Journal of System Simulation

Volume 30 | Issue 5 Article 13

1-3-2019

Key Frame Extraction for SLAM Based on Spatio-temporal Slices

Xinliang Zhang
School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China;

Li Yang
School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China;

Yunji Zhao
School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China;

Follow this and additional works at: https://dc-china-simulation.researchcommons.org/journal

Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Key Frame Extraction for SLAM Based on Spatio-temporal Slices

Abstract

Abstract: To deal with the data redundancy and computational complexity of the mobile robot 3D visual SLAM (simultaneous localization and mapping), a key frame extraction technique based on spatio-temporal slices is proposed. Based on the analysis of the texture characteristics of the spatio-temporal slices, it yields that *the motion state of the image acquisition device can be represented by the slope change of the texture.* T*he similarity degree between two adjacent slices is measured to find out the key frame by use of the nearest neighbor pixel-matching algorithm*, which can accurately describe the motion state of the device with the reduction of the redundant frames. The experimental results show that the proposed method can reduce the data redundancy in frames effectively, and presents an advantage in the localization capability and processing speed in SLAM due to the exclusion of the unnecessary computation.

Keywords

SLAM, key frame, spatio-temporal slices, nearest neighbor pixel-matching algorithm

Recommended Citation

Zhang Xinliang, Li Yang, Zhao Yunji. Key Frame Extraction for SLAM Based on Spatio-temporal Slices[J]. Journal of System Simulation, 2018, 30(5): 1724-1729.

This paper is available in Journal of System Simulation: https://dc-china-simulation.researchcommons.org/journal/vol30/iss5/13

第 30 卷第 5 期
2018 年 5 月

系统仿真学报©
Journal of System Simulation

Vol. 30 No. 5
May, 2018

一种基于时空切片的 SLAM 关键帧提取方法

张新良，李阳，赵运基
(河南理工大学电气工程与自动化学院，河南 焦作 454000)

摘要：针对移动机器人三维视觉同步定位与地图构建 SLAM (Simultaneous Localization And Mapping) 过程存在的数据冗余和计算复杂度高的问题，提出了一种基于时空切片的关键帧提取技术。基于对时空切片纹理的特征分析，*将图像采集设备的运动状态反映为时空切片纹理的倾斜度变化，进而通过度量相邻时空切片间像素的相近程度，利用最邻像素匹配法，确定能够准确描述图像采集设备运动状态的关键帧*。实现了在定位精度要求下的关键帧选取，降低 SLAM 过程中的计算量。实验结果表明，该方法能有效减少冗余帧，在定位性能和处理速度上具有优势。

关键词：同步定位与地图构建；关键帧；时空切片；最邻像素匹配法
中图分类号：TP24　　文献标识码：A　　文章编号：1004-731X (2018) 05-1724-06
DOI: 10.16182/j.issn1004731x.joss.201805013

Key Frame Extraction for SLAM Based on Spatio-temporal Slices

Zhang Xinliang, Li Yang, Zhao Yunji
(School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: To deal with the data redundancy and computational complexity of the mobile robot 3D visual SLAM (simultaneous localization and mapping), a key frame extraction technique based on spatio-temporal slices is proposed. Based on the analysis of the texture characteristics of the spatio-temporal slices, it yields that *the motion state of the image acquisition device can be represented by the slope change of the texture*. T*he similarity degree between two adjacent slices is measured to find out the key frame by use of the nearest neighbor pixel-matching algorithm*, which can accurately describe the motion state of the device with the reduction of the redundant frames. The experimental results show that the proposed method can reduce the data redundancy in frames effectively, and presents an advantage in the localization capability and processing speed in SLAM due to the exclusion of the unnecessary computation.

Keywords: SLAM; key frame; spatio-temporal slices; nearest neighbor pixel-matching algorithm

引言

同步定位与地图创建(SLAM)问题是指把移动机器人放在未知环境中，机器人增量式地构建未知环境的连续地图，同时精确定位自己在地图上的位置，SLAM 问题的解决是实现移动机器人真正自主的核心技术[1-3]。机器人的地图建模与定位依赖于其自身的传感器，由于 RGB-D 传感器(如 Kinect)可同时提供场景的彩色与深度信息，因此近年来其 3 维建模与定位技术引起了研究者们的广泛关注。

利用 RGB-D 传感器获得的图像构建拓扑地图

收稿日期：2016-06-23　　修回日期：2016-09-09；
基金项目：国家自然科学基金(U1404612)，教育部博士点基金-新教师类(20134116120003)；
作者简介：张新良(1978-)，男，山东潍坊，博士，副教授，研究方向为系统建模与控制、检测技术与自动化装置等；李阳(1990-)，男，河南信阳，硕士生，研究方向为智能仪器。

Published by Journal of System Simulation, 2018

环境的过程，本质上还是增量式的地图创建过程，即根据机器人的运动模型和传感器的观测模型，实现系统状态的预测更新和观测更新。因此，如果在建模过程中对采集到的每幅图片都进行处理，则不可避免地会产生数据冗余，影响定位与构图的实时性。研究 RGB-D 图像的关键帧选取方法，在保证定位精度的要求下减少冗余帧，是提高移动机器人的建模效率的有效方法。如基于空域特征的关键帧提取算法[4]，通过度量当前帧与前一帧的匹配距离如旋转角度或平移距离来产生新的关键帧；基于时域特征的时间均匀采样法[5]，利用连续采样的图片外观与采样时间之间的相关性来确定关键帧的选取位置等。但空域关键帧提取算法需要对 RGB-D 传感器的每一帧图像进行比对，因此其计算复杂度仍然较高，而时域关键帧提取方法由于采用的是时域上的均匀采样，因此其无法区分移动机器人的加速、减速及停止等复杂的运动状态。本文考虑图像在空域和时域上的特征，引入时空切片技术来分析 RGB-D 传感器图像的三维结构特性，然后利用最近邻像素匹配法来确定时空域中连续帧的内容稳定区域，进行 SLAM 关键帧的选取。

1 时空切片的特性分析

时空切片是将连续帧上相同位置的像素条带，在时间轴上按时序展开后形成的一副图像，如图 1 所示，根据时空切片提取的方式不同，时空切片也具有多样性[6]，水平时空切片表示沿着图像水平方向 X 切割得到的像素，垂直切片则是沿着垂直方向 Y 得到的像素。对于一个含有 N 帧的图像数据集而言，其水平时空切片即为固定 Y 坐标连续提取到的一行像素点的组合，其中第 i 帧图像中提取的水平切片 h_i 可表示为：

$$h_i = \left(P_i(1, y_k), \ldots P_i(x, y_k), \ldots P_i(m, y_k)\right)$$
$$0 < i < N, 0 < k < n$$

其中：(x, y)表示图像维，每帧图像的大小为 $m\times n$，$P_i(x, y_k)$表示在第 i 帧的(x, y_k)处的像素值。每一帧的同一处提取到的时空切片按帧顺序合并在一起，即可得到一个二维的时空切片图像[7]。

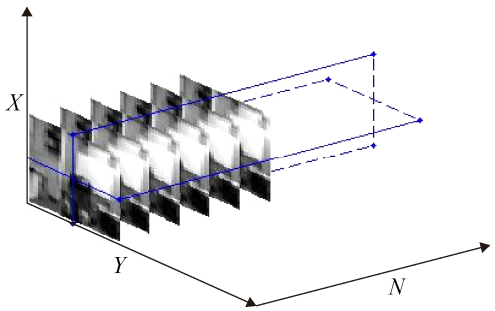
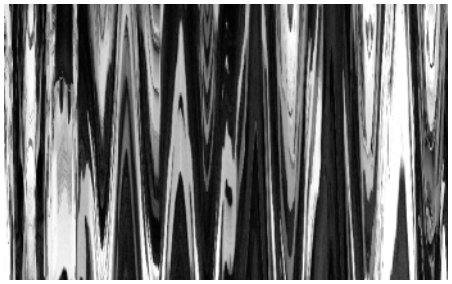


图 1 时空切片的选取
Fig. 1 Extraction of spatio-temporal slices

下面使用数据集来分析时空切片的特性，进而证明时空切片纹理特征能够反映图像采集设备的运动情况，可以作为提取关键帧的一种度量方法。

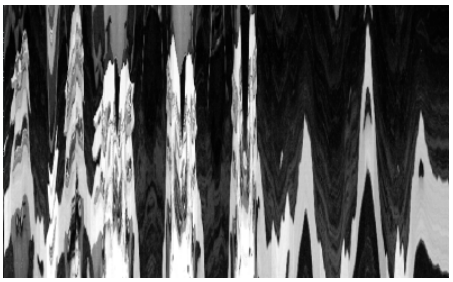
图 2 所示为从不同数据集提取出来的时空切片图像[8]，其中图 2(a)时空切片的数据集 freiburg1_rpy 对应场景是结构和纹理较为复杂的办工桌，图像采集设备位置固定、同时沿着主轴做旋转运动，可以看出，其切片纹理和色调均体现出连续变化的特征，且纹理变化非常明显。图 2(b)时空切片的数据集 freiburg1_xyz 对应场景与图 2(a)类似，但其图像采集设备仅做平移运动，其时空切片纹理和色调同样体现出了交替变化的特征。图 2(c)时空切片数据集 freiburg1_floor 对应场景是大部单一颜色的室内地板，图像采集设备作平移和旋转运动，其时空切片纹理也出现不断变化的现象，但颜色大多为单一色调。观察图 2(a)~(c)切片图像的变化，可以得到结论：图像采集设备旋转或平移运动时，都会引起时空切片纹理斜率的变化，而时空切片色调的变化则反映了场景背景的变化。



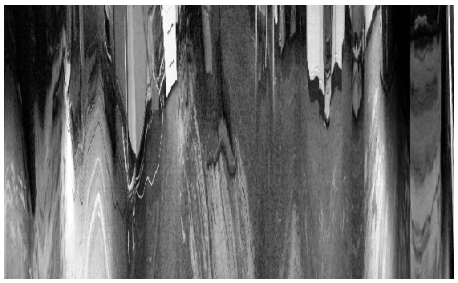
(a) 数据集 freiburg1_rpy 的时空切片

https://dc-china-simulation.researchcommons.org/journal/vol30/iss5/13
DOI: 10.16182/j.issn1004731x.joss.201805013

第 30 卷第 5 期 系统仿真学报 Vol. 30 No. 5
2018 年 5 月 Journal of System Simulation May, 2018

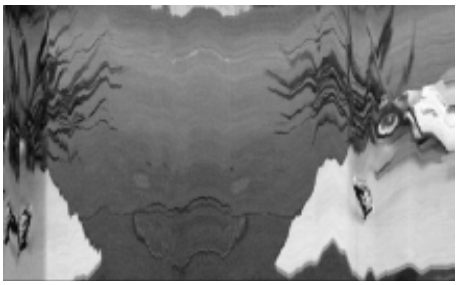


(b) 数据集 freiburg1_xyz 的时空切片



(c) 数据集 freiburg1_floor 的时空切片

(d) 数据集 freiburg2_rpy 的时空切片



(e) 数据集 freiburg2_xyz 的时空切片

图 2 不同数据集的切片纹理变化

Fig. 2 Texture changes of slices from different data sets

进一步，分析图像采集设备不同的运动状态对时空切片纹理特征的影响。采用与图 2(a)数据集类似的场景，图像采集设备同样位置固定只做旋转运动，但运动速度降低，对应的时空切片图像如图 2(d)所示。可以看出，对应图像采集设备缓慢的运动速度，其纹理和色调的变化趋于平缓。对于图像采集设备做平移运动的场合，采用与图 2(b)类似的场景，通过降低图像采集设备的平移速度得到对应的时空切片图像如图 2(e)所示，可以看出所得时空切片较之图 2(b)其纹理和色调的变化也趋于平缓。因此，得到结论：图像采集设备运动越剧烈，则其时空切片纹理的变化越明显，体现为纹理的倾斜度越大。

2 最近邻像素匹配法

最近邻像素匹配法[9]是将当前切片上每一个像素与前一帧对应像素及其临近点相比较，通过度量两像素点的相近程度来判断连续两帧之间图像采集设备的运动状态，对应于时空切片上，即摄像机运动较为剧烈时体现为纹理倾斜的情况。如图 3 所示，设 $I(N, m)$为第 N 帧切片上的第 m 个像素点，取其前一帧切片对应从 $I(N\text{-}d,\ m\text{-}d)$至 $I(N\text{-}1,\ m+d)$的 $2d+1$ 个像素点，分别度量 $2d+1$ 个像素点与 $I(N, m)$相近程度，即利用两个像素点的像素值之差作为像素点的相近程度，差值越小，则相近度越高；反之则越低，如果其相近程度超出设定的阈值，则认为该点处发生了突变，将产生新的关键帧。

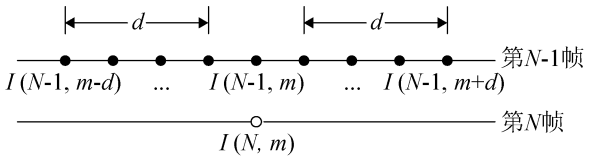


图 3 最近邻像素匹配法

Fig. 3 Nearest neighbor pixel-matching algorithm

最近邻像素匹配法提取关键帧的算法步骤描述如下。

① 将采集到的当前帧 RGB 三通道处理，取各单通道图片上 1/2 高度处的像素，构造水平切片；

② 利用最近邻像素匹配法，度量当前切片上的各像素与前一帧对应位置邻域内的像素的相近程度，若当前像素无历史相近颜色像素，则该像素点为突变像素点；分别统计三个单通道图像切片上的突变像素个数并求和；

③ 分别取当前图片上 1/4、3/4 高度处的水平切片，按步骤①和②，求出相应高度的突变像素之和；

④ 将 1/2、1/4、3/4 高度处切片的突变像素个数求和，若超过指定的某一阈值，则当前帧更新为关键帧，否则，关键帧不变。

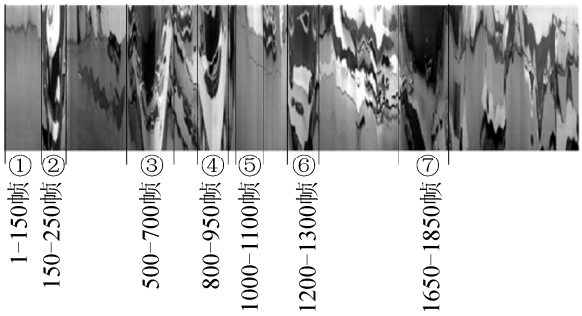
3 仿真实验

3.1 算法可靠性验证

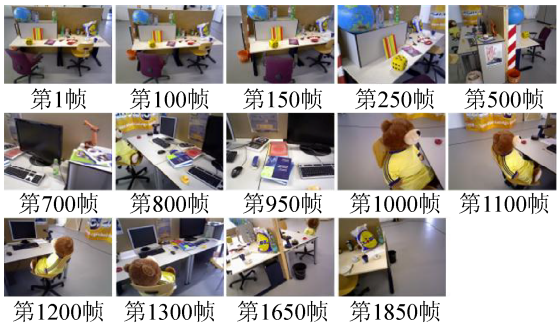
为了验证算法的可靠性，取文献[8]中的数据集“freiburg3_long_office_household”作为分析对象，其表示的是图像采集设备绕室内的一张有复杂纹理的办工桌运动一周回到原点的过程。其运动轨迹中包含了图像采集设备的旋转、平移、速度的变化多种不同的运动姿态，因此依据其提取到的时空切片纹理具有多样性，更能体现图像采集设备在应用中的实际运行状态。

数据集的连续时空切片图像如图 4(a)所示，图 4(b)是数据集的几个关键时间点上的图像帧，图 4(c)是利用最近邻像素匹配法(突变阈值为 100)提取的关键帧按时序排列得到的关键帧时空切片图像。由图 4(b)中的图像可以看出，对应第 1-150 帧和第 1 000-1 200 帧图像，其图像内容变化不大，即图像采集设备运动的幅度较小，对应图 4(a)中的时空切片，其纹理趋于水平且变化平缓，而在图 4(c)中则表现为构造的关键帧数量极少，其形成的切片图像要比图 4(a)中①处和⑤处要少得多，分别为 15 帧和 23 帧。对应第 150-250 帧、500-700 帧、800-950 帧、1 200-1 300 帧和第 1 650-1 850 帧之间的图像，图像差异较大，即图像采集设备产生剧烈运动，反映在图 4(a)的时空切片纹理上，则表现为明显的倾斜度变化，此时构造的关键帧的数目也急剧增加，如图 4(c)的③、④、⑥处的关键帧数目分别为 153、142 和 75，与提取之前的数目变化不大。

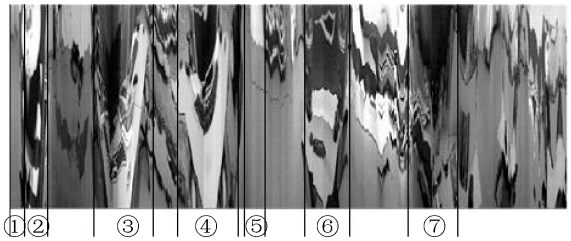
由此可以得出结论，基于时空切片，利用最近邻像素匹配方法选取关键帧，能有效去除图像采集设备缓慢运动时的冗余图像信息，保留含有关键信息的片段。



(a) 数据集的时空切片



(b) 关键时间点上的图像帧



(c) 提取关键帧后的时空切片

图 4 时空切片与关键帧

Fig. 4 Spatio-temporal slices and the key frame

3.2 实验验证

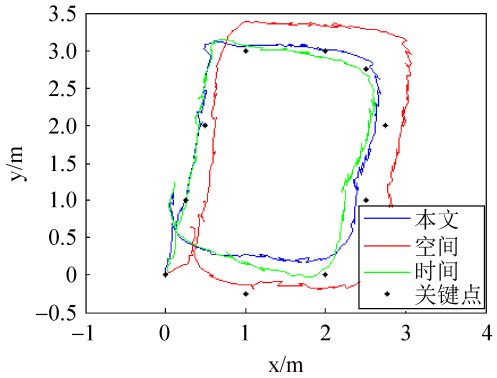
为了验证本文的时空切片关键帧选取算法的性能，与文献[7]的时间均匀采样法及文献[6]的空间均匀采样法进行比较，算法均在 RGB-D SLAM 平台上运行。实验采用华硕 Xtion RGB-D 图像传感器，RGB 图像分辨率 640×480，最高帧率 30fps，硬件平台为 Lenovo PC，2.6GHz Intel Core i5 CPU，8G RAM。

在室内进行两次实验，实验中图像传感器均在同一水平面上、按同一路径、围绕同一对象(一张办公桌)，环绕一周回到起始点位置。第一次实验，图像传感器匀速运动，速度约为 10 cm/s，对应传感器的简单运动状态；第二次实验，图像传感器分别做

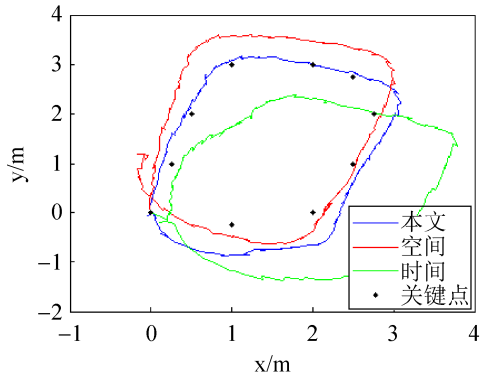
https://dc-china-simulation.researchcommons.org/journal/vol30/iss5/13
DOI: 10.16182/j.issn1004731x.joss.201805013

第 30 卷第 5 期 系统仿真学报 Vol. 30 No. 5
2018 年 5 月 Journal of System Simulation May, 2018

加速、减速至停止运动，对应传感器的复杂运动状态。分别使用上述三种选取算法来提取关键帧，所取关键帧数目为总帧数的 1/4，由关键帧所描绘的图像采集设备运动轨迹如图 5(a)和(b)所示。



(a) 图像采集设备简单运动模式的轨迹



(b) 图像采集设备复杂运动模式的轨迹

图 5 关键帧提取算法的图像采集设备定位性能

Fig. 5 Localization performance of the image acquisition device for the key frame extraction algorithm

此处我们使用均方意义下的距离来评价三种路径估计结果的优劣，对每条路径，将所有取自路径上的点(在 y 轴坐标相等的情况下与关键点相对应的点)与关键点求距离的平方和。由图 5(a)可以看出，当图像采集设备只做简单的近似匀速运动时，时间均匀采样法、空间均匀采样法和本文算法提取的关键帧所描述路径的逼近误差分别是 1.17、4.18 和 1.33。在该评价指标下，本文提出的时空切片算法与时间均匀采样法定位效果相近，但优于空间均匀采样法。

当图像采集设备做复杂的变速运动时，如图 5(b)所示。时间均匀采样法、空间均匀采样法和本文算法提取的关键帧所描述路径的逼近误差分别是 11.09、5.32 和 1.87。可以看出，本文提出的时空切片算法精度要优于时间均匀采样法和空间均匀采样法。即本文的时空切片关键帧选取算法在图像采集设备做低速简单运动和复杂运动时，均能表现出很好的定位精度，尤其在复杂运动情况下，本文算法计算精度要远优于其他两种算法。

作为对比，以文献[8]数据集 freiburg1 中的 8 个典型室内视频序列为例比较三种算法的定位性能，其中<frl/xyz>、<frl/rpy>包含较简单的相机运动，其他序列包括大规模室内场景与不规则的相机运动。表 1 给出了每个数据集的总帧数，3 种算法采样得到的关键帧数及各算法的 ATE(绝对路径误差)和 RPE(相对相机姿态误差)误差对比。

表 1 关键帧提取算法的性能比较

Tab. 1 Performance comparison of the key frame extraction methods

数据集	总帧	本文				时间均匀采样法[7]				空间均匀采样法[6]			
		关键帧数	FPS	ATE	RPE	关键帧数	FPS	ATE	RPE	关键帧数	FPS	ATE	RPE
fr1/360	745	180	4.26	**0.120**	**0.155**	186	4.63	0.177	0.224	190	4.38	0.152	0.208
fr1/desk2	621	160	4.53	**0.085**	0.141	155	4.60	0.361	0.520	153	4.06	0.123	0.148
fr1/desk	577	139	3.90	0.052	0.098	144	3.85	**0.029**	0.048	140	3.75	0.033	**0.047**
fr1/plant	1 122	283	2.11	0.087	0.216	280	2.24	**0.078**	**0.174**	281	1.77	0.079	0.176
fr1/room	1 351	340	1.88	**0.177**	**0.278**	337	2.13	0.310	0.491	332	1.94	0.271	0.353
fr1/rpy	697	173	2.90	0.029	0.036	174	2.95	**0.027**	0.036	180	2.46	**0.034**	0.050
fr1/teddy	1 403	344	1.66	0.125	0.148	350	1.75	**0.098**	**0.147**	342	1.63	0.270	0.365
fr1/xyz	793	207	2.35	0.020	0.025	198	2.44	0.019	0.025	203	1.75	**0.018**	**0.024**
平均	——	——	2.95	0.087	0.137	——	3.07	0.137	0.208	——	2.72	0.122	0.174

从表 1 中可以看出，在简单运动的<frl/xyz>、<frl/rpy>处，本文提出的时空切片算法，其 ATE 和 RPE 误差与时间均匀采样算法及空间均匀采样算法相比性能相当，但在包含相机复杂运动的数据集尤其是<fr1/360>、<fr1/desk2>和<fr1/room>中的相机姿态估计效果，精确度明显优于其他两种算法，平均 ATE 误差达到 0.087，RPE 误差达到 0.137。从处理帧的速度来看，在关键帧数目大致相同的情况下，本文算法的平均帧率是 2.95FPS，而空间均匀采样法的平均帧率为 2.72FPS，即平均帧率提高 8.5%。因此，无论从误差精度还是处理速度上来看，本文采用的算法都具有非常良好的效果。

4 结论

本文使用了一种基于时空切片的关键帧提取技术，对传统的 SLAM 算法寻找关键帧过程中时间复杂度高，计算量大的问题进行了优化，在保证一定精度的情况下，提取合适数目的关键帧进行处理，有效地减少了冗余帧。与常用的时间均匀采样算法和空间均匀采样算法对比结果显示，本文提出的时空切片关键帧选取算法无论在计算精度上还是在处理速度上，都具有良好的特性。

参考文献:

[1] Durrantwhyte H, Bailey T. Simultaneous Localization and Mapping: Part I[J]. IEEE Robotics & Automation Magazine(S1070-9932), 2006, 13(2): 99-110.

[2] 梁明杰，闵华清，罗荣华. 基于图优化的同时定位与地图创建综述[J]. 机器人, 2013, 35(4): 500-512.
Liang Mingjie, Min Huaqing, Luo Ronghua. Graph-based SLAM: A Survey[J]. ROBOT, 2013, 35(4): 500-512.

[3] 许亚芳，孙作雷，曾连荪，等. 基于多次测量更新的移动机器人 SLAM 仿真[J]. 系统仿真学报，2015, 27(6): 1288-1293.
Xu Yafang, Sun Zuolei, Zeng Liansun, et al. Mobile Robot SLAM Simulation with Multi Measurement Update[J]. Journal of System Simulation, 2015, 27(6): 1288-1293.

[4] Stuckler J, Behnke S. Integrating Depth and Color Cues for Dense Multi-Resolution Scene Mapping Using RGB-D Cameras[C]// IEEE International Conference on Multisensor Fusion and Information Integration. USA: IEEE, 2012: 162-167.

[5] Ho K L, Newman P. Loop closure detection in SLAM by combining visual and spatial appearance[J]. Robotics & Autonomous Systems(S0921-8890), 2006, 54(9): 740-749.

[6] Ngo C W, Pong T C, Zhang H J. Motion analysis and segmentation through spatio-temporal slices processing[J]. IEEE Transactions on Image Processing (S1057-7149), 2003, 12(3): 341-355.

[7] Ngo C W, Pong T C, Zhang H J. On clustering and retrieval of video shots through temporal slices analysis[J]. IEEE Transactions on Multimedia (S1520-9210), 2002, 4(4): 446-458.

[8] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]// IEEE/RSJ International Conference on Intelligent Robots & Systems, USA: IEEE, 2012: 573-580.

[9] 李勇，刘雨，唐波，等. 一种基于时空切片的镜头突变检测方法[J]. 计算机工程与科学，2006, 28(3): 53-54.
Li Yong, Liu Yu, Tang Bo, et al. An Algorithm for Abrupt Video Shot Change Detection Based on Spatio-Temporal Slices[J]. Computer Engineering & Science, 2006, 28(3): 53-54.

https://dc-china-simulation.researchcommons.org/journal/vol30/iss5/13
DOI: 10.16182/j.issn1004731x.joss.201805013