运动场的像素轮廓

基于二维运动场的单目视频稳像算法^[19-20]将运动向量按照时序排列获得如图 1a 所示像素轮廓,通过对像素轮廓作平滑处理达到视频稳像效果.基于三维运动场的像素轮廓则将三维运动向量按照时序排列,因为立体视频在像平面和深度方向的运动对视觉感知都有影响,而基于二维运动场的像素轮廓无法反映深度方向的运动,所以本文针对立体视频应用场景流计算三维运动场,使用基于三维运动场的像素轮廓表述视频运动.即在立体视频 (x, y) 处像素的三维像素轮廓为 $\mathbf{M}(x, y) = \{\mathbf{M}_1(x, y), \mathbf{M}_2(x, y), \dots, \mathbf{M}_N(x, y)\}$, 其中, N 表示视频的帧数, $\mathbf{M}(x, y)$ 表示一个(如图 1b 所示)三维向量的序列.

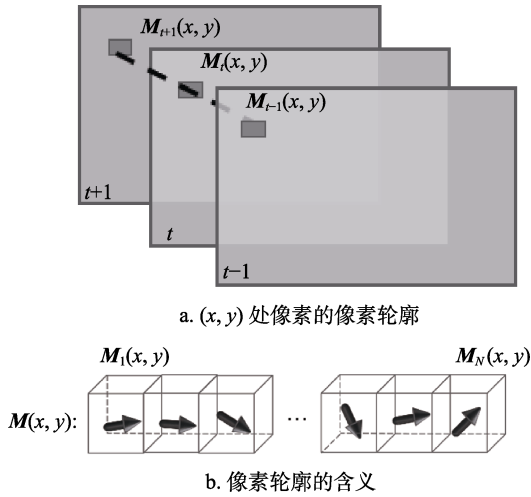


图 1 像素轮廓

2.1.3 平滑度计算

一阶变化量和二阶变化量常用于反映运动的平滑程度.许多视频稳像算法假设平滑运动是速度和加速度较小的运动,所以在路径平滑处理时,须对运动序列的一阶变化量和二阶变化量进行平滑^[23-24].由此,本文定义 (x, y) 处像素轮廓的平滑度为

$$F_{ps}(x, y) = \lambda_1 \sum_{t=1}^N \|\mathbf{M}_t(x, y)\|^2 + \lambda_2 \sum_{t=1}^N \|\mathbf{M}_t(x, y) - \mathbf{M}_{t-1}(x, y)\|^2.$$

其中,权重系数 λ_1 和 λ_2 的作用是调节一阶变化量和二阶变化量对平滑度的影响; t 表示视频帧的序号.视频 V 的平滑度指视频上各个像素轮廓的平滑度均值

$$F_{vs}^V = \frac{1}{H \times W} \sum_{x=1}^W \sum_{y=1}^H F_{ps}(x, y).$$

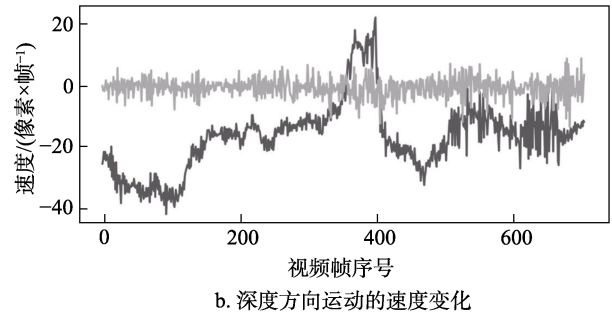
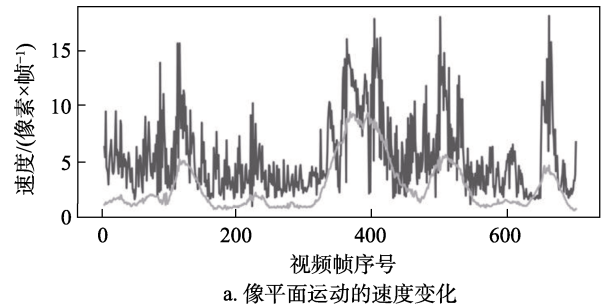
其中, H 和 W 分别表示视频的长度像素数和宽度像素数.

2.2 运动视觉舒适度变化

如图 2 所示,无论在成像平面还是深度方向,立体视频稳像都会显著地改变原始视频的运动. Du 等^[21]给出某一时刻视觉舒适度 C 和立体视频内容的运动速度之间的关系为

$$C(v_{xy}, v_d, d) = p_1 v_{xy} + p_2 v_z + p_3 d + p_4 v_{xy} v_z + p_5 v_{xy} d + p_6 v_z d + p_7.$$

其中, v_{xy}, v_d 分别表示像素在成像平面和深度方向的运动速度; d 表示视差; $p_1, p_2, p_3, p_4, p_5, p_6$ 和 p_7 表示参数.



——原始视频; —— 稳像视频

图 2 稳像前后速度对比

因为人眼对局部差异明显的区域会更敏感,所以计算运动视觉舒适度时依照光照强度的局部差异调节像素权重^[25].最终,稳像视频 V 的运动视觉舒适度为

$$F_{mc}^V = \sum_{t=1}^N \sum_{x=1}^W \sum_{y=1}^H C(v_{xy}^V(x, y, t), v_d^V(x, y, t), d^V(x, y, t)) \cdot \frac{I^V(x, y, t)}{\sum_{(u, v) \in \Omega(x, y)} I^V(u, v, t)}.$$

其中, $I^V(x, y, t)$ 表示稳像视频的第 t 帧在 (x, y) 处的光照强度; $\Omega(x, y)$ 表示 (x, y) 处邻域的 8 个像素.视觉舒适度变化为稳像后视频 V 和对应原始视频 V^* 的运动视觉舒适度的差,公式为

$$F_{\text{vmc}}^V = F_{\text{mc}}^V - F_{\text{mv}}^{V*}.$$

2.3 形变视觉舒适度变化

为了使得视频内容符合稳像算法所规划的平滑运动轨迹, 通常需要对视频图像进行变形, 而立体图像的变形会改变视觉舒适度. 当立体视频图像的放缩比例超过一定阈值时, 会严重降低视觉舒适度^[22]. 针对阈值判别问题, 一些工作采用平方对数的形式进行建模^[25]. 本文参照该形式对稳像引起的视觉舒适变化进行建模, 即

$$F_d(e) = p_8 \log^2(e) + p_9 \log(e) + p_{10}.$$

其中, e 表示图像发生的形变量; p_8 , p_9 和 p_{10} 表示参数. 当形变量较小时, 形变视觉舒适度的变化较小; 当形变量增大到一定程度时, 形变视觉舒适度则会明显降低.

通常, 稳像算法基于网格对视频图像进行扭曲变形. 在计算形变量时, 对于第 t 帧视频, 本文将其等分成如图 3 所示 4×4 块区域, 分别计算每个区域稳像前后的仿射矩阵 E , 并以仿射矩阵最大的 2 个特征值 e_1 和 e_2 之比作为该区域的形变量^[10], 即 $e = e_1/e_2$. 图像的形变量为各区域形变量的均值, 即 $\bar{e} = \frac{\sum \sum e}{16}$.

视频 V 的形变视觉舒适度变化为所有帧的形变视觉舒适度变化的均值, 即

$$F_d^V = \sum_{t=1}^N F_d(e_V(t));$$

其中, $e_V(t)$ 表示视频 V 的第 t 帧的形变量.

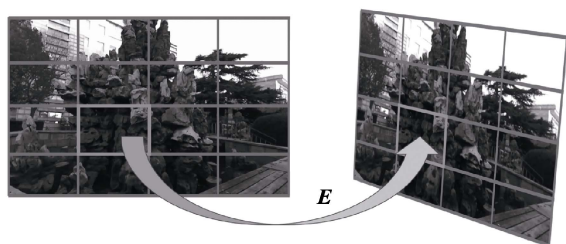


图 3 划分网格区域计算形变量

3 仿真立体视频数据集

一个高质量的数据集对算法评估是至关重要的, 实际环境中很难获取一个包含精确深度信息且剧烈抖动的立体视频. 因此, 本文在物理引擎中

合成仿真视频, 生成深度图.

3.1 数据制作方法

本文的视频内容通过 Unreal 4^①合成. 借助 Unreal 4 提供的虚拟拍摄功能, 可以将相机运动传感器接入到物理引擎中, 使得其中预设的虚拟相机和现实世界中的相机保持一致地运动, 由此, 拍摄过程中的抖动能直接引入合成的仿真视频中. 因为物理引擎中各个组件的坐标和形状都是明确的, 所以可以方便地计算场景到相机的距离. 本文借助 UnrealCV^[26]计算场景的深度图, 并随后依据指定的双目相机的基线距离将深度图转化为视差图.

使用以上 2 个工具可以获得深度信息精确且带有实际拍摄抖动的立体仿真视频. 如图 4 所示, 本文制作的数据集包含丰富且逼真的场景, 在复杂场景下也可以得到精确的深度图.

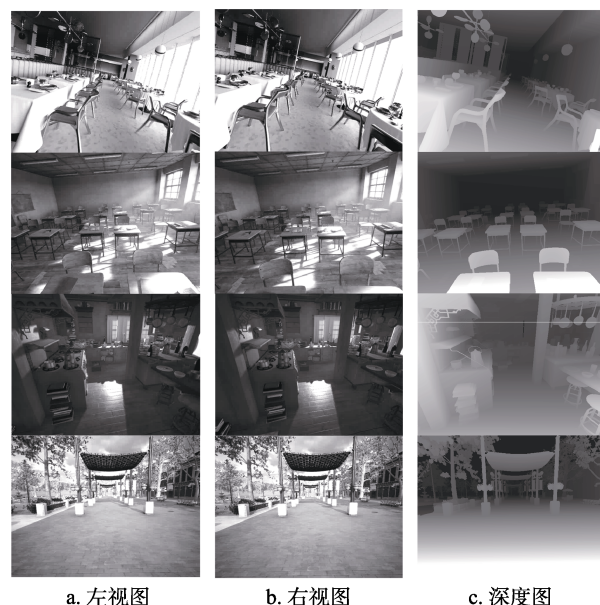


图 4 数据集视频

3.2 数据属性

日常生活中, 随意拍摄的视频包含复杂的运动, 单一运动模式无法测试稳像算法的鲁棒性. 为了全面地评价立体视频稳像算法, 本文将常见的运动划分成简单抖动、行走、旋转、奔跑和攀爬 5 种类型, 其属性如表 1 所示. 各个类型的含义如下:

(1) 简单抖动. 在自然拍摄过程中拍摄者长期手持相机, 难以长时间保持稳定, 导致视频出现轻微的高频抖动.

① <https://www.unrealengine.com>

(2) 行走. 边走边拍是一种最为常见的拍摄方式, 视频往往抖动幅度大而抖动频率适中.

(3) 旋转. 为了展示周围的环境, 拍摄者通常会旋转拍摄四周, 虽然这类视频的运动幅度大, 但是抖动频率低.

(4) 奔跑. 奔跑拍摄的视频运动幅度大、频率高, 通常无法提取充足的特征点, 所以对于稳像算法极具挑战.

(5) 攀爬. 攀爬楼梯过程中拍摄的视频在垂直方向有大幅移动, 能测试稳像算法在垂直方向的稳像能力.

表 1 数据表属性

运动模式	数量/条	时长/s
简单抖动	16	30~60
行走	12	12~15
旋转	10	5~7
攀爬	10	4~6
奔跑	7	6~10

3.3 视频的稳像处理

当前已公开的立体视频稳像算法主要有 Subspace 立体视频稳像(subspace stereo video stabilization, SS)算法^[14], Joint-BundlePath 立体视频稳像(Joint-BundlePath stereo video stabilization, J-BP)算法^[15], MeshFlow 立体视频稳像(MeshFlow stereo video stabilization, MF)算法^[27], 基于三维电影制作原则的立体视频稳像(3D cinematography principles video stabilization, 3DVS)算法^[28], 基于三维移动设备的立体视频稳像算法(3D mobile video stabilization, 3DMVS)^[29]等. 其中, 由于 3DMVS 算法需要陀螺仪辅助, 因此不适用于本文的实验; MF 算法是在 J-BP 算法的基础上针对视差偏移问题进行了优化. 本文最终选取 SS, MF 和 3DVS 这 3 种立体视频稳像算法对抖动视频进行稳像处理, 每条视频经过不同的稳像算法处理后得到一组视频: 原始版本、SS 稳像版本、MF 稳像版本和 3DVS 稳像版本.

3.4 主观评价

依据评价者的主观感受对稳像前后的视频进行评分. 本文设计了如图 5 所示的交互界面, 便于评价者观看和评价视频, 评价者通过色差式立体眼睛观看立体视频. 为了保证实验的公平性, 每组 4 条不同版本的视频被连续播放, 但各个版本的播放顺序是随机的. 由于立体视频很容易引起视频疲劳, 因此为了不影响视频视觉舒适度的评分, 要

求评价者在观看视频后休息一段时间, 休息时间与上一段视频的时长相同.



图 5 主观评价交互界面

对每段视频, 评价者都需要回答 3 个问题:

(1) 你的眼睛感受如何? 即视频的视觉舒适度 S_{vc} . 其中, 1~5 分分别表示疲劳得难以忍受、疲劳且有眼部不适、明显疲劳但在耐受范围内、稍微疲劳和无疲劳感.

(2) 视频抖动程度如何? 即平滑质量 S_{sc} . 其中, 1~5 分分别表示非常抖、比较抖、抖动、稍微抖和平稳.

(3) 请结合视觉舒适度和视频抖动对视频做综合质量评价. 即综合稳像质量 S . 其中, 1~5 分分别表示非常差、较差、一般、好和非常好.

为了防止评价者对视觉舒适度和视频平滑质量的概念产生误解, 本文在主观实验前告知评价者以下原则:

(1) 视频的平滑质量和眼睛的舒适度都没有明确的定义, 完全依据人的主观感受.

(2) 视觉的不舒适主要指眼睛出现视力模糊、眼痒、眼胀、头晕和恶心等症状.

(3) 为了能更好地区分抖动的程度, 评价者可以更多地关注场景中的动态物体和场景边界的运动.

最终, 本文共计收到 80 份问卷, 取主观评价的平均值作为视频各项指标的最终值.

4 回归模型训练与推理

支持向量回归(support vector regression, SVR)模型被广泛应用在视频图像质量评价^[4,30]. 为了评价立体视频稳像质量, 利用第 3 节介绍的 3 个客观特征和主观评价结果训练 SVR. 本文使用立体视频运动平滑度变化、运动视觉舒适度变化和形变视

觉舒适度变化作为稳像视频的变化特征, 以主观评价结果作为视频的标签进行训练.

4.1 变化特征参数计算

Du 等^[21]使用其设置的主观实验求解运动视觉舒适度 C 的参数. 因为本文实验中使用视频与文献[21]不同, 所以需要应用本文的主观实验结果计算参数. 运动视觉舒适度包含参数 $\mathbf{P}_{mc} = [p_1, p_2, p_3, p_4, p_5, p_6, p_7]$, 使用模型训练视频的主观视觉舒适度评分 S_{vc} 确定 $\mathbf{P}_{mc}$. 使用所有样本视频求解二次优化问题

$$\arg \min_{\mathbf{P}_{mc} \in \mathbb{R}^{10}} \sum_{i=1}^Z (F_{mc}^i - S_{vc}^i)^2.$$

其中, i 表示样本编号; Z 表示样本数量. 满足上式的 $\mathbf{P}_{mc}$ 使得样本视频的运动视觉舒适度变化特征与对应的主观视觉舒适度评价结果接近一致.

对形变视觉舒适度变化 F_d 的参数 $\mathbf{P}_d = [p_8, p_9, p_{10}]$ 也按照相似的形式进行求解, 即

$$\arg \min_{\mathbf{P}_d \in \mathbb{R}^3} \sum_{i=1}^Z (F_d^i - S_{vc}^i)^2.$$

立体视频运动平滑度变化 F_{ps} 的参数 $\lambda = [\lambda_1, \lambda_2]$ 则通过 S_{sc} 确定, 公式为

$$\arg \min_{\lambda \in \mathbb{R}^2} \sum_{i=1}^Z (F_{ps}^i - S_{sc}^i)^2,$$

使样本的立体视频运动平滑度变化特征符合主观的运动平滑评价.

4.2 模型训练

模型的训练过程如图6所示, 使用立体视频运动平滑度变化、运动视觉舒适度变化和形变视觉舒适度变化, 作为第 i 号样本的变化特征 $\mathbf{F}_i = [F_{vs}^i, F_{vmc}^i, F_d^i]^T$, 主观实验的综合评分 S_i 作为标签. SVR 的目标函数为

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^Z (c_{\varepsilon}(\xi_i^+) + c_{\varepsilon}(\xi_i^-)) \\ \text{s.t. } S_i - (\mathbf{w}^T \phi(\mathbf{F}_i) + b) \leq \varepsilon + \xi_i^+ \\ S_i - (\mathbf{w}^T \phi(\mathbf{F}_i) + b) \geq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0, i=1, 2, \dots, Z \end{cases}.$$

其中, $\mathbf{w}$ 表示超平面; C 表示惩罚损失函数; b 表示偏置; ξ_i^+, ξ_i^- 表示松弛变量; $\phi(\cdot)$ 表示高维特征空间的映射; $c_{\varepsilon}(\cdot)$ 表示损失函数; ε 表示不敏感损失因子; Z 表示松弛标量的数量. 因为样本数量较少, 所以本文基于经验选择将变化特征映射到高斯特征空间, 则 $\phi(\cdot)$ 对应的核函数为

$$K(\mathbf{F}_i, \mathbf{F}_j) = \exp\left(-\frac{\|\mathbf{F}_i - \mathbf{F}_j\|^2}{2\sigma^2}\right).$$

其中, $\sigma > 0$ 表示高斯核参数. 当误差超过 ε 时, 计算损失 $c_{\varepsilon}(\xi) = |\xi|_{\varepsilon} = \max\{0, |\xi| - \varepsilon\}$.

每次训练过程中, 模型参数都通过网格搜索选取.

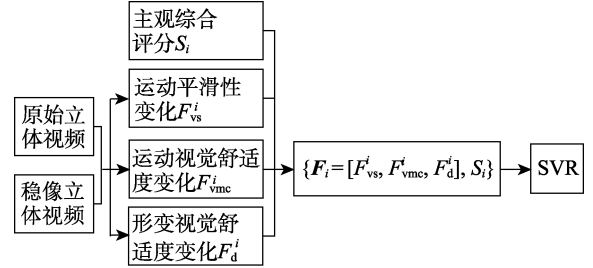


图6 模型训练过程

4.3 模型推理

在利用训练得到的评价模型 $f(\cdot)$ 对立体视频的稳像效果进行评价前, 需要使用已经确认的特征参数计算立体视频的变化特征. 对于变化特征为 $\mathbf{F}$ 的立体视频, 其客观评价结果计算公式为

$$f(\mathbf{F}) = \mathbf{w}^T \phi(\mathbf{F}) + b.$$

其值越高, 表示立体视频的稳像效果越好.

5 实验及结果分析

立体视频稳像效果评价的目标是使得评价结果与主观感受相符, 可以通过主客观评价结果的一致说明评价模型的有效性. 使用第3节提出的数据集训练评价模型, 并使用评价模型对如图7所示的10组真实场景视频的视频稳像效果进行评价; 然后分别对主观评价结果和模型评价结果进行排序; 最终用统计学中常用的斯皮尔曼相关系数和肯德尔系相关系数, 表示主观评价结果与模型评价结果之间的一致性.



图7 部分真实视频示例

5.1 模型评价指标

斯皮尔曼相关系数的计算过程如下: 对稳像视

频的主观评价结果 $\{A_k\}$ 和客观评价结果 $\{f(F)_k\}$ 分别进行排序, 得到各个稳像算法的排序位次 $\{a_k\}$ 和 $\{b_k\}$, 则斯皮尔曼相关系数为

$$\rho = 1 - \frac{6 \sum_{k=1}^L \text{abs}(a_k - b_k)}{L(L^2 - 1)}.$$

其中, L 表示稳像算法的数量. 显然, 斯皮尔曼相关系数越高, 则 2 个集合的相关程度就越大.

肯德尔相关系数则为

$$\tau = \frac{n_c - n_d}{L(L-1)/2}.$$

其中, n_c 表示一致元素对 $\{(k, l) | a_k < a_l, b_k < b_l\}$ 的数量; n_d 表示不一致元素对的数量.

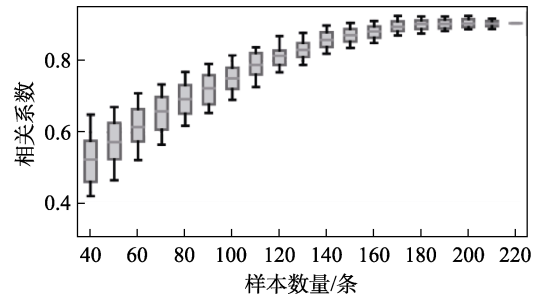
5.2 模型稳定性分析

为了评估本文模型的稳定性, 探究样本数大小对模型评价效果的影响. 训练数据的样本数量步长为 10, 从 40 条逐渐增加到 220 条. 除了样本数量为 220 条的数据只能选取一组外, 其余各个规模的训练数据都随机选取 1 000 组. 单次实验中, 先从 55 组仿真视频中随机选取视频组, 然后从组内选取原始视频和(随机选取)至少一条稳像视频加入训练数据中, 再使用训练数据计算 3 个特征的特征, 计算样本的特征并训练 SVR 模型, 最后用训练出的模型对 10 组真实场景视频进行评价. 实验结果如图 8 所示. 其中, 对同样本数量进行多次实验, 使用盒式图统计斯皮尔曼相关系数和肯德尔相关系数. 可以看出, 随着样本数量的增加, 相关系数的中位数逐渐增大, 并且相关系数的上下限也在逐步变小, 表明模型的评价精确度越来越高且波动越来越小; 当样本数量大于 180 条后, 模型的评价能力趋于稳定, 且达到较高的评价精度.

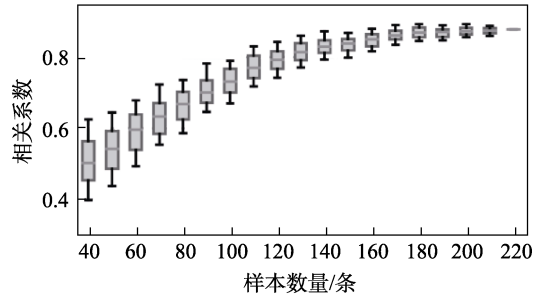
5.3 立体视频稳像算法效果评价

使用由 55 组仿真视频训练得到的模型对立体视频稳像算法的效果进行评估, 所用数据为 10 组真实拍摄的视频, 这些视频的处理和主观评价与第 3 节相同. 图 9 所示为 10 组真实场景视频的主观评价结果与本文模型的预测结果的对比. 可以看出, 模型评价结果的稳像效果排名和主观评价所呈现的结果基本一致; 10 组视频评价结果的平均斯皮尔曼相关系数 $\rho = 0.9207$, 平均肯德尔系数为 $\tau = 0.9317$, 说明模型评价结果与主观评价结果有很强的相关性.

实验结果表明, MF 的稳像效果整体优于 SS 和 3DVS, 这是因为 SS 和 3DVS 都依赖长轨迹, 而



a. 斯皮尔曼相关系数随样本数量变化



b. 肯德尔相关系数随样本数量变化

图 8 样本数量与模型评价精度关系

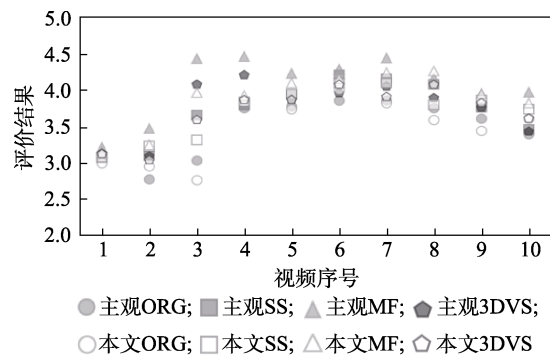


图 9 真实视频主客观评价结果

MF 不依赖长轨迹. 视频 3 是在奔跑过程中拍摄的, 由于运动剧烈, 因此无法提取充足的长轨迹, 导致 SS 和 3DVS 的稳像效果抖动依然明显, 且有明显形变, 所以在视频 3 的主客观评价中, MF 的稳像效果都是最优的; 视频 4 是对远景进行拍摄, 当场景的深度较大时, 场景流算法通常无法得到准确的结果, 导致计算出的立体视频运动平滑度变化和运动视觉舒适度变化都不可靠, 使得模型对 4 条不同版本视频的评价结果没有明显差异; 视频 8 的 3DVS 稳像效果在模型评价中优于 SS, 而主观评价中稍逊于 SS, 这是因为拍摄内容是灌木丛, 大量密集重复的纹理容易引起特征点的误匹配, 导致形变量计算不准确.

6 结 语

本文结合立体视频的视觉舒适度提出立体视

频稳像效果的无参考评价方法,并建立了一个针对立体视频稳像的立体视频数据集.本文方法同时考虑稳像算法带来的运动改变和图像形变,更符合立体视频稳像效果的主观评价.最后通过实验,证明了本文方法的有效性.未来将探究现实场景中立体视频的性质,如运动模糊、视频裁剪等对观感的影响.运动模糊会显著地影响视觉舒适度,且人眼对立体视频边界的运动十分敏感^[28].因此,构建一个结合视频裁剪、运动模糊等因素的综合立体视频稳像效果评价算法,是今后的工作方向.

参考文献(References):

- [1] Galkandage C, Calic J, Dogan S, *et al.* Full-reference stereoscopic video quality assessment using a motion sensitive HVS model[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(2): 452-466
- [2] López J P, Rodrigo J A, Jiménez D, *et al.* Stereoscopic 3D video quality assessment based on depth maps and video motion[J]. *EURASIP Journal on Image and Video Processing*, 2013, 2013: Article No.62
- [3] Yang J C, Ji C Q, Jiang B, *et al.* No reference quality assessment of stereo video based on saliency and sparsity[J]. *IEEE Transactions on Broadcasting*, 2018, 64(2): 341-353
- [4] Chen Z B, Zhou W, Li W P. Blind stereoscopic video quality assessment: from depth perception to overall experience[J]. *IEEE Transactions on Image Processing*, 2018, 27(2): 721-734
- [5] Yang J C, Zhao Y, Jiang B, *et al.* No-reference quality assessment of stereoscopic videos with inter-frame cross on a content-rich database[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(10): 3608-3623
- [6] Feng Y J, Li S M, Chang Y L. Multi-scale feature-guided stereoscopic video quality assessment based on 3d convolutional neural network[C] //*Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Los Alamitos: IEEE Computer Society Press, 2021: 2095-2099
- [7] Qu H, Song L, Xue G J. Shaking video synthesis for video stabilization performance assessment[C] //*Proceedings of the Visual Communications and Image Processing*. Los Alamitos: IEEE Computer Society Press, 2013: 1-6
- [8] Niskanen M, Silven O, Tico M. Video stabilization performance assessment[C] //*Proceedings of the IEEE International Conference on Multimedia and Expo*. Los Alamitos: IEEE Computer Society Press, 2006: 405-408
- [9] Dong Changqing, Cheng Xuemin, Hao Qun. Speed and precision's improvement in electronic image stabilization[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2018, 30(8): 1543-1553(in Chinese)
(董常青, 程雪岷, 郝群. 电子稳像算法的速度与精度改进[J]. *计算机辅助设计与图形学学报*, 2018, 30(8): 1543-1553)
- [10] Liu S C, Yuan L, Tan P, *et al.* Bundled camera paths for video stabilization[J]. *ACM Transactions on Graphics*, 2013, 32(4): Article No.78
- [11] Zhang L, Zheng Q Z, Huang H. Intrinsic motion stability assessment for video stabilization[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 25(4): 1681-1692
- [12] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The Kitti vision benchmark suite[C] //*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2012: 3354-3361
- [13] Yang G R, Song X, Huang C Q, *et al.* DrivingStereo: a large-scale dataset for stereo matching in autonomous driving scenarios[C] //*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2019: 899-908
- [14] Liu F, Niu Y Z, Jin H L. Joint subspace stabilization for stereoscopic video[C] //*Proceedings of the IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2013: 73-80
- [15] Guo H, Liu S C, Zhu S Y, *et al.* Joint bundled camera paths for stereoscopic video stabilization[C] //*Proceedings of the IEEE International Conference on Image Processing*. Los Alamitos: IEEE Computer Society Press, 2016: 1071-1075
- [16] Mayer N, Ilg E, Häusser P, *et al.* A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C] //*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2016: 4040-4048
- [17] Lin L Q, Liu Y L, Hu Y G, *et al.* Capturing, reconstructing, and simulating: the UrbanScene3D dataset[C] //*Proceedings of the 17th European Conference on Computer Vision*. Heidelberg: Springer, 2022: 93-109
- [18] Katok A, Hasselblatt B. Introduction to the modern theory of dynamical systems[M]. Cambridge: Cambridge University Press, 1995
- [19] Liu S C, Yuan L, Tan P, *et al.* SteadyFlow: spatially smooth optical flow for video stabilization[C] //*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2014: 4209-4216
- [20] Liu S C, Tan P, Yuan L, *et al.* MeshFlow: minimum latency online video stabilization[C] //*Proceedings of the 14th European Conference on Computer Vision*. Heidelberg: Springer, 2016: 800-815
- [21] Du S P, Masia B, Hu S M, *et al.* A metric of visual comfort for stereoscopic motion[J]. *ACM Transactions on Graphics*, 2013, 32(6): Article No.222
- [22] Kooi F L, Toet A. Visual comfort of binocular and 3D displays[J]. *Displays*, 2004, 25(2/3): 99-108
- [23] Wang Y S, Liu F, Hsu P S, *et al.* Spatially and temporally optimized video stabilization[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(8): 1354-1361
- [24] Grundmann M, Kwatra V, Essa I. Auto-directed video stabilization with robust L1 optimal camera paths[C] //*Proceedings of the CVPR*. Los Alamitos: IEEE Computer Society Press, 2011: 225-232
- [25] Didyk P, Ritschel T, Eisemann E, *et al.* A luminance-contrast-aware disparity model and applications[J]. *ACM Transactions on Graphics*, 2012, 31(6): Article No.184

- [26] Qiu W C, Yuille A. UnrealCV: connecting computer vision to unreal engine[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2016: 909-916
- [27] Guo H, Liu S C, Zhu S Y, *et al.* View-consistent meshFlow for stereoscopic video stabilization[J]. IEEE Transactions on Computational Imaging, 2018, 4(4): 573-584
- [28] Liu C W, Huang T H, Chang M H, *et al.* Stereoscopic media editing based on 3D cinematography principles[C] //Proceedings of the ACM SIGGRAPH Posters. New York: ACM Press, 2011: Article No.38
- [29] Chu C H. Video stabilization for stereoscopic 3D on 3D mobile devices[C] //Proceedings of the IEEE International Conference on Multimedia and Expo. Los Alamitos: IEEE Computer Society Press, 2014: 1-6
- [30] Zou Liangtao, Jiang Gangyi, Yu Mei, *et al.* No-reference image quality assessment of high dynamic range image based on tensor domain perceptual features[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(10): 1850-1858(in Chinese)
(邹良涛, 蒋刚毅, 郁梅, 等. 基于张量域感知特征的无参考高动态范围图像质量评价[J]. 计算机辅助设计与图形学学报, 2018, 30(10): 1850-1858)