

基于仿射变换的图像序列拼接方法

王建华

(哈尔滨师范大学计算机科学系 哈尔滨150080)

Affine Transformation Based Image Mosaics

WANG Jian-Hua

(Department of Computer Science, Harbin Normal University, Haerbin 150080)

Abstract When Image Mosaics is produced, we optimally solve the registration transformation for adjacent frames with traditional ways. The way is slow, heavy and sometimes gets stuck in local minima. The paper provides the way that solves the transformation with a affine transformation model. The way automatically produces the answer for the frames with larger bias. It greatly speedups the process of image mosaics. It plays important role in quick and real-time making image mosaics.

Keywords Affine transformation model, Virtual reality, Image mosaics

1 引言

基于几何的虚拟现实技术有许多局限和不足。尽管在真实感图像生成方面作了大量工作,发展了多种造型技术来模拟自然界物体的形状。但遗憾的是,与拍摄到的图像相比,无论采用的模型多么复杂,都只能在一定程度上减少与真实图像之间的差别。其次建模繁琐,绘制速度与景物的复杂度有关,需专门昂贵的绘制硬件。为此人们提出了一种直接从已存在的场景或物体的图像(照片/录像等)中构造其虚拟体的方法,这样得到的效果更接近于真实的观察结果,我们把这种自然真实感称为“照片真实感”。而基于图像的虚拟现实就是从真实世界的图像中构造其虚拟模型,使得操作者能够看到具有照片真实效果的视图。从一个视频序列创建无缝全景图像是近来的研究趋势,它的主要研究领域是利用廉价PC摄像机和适当约束的摄像机运动来自动建立大型、高分辨的图像。它具有广泛的应用领域:视频浏览,视频监视及虚拟现实等。它的快速实现具有更重要的意义,可以使我们在浏览全景图像,也可对宽广景物的采样起指导作用。但现有基于视频的图像拼接技术^[1-3]都是要对视频序列的每一对相邻帧进行处理,即要优化计算它们的平面投影变换。该算法先采用传统的有8个未知参数的矩阵来描述两幅图像之间的变换关系,并用最小化算法估计该矩阵;但该方法有一定的局限性,即优化参数较多,易导致陷入局部极小问题。其次由于视频数量大,且计算这些变换的运算量亦很大,因此导致整个全景图像的生成速度缓慢。我们采用六参数的仿射变换来减少优化参数,以避免上述问题。考虑到视频序列冗余性高,我们分步优化求解这些变换。它可对重叠部分较小的帧进行自动处理,对此我们对像机的运动做适当约束即可满足上述自动求解的条件。因此使用该模型不必逐帧处理,而是距离较远且偏差数百个像素的帧进行对齐融合,从而使工作量大大减少。本文算法首先推导出仿射变换,然后给出优化迭代公式和算法。

2 仿射变换模型

对于平面景物或绕视点旋转所采集的图像,它们之间存在如下的整合变换 $H_{3 \times 3}$,设图像 I_1 和 I_2 它们的对应点的齐次坐标 $\bar{x}_1 = (x_1, y_1, 1)^T$, $\bar{x}_2 = (x_2, y_2, 1)^T$,则

$$\bar{x}_2 = R \cdot \bar{x}_1 \quad \text{其中 } R = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & 1 \end{bmatrix} \quad (1)$$

R矩阵的求解方式为二类:一类是指定对应点,该方法速度慢,需人工干预。另一类是基于SSD强度(Sum of Square Difference Intrlsitg)的匹配测度来优化求解。当处理视频序列的相邻帧时,可自动求解R。但要求二帧差距小,重叠大。在制作平旋转拍摄景物的图像拼接时,摄像机一般做旋转角度小运动,且PC摄像机的视野普遍较小。如果我们把上述条件变成一般的约束限制,我们使用如下的仿射变换模型即公式(2)来分步优化求解变换 H_2 ,可自动处理相距较远,且重叠较小的帧。

$$\bar{x}_2 = H \cdot \bar{x}_1 \quad \text{其中 } H = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

仿射变换模型:在制作旋转拍摄景物的图像拼接时,如果相机受如下限制:①旋转角度小;②且相机基本垂直于景物时,则优化生成的过程可分解如下步骤:①估计H的平移参数 h_2 和 h_5 ;②由 h_2 和 h_5 做为初始值,用基于SSD强度的匹配测度方法可自动优化生成 H_2 的仿射参数 (h_0, h_1, \dots, h_5) 。

证明:

令 $Q = (X, Y, Z)^T$, $Q' = (x', y', z')^T$ 为景物中一点但对二个不同的相机坐标系。

令 $q = (x, y, 1)^T$, $q' = (x', y', 1)^T$ 表示两图像的对应点。

$q \cong MQ$, $q' \cong M'Q'$, 其中 \cong 表示仅差一个常量因子, M , M' 为投影矩阵。

令 Π 是拥有单位法向量 n 的平面景物, d 是平面距相机中心的距离, R, t 为 Q' 坐标系相对 Q 的平移和旋转矩阵,则 q 与 q' 有如下关系^[5]:

$$q' \cong Hq$$

其中 $H = V'(R + t \cdot n^T/d)V^{-1} \dots (2)$; V 为相机的内部参数矩阵。

为了讨论方便,我们假设 $V' = V = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 即图像坐

标轴是正交的,纵横比为1,且图像坐标系的原点为图像中央,即摄像机的光轴穿越图像的点。

设 $R = \begin{bmatrix} 1 & -\Omega_x & \Omega_y \\ \Omega_x & 1 & -\Omega_x \\ -\Omega_y & \Omega_x & 1 \end{bmatrix}$, $(\Omega_x, \Omega_y, \Omega_z)$ 为绕各轴的旋轴弧度, $n = (\alpha, \beta, \gamma)^T$, 则 $H = [[H]_1, [H]_2, [H]_3]^T$;

$$[H]_1 = [1 + \frac{tx \cdot \alpha}{d}, \frac{tx \cdot \beta}{d} - \Omega_z, \frac{f \cdot Dx \cdot \gamma}{d} + f \cdot \Omega_y]$$

$$[H]_2 = [\frac{ty \cdot \alpha}{d} + \Omega_z, 1 + \frac{ty \cdot \beta}{d}, \frac{f \cdot ty \cdot \gamma}{d} - f \cdot \Omega_x]$$

$$[H]_3 = [\frac{1}{f}(\frac{tz \cdot \alpha}{d} - \Omega_y), \frac{1}{f}(\frac{tz \cdot \beta}{d} - \Omega_x), 1 + \frac{tz \cdot \gamma}{d}]$$

由(2)式得

$$x' = \frac{[H]_{1q}}{[H]_{3q}}; y' = \frac{[H]_{2q}}{[H]_{3q}} \quad \text{令 } D = [H]_3 \cdot q = \frac{x}{f}(\frac{tz \cdot \alpha}{d} - \Omega_y) + \frac{y}{f}(\frac{tz \cdot \beta}{d} + \Omega_x) + 1 + \frac{tz \cdot \gamma}{d}$$

由于 PC 摄像机的视野小, 因此 f 相对于像素单位值较大, 且前后运动小, 则 $D \approx 1$,

又 $\because \Omega_x, \Omega_y, \Omega_z$ 的值很小, 且相机基本垂直于平面景物 $\therefore \alpha, \beta$ 很小, 而 γ 近似为 1,

$\therefore [-1]_{1q} [H]_{2q}$ 的后二项, $h_2 \approx \frac{fDx}{d} \quad h_3 \approx \frac{fDy}{d}$ 正好是二图像的平移视差,

\therefore 用二图像的平移量作为 h_2 和 h_3 的初始值估计是充分的。

同理由于 PC 摄像机的视野小, 且相机的平移相对于 d 是很小的。

$\therefore [H]_{1q}$ 的第 1 项和第 2 项 $x(1 + \frac{tx \cdot \alpha}{a}), (\frac{tx \cdot \beta}{a} - \Omega_z)$ 近似为 1, 0。

同理 $[H]_{2q}$ 第 1 项和第 2 项 $\Omega_z + \frac{ty \cdot \alpha}{a}, 1 + \frac{ty \cdot \beta}{a}$ 可近似为 0, 1。

\therefore 我们有以下的仿射变换:

$$x' = xh_0 + yh_1 + h_2$$

$$y' = xh_3 + yh_4 + h_5$$

$$\text{其中 } [H]_1 = [h_0, h_1, h_2] = [1 + \frac{tx \cdot \alpha}{d}, \frac{tx \cdot \beta}{d} - \Omega_z, \frac{f \cdot Dx \cdot \gamma}{d} + f \cdot \Omega_y]$$

$$[H]_2 = [h_3, h_4, h_5] = [\frac{ty \cdot \alpha}{d} + \Omega_z, 1 + \frac{ty \cdot \beta}{d}, \frac{f \cdot ty \cdot \gamma}{d} - f \cdot \Omega_x]$$

3 图像整合

在制作平面景物的全景图像时, 手持 PC 摄像机做逐行扫描景物, 基本上符合仿射变换模型的约束要求。因此我们只需对距离较远且重叠较小部分的帧做拼接对齐运算, 可使制作全景图像的速度大大加快。我们可以从视频序列推算出二重叠帧的距离, 使分步计算平面投影的过程自动完成, 具体算法如下:

(1) 由视频序列推算出二帧的平移量, 我们采用重叠面积至少为二分之一图像大小左右, 以确保拼接质量。具体操作为取若干个小窗口来推算出相邻两帧的平移距离, 然后累计出偏差较大帧的平移量。

(2) 用相关法来计算二帧的平移量 h_2 和 h_3 , 仅使用重叠

部分图像中央附近面积的三分之二。

(3) 根据上述结果作为仿射参数 (h_0, h_1, \dots, h_5) 的初始估计, 用阻尼最小二乘法来计算 H 的仿射参数, 使用整幅图像面积来计算。

(4) 根据上述结果, 对图像进行拼接融合制作图像 Mosaics。

相关性方法是图像匹配的常用方法。它的特点是运算稳定收敛, 但计算量大。为此我们可以采用多分辨层次模型方法, 在低分辨率图像进行运算, 然后用在较高分辨率图像上的初始值, 这样可减少相关匹配的运算量。

阻尼最小二乘法求解 H 系数的特点是迭代次数少精度高, H 在计算的开始时为单位矩阵, 然后逐步加入它的平移量, 仿射参数作为最终求解 H 的初始参数, 它的公式及求解过程如下:

设两图像 $I_1(x, y'), I_0(x, y)$, 则

$$x' = xh_0 + yh_1 + h_2; y' = xh_3 + yh_4 + h_5 \quad (3)$$

$$E\{H\} = \min_H \sum_i (I_1(x'_i, y'_i) - I_0(x_i, y_i))^2 = \min_H \sum_i e_i^2 \quad (4)$$

其中 $e_i = I_1(x'_i, y'_i) - I_0(x_i, y_i)$, (x'_i, y'_i) 为 (x_i, y_i) 的对应点,

$$\text{则 } \begin{cases} \Delta H_R = (A + \lambda I)^{-1} b \\ H_{k+1} = H_k + \Delta H_k \end{cases} \quad (5)$$

其中 A, b 为近似 Hessian 矩阵和加权梯度向量, $H = (h_0, h_1, \dots, h_5)$

$$A \text{ 中的元素 } a_{uv} = \sum_i \frac{ae_i}{ah_u} \cdot \frac{ae_i}{ah_v} \quad b_k = - \sum_i e_i \cdot \frac{ae_i}{ah_k}$$

$$\frac{ae_i}{ah_k} = \frac{aI'}{ax'} \cdot \frac{ae_i}{ah_k} + \frac{az' ay'}{ay' ah_k}$$

完成该算法的主要步骤是将 H_{k+1} 代入 (4) 使其逐渐减少到小于某一个阈值, 或迭代次数达到某固定数, 否则增加重新计算 (5) 式。迭代完成上述过程, 详细请看文 [2]。

当 H 求解后, 我们将二幅图像拼接成较大图像。为了减少图像重叠之间的光强不连续性, 我们采用一种简单的帽函数 [2], $W(x, y) = W(x) * W(y)$ 其中 $W(x)$ 为三角状函数, 该函数在图像边缘处为 0, 而中间值为最大值, 在融合时将权加到重叠图像的像素强度上:

$$C(x, y) = \sum_K w_K(x, y) \cdot I_k(x, y) / \sum_K w_K(x, y)$$

其中 k 为帧号, C(x, y) 为融合图像的像素强度值, 使用该方法图像拼接融合效果好。

4 实验结果

我们从一个平面景物拍摄的视频序列中抽取了三个帧, 它们偏差在 100 多个像素和 80 多个像素。平面景物距离相机 25 厘米左右。PC 摄像机为天敏 2000PC-视频卡。图 1 表示了这三个帧, 图 2 是我们运用仿射变换模型产生的图像 Mosaics。图像拼接的质量是较高的。

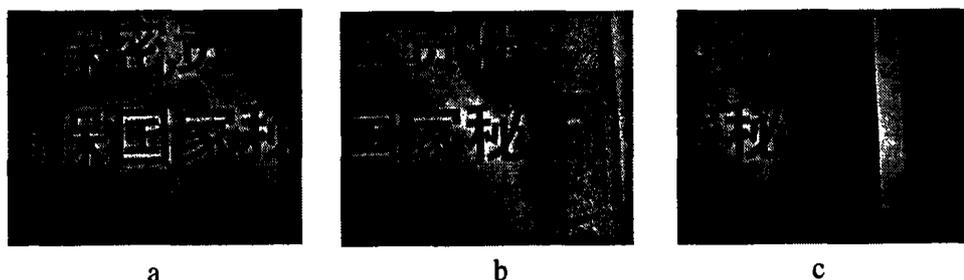


图 1 帧序列中的三帧

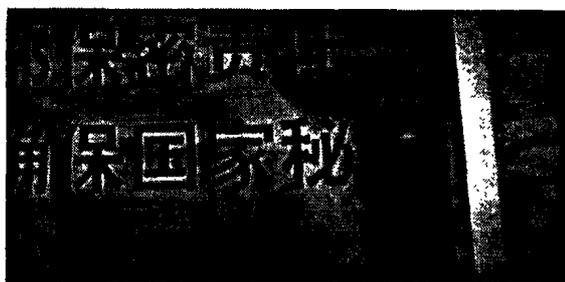


图2 全景图像

结束语 本文提出的仿射变换模型,能够在图像偏差较大的情况下实现图像的自动整合。如果只采用层次模型直接对齐图像,则两图像的偏差仅能在几个像素范围内。在实际实现中,工作量太大,算法无实际意义。该方法速度快,工作量

小,优化参数少,可避免陷入局部极小的问题。在实际实现过程中,一般图像偏差控制在几十个像素内,即可快速高质量地建立全景图像。

参考文献

- 1 Zelinik-Manor. 2, Irani. M. Multi-Trame Estimate of Planar Motion. IEEE Transaction on PAMI, 2000, 22(10)
- 2 Szeliski R. Video Mosaics for Virtual Environment. IEEE CGQA, March. 1996
- 3 Sawhney S H. True Multi-Trame Alignments and its Application to Mosaicing and Lens Distortion Connection. IEEE Transactcion on PAMI, 1999, 21(3)
- 4 Shum H, Szeliski R. System and Experiments Paper : Construction of Panoramic Image Mosaics with Global and Local Alignment. International Journal of Computer Vision, 2000, 36 (2): 101~130
- 5 马颂德. 计算机视觉. 科学出版社, 1997
- 6 Anadan K R P. Representation of scene from collection of images. IEEE CVPR 1995

(上接第35页)

为真,初始状态有 H 提出问题,终止的时候问题得到了解决, H 和 M 都相信 p 是假的。

首先 H 询问 M 是否 p 是真的, M 接到问题后没有直接的 belief 给出答案。所以只有先去询问自身的 p 祖先命题集,但从祖先也无法得出答案,只好转向兄弟命题集去求解问题,由于 q 是 p 的兄弟命题集的一个元素, q 是成立的,根据封闭世界假设,得出 p 不成立,此时 H 提出的问题得到了解决,成功退出。

为了让这个例子更加具体实用,来看下面的实例:

机器具有知识:奶牛会产奶。

奶牛和黄牛都是牛的儿子结点。

H 询问 M: 黄牛会产奶吗?

经过上面的推理过程 M 能够得出答案: 黄牛不会产奶。满足了 H 的愿望,双方的愿望达到了新的平衡,问题得到了解决。

习的能力,虽然所能处理的问题还比较简单,但有理由相信,规则和知识库的逐渐丰富将会带来可喜的效果。

语境在话语理解中是非常重要的参考因素,并且它是个随着时间、空间、周围环境的变化以及交流的进行不断变化着的量,这正是我们话语分析中遇到的最大困难。难以做到对语境的精确把握,也就难以做到对话的连贯和自然。纵观科学发展的历史,但凡一门技术和科学,如果没有一个可靠的理论作保证,就像空中楼阁虚幻而不会长久。尽管我们见到了不少对话程序成功地通过了各类测试,但是这并不能说明我们找到了机器理解自然语言的有效办法。话语理解缺乏的不但是对各类信息的形式化描述,更缺乏扎实可靠有效的理论基础。我们期待不久的将来能够努力建立一套更加完整、实用且可计算的话语解释理论。

参考文献

- 1 Lu Ruqian, Zhang Songmao. PANGU - An agent-oriented knowledge base. In: Processing of Conference on Intelligent Information Processing (16th wcc2000), 486~493
- 2 Beun R-J, On the generation of coherent dialogue. Pragmatics & Cognition, 2001, 9: 1, 37~68. ISSN 0929-0907
- 3 Saygin A P, Cicekli I. Pragmatics in human-computer conversations. Journal of Pragmatics, 2002, 34: 227~258
- 4 周月娇. 基于常识的二义性排除; [硕士论文]. 中科院数学所, 1998
- 5 Brown G, Yule G. Discourse Analyse. Cambridge University Press, 1983
- 6 Peccei J S. Pragmatics. Routledge Press, 1999
- 7 Wang Hui, Zhan Weidong, Yu Shiwen. The Development of Semantic Knowledge-base of Contemporary Chinese and its Applications. The Second China-Japan Natural Language Processing Joint Research Promotion Conf. CJNIPL, 2002. 211~225
- 8 杨帆. 图灵测试中对话处理研究; [硕士论文]. 中科院数学所, 2001
- 9 何自然, 冉永平. 语用和认知-关联理论研究. 外语教学和研究出版社, 2001
- 10 Zheng Hong, Lu Ruqian, et al. Ontology-Based Semantic Cache in AOKB. Journal of Computer Science & Technology, 2002, 17(5): 657~664
- 11 Bezuidenhout A, Cutting J C. Literal meaning, minimal propositions, and pragmatic processing. Journal of Pragmatics, 2002, 34: 433~456
- 12 Krahmer E, Piwek P. Presupposition projection as proof construction. In: H. Bunt, R. Muskens, eds. Computing Meaning, 1999, 1: 281~300
- 13 Halpern J, Moses Y. Knowledge and common knowledge in a distributed environment. Journal of the ACM, 1990, 37: 549~587
- 14 Bunt H. Context Representation for Dialogue Management. In: P. Bouquet, L. Serafini, eds. Modeling and Using Context. Lecture Notes in Artificial Intelligence Vol. 1668. Berlin, Springer
- 15 Lu Ruqian, et al. Agent oriented commonsense knowledge base. Acta Scientia, 2000, 43(6): 641~652

表1

	M			ACTIONS	H	
	B _M	B _M I _H	B _M →B _H		B _H I _M	B _M →B _H
初始 S ₀	∅	∅	∅	Question [H, p]	∅	∅
S ₅		(p)	∅	Question [M, Fathers(p)]	∅	∅
S ₈		(p)	∅	Ignorance [M, Fathers(p)]	Fathers(p)	∅
S ₆		(p)	∅	Question [M, Brothers(p)]	∅	Fathers(p)
S ₂	→p	(p)	∅	State [M, →p][0][0]	∅	Fathers(p)
S ₁₀	→p	∅	∅	Success []	∅	Fathers(p)

结束语 通过分析当前的认知语境,而获得对话语的适当解释,很多人做过这种方法的尝试,但都没有收到很好的效果。本文设计的 CBT 提供了一套话语理解过程中认知语境信息的形式化生成和更新规则,并且具备了常识检查和机器学