

基于优化三维均值坐标的混合式视频融合

刘奇松^a, 马利庄^b

(上海交通大学 a. 软件学院; b. 电子信息与电气工程学院, 上海 200240)

摘要: 基于二维均值坐标插值的视频融合方法会产生局部色彩失衡和画面闪烁等失真现象。为此, 提出一种基于优化三维均值坐标的视频融合方法。在连续的视频序列上建立三维均值坐标, 根据不同的边界情况优化均值坐标的计算, 并利用图像 matting 技术提高融合效果。实验结果表明, 与现有方法相比, 该方法可以更加有效地提高融合质量。

关键词: 均值坐标; 混合式融合; 视频融合; 视频处理; 时空一致性

Hybrid Video Fusion Based on Optimized 3D Mean-value Coordinates

LIU Qi-song^a, MA Li-zhuang^b

(a. School of Software; b. School of Electronic, Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China)

【Abstract】 Interpolation-based fusion method using 2D mean-value coordinates may product some unnatural effects. Aiming at this problem, this paper presents a new method of video fusion to avoid these flaws. Based on 3D mean-value coordinates, a boundary-aware algorithm is provided to optimize the determination of 3D mean-value coordinates. Matting-based methods are used to refine produced video. Experimental results show that this method can enhance the quality of composition more effectively compared with the existing approaches.

【Key words】 mean-value coordinates; hybrid fusion; video fusion; video processing; temporal and spatial consistency

DOI: 10.3969/j.issn.1000-3428.2011.17.096

1 概述

无缝视频融合是视频编辑中的一项重要技术。它将源视频序列中的一块区域尽可能自然地粘贴到目标视频序列中。目前主要有 2 类可以实现高质量视频融合的方法: 基于 alpha 值的融合和基于梯度域的融合。文献[1]通过对局部区域求闭合形式解来获得更加自然的 α 值。基于 alpha 值方法在融合区域的边界处可以得到比较平滑的过渡。但当前景和背景色彩差别较大时, 融合的结果在整体上会显得不太协调, 而且求取 α 往往消耗大量的时间。

梯度域融合方法要求融合图像及视频在梯度域上尽可能平滑。文献[2]将泊松方程用于图像的梯度域融合, 但这种方法对边界条件依赖较高。文献[3]对边界建立特征向量, 并根据不同的特征采用不同的融合方法, 但这种方法需要大量的交互。文献[4]提出一种基于加权双正交自适应小波变换的图像融合算法, 但这种方法并不适用于视频融合。

一些基于调和坐标的融合方法已被应用到实际问题中。文献[5-6]通过使用二维均值坐标来逼近求解泊松方程, 可以得到很高的效率, 但应用在视频融合上时可能会导致画面闪烁。文献[7]说明了将 alpha 融合和梯度域融合相结合的可能性, 但主要应用在图像编辑方面。

本文提出一种基于优化的三维均值坐标的混合式视频融合方法。该方法具有如下特点: (1) 将现有的二维处理方法扩展到三维上, 从而可以充分利用视频的帧间信息; (2) 结合图像 matting 算法, 在局部处理上进行改善, 从而得到高质量的视频融合效果。

2 优化的三维均值坐标

2.1 三维均值坐标

二维均值坐标是将权值定义为一阶邻域的组合。它只依

赖于边界点和内部点之间的角度和距离。三维均值坐标^[8]与此不同。

设 Ω 是 R^3 上的一个由三角面片构成的闭合的凸多面体。设点 V 是多面体内部的一个点, 多面体表面上的任意一个三角面片都可以和点 V 构成一个四面体, 如图 1(a)所示。需要得到的是点 V 处的均值坐标。四面体可以被映射到一个球面上从而变成一个球面四面体, 如图 1(b)所示。

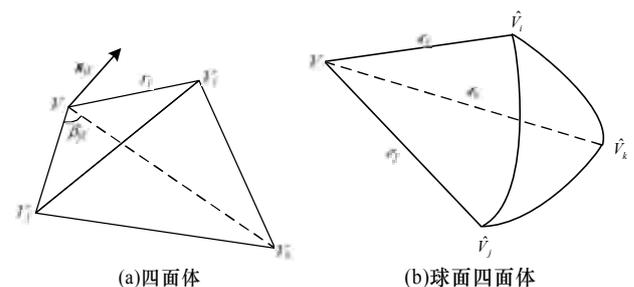


图 1 三维均值坐标

如果将点 V 看作球的中心, \hat{V}_i 则是 V_i 的映射, \hat{V}_j 、 \hat{V}_k 也是如此。定义单位向量 $e_i = (V_i - V)/r_i$, 其中, $r_i = |V_i - V|$; $V_i = \hat{V} + e_i$ 。有:

$$V = \sum_{i=1}^n \lambda_i(V) V_i \quad (1)$$

其中, $\lambda_i(V)$ 是插值的权重, 它可以用如下公式描述:

基金项目: 国家“863”计划基金资助项目(2009AA01Z334); 国家自然科学基金资助项目(61073089)

作者简介: 刘奇松(1985—), 男, 硕士研究生, 主研方向: 视频融合, 图像增强技术; 马利庄, 教授、博士生导师

收稿日期: 2011-03-15 **E-mail:** richieliuse@sjtu.edu.cn

$$\lambda_i = \omega_i / \sum_i \omega_i \quad (2)$$

$$\omega_i = \frac{1}{r_i} \sum_{V_j \in N(V_i)} \mu_j \quad (3)$$

其中, $N(V_i)$ 表示点 V_i 的邻接三角形。

根据图 1(a), 令 β_{jk} 是线段 (V, V_j) 和 (V, V_k) 的夹角, 令 \mathbf{n}_{jk} 是平面 (V, V_j, V_k) 的单位法向量, $\mathbf{n}_{jk} = (\mathbf{e}_j \times \mathbf{e}_k) / \|\mathbf{e}_j \times \mathbf{e}_k\|$ 。有:

$$\mu_i = \frac{\beta_{jk} + \beta_{ij} \mathbf{n}_{ij} \cdot \mathbf{n}_{jk} + \beta_{ki} \mathbf{n}_{ki} \cdot \mathbf{n}_{jk}}{2\mathbf{e}_i \cdot \mathbf{n}_{jk}} \quad (4)$$

一旦得到均值坐标, 就可以使用它们来对定义在边界顶点上的任意方程 φ 进行平滑插值:

$$\hat{\varphi}(x) = \sum_{i=0}^{N-1} \lambda_i(x) \varphi(V_i) \quad (5)$$

其中, x 是 Ω 内的任意点。

2.2 边界适应的网格生成算法

将一段视频分成小的连续视频序列(通常是 3 帧~7 帧), 称之为帧立方体。立方体中间的一帧是需要处理的帧, 其他的帧用来提供帧间信息。由于三维均值坐标需要在的一个闭合的多面体中进行计算, 但采样的边界点全部在平行的平面上, 因此, 并不能形成这样的一个闭合多面体。这里添加 2 个额外的辅助点来解决这个问题。

如图 2 所示, 上面一个多边形代表帧立方体中顶部(底部)的帧, $B_i(i=1, 2, \dots, n)$ 表示边界上的采样点; V 表示添加的辅助点。

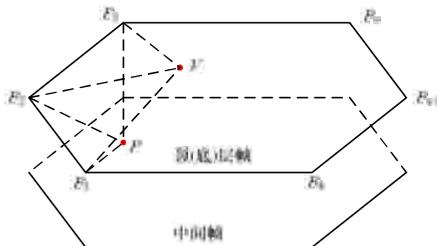


图2 边界适应的网格生成示例

点 V 不应该离点 P 太远, 否则由式(4)计算出的 μ_i 会变得太小。为实现本文目标, 点 V 的位置按以下步骤计算: (1)计算点 P 和 B_i 之间的距离; (2)选择距离点 P 最近的 m 个边界点; (3)从剩下的边界点中均匀选取 n 个点; (4)将点 V 放到这 $m+n$ 个边界点的重心。这里要求 $m>n$, 因为需要将点 V “拽”到离点 P 较近的位置。

通常这 2 个辅助点的 λ 值都比较小, 因此, 在式(2)的计算中可以忽略它们, 而不会对最终的结果产生较大的影响。

3 混合式视频融合

3.1 三维均值融合

令 $S_i \subset R^2$ 表示源视频的第 i 帧, $T_i \subset R^2$ 表示目标视频的第 i 帧($i=1, 2, \dots, k$)。 $\psi_i: S_i \rightarrow R$ 和 $\varphi_i: T_i \rightarrow R$ 是定义在 S_i 和 T_i 上的亮度方程。令 $P_s \subset S$ 表示用户标记的前景区域, $P_t \subset T$ 是目标视频中需要融合的区域, P_s 和 P_t 的边界具有相同的形状。假设它们的边界 ∂P_s 和 ∂P_t 是由线段构成, 并有相同数目的边界点。现在的任务是无缝的将 P_s 和 T_i 融合在一起。为得到令人满意的结果, 笔者希望前景和背景区域的梯度尽可能保持平滑, 特别是在 ∂P_s 和 ∂P_t 上。一般使用泊松方法:

$$\Phi = \min \int F(\nabla L, \nabla \psi) dx dy dt \quad (6)$$

其中, L 是融合后视频中的一帧; $F(\nabla L, \nabla \psi) = \|\nabla L - \nabla \psi\|^2$ 。根据变分原理, 问题可以转化成三维泊松方程的求解:

$$\Delta L = \text{div} \nabla \psi \quad (7)$$

令 $\hat{L} = L - \psi$, 有 $\Delta \hat{L} = 0$, 边界条件为 $\hat{L}|_{\partial P_t} = \sum_i \varphi_i - \psi_i$, 则

最终需要计算的方程可以简单的表示为:

$$L = \hat{L} + \psi \quad (8)$$

假设在边界条件 $\partial P_t = P_{t_0}, P_{t_1}, \dots, P_{t_m} = P_{t_0}$ 下, P_t 是 P_s 需要被粘贴到的目标区域, $x \in P_t$ 是一个内部点, 可以通过三维均值坐标来插值得到 \hat{L} , 而不用计算式(8)。此时,

$$\hat{L}(x) = \sum_{j=0}^k \sum_{i=0}^{n-1} \lambda_{ji}(x) (\varphi - \psi)(P_{ji}) \quad (9)$$

其中, $\lambda_{ji}(x)$ 通过三维均值坐标得到。

在三维均值坐标下, 本文方法考虑与 L 帧相邻的多个帧的边界条件, 这样就保持了视频处理中的时空一致性。

3.2 混合式融合

笔者注意到类似于其他梯度域融合方法, 当前景区域和背景区域的纹理比较复杂时, 由式(9)直接提供的融合结果会产生一些局部模糊的效果。为消除这些模糊的部分, 可以在融合之前先进行 matting 处理。但融合区域边界形状的改变会增加网格生成的时间消耗, 更严重的是, 如果融合的边界是极凹的, 可能会产生错误的融合结果。

本文提出一种混合式的融合方法来减少模糊。修改式(8)来限制局部融合的程度, 公式表示为:

$$L = \tau \times \hat{L} + \psi, \tau = \frac{1-\alpha}{k_1 \times e^{\frac{1-\alpha}{k_2}}} \quad (10)$$

其中, $\tau \in [0, 1]$; α 是图像的一个透明值; $k_1 \in (0, 1]$; $k_2 < 1$ 。 k_1 用来约束融合的程度。

4 实验结果

比较二维均值坐标融合方法和本文方法的融合生成结果, 选取生成视频中的一帧, 如图 3 所示。

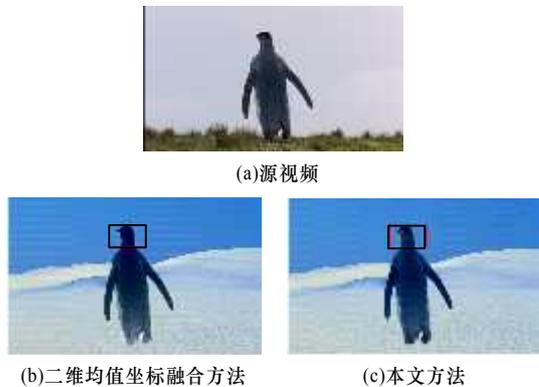


图3 融合结果比较

图 3(b)和图 3(c)中的背景由两部分构成, 上部的蓝天部分和下部的雪地部分。使用二维均值坐标克隆的方法生成的结果(图 3(b))存在明显的局部色彩失调现象, 特别是在企鹅的头部区域, 一些原本是黄色的皮毛变得过于偏蓝。另外, 企鹅脚的部分也过于偏白, 整只企鹅的色调显得不太协调。图 3(c)是通过本文方法生成的结果, 相比之下, 整体的色调保持的比较一致, 同时也不存在局部颜色明显失调的现象。

5 结束语

本文提出一种新的基于优化的三维均值坐标的混合式的视频融合方法, 该方法在实现无缝均值克隆的同时, 改善了先前方法中常出现的模糊失真。下一步研究工作是考虑将更高效的 alpha matting 技术(如共享式软抠取)结合到视频融合中, 以得到更高效更稳定的结果。(下转第 290 页)