

一种视频序列拼接的新方法*

谢绍霞¹, 郭三华², 初玲², 曹立娟³, 王万君³

(1.烟台汽车工程职业学院 汽车工程系, 山东 烟台 265500;

2.烟台汽车工程职业学院 电子工程系, 山东 烟台 265500;

3.烟台汽车工程职业学院 科研处, 山东 烟台 265500)

摘要: 提出了一种基于视频序列拼接的新方法。首先, 利用 KLT 算法对视频序列中特征点进行提取和跟踪, 实现关键帧粗略选取; 其次, 在选取的关键帧中利用 SURF 算法进行特征提取, 利用最近邻距离比进行特征点匹配, 通过 RANSAC 估计算法求精单映矩阵, 并结合关键帧选取判定寻找最佳关键帧; 最后, 利用级联单映矩阵和加权融合算法实现视频序列拼接。实验验证了该方法的有效性。

关键词: KLT 算法; 关键帧; SURF 算法; 单映矩阵

中图分类号: TP391.41

文献标识码: A

文章编号: 1674-7720(2011)21-0040-03

The new mosaic algorithm of video sequence

Xie Shaoxia¹, Guo Sanhua², Chu Ling², Cao Lijuan³, Wang Wanjun³

(1.Department of Automotive Engineering, Yantai Vocational College of Automotive Engineering, Yantai 265500, China;

2.Department of Electronic Engineering, Yantai Vocational College of Automotive Engineering, Yantai 265500, China;

3.Department of Scientific Research, Yantai Vocational College of Automotive Engineering, Yantai 265500, China)

Abstract: This paper proposes a new mosaic algorithm of video sequence. Firstly, the rough key frames are extracted from the video sequence using KLT feature extraction and tracking. Then, the features of the rough key frames extracted by SURF algorithm are matched by the nearest neighbor. With the RANSAC robust estimation for homography matrix and key frames selection judgment, the final key frames are extracted. Finally, with the cascade of homography matrix, the seamless mosaic is rendered by the image blending. The experiment demonstrates this algorithm's efficient.

Key words: KLT algorithm; key frames; SURF algorithm; homography matrix

图像拼接是构建高分辨率大场景的关键技术, 在虚拟现实场景表达、计算机视觉、全景图绘制中有着重要应用, 也是计算机图形学领域中研究热点^[1]。视频序列拼接是由多帧视频序列拼接而成的完整全景图像, 在视频监控、医学图像处理、遥感图像处理等方面得到广泛应用^[2]。

在视频序列的拼接中, 相邻帧间重叠部分较大, 若每相邻帧都做一次拼接, 不仅耗费大量时间, 而且随着所需拼接帧数量增多, 匹配误差会增大, 造成拼接效果不佳。利用关键帧拼接表示整个视频序列拼接成为有效的方法。参考文献[3]提出 SIFT 结合 Kalman 跟踪算法进行关键帧的提取及拼接的实现, 由于视频序列帧数量较

大, 而 SIFT 算法本身复杂, SIFT 对每一帧都进行处理, 累积造成运算时间比较长。参考文献[4]提出利用分层式自适应帧采样的视频拼接, 但这种方法的计算量大, 算法限制条件较多, 容易导致拼接失败。参考文献[5]提出采用二叉树方式来解决重叠区域大小确定问题, 但是没有从根本上解决大量视频序列拼接时匹配误差增大的问题。针对以上问题, 提出了一种新的视频序列拼接算法, 可以有效提取关键帧, 利用关键帧拼接表示整个视频序列拼接, 从而节省视频拼接时间。首先, 利用 KLT 算法对视频序列中每帧特征点进行提取, 并通过特征点的跟踪实现进行关键帧粗略选取。其次, 在选取的关键帧中利用 SURF 算法进行纹理特征提取, 并利用最近邻距离比进行特征点匹配^[6], 通过估计算法求精单映矩

* 基金项目: 山东省高等学校科技计划项目(J10LG78)

阵,结合关键帧选取判定寻找最佳关键帧。最后,通过RANSAC级联单映矩阵和加权融合算法实现拼接,取得较好效果。

1 帧间的单映变换矩阵及关键帧选取判定

单映矩阵变换是一种常用的帧间变换模型,主要用于任意场景空间摄像机为旋转或者缩放运动,或者空间为平面场景和任意摄像机的运动。单映矩阵变换表示为:

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \mathbf{H} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

其中, $\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & h_8 \end{bmatrix}$, 一般 \mathbf{H} 有 8 个自由度, $h_8=1$ 。

对于相邻的关键帧,可以直接采用上述帧间变换模型,但是对于非相邻的关键帧,可以利用单映矩阵的级联性质,得到非相邻关键帧之间的单映变换矩阵。

假设第 k 帧、第 h 帧为非相邻的关键帧,以第 k 帧作为参考帧,第 h 帧为目标帧,利用单映矩阵的级联性质,可以得到两非相邻关键帧的单映变换矩阵:

$$\mathbf{H}_{k,h} = \mathbf{H}_{k,l} \mathbf{H}_{l,n} \cdots \mathbf{H}_{m,t} \mathbf{H}_{t,h} \quad (1)$$

其中, $\mathbf{H}_{k,l}, \mathbf{H}_{l,n}, \cdots, \mathbf{H}_{m,t}, \mathbf{H}_{t,h}$ 分别是第 k 帧、第 h 帧之间的相邻关键帧的单映变换矩阵,如图 1 所示。

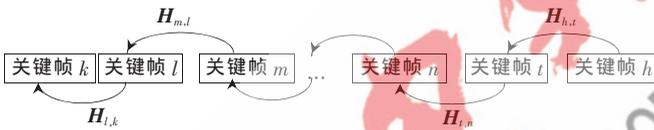


图 1 单映矩阵的级联示意图

在关键帧选取效果不好的情况下,非相邻关键帧数量比较多,容易造成匹配误差增大。为了进一步减少关键帧的数量,减小单映矩阵级联时造成的误差,通过如下比较进行运算,进一步提取关键帧,从而完成关键帧的选取判定。具体步骤为:

(1) 设定第 k 帧、第 h 帧为非相邻的关键帧,直接计算两关键帧单映矩阵 $\mathbf{H}_{k,h}$;

(2) 将单映矩阵级联方法和直接计算单映矩阵方法计算出来的结果进行比较,比较 $h_0, h_1, h_2, h_3, h_4, h_5, h_6, h_7$ 值的误差大小,只要有一项数值超过预定的阈值,则认为第 h 帧的前一项关键帧是须保留的,第 k 帧和第 h 帧前一项关键帧作为保留关键帧,两者之间的其他关键帧可以省略。

2 关键帧的提取方法

2.1 KLT 特征点跟踪算法

由于视频序列帧与帧之间的冗余较大,考虑采用 KLT 特征点跟踪算法来实现关键帧的粗略选取。KLT 算法是以待跟踪窗口在视频图像帧间的灰度差平方和作为度量的跟踪算法^[8]。对于相邻视频帧 I 和视频帧 J 中的两个窗口,直接的 SSD 为:

$$\varepsilon = \iint_W \left[J\left(X + \frac{d}{2}\right) - I\left(X - \frac{d}{2}\right) \right] W(X) dX \quad (2)$$

其中, $X = [x, y]^T$, $d = [d_x, d_y]^T$, 窗函数 $W(x, y)$ 通常是常数 1,对图像 J 进行泰勒公式展开,可以近似得出如下公式:

$$J\left(X + \frac{d}{2}\right) \approx J(X) + \frac{d_x}{2} \frac{\partial J}{\partial x}(X) + \frac{d_y}{2} \frac{\partial J}{\partial y}(X) \quad (3)$$

$$\text{同理可得: } I\left(X - \frac{d}{2}\right) \approx I(X) - \frac{d_x}{2} \frac{\partial I}{\partial x}(X) + \frac{d_y}{2} \frac{\partial I}{\partial y}(X) \quad (4)$$

为了求 SSD 最小值,则有:

$$\begin{aligned} \frac{\partial \varepsilon}{\partial d} &= 2 \iint_W \left[J\left(X + \frac{d}{2}\right) - I\left(X - \frac{d}{2}\right) \right] \left[\frac{\partial J\left(X + \frac{d}{2}\right)}{\partial d} - \frac{\partial I\left(X - \frac{d}{2}\right)}{\partial d} \right] W(X) dX \\ &\approx \iint_W \left[J(X) - I(X) + \frac{1}{2} g^T d \right] g(X) W(X) dX \end{aligned} \quad (5)$$

其中, $g = \left[\frac{\partial}{\partial x}(I+J), \frac{\partial}{\partial y}(I+J) \right]$ 。

为得到偏移量 d ,令偏分方程为零,得:

$$\mathbf{Z}d = e \quad (6)$$

其中, \mathbf{Z} 为 2×2 矩阵, $\mathbf{Z} = \iint_W g(X) g^T(X) W(X) dX$, $g \cdot g = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix}$;

e 为 2×1 矩阵且 $e = 2 \iint_W [I(X) - J(X)] g(X) W(X) dX$ 。

利用计算 \mathbf{Z} 的特征值,选取两个特征值比较大的点作为跟踪的特征点,在选定的特征点基础上再利用 $\mathbf{Z}d = e$ 可以计算出特征点的漂移。跟踪停止条件是特征值小于规定的最小特征值以及迭代次数超过设定次数。

2.2 SURF 特征提取算法

KLT 特征点跟踪算法有较高的求解效率,但是对纹理变化复杂的情况,常由于误匹配而造成被跟踪点的丢失。因此,利用上述 KLT 特征点跟踪算法提取关键帧后,对关键帧再利用 SURF 算法进行特征点的提取,为后续提取最佳关键帧提供有效方法。

SURF 算法利用快速 Hessian 检测算法提取特征点^[9], Hessian 矩阵具有良好的计算时间和精度表现。

SURF 特征描述子的提取可以分为两步:(1) 根据特征点周围的一个圆形区域找到特征点的主方向;(2) 在选定的主方向上构建一个矩形区域,并提取所有的特征描述点信息。在主方向上构建一个大小为 20σ 的窗口(σ 表示尺度),并将该窗口区域分为 4×4 的子区域,对于每一个子区域,分别计算相对于主方向的水平和垂直方向 Haar 小波响应,每个子区域得到 4 维向量,因此 4×4 的子区域得到 64 维特征点描述子,它可以扩展到 128 维的特征点描述子,一般采用 128 维特征点描述子。

3 拼接算法的实现及实验结果

3.1 拼接方法的实现

按照上述所述,拼接的具体实现步骤如下:

(1) 为了选取关键帧子序列,使全景图内容丰富,第一帧和最后一帧为必选关键帧,选取第一帧视频序列关键帧为基准帧,提取基准帧的特征点。

(2) 利用 KLT 算法进行特征点的提取并跟踪,从而确定粗略的关键帧,具体如下:

① 假设选取的特征点个数为 N ,比例因子为 α ,对视频序列经过特征点跟踪,当特征点个数减至 αN 时,停止跟踪,选取当前帧为关键帧,并作为后续跟踪的基准帧;

② 重复上述过程,直至视频序列跟踪完毕,最后获取粗略视频关键帧,对原始视频帧图像进行跟踪,计算量较大,为减少计算量,利用高斯图像金字塔,并通过插值获取原始视频帧中特征点^[10]。

(3) 利用 SURF 算法对关键帧提取特征点,采用最近邻距离比进行特征点匹配,并利用帧间单映矩阵模型和关键帧选取判定方法进行优化的关键帧选择。具体如下:

① 利用 SURF 特征点提取算法对步骤 (2) 选定的关键帧进行特征提取;

② 对相邻关键帧利用帧间单映矩阵模型进行匹配计算。为了使单映矩阵 H 的估计准确,利用 RANSAC 鲁棒估计方法得到相邻关键帧之间单映矩阵 H 的估计,具体步骤为:

(a) 随机抽取 $n \geq 4$ 对匹配特征点来估计矩阵 H 的参数;

(b) 对于步骤 (2) 中的每一对匹配点,计算对单映矩阵 H 的拟合误差;

(c) 设定一个门限值,若拟合误差小于此门限值,表示匹配点对是一致点,并统计一致点的数目;

(d) 重复步骤 (a)~(c),直到所有的一致点集中至少有一个有效表征集的概率大于一定的数值为止;

(e) 选择具有最大一致点集的单映矩阵 H 。

③ 对非相邻关键帧利用单映矩阵的级联性进行计算,利用关键帧选取判定方法进一步得到选定关键帧。

(4) 将步骤 (3) 选定的关键帧作为最终拼接的关键帧,利用单映矩阵级联和加权融合算法完成视频序列的拼接。

3.2 实现结果

实验采用自拍的两段视频,利用上述方法完成了视频序列的拼接,效果比较好。

图 2 所示是将拍摄的一段 200 帧的视频利用上述方法获取的最终关键帧,其拼接效果图如图 3 所示,剪切处理后的视频序列拼接最终效果图如图 4 所示。

图 5 是自拍的一段 350 帧的视频,利用上述算法获取的关键帧,视频拼接效果图如图 6 所示,剪切处理后视频序列最终拼接效果图如图 7 所示。

本文采用了一种新的视频序列拼接方法,利用 KLT 特征点跟踪算法实现粗略关键帧的选取,再次利用 SURF 特征点提取算法结合最近邻距离比匹配方法、关键帧判定准则,对关键帧进行进一步提取,并利用

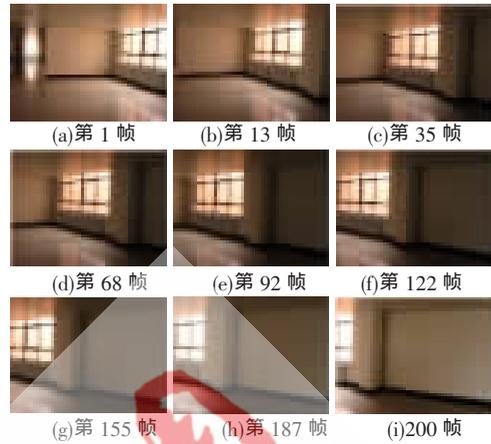


图 2 获取视频序列关键帧



图 3 关键帧表示的视频序列拼接



图 4 关键帧表示的视频序列最终拼接效果

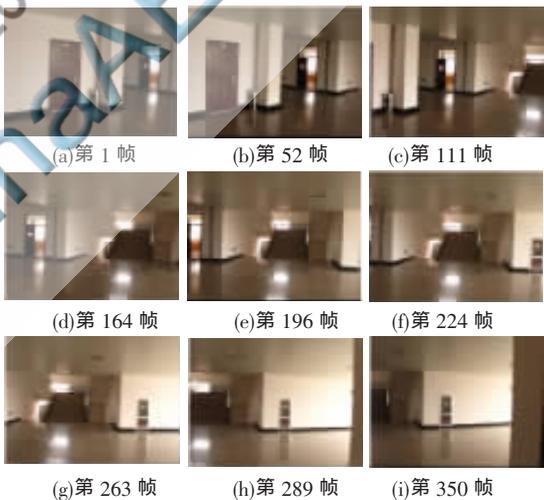


图 5 获取视频序列关键帧



图 6 关键帧视频序列拼接图



图 7 关键帧表示的视频序列最终拼接效果

RANSAC 估计算法对单映矩阵进行求精,通过级联单映矩阵和加权融合算法实现视频序列拼接,取得了较好效果。

参考文献

- [1] KIM D H, YOON Y I, CHOI J S. An efficient method to build panoramic image mosaics [J]. Pattern Recognition Letters, 2003,24 (1): 2421 - 2429.
- [2] SHUM H Y, SZELISKI R. Panoramic image mosaics [R]. Technical Report, MSR -TR -97 -23, Microsoft Research, Redmond, WA, USA, 1997:1-3.
- [3] FADAEIESLAM M J, FATHY M, SORYANI M. Key frames selections into panoramic mosaics [C]. Proceedings of the 7th International Joint Conference on Information, Communication and signal, Macau, 2009.
- [4] 刘永,王贵锦,姚安邦,等.基于自适应帧采样的视频拼接[J].清华大学学报(自然科学版),2010,50(1):108-112.
- [5] BABU D R R, RAVISHANKAR M. Automatic seamless image mosaicing: an approach based on quad-tree technique[C]. Proceedings of the World Congress on Engineering, University of Oxford, UK, 2010,London,UK.
- [6] LOWE D G. Distinctive image features from scale-invariant key points [J]. International Journal of Computer Vision, 2004,60(2):91-110.
- [7] HARTLEY R, AISSENRMAN A. Multiple view geometry in computer vision [M]. Cambridge, UK: Cambridge University Press, 2000.
- [8] TOMASI J S C. Good features to track [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA,1994:593-600.
- [9] BAY H, TUYTELAARS T, GOOL L V. SURF:speed up robust features[J]. Computer Science, 2006,3951(1):404-417.
- [10] SINHA S N, FRAHM J M, POLLEFEYS M, et al. Feature tracking and matching in video using programmable graphics hardware [J]. Machine Vision and Applications, 2007, 22(1): 207-217.

(收稿日期:2011-08-19)

作者简介:

谢绍霞,女,1972年生,讲师,主要研究方向:数字图像处理。

郭三华,男,1982年生,硕士研究生,讲师,主要研究方向:数字图像处理、计算机图形学。

初玲,女,1966年生,副教授,主要研究方向:数字图像处理。