doi:10.3969/j.issn.1001-4616.2025.02.009

VR 中基于改进 SIFT 特征匹配的手势检测与跟踪

苏　琳[1],邹　静[2],许　媚[3]

(1.闽南科技学院艺术设计学院,福建 泉州 362332)
(2.贵州大学计算机科学与技术学院,贵阳 贵州 550025)
(3.西北师范大学教育科学学院,甘肃 兰州 730070)

[摘要]　随着虚拟现实(virtual reality,VR)技术的迅猛发展,对自然人机交互的需求日益增长,其中手势识别技术扮演着至关重要的角色. 它不仅要求高度的准确性,还必须保证实时响应,以确保用户能够获得流畅的交互体验. 本研究提出了一种创新的手势检测与跟踪方法,该方法基于改进的尺度不变特征变换(scale invariant feature transform,SIFT)特征匹配技术,专门针对 VR 环境中的手势识别进行了优化. 首先,本文对 SIFT 算法进行了深入的改进,通过引入先进的描述子来增强特征的区分度,这使得算法能够更准确地捕捉到手势的关键特征. 然后,为了进一步提升匹配的准确性,我们精心设计了特征匹配策略,优化了特征点之间的对应关系,确保了在复杂场景下也能实现高效匹配. 最后,针对实时性的需求,本文开发了一套算法优化策略,通过调整算法流程和计算方式,确保了算法即使在动态和多变的环境中也能保持高效稳定的运行,从而满足了实时手势跟踪的应用需求. 实验结果表明,所提模型的预测准确率为 0.926,表现出了优异的预测性能.

[关键词]　虚拟现实,手势检测,改进 SIFT 算法,特征匹配,人机交互
[中图分类号]TP391.9　[文献标志码]A　[文章编号]1001-4616(2025)02-0083-08

Gesture Detection and Tracking Based on Improved SIFT Feature Matching in VR

Su Lin[1],Zou Jing[2],Xu Mei[3]

(1.School of Art and Design,Minnan University of Science and Technology,Quanzhou 362332,China)
(2.School of Computer Science and Technology,Guizhou University,Guiyang 550025,China)
(3.College of Education Sciences,Northwest Normal University,Lanzhou 730070,China)

Abstract:With the rapid development of virtual reality(VR) technology,the demand for natural human-computer interaction is increasing,and gesture recognition technology plays a crucial role. It not only requires high accuracy,but also must ensure real-time response to ensure that users can have a smooth interactive experience. This study proposes an innovative gesture detection and tracking method based on an improved Scale Invariant Feature Transform(SIFT) feature matching technique,specifically optimized for gesture recognition in VR environments. Firstly,this article has made in-depth improvements to the SIFT algorithm by introducing advanced descriptors to enhance feature discrimination,which enables the algorithm to more accurately capture key features of gestures. Then,in order to further improve the accuracy of matching,we carefully designed a feature matching strategy,optimized the correspondence between feature points,and ensured efficient matching even in complex scenes. Finally,in response to the real-time requirements,this article developed an algorithm optimization strategy that ensures efficient and stable operation of the algorithm even in dynamic and changing environments by adjusting the algorithm flow and calculation methods,thus meeting the application needs of real-time gesture tracking. The experimental results show that the prediction accuracy of the proposed model is 0.926,demonstrating excellent predictive performance.

Key words:virtual reality,gesture detection,improve SIFT algorithm,feature matching,human-computer interaction

在艺术数字化浪潮的推动下,虚拟现实(virtual reality,VR)技术正以前所未有的速度发展,为人机交互领域带来了革命性的机遇. 在这一背景下,手势识别技术,作为 VR 环境中最为直观和自然交互方式之

收稿日期:2024-11-26.
基金项目:福建省自然科学基金面上项目(2023J011406).
通讯作者:苏琳,讲师,研究方向:虚拟现实,人工智能. E-mail:3051700287@qq.com

一,已经成为艺术设计研究和应用的热点[1]. 它不仅极大地提升了 VR 系统的交互效率和用户体验,也为数字艺术辅助设计、机器人控制等其他领域提供了创新的解决方案. 尽管如此,手势的多样性和复杂性,以及 VR 环境中光照、背景等干扰因素,使得手势检测与跟踪技术面临着严峻的挑战.

传统的基于计算机视觉的手势检测方法,例如基于肤色、轮廓等底层特征的分割方法,虽然在特定条件下取得了一定的成效,但往往难以适应复杂多变的环境. 随着深度学习和人工智能技术的进步,基于数据驱动的方法开始被引入手势识别领域. 然而,这些方法通常需要大量的标注数据,且在实时性方面存在不足[2]. 目前,研究人员已经提出了大量的人工智能手势检测和跟踪算法,对于手势检测,Kunz 等[3]提出了一种基于 Kinect 的定向手势检测算法,该算法通过多种算法的结合达到了更好的精度. Parada 等[4]设计了一种基于射频识别(radio frequency identification,RFID)可穿戴手势检测系统,可以有效地满足物联网系统的要求. 然而,以上研究只能检测手势,无法有效实现手势跟踪. 针对手势跟踪问题,Kim 等[5]提出了一种基于深度摄像头的沉浸式触觉反馈三维手势控制算法,该算法具有较好的实时性. Song 等[6]提出了一种基于背景减法和 Kinect 的手势检测与跟踪系统,可以从原始输入图像中检测出手势运动区域. Chen 等[7]运用 Haar-Like 特征结合 AdaBoost 框架,开发了一种能够识别 4 类手势的分类器,该分类器在多变光照及复杂背景条件下依然保持了较高的识别率. Sherly 等[8]通过 SIFT 算子提取出对旋转、缩放和亮度变化具有不变性的特征,并利用这些特征通过 AdaBoost 分类器进行手势分类,特别是在多角度手势识别方面取得了显著成效. Meanshift 算法通过迭代寻找特征点的局部密度峰值来进行跟踪,能够处理包括颜色、纹理和梯度在内的多种视觉特征. Wang 等[9]对 Meanshift 算法进行了改进,提出了 Camshift 算法,该算法通过颜色直方图来确定目标的中心和尺寸. Fang[10]将粒子滤波与 Meanshift 相结合,提高了跟踪的实时性,尤其是在处理快速手势变化和遮挡场景时表现出色,实验结果证明了这种组合方法的有效性.

尺度不变特征变换(SIFT)算法,作为一种经典的特征提取和匹配方法,在手势识别中确实发挥着重要作用. SIFT 算法能够在不同尺度和方向下提取图像的关键点,并保持特征的不变性. 然而,SIFT 在处理高动态场景和复杂背景时存在局限性,例如对快速运动的鲁棒性不足和对光照变化敏感[11]. 此外,SIFT 算法在计算效率上也存在挑战,尤其是在大规模数据集上. 为了提高手势识别的准确性和鲁棒性,同时考虑到计算效率,对 SIFT 算法进行改进显得尤为重要. 改进后的 SIFT 算法通过优化特征提取过程和增强对动态变化的适应性,能够在保持较低计算成本的同时,提供与深度学习方法相媲美的性能,特别是在小样本学习和实时应用场景中. 这些改进使得 SIFT 算法在手势识别领域更具竞争力,尤其是在需要快速响应和处理大量数据时.

针对上述挑战,本文提出了一种基于改进 SIFT 特征匹配的手势检测与跟踪方法,旨在提升 VR 环境中手势识别的性能. 首先,对 SIFT 算法进行了深入的改进,引入了新的描述子以增强特征的区分度. 这一改进使得算法能够更准确地捕捉到手势的关键特征,即使在复杂光照和背景下也能保持较高的识别准确率. 其次,我们精心设计了特征匹配策略,优化了特征点之间的对应关系,从而显著提升了匹配的准确性. 此外,为了满足实时性的需求,我们开发了一套算法优化策略,通过调整算法流程和计算方式,确保了算法即使在动态和多变的环境中也能保持高效稳定的运行.

1　基础知识

1.1　Gaussian 背景建模

在虚拟现实领域,手势检测与跟踪技术面临的一个主要挑战是精确地区分前景手势对象与背景环境. Gaussian 背景模型的核心理念是为图像中每个像素点的亮度赋予一个高斯概率分布,这使得模型能够精确地模拟并区分前景与背景,如公式(1)所示:

$$P(I(x,y)=v)=\mathrm{e}^{-\frac{(v-\mu)^2}{2\sigma^2}}, \tag{1}$$

式中,μ 和 σ^2 分别代表像素点亮度的期望值和方差,由下面公式给出:

$$\mu(x,y)=\frac{1}{T}\sum_t I_t(x,y), \tag{2}$$

$$\sigma^2(x,y)=\frac{1}{T}\sum_t (I_t(x,y)-\mu)^2, \tag{3}$$

模型建立过程,首先根据视频序列的前几帧连续图像对每个像素点计算和参数,然后每个像素点都得到了对应的概率密度分布函数,此时即可对下面的图像帧进行前后景的分离,具体通过如下公式:

$$|I(x,y)-\mu(x,y)|>2\sigma. \tag{4}$$

通过这种方式,Gaussian 背景模型能够有效地从视频序列中分离出前景对象,即手势. 本文利用该模型对 VR 环境中的手势进行建模和识别,通过计算每个像素点属于背景的概率,并设置阈值来判断是否为前景对象.

1.2　运动目标检测方法

帧间差分法是一种广泛应用于运动检测领域的算法,其核心思想在于利用视频序列中相邻帧之间的像素差异来识别运动目标. 由于视频具有时间连续性,在没有运动目标的情况下,相邻帧之间的变化微乎其微;而在有运动目标出现时,相邻帧之间的像素差异会显著增加. 该算法通过计算连续两帧或三帧之间的像素差,并获取这些差值的绝对值,进而通过设定阈值来判断是否存在运动目标[12]. 图 1 展示了帧间差分法的计算步骤.

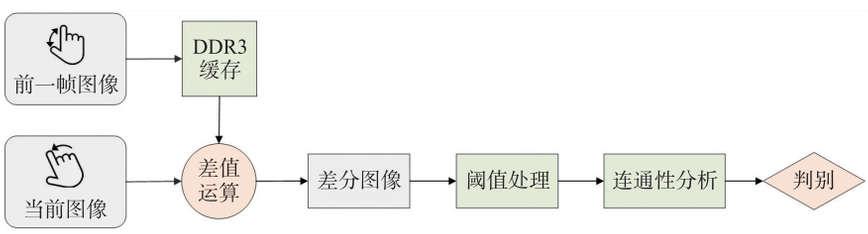


图 1　帧间差分法的计算步骤

Fig. 1　Calculation steps of inter frame difference method

帧间差分法的数学表达式如式(5)和(6)所示:

$$D_n(x,y)=|f_n(x,y)-f_{n-1}(x,y)|, \tag{5}$$

$$R_n(x,y)=\begin{cases}1, & D_n(x,y)>T\\ 0, & D_n(x,y)<T\end{cases}, \tag{6}$$

式中,$D_n(x,y)$表示差分图像,$f_n(x,y)$、$f_{n-1}(x,y)$表示第 n 帧图像以及第 $n-1$ 帧图像,T 表示阈值,$R_n(x,y)$表示输出的二值化图像.

2　SIFT 特征匹配算法

尺度不变特征变换(SIFT)是一种用于提取图像内局部特征的描述子,它能够提供对光照变化、视角差异、缩放和旋转的稳定性,同时在尺度空间中实现有效定位,以降低噪声的干扰. SIFT 特征提取过程的改进主要包括以下几个关键步骤:首先,通过优化差分高斯(DoG)函数在不同尺度上探测潜在的特征点,改进后的 DoG 函数能够更有效地识别出对尺度缩放和旋转变化具有不变性的特征点;其次,对这些候选点进行精确定位,改进后的定位算法能够更准确地确定它们在图像中的具体位置和尺度,提高了定位的精度;然后,为每个关键点分配一个或多个主方向,改进后的方向分配算法增强了对图像旋转的不变性,使得特征点在旋转后仍能被准确匹配;最后,围绕每个关键点的邻域区域计算梯度的统计特性,改进后的梯度统计算法能够生成更具区分度的 SIFT 特征向量,提高了特征的稳定性和唯一性. 这些改进点共同作用,使得改进的 SIFT 算法在手势检测与跟踪任务中表现出更高的准确性和鲁棒性[13].

2.1　DoG 尺度空间生成

图像的尺度空间可以通过函数 $L(x,y,\sigma)$来表示,该函数是通过将可变尺度的高斯函数 $G(x,y,\sigma)$与图像 $I(x,y)$进行卷积运算得到的. 这种表示方法使得我们能够在多个尺度上分析图像,捕捉从细节到全局的不同层次的特征. 即

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y), \tag{7}$$

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}. \tag{8}$$

在特定尺度层面,通过计算两个连续的高斯尺度空间图像的差值生成差分高斯(DoG)图像 $D(x,y,\sigma)$,

这一过程有效地突出了图像中的特征点. 在处理响应值图像时,采用了非最大值抑制技术来执行局部极值检测,这一过程能够精确地在二维空间和尺度维度上识别出显著的特征点. 通过这种方式,只有那些在邻域内最为突出的点被保留,从而确保了特征点的精确定位. 其中:

$$D(x,y,\sigma)=L(x,y,k\sigma)-L(x,y,\sigma). \tag{9}$$

该过程排除了那些在局部区域内不够显著的响应,只保留了真正重要的特征点,增强了特征检测的准确性和效率.

2.2　特征点搜索

DoG(difference of gaussian)图像的生成基于图像金字塔的构建过程. 在这个过程中,金字塔被划分为多个组,每组内包含多个层次[14]. 新一组的图像数据通过采用降采样技术从上一组获得,这种方式有效地减轻了卷积操作的计算量. 在每个分组内,通过将相邻两层的高斯模糊图像进行相减,生成了 DoG 图像,具体实现见图 2.

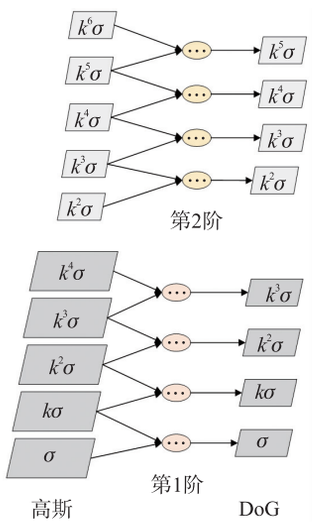


图 2　高斯金字塔与 DoG 的实现

Fig. 2　Implementation of Gaussian pyramid and DoG

在构建高斯金字塔的过程中,关键步骤不仅包括设定尺度空间的参数,以决定图像在不同分辨率下的表示,还涉及到如何通过高斯模糊捕捉图像在各个尺度上的特征. 为了处理快速运动手势引起的模糊问题,本文采用了多尺度高斯金字塔来增强对快速运动的适应性,允许算法在不同层次上捕捉到更加清晰和稳定的手势特征. 此外,通过引入自适应阈值和局部窗口跟踪技术,算法能够在动态环境中有效抑制由于手势快速移动引起的模糊,并保持跟踪的精度. 这些策略使得算法能够在变化的环境中准确捕捉手势的动态变化,即使在手势快速移动或背景复杂的情况下也能有效跟踪. 具体来说,算法通过动态调整高斯金字塔的尺度参数和梯度计算的灵敏度,以适应不同速度的手势运动,并通过局部窗口跟踪策略,减少环境噪声对跟踪精度的影响. 通过恰当选择这些参数,可以确保在后续的 DoG 计算中有效地突出图像中的关键视觉信息,其中,σ 和 o、s 的关系:

$$\sigma(o,s)=\sigma_0 2o+s/S,o\in o_{\min}+[0,\cdots,o-1],s\in[0,\cdots,S-1], \tag{10}$$

位置空间坐标 $X(x,y)$ 是组 Octave 的函数,设 X_0 是 o 组的空间坐标,则

$$X=2^oX_0,o\in Z,X_0\in[0,\cdots,N_0-1]\times[0,\cdots,M_0-1], \tag{11}$$

如果(M_0,N_0)是基础组 $o=0$ 的分辨率,则其他组的分辨率为

$$M_0=\frac{M_0}{2^o},N_0=\frac{N_0}{2^o}. \tag{12}$$

2.3　特征点方向分配

为了使特征点的描述子对图像旋转具有不变性,本方法为每个特征点定义了一个主方向,这个方向是通过分析其周围图像区域的梯度来确定的. 具体来说,对于每个检测到的特征点,需要计算以该点为中心、半径为 $3\times1.5\sigma$ 的圆形区域内的梯度方向和大小. 梯度的方向(幅角)和大小(幅值)是通过有限差分方法得出的,如下所示:

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2}, \tag{13}$$

$$\theta(x,y)=\arctan\left(\frac{L(x+1,y)-L(x-1,y)}{L(x,y+1)-L(x,y-1)}\right). \tag{14}$$

一旦确定了图像中特征点的主导方向,便能够为每个特征点归纳出 3 个关键属性:位置、尺度和方向,分别用坐标(x,y)、尺度 σ 和方向 θ 来表示. 这 3 个参数共同定义了特征点的 SIFT 特征区域,为后续的特征描述和匹配提供了必要的上下文信息.

2.4　特征点特征矢量生成

SIFT 描述子 $F(x,y,\theta)$通过汇总关键点周围的高斯尺度图像的梯度信息来构建,本质上是一个三维

数组,通常被展平成一个向量. 描述子的构建与关键点所在的尺度层密切相关,因此必须基于该尺度层的高斯图像来确定梯度,以确保描述子能够准确地反映特征点的局部特性[15]. 这种计算方法使得描述子能够精确地捕捉到特征点周围的细节变化,从而为图像特征的匹配和识别提供了可靠的基础.

本方法把关键点的邻域划分成 $B_p \times B_p$ 的子区域,每个子区域的大小为 $m\sigma$ 像素,这里 $m=3$,$B_p=4$,而 σ 是关键点的尺度参数,每个关键点还包含 8 个方向的梯度信息,从而形成了一个 128 维的 SIFT 特征向量. 这一过程有效地捕捉了关键点周围的局部结构信息,为图像特征的描述提供了一种鲁棒的表示方法.

3　基于改进 SIFT 特征匹配的手势检测与跟踪

3.1　RIT-SIFT 特征描述子的匹配

利用 SIFT 算法检测到的关键点 $f_i(x,y,\sigma,\theta)$ 位于高斯金字塔的特定层上,其空间尺度因子 σ 和旋转角度 θ 确定了关键点的位置和方向. 以 f_i 为中心选取一个 8×8 的图像区域作为描述子提取的局部区域,该区域经过旋转和尺度变换后,形成了 64 维的 RIT 特征描述子 S_i. RIT-SIFT 描述子综合了 SIFT 的 128 维局部特征向量 $\boldsymbol{Q}$ 和 RIT 的 64 维特征向量 $\boldsymbol{S}$,构成一个 192 维的描述子,用于描述关键点周围的图像信息. 在进行特征匹配时对向量 $\boldsymbol{Q}$ 和 $\boldsymbol{S}$ 采取了差异化的处理方法,分别计算了它们之间的相似度指标,并将这些指标综合起来,从而去衡量两幅图像中关键点的匹配程度. 针对 SIFT 的局部特征向量 $\boldsymbol{Q}$,本文选择了欧几里得距离作为衡量其相似度的准则.

$$d_{\boldsymbol{Q}}=|Q_i-Q_j|=\sqrt{\sum\nolimits_n (Q_{i,n}-Q_{j,n})^2} \tag{15}$$

在特征匹配过程中,针对 SIFT 局部特征向量的每一维度进行了比较,其中 $Q_{i,n}$ 和 $Q_{j,n}$ 分别代表第 i 个和第 j 个关键点的第 n 个特征维度(n 不超过 128). 与此同时,对于 RIT 特征向量 $\boldsymbol{S}$,本文采用了 χ^2 统计量来衡量其匹配度.

$$d_{\boldsymbol{S}}=\chi^2=\frac{1}{2}\sum\nolimits_m \frac{(S_{u,m}-S_{v,m})^2}{S_{u,m}+S_{v,m}}, \tag{16}$$

式中,$S_{u,m}$ 和 $S_{v,m}$ 分别代表第 u 个和第 v 个关键点的特征向量的第 m 个维度,且 m 的值不超过 128.

在完成了局部特征向量与 RIT 特征向量间的匹配度评估之后,采用加权的距离度量 $d=\lambda d_{\boldsymbol{Q}}+(1-\lambda)d_{\boldsymbol{S}}$ 来衡量 RIT-SIFT 特征描述子的相似性. 在此过程中,权重系数 λ 被用于调节不同特征的影响力,确保特征匹配过程的均衡性. 对于给定图像中的任一关键点,本方法在目标图像中筛选出两个与该点描述子距离最小的关键点. 通过比较这些距离,选择距离最小的关键点作为潜在的匹配点. 这一过程确保了匹配的精确性和可靠性,从而提高了整体手势跟踪算法的性能.

3.2　基于 RIT-SIFT 的手势检测与跟踪

手势跟踪任务可以类比于在马尔可夫框架内预测隐含状态的挑战,简单来说,特定时刻 t 下手势在连续图像中的位移可以通过仿射变换参数集合来捕捉. 在本文提出的手势跟踪方法中,这些参数定义了手势动态状态的空间,标记为 X_t. 这个状态空间 X_t 反映了手势在图像平面内的位置、大小和方向变化,为手势的实时跟踪提供了数学基础. 给定一系列观测到的手势图像 $I_t=\{I_1,\cdots,I_t\}$,目标是预测手势的隐含运动状态,即求解条件概率 $p(X_t|I_t)$,如公式(17)所示.

$$p(X_t|I_t)=p(I_t|X_t)*\int p(X_t|X_{t-1})*p(X_t|I_{t-1})\mathrm{d}X_{t-1}. \tag{17}$$

在手势的动态变化中,手势在视觉画面中的位置和尺寸等关键运动特征,可以通过一组仿射变换参数来捕捉. 这些参数能够精确描述手势在图像序列中的平移、旋转、缩放等变化,从而为手势跟踪提供了一个有效的数学工具. 通过这些参数,可以量化手势在二维平面上的运动轨迹,并实时更新手势的状态. 这种方法不仅能够处理手势的全局变换,还能够适应手势局部形状的变化,为手势识别和跟踪提供了一种灵活而强大的解决方案. 各种情况下手势变化的缩放矩阵如下:

$$\boldsymbol{T}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ x & y & 1\end{bmatrix}, \tag{18}$$

$$\boldsymbol{T}=\begin{bmatrix}\sigma & 0 & 0\\ 0 & \sigma & 0\\ 0 & 0 & 1\end{bmatrix}, \tag{19}$$

$$\boldsymbol{T}=\begin{bmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}. \tag{20}$$

在特定的时间点 t,手势在空间中的运动状态由多个仿射变换矩阵的累积效果来定义. 这些矩阵共同作用,形成了描述手势运动的完整仿射变换. 简而言之,手势在这一时刻的运动特征是由一系列仿射变换矩阵的叠加效应所决定的,这种复合仿射变换提供了一种全面而精确的方式来捕捉手势在图像中的位置、大小和方向的变化.

$$\boldsymbol{T}=\begin{bmatrix}t_{11} & t_{12} & 0\\ t_{21} & t_{22} & 0\\ t_{31} & t_{32} & 1\end{bmatrix}=\sigma\begin{bmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ x & y & 1\end{bmatrix}. \tag{21}$$

因此,在 t 时间构建的手部轮廓模型如图 3(a)所示,在时间$(t+\Delta t)$的手部轮廓模型如图 3(b)所示.

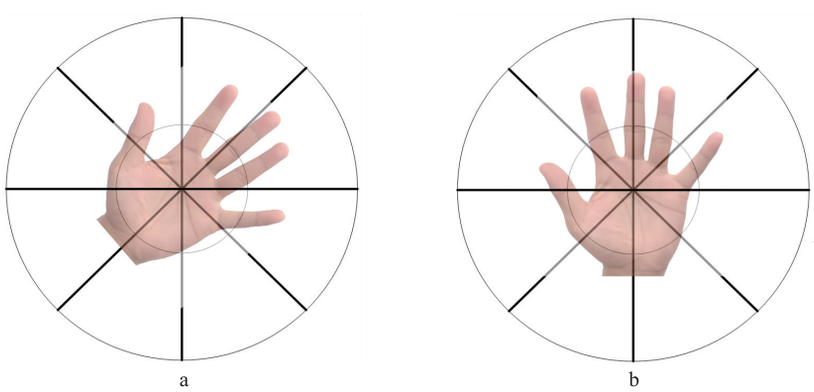


图 3　手部轮廓模型构建

Fig. 3　Construction of hand contour model

4　实验与结果分析

为了评估本文提出的改进 SIFT 算法在手势识别与跟踪领域的有效性,本文将其与几种广泛使用的跟踪技术(包括传统 SIFT、基于深度神经网络的 CNN,以及基于光流的方法)进行了对比实验[16-18]. 实验目的是分析各方法在跟踪手势视频序列时的性能,并对比它们的中心定位误差. 实验环境搭建在 Windows 系统上,使用 VS2010 和 OpenCV 2.4.3 进行开发,测试所用硬件配置为 2.30GHz 处理器和 2GB 内存. 在实验中,手势的特征表示由局部二值模式(LBP)特征直方图和 RGB 颜色空间的 3 个通道构成,其中 LBP 特征是基于 3×3 像素邻域计算得到的. 在训练阶段设定了 9 个弱分类器,而在更新过程中保留了其中 3 个表现最佳的弱分类器. 此外,手势检测的初始网格点被设定为 100 个,以确保算法能够有效覆盖手势区域. 通过这些设置,我们旨在为改进后的 SIFT 算法提供一个严格的测试环境,并与现有技术进行公平比较.

首先,本文构建了一个包含 1 000 个视频序列的综合数据集,这些视频序列经过精心设计,以确保手势类型的多样性和场景的真实性. 数据集中的手势动作包括但不限于抓取、挥动、指点、滑动等,覆盖了日常生活中的常见手势以及特定应用场景下的手势动作. 为了模拟不同的拍摄条件,视频序列在多种光照环境下进行录制,包括室内自然光、室内人造光、室外阳光以及阴天环境. 此外,背景复杂度也经过特别考虑,包含了从简单纯色背景到复杂的室内外场景. 在数据集的标注过程中,本文采用了严格的标注协议,确保每个手势动作都能被准确地识别和分类. 每个视频序列都被标注了起始帧和结束帧,以及手势的关键点位置. 这些标注信息将用于后续的特征提取和算法训练. 对这些视频序列进行预处理,包括帧提取、去噪和灰度化,以减小计算复杂度并提高后续处理的准确性. 帧提取确保了视频中的关键动作帧被选取

用于分析;去噪处理通过高斯滤波器减少图像噪声,提高图像质量;灰度化则将彩色图像转换为灰度图像,减少计算量并专注于手势的轮廓和形状信息. 这些预处理步骤为算法提供了清晰、一致的输入数据,从而提高了最终检测结果的准确性. 如表 1 所示,这些改进显著提升了检出率结果.

从上述数据可以看出,本文提出的方法在所有性能指标上都取得了最佳结果. 准确率和召回率的提升表明本文方法在识别手势的准确性和完整性方面都有显著优势,F1 分数的提高进一步证实了这一点. 此外,尽管本文方法的计算复杂度略高于传统 SIFT 方法,但其处理时间仍然保持在可接受的范围内,且远低于基于深度学习的 CNN 方法,这证明了本文方法在实际应用中的可行性和高效性.

表 1　4 种模型对手势检测与跟踪的检出率结果

Table 1　The detection rate results of four models for gesture detection and tracking

模型	Accuracy	Recall	F1-score
SIFT	0.852	0.805	0.828
CNN	0.897	0.872	0.884
光流场	0.783	0.751	0.767
RIT-SIFT	0.926	0.903	0.912

经过 100 次实验,4 种不同算法在 200 帧视频序列上的手势跟踪结果如图 4 所示.

由图 4 可以看出,随着视频帧数的增加,4 种算法得到的中心点误差值(像素数)都在不断增加,且本文提出的算法可以连续跟踪运动手势. 同时,在相同条件下,该算法具有较小的跟踪误差,4 种算法的手势识别结果如表 2 所示.

与其他算法相比,本文算法的手势识别准确率可以稳定在 90%左右,可以有效地应用于虚拟现实人机交互系统.

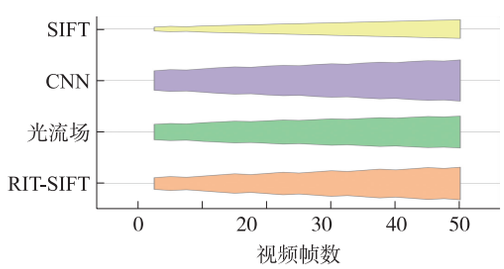


图 4　4 种手势算法的检测与跟踪结果

Fig. 4　The results of four models for gesture detection and tracking

表 2　4 种模型对手势检测与跟踪的预测结果

Table 2　Prediction results of four models for gesture detection and tracking

手势动作	SIFT	CNN	光流场	RIT-SIFT
向上	0.83	0.82	0.86	0.91
向下	0.85	0.87	0.87	0.92
向左	0.83	0.85	0.88	0.91
向右	0.85	0.89	0.85	0.93
旋转	0.81	0.86	0.84	0.90

5　结论

在本研究中,我们成功开发了一种基于改进 SIFT 特征匹配的手势检测与跟踪方法,专注于提升 VR 环境中手势识别的精确度和响应速度. 通过融入创新的描述子和优化特征匹配策略,我们的方法显著提高了对手势关键特征的捕捉能力,即便在复杂多变的环境中也能维持卓越的识别准确率. 此外,通过一系列算法优化措施,确保了系统的高效稳定运作,满足了实时手势跟踪的严苛要求. 实验结果充分证明了我们的方法在准确率、召回率和 F1 分数等关键性能指标上超越了现有技术,显示出广泛的应用潜力. 这些研究成果不仅为 VR 领域人机交互技术提供了创新的解决方案,也为手势识别技术的未来发展奠定了坚实的基础.

[参考文献]

[1]　高方远,何立火. 基于深度网络和视觉特性的无参考图像质量评价方法[J]. 南京师大学报(自然科学版),2019,42(3):20-26.

[2]　龚新儒,张德育,谢少威. 虚拟现实中基于轻量级神经网络的手势识别[J]. 南京理工大学学报(自然科学版),2024,48(3):367-373.

[3]　KUNZ A,ALAVI A,SINN P. Integrating pointing gesture detection for enhancing brainstorming meetings using kinect and pixelsense[J]. Procedia cirp,2014,25:205-212.

[4]　PARADA R,MELIA S J. Gesture detection using passive rfid tags to enable people-centric IoT applications[J]. IEEE

communications magazine,2017,55(2):56-61.
[5] KIM K,KIM J,CHOI J,et al. Depth camera-based 3D hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions[J]. Sensors,2015,15(1):1022-1034.
[6] SONG W,LU Z,LI J,et al. Hand gesture detection and tracking methods based on background subtraction[J]. Lecture notes in electrical engineering,2014,309:485-490.
[7] CHEN Q,GEORGANAS N D,PETRIU E M. Hand gesture recognition using haar-like features and a stochastic context-free grammar[J]. Instrumentation and measurement,2008,57(8):1562-1571.
[8] SHERLY L A,JAYA T. An efficient indoor scene character recognition using Bayesian interactive search algorithm-based adaboost-CNN classifier[J]. Neural computing and applications,2021,33(22):15345-56.
[9] WANG J,JIA Z,LAI H,et al. A real time target face tracking algorithm based on saliency detection and Camshift[J]. Multimedia tools and applications,2023,82(28):43599-624.
[10] FANG B. Vehicle rear end collision warning method based on MeanShift and Kalman filter tracking[J]. International journal of vehicle design,2024,95(1):1-21.
[11] ALI A,GHOSH A,CHAUDHURI S S. Real-time tracking of moving objects through efficient scale space adaptation and normalized correlation filtering[J]. Signal,image and video processing,2024,18(1):679-89.
[12] 邬忠萍,刘新厂,郝宗波. 基于并行 CNN 和识别策略优化的车牌识别方法研究[J]. 南京师大学报(自然科学版),2023,46(3):98-104.
[13] 鲁东生,龙华. 一种基于谱图 SIFT 的同源频谱监测数据判定方法[J]. 计算机科学,2024,51(S1):777-783.
[14] PIRAN Z,NITZAN M. SiFT:uncovering hidden biological processes by probabilistic filtering of single-cell data[J]. Nature communications,2024,15(1):760-772.
[15] ZHANG Y,ZHU D,WANG Y L. Improved fast feature matching method of SIFT[J]. Control and automation pubilcation group,2008,24(11):220-222.
[16] 王锟,宋永红,郑斐,等. 基于卷积神经网络的嵌入式手势检测算法[J]. 计算机工程与应用,2019,55(4):137-141,178.
[17] 姬晓飞,张旭,李俊鹏. 基于多域卷积神经网络跟踪的动态手势识别[J]. 沈阳航空航天大学学报,2021,38(5):51-57.
[18] 陈威. 一种基于光流的动态手势识别方法[J]. 哈尔滨师范大学自然科学学报,2015,31(6):81-86.

[责任编辑:陆炳新]

(上接第 82 页)

[25] LIU B Q,JIA B B,ZHANG M L. Towards enabling binary decomposition for partial multi-label learning[J]. IEEE transactions on pattern analysis and machine intelligence,2023,45(11):13203-13217.
[26] ZHANG M L,FANG J P. Partial multi-label learning via credible label elicitation[J]. IEEE transactions on pattern analysis and machine intelligence,2020,43(10):3587-3599.
[27] ZOU Y,HU X,LI P,et al. Learning shared and non-redundant label-specific features for partial multi-label classification[J]. Information sciences,2024,656:119917.
[28] CHEN Z S,WU X,CHEN Q G,et al. Multi-view partial multi-label learning with graph-based disambiguation[C]// Proceedings of the AAAI Conference on artificial intelligence,2020,34(4):3553-3560.
[29] XU N,WU Y D,QIAO C,et al. Multi-view partial multi-label learning via graph-fusion-based label enhancement[J]. IEEE transactions on knowledge and data engineering,2022,35(11):11656-11667.
[30] YUAN J,LIU W,GU Z,et al. A unified framework for graph-based multi-view partial multi-label learning[J]. IEEE access,2023,11:49205-49215.
[31] 耿新,徐宁. 标记分布学习与标记增强[J]. 中国科学:信息科学,2018(5):5.

[责任编辑:陆炳新]