多频飞行时间相机实时深度补偿算法研究

王森¹, 王润孝¹, 左欣欣², 于薇薇¹

(1.西北工业大学 机电学院, 陕西 西安 710072; 2.西北工业大学 计算机学院, 陕西 西安 710072)

摘 要:针对多频飞行时间相机在曝光时间内物体运动以及相位融合中产生的偏差,提出一种基于光流以及空间核密度估计的实时补偿算法。首先获取原始的多频相位图像,计算出不同频率之间的相对光流,获得每种频率下的补偿相位;在此基础上针对不同频率的相位做出可信度假设并利用空间核密度估计对其进行排名,获得最终的相位融合图并生成对应的深度图。实验部分,使用 Kinect V2 相机获取的原始多频相位图对算法进行验证,同时利用 GPU 进行并行加速;结果表明该算法可以实时对物体平行于光轴运动的相位融合偏差进行补偿,有效地提高了深度图像的精度和稳定性。

关 键 词:飞行时间相机;多频;光流;空间核密度估计;Kinect V2

中图分类号:TP391

文献标志码:A

文章编号:1000-2758(2019)01-0152-08

消费级别的飞行时间相机由于不需要复杂的机械安装与校准,并且可方便的提供物体的深度信息成为学术界和工业界研究的热点^[1]。但是由于其测距原理限制,会产生一些系统和非系统的误差^[2]。当物体运动时,对应的相位图像也会移动,在利用多帧相位图进行深度融合计算时,由于数据差异较大,会造成边缘数据的不一致和不稳定,严重影响后续使用。从文献[1]中可以明显看出快速运动人体深度图像的边缘缩减问题。

为了解决移动物体单帧中的相位图移动问题,Hussmann 等^[3]利用相位图差标记出物体移动区域,然后根据物体线性移动的特点计算出移动距离,最终利用 FPGA 进行加速,可达到实时效果。但是该算法仅限于线性平动的物体。Lindner 等^[4]在传统四相位漂移算法的基础上,计算任意连续 2 个相位之间稠密的光流场并对其他相位图进行补偿。Lefloch^[5]在文献[4]的基础上,只计算一组频率中首尾 2 帧的光流场,其余光流场则由线性插值而来,该算法提高了整体的计算效率,但是其严重依赖运动的连续性,当首尾 2 帧出现周期性运动时会出现较大的系统偏差。Crabb 等^[6]利用概率框架输入多张同频率相位图对深度图像进行全局优化,但是无法

很好地解决物体运动的问题。

为了增加深度测距的稳定性以及距离,仿照多普勒雷达效应,在单一频率的基础上产生了多频飞行时间相机,利用不同频率在同一位置返回的距离,采用经典的中国剩余定理^[7]可以较大地提升精度并扩大无歧义的距离。但是中国剩余定理仅适用于系统误差与噪声较小的情况。由于相位的测定与物体的距离及表面的反射率有关,因此可认定物体表面的反射率局部恒定,在该假设下,利用马尔科夫随机场、数据项以及反射平滑项作为目标函数进行优化成为一种通用的方法。Droeschel 等^[8]利用 2 个不同的时间飞行相机模拟双频的相位图,采用马尔科夫随机场加入频率和反射率的数据项作为优化目标。Ti 等^[9]则模拟间隔频率的时间飞行相机相位图生成方式,只利用一次生成的多频相位图对整体进行优化。Lawin^[10]利用空间核密度估计对每种频率的可信度进行对比,有效地提高了整体测量距离和准确度。

为了提高多频飞行时间相机在物体运动时的测距精度,本文结合单一频率的运动补偿以及多频率的空间核密度预估相位融合算法(以下简称为 MCO 算法),分别利用大范围偏移的光流进行运动补偿

收稿日期:2018-03-06

基金项目:国家自然科学基金(51475373, 61603302)、陕西省科技统筹创新工程项目(2016KTZDGY06-01)、陕西省自然科学基金(2016JQ6009)与 111 引智计划(B13044)资助

作者简介:王森(1988—),西北工业大学博士研究生,主要从事机器人视觉研究。

及空间核密度对相位进行融合。最终利用 Kinect V2 相机验证算法的有效性。结果表明,该算法可以针对物体平行于光轴的运动,可实时有效地提高深度图像的精度和稳定性。

1 背景

1.1 飞行时间测距原理

传统的单一频率飞行时间相机利用传感器发出经调制的近红外光,遇物体后反射,计算光线发射和反射的相位差来计算被拍物体的距离。针对单一频率的测距,学者们为了简化计算,利用4张1组的相位图进行计算,其距离计算如公式(1)~(2)所示:

$$D = c\varphi/2\pi\omega \quad (1)$$

$$\varphi = \arctan(P_0 - P_2, P_3 - P_1) \quad (2)$$

式中, c 为光速, φ 为相位差, ω 为角速度, $P_i (i=0,1,2,3)$ 表示对应时刻的相位图,其差值为2张图的标量之差,为对应像素的直接相减。

当一组相位图包含任意 N 张图时,其相位计算如公式(3)^[11] 所示

$$\varphi = \arg \left(\frac{2}{N} \sum_{n=1}^N v_n e^{-i(p_0 + 2\pi(n-1)/N)} \right) \quad (3)$$

式中, p_0 表示相位的补偿, v_n 表示对应频率下的测量电压。

由以上测距原理可以推出对应调制频率可探测到的无歧义的距离最大距离为

$$D_{\max} = c/4\pi f \quad (4)$$

式中, f 为信号的频率, $f = \omega/2\pi$ 。

1.2 多频飞行时间原理

为了解决单一频率测量距离有限以及稳定性不高的问题,人们提出了多频率的飞行时间测距方法。利用不同频率在同一位置的返回距离进行融合,其经典的方法为中国剩余定理,如等式(5)和(6)所示:

$$d_i = D \bmod r_i \quad (5)$$

$$D = k_i r_i + d_i \quad (6)$$

式中, i 表示频率, r_i 表示第 i 个频率下的无歧义最大距离, $\bmod$ 表示取余, D 表示实际的距离。该算法适合在无噪音下的多频率融合。 d_i 表示取余后的实际距离, k_i 表示该频率下可能取到的整数倍数。

Kinect V2 是目前消费级别深度相机中比较常用的一种多频率的飞行时间相机,该相机在1张深度图像的生产中采集了3种不同频率的相位图(如

图1^[10]所示),每组频率包含3张相位图。

图1 Kinect V2 在不同频率及相位下的对应距离

1.3 光流原理

假设2帧图像间的亮度恒定而且保持空间的一致性,那么光流可以评估了2帧图像之间的变化光流场的目的是找到连续帧之间每一个像素的运动信息($\delta x, \delta y$),假设2帧图像的时间差为 δt ,则约束方程可以写为:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (7)$$

本文的目的是求出连续的相位图之间的运动信息,当连续相位图的运动基本平行于相机光轴时,此时同一位置的亮度相近,可认为其满足光流运算的基本假设。

2 算法

如图2所示,采用第三方的驱动 libfreenect2^[13] 获取原始的相位图像,每次获取3种频率各3幅图(在框架图中我们只选取其中的1幅作为代表),进行整体预处理后进入框架图进行计算。以频率2(16 MHz)为基准,使用大范围偏移的光流分别对频率1和频率3进行光流计算,然后对其分别进行补偿。在补偿相位图基础上利用空间核密度对不同频率相位的可信度进行排名,在此基础上进行相位融合,利用最终融合的相位图生成深度图。

2.1 预处理

在进行整体的算法之前,对相位图进行预处理。为了降低相位图整体的噪声但不影响高频的边界部分,对每张相位图进行双面滤波。为了加速后续的整体算法,对运动部分进行粗略检测并进行标记,我们采用相位图差(灰度图差)的方法,计算2张相位图的数值差,如图3a)所示;将图3a)中阈值设定为

3 后提取二值图像 (见图 3b)), 在后面的光流计算中只对标记区域进行运动补偿。

图 2 算法整体框架图

图 3 相位图差及对应二值图

2.2 大范围偏移光流计算

设定 I_1 和 I_2 为需要处理的 2 帧连续图像。 $\mathbf{x}:(x,y)$ 则表示图像上的任意点, $\mathbf{w}:(u,v)$ 表示对应的光流场。在遵守对应点的颜色值一样的假设下, 传统的光流方法一般建立如下 3 个约束来优化整个问题。

$$E_{\text{color}}(\mathbf{w}) = \int \Psi(|I_2(\mathbf{x} + \mathbf{w}(\mathbf{x})) - I_1(\mathbf{x})|^2) d\mathbf{x} \tag{8}$$

式中, Ψ 表示惩罚函数, $\Psi(t^2) = \sqrt{t^2 + 10^{-6}}$ 用来处理一些外点。该项为颜色一致约束项。

考虑到光照的影响, 因此加入 (9) 式的约束, 表示在颜色梯度上的一致性:

$$E_{\text{grad}}(\mathbf{w}) = \int \Psi(|\nabla I_2(\mathbf{x} + \mathbf{w}(\mathbf{x})) - \nabla I_1(\mathbf{x})|^2) d\mathbf{x} \tag{9}$$

为此加入整个计算空间的平滑项保证整体计, 结果的平滑减少二义性

$$E_{\text{grad2}}(\mathbf{w}) = \int \Psi(|\nabla u\mathbf{x}|^2 + |\nabla v\mathbf{x}|^2) d\mathbf{x} \tag{10}$$

以上的三项所组成的光流模型可以处理移动较小的结果。经过实验发现, 在高速运动中相同频率相邻相位图之间将会出现多达 10 个像素的偏移, 所以需要大范围偏移的光流方法并且针对平滑项进行修改。受到文献 [10] 的启发, 在上面的 3 项外, 本章加入最近邻的加权平均项 (见公式 (11)) 保证在局部运动范围内有着较好的平滑度。

$$E_{\text{nn}}(\mathbf{w}) = \sum_{j=1}^J \rho_j(\mathbf{x}) \int \Psi((u(\mathbf{x}) - u_j(\mathbf{x}))^2 + (v(\mathbf{x}) - v_j(\mathbf{x}))^2) d\mathbf{x} \tag{11}$$

该项表示 J 个最近邻区域匹配的插值的加权平均, $\rho_j(\mathbf{x})$ 是其中的权重, 当无相应匹配为 0, 有对应匹配是其对应距离。在本章中选取 $J = 5$ 。

以上 4 项加起来为非凸函数, 因此无法显式地求出最优解。为了提高算法的速度, 需要尽可能采用比较接近结果的初始值, 因此采用特征算子匹配对 2 帧图片项进行初始匹配, 可以将其解作为初始值带入整个目标函数中。因此加入第五项 $E_{\text{desc}}(\mathbf{w}_1)$, 既可以作为约束项, 也可以作为整个模型的初始解。

$$E_{\text{desc}}(\mathbf{w}_1) = \int |f_2(\mathbf{x} + \mathbf{w}_1(\mathbf{x})) - f_1(\mathbf{x})|^2 d\mathbf{x} \tag{12}$$

最终光流的目标函数为公式 (13), 初始化时, 采用 HOG 算子算出每张图的特征算子并对其进行

匹配,在此基础上将其转化为稀疏光流场。将其作为初始解带入整个能量模型中,利用逐次超松弛(SOR)迭代进行计算直到整个能量模型收敛。

$$E = E_{\text{color}}(\mathbf{w}) + \beta_1 E_{\text{grad}}(\mathbf{w}) + \beta_2 E_{\text{grad}2}(\mathbf{w}) + \beta_3 E_{\text{nn}}(\mathbf{w}) + E_{\text{desc}}(\mathbf{w}_1) \quad (13)$$

在本文中,选取 $\beta_1 = 5, \beta_2 = 5, \beta_3 = 25$ 。

图 4 为通过公式(13)计算得出的连续 2 帧相位图及其对应的可视化光流。

图 4 连续 2 帧相位图及其对应的光流图

2.3 基于空间核密度估计的相位融合

在相位融合中,为了提高相位融合的精度和整体的测量距离,受到文献[10,15]的启发,对每种频率补偿后的相位图分别做出可信度假设并利用空间核密度估计对其进行排名,最终获得精确稳定的融合相位。

针对每一个像素 x 的可信度假设 $p(t^i(x))$,利用空间核密度的计算如下:

$$p(t^i(\mathbf{x})) = \frac{\sum_{j=1}^2 \sum_{k \in N(\mathbf{x})} \omega_{jk} K(t^i(\mathbf{x}) - t^j(\mathbf{x}_k))}{\sum_{j=1}^2 \sum_{k \in N(\mathbf{x})} \omega_{jk}} \quad (14)$$

式中, j 表示不同的假设数,在本文中使用双假设; $N(\mathbf{x})$ 表示其邻域像素,其定义为半径 r 内的所有像素, r 在本文中设定为 3; K 是内核函数,本文采用高斯内核,其中 $K(\mathbf{x}) = e^{-x^2/2h^2}$, $h = r/2$ 为内核尺度。

ω_{jk} 是相邻像素的对应权重,如(15)式所示,主要包含 3 部分:① $G(\mathbf{x} - \mathbf{x}_k, \sigma)$ 利用高斯函数表示空间权重;② $p(t^j(\mathbf{x}_k) | \mathbf{n}_j(\mathbf{x}_k))$ 表示的多频解算的概率,其中 $\mathbf{n}_j = (n_1, n_2, n_3)$ 分别表示解算向量的系数;③ $p(t^j(\mathbf{x}_k) | \mathbf{a}_j(\mathbf{x}_k))$ 表示相位的概率,其中 $\mathbf{a}_j = (a_1, a_2, a_3)$ 分别表示 3 种对应频率下的振幅。

$$\omega_{jk} =$$

$$G(\mathbf{x} - \mathbf{x}_k, \sigma) p(t^j(\mathbf{x}_k) | \mathbf{n}_j(\mathbf{x}_k)) p(t^j(\mathbf{x}_k) | \mathbf{a}_j(\mathbf{x}_k)) \quad (15)$$

整体平滑 经过如上计算之后,可以获取一张比较准确的相位图。在此基础上,针对空间中的每一个像素点与其周围 N 个点的关系,提出保持边缘 Laplacian 平滑的能量函数:

$$E_{\text{lap}} = \sum_{\mathbf{x}} \exp \left(P(\mathbf{x}) - \frac{\sum_{k \in N(\mathbf{x})} \omega(\mathbf{x}, \mathbf{x}_k) \cdot P(\mathbf{x})}{\sum_{k \in N(\mathbf{x})} \omega(\mathbf{x}, \mathbf{x}_k)} \right) \quad (16)$$

式中, $P(\mathbf{x})$ 表示当前的相位值。

$$\omega(\mathbf{x}, \mathbf{x}_k) = \exp \left(- \frac{\| P(\mathbf{x}) - P(\mathbf{x}_k) \|^2}{\sigma_{\text{lap}}^2} \right) \quad (17)$$

在本章中, $\sigma_{\text{lap}} = 0.01$,与可信度假设类似,取半径 $r = 3$ 以内的所有像素点作为其邻域点计算边缘的有效性。此后对(16)式进行优化,获得最终的相位图。最后对相位图进行处理,获取对应的深度图像。

3 实验验证

为了验证算法的有效性,我们分别使用公有数据集以及自主采集的数据对算法进行验证。其中自主数据利用 Kinect V2 采集了多组不同环境中运动的原始相位数据,所有图像大小均为 512×424。

3.1 数据集验证

采用论文[10]中提供的数据集作为空间核密度估计算法的验证部分,其中主要验证静态状态下文献[10]中的算法(下称 Lawin 算法)与本文 MCO 算法的对比。

其中图 5 为某教堂的部分,图 5a)为数据集提供的标准深度图像,图 5d)为在固定角度下的彩色图像。图 5b)为 Lawin 算法计算出来的结果,图 5c)为 MCO 算法计算出来的结果。图 5e)和 5f)则分别为 Lawin 和 MCO 算法与标准深度图像的误差图。按照公式(18)对误差比例进行量化,其中 MCO 算法的误差比例为 15.87%, Lawin 算法的则为 22.14%。本文还利用公式(19)对误差的具体数值进行统计,其中 Lawin 算法的整体平均误差为 1 998 mm, MCO 算法的整体平均误差为 1 640 mm。

图 5 某教室的数据对比(图中所有单位为 mm)

与图 5 的图例相同,图 6 为某图书馆的一角。分别利用公式 (18) 和 (19) 对误差进行量化后, Lawin 算法的误差比例 E_p 为 24.54%, MCO 算法的误差比例为 18.67%, Lawin 算法的整体平均误差 E_d 为 2 037 mm, 本文算法为 1 720 mm。

$$E_p = \left(\sum_{i=1}^N \frac{\|g_i - c_i\|}{g_i} \right) / N \tag{18}$$

$$E_d = \sum_{i=1}^N \|g_i - c_i\| / N \tag{19}$$

式中, N 表示像素点的个数, g_i 表示该点的真正距离, c_i 表示计算出来的该点的距离。

图 6 某图书馆的数据对比(图中所有单位为 mm)

3.2 实际数据验证

利用 Kinect V2 采集了一些实际的运动数据对整体算法进行对比验证。

图 7 和 8 分别为高速挥手和挥动箱子下的深度图对比,并分别与 libCRT 算法进行比对。从结果图中可以看出,libCRT 算法会不同程度地减小运动部分的面积,主要是因为运动造成了连续的相位图的边缘差比较大,当采用中国剩余定理计算出的不同

频率下的融合相位差距比较大时,系统会自动按照算法取周围的平均或者取 0,这样会使得边缘数据产生较大的问题。本文算法则将所有的相位数据利用光流转化到同样频率下,这样就会减少融合时的误差,尤其减少了运动时边缘的数据的差距,提高了边缘数据的稳定性;而后利用空间核密度估计有效地提高了整体的精确性。

图 7 挥动手臂时对比(图中所有单位为 mm)

图 8 挥动箱子时对比(图中所有单位为 mm)

3.3 运算环境与时间

本文实验环境为 Visual studio 2015, CPU 为

Intel Core i5 处理器,主频为 2.2 GHz,显卡为 Intel HD Graphics 5500,内存为 8 GB。整体的算法主要

利用 C++实现,因为空间核密度估计部分是局部进行的,因此可以进行并行计算,在实验中,我们使用 GPU 进行加速。经过测算,1 幅图的耗时如表 1 所示,可以看出光流计算是最耗时的部分,但是由于我们只计算运动部分的光流,极大减小了运算量,算法的时效性得以保证,可达到每秒 10 帧,满足实时应用的要求。

表 1 耗时测算

计算流程	耗时/ms
光流计算	~70
相位融合	~20
其他	<10

4 结 论

本文针对飞行时间相机在产生的系统以及运动

产生的偏差,提出一种基于光流和空间核密度估计相结合实时补偿算法。首先获取原始的多频相位图像,计算出不同的频率之间的相对光流,获得每种频率下的补偿相位;在此基础上针对不同频率做出可信度假设并利用空间核密度估计对其进行排名,获得最终的相位融合图并生成对应的深度图。在实验阶段利用 Kinect V2 相机算法进行验证并利用 GPU 对空间核密度的相位融合部分进行优化,实验结果表明,与 libCRT 算法和 Lawin 算法相比,本文算法在取得了较好的实验效果并可以达到实时效果。

本文算法适合处理大范围偏移下的物体运动补偿。在下一步的应用中,期望可以处理相机运动所造成的补偿,并加入更多的传感器信息,如 IMU 的信息或者彩色图像的信息,建立一个由粗到精细的框架,利用 IMU 或者彩色图像进行粗略的补偿处理,减少光流的计算,提高算法的整体效率。

参考文献:

[1] ZUO X, WANG S, ZHENG J, et al. High-Speed Depth Stream Generation from a Hybrid Camera[C]//ACM on Multimedia Conference, 2016: 878-887

[2] LINDNER M, KOLB A. Lateral and Depth Calibration of PMD-Distance Sensors[C]//International Symposium on Visual Computing, 2006: 524-533

[3] HUSSMANN S, HERMANSKI A, EDELER T. Real-Time Motion Artifact Suppression in TOF Camera Systems[J]. IEEE Trans on Instrumentation & Measurement, 2011, 60(5):1682-1690

[4] LINDNER M, KOLB A. Compensation of Motion Artifacts for Time-of-Flight Cameras[C]//Dagm 2009 Workshop on Dynamic 3D Imaging, Springer-Verlag, 2009:16-27

[5] LEFLOCH D, HOEGG T, KOLB A. Real-Time Motion Artifacts Compensation of ToF Sensors Data on GPU[C]//Three Dimensional Imaging, Visualization, and Display International Society for Optics and Photonics, 2013: 87380U-87380U

[6] CRABB R, MANDUCHI R. Probabilistic Phase Unwrapping for Single-Frequency Time-of-Flight Range Cameras[C]//International Conference on 3D Vision, 2014:577-584

[7] JONGENELEN A P P, BAILEY D G, PAYNE A D, et al. Analysis of Errors in ToF Range Imaging with Dual-Frequency Modulation[J]. IEEE Trans on Instrumentation & Measurement, 2011, 60(5):1861-1868

[8] DROESCHEL D, HOLZ D, BEHNKE S. Multi-Frequency Phase Unwrapping for Time-of-Flight Cameras[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010:1463-1469

[9] TI C, YANG R, DAVIS J. Single-Shot Time-of-Flight Phase Unwrapping Using Two Modulation Frequencies[C]//International Conference on 3D Vision, 2016: 667-675

[10] LAWIN F J, FORSSÉN P E, OVRÉN H. Efficient Multi-Frequency Phase Unwrapping Using Kernel Density Estimation[C]//European Conference on Computer Vision, Springer, 2016:170-185

[11] JÄHNE B. Theoretical and Experimental Error Analysis of Continuous-Wave Time-of-Flight Range Cameras[J]. Optical Engineering, 2009, 48(1): 3602

[12] LUCAS B D, KANADE T. An Iterative Image Registration Technique with an Application to Stereo Vision(DARPA)[J]. Nutrient Cycling in Agroecosystems, 1981, 83(1): 13-26

[13] XIANG L, ECHTLER F. libfreenect2: Release 0.2 [EB/OL]. (2016-04-28)[2018-02-22]. <http://github.com/OpenKinect/>

libfreenect2

- [14] BROX T, MALIK J. Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2011, 33(3):500-13
- [15] FELSBERG M, FORSSEN P E, SCHARR H. Efficient Robust Smoothing of Low-Level Signal Features[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2009, 28(2):209-22

Real-Time Artifact Compensation for Depth Images of Multi-Frequency ToF

WANG Sen¹, WANG Runxiao¹, ZUO Xinxin², YU Weiwei¹

(1.School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China;
2.School of Computer Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: During the last few years, Time-of-Flight(TOF) sensor achieved a significant impact onto research and industrial fields due to that it can capture depth easily. For dynamic scenes and phase fusion, ToF sensor's working principles can lead to significant artifacts, therefore an efficient method to combine motion compensation and kernel density estimate multi-frequency unwrapping is proposed. Firstly, the raw multi-phase images are captured, then calculate the optical flow between each frequency. Secondly, by generating multiple depth hypotheses, uses a spatial kernel density estimation is used to rank them with wrapped phase images. Finally, the accurate depth from fused phase image is gotten. The algorithm on Kinect V2 is validated and the pixel-wise part is optimized using GPU. The method shows its real time superior performance on real datasets.

Keywords: ToF; multi-frequency; optical flow; kernel density estimation; Kinect V2