

# 基于 Kanade-Lucas-Tomasi 算法的人脸特征点跟踪方法

段 鸿 程义民 王以孝 蔡尚书

(中国科学技术大学电子科学与技术系 合肥 230026)

**摘 要** 与传统的在人面部画上标识点的特征点跟踪方法不同, KL T (Kanade-Lucas-Tomasi) 算法可以从未加标识点的正面人像视频系列中通过特征纹理信息直接获取面部某些特征点的位移. 在 KL T 算法中加入了基于人脸统计信息的经验约束, 使 KL T 算法更加合理有效.

**关键词** 虚拟人; 人脸特征点; 视频跟踪; 视频分析  
**中图法分类号** TP391.9

## Tracking Facial Feature Points Using Kanade-Lucas-Tomasi Approach

Duan Hong Cheng Yimin Wang Yixiao Cai Shangshu

(Department of Electronic Science & Technology, University of Science & Technology of China, Hefei 230026)

**Abstract** Compared to normal methods that need to make marks on person's face, this method can track displacements of facial feature points from frontal face video through texture information of the face. At the same time, in order to improve the method, we give the static angle and distance limitation of facial organs to the KL T (Kanade-Lucas-Tomasi) algorithm.

**Key words** virtual person; facial feature points; video tracking; video analysis

## 1 引 言

动态人脸仿真技术在虚拟人物设计、电影/动画制作、远程教学等诸多方面都有着广泛的应用,也是计算机图形学领域一个具有挑战性的课题. 这其中的人脸特征跟踪技术对于动画人脸重建、人类感情信息分析、特定人脸的识别等都具有重要的意义. 如在人脸肌肉模型的单视频驱动<sup>[1]</sup>方法中,通过跟踪视频中人脸特征点的二维位移,来获取面部肌肉向量的收缩量,最终达到驱动三维人脸模型的目的.

Kouadio 等提出了一种通过加标识点来跟踪视频中人脸特征点的方法<sup>[2]</sup>,文献[3]则采用了一种分析人面部曲线的方法来跟踪人脸视频. 这些特征

跟踪方法的局限是要在被摄像者面部加上标识点或者曲线. 如果能够从未加标识点的正面人像视频系列中直接获取面部某些特征点的位移,将会更有意义. 本文给出了一种在首帧中确定搜索特征点,采用改进的 KL T (Kanade-Lucas-Tomasi) 算法对未加标识点的人脸正面视频图像进行特征点跟踪,进而求得人脸特征点运动参数的方法.

文献[4]也对人脸特征约束进行过研究并提出过相应的改进跟踪算法. 本文对其 FC-KL T 算法的改进主要在于:文献[4]将所有特征点均采用单一基点约束,本文改成对面部各个器官进行基点约束,而器官内部采用近邻约束,化解了基点出现跟踪误差时距离基点较远的点漂移的风险;同时不采用对特征点直接跟踪,而是选择以特征点为中心的矩形窗

原稿收到日期:2003-01-16;修改稿收到日期:2003-05-13. 段 鸿,男,1976 年生,博士研究生,主要研究方向为计算机视觉、图像处理、多媒体技术. 程义民,男,1945 年生,教授,博士生导师,主要研究方向为计算机视觉、图像处理、深度图像分析. 王以孝,男,1946 年生,副教授,主要研究方向为计算机视觉、图像处理、视频多媒体技术. 蔡尚书,男,1980 年生,硕士研究生,主要研究方向为计算机视觉、图像处理、三维物体建模.

口中最适合跟踪的点来进行 KL T 跟踪,提高了跟踪的效率。

本文方法已经在一台 PIII 800 PC 机上进行了模拟,并获得了令人满意的跟踪效果。

## 2 KL T 算法计算特征点位移

KL T 算法<sup>[5-6]</sup>是一种以待跟踪窗口  $W$  在视频图像帧间的灰度差平方和 (Sum of Squared intensity Differences, SSD) 作为度量的跟踪算法。

对于灰度图像, KL T 算法假设一个包含了特征纹理信息的特征窗口  $W$ , 在  $t$  时刻图像帧  $I(x, y, t)$  和在  $t +$  时刻图像帧  $I(x, y, t + )$  中的位置满足

$$I(x, y, t + ) = I(x - x, y - y, t) \quad (1)$$

即在  $I(x, y, t + )$  中的每个像素点,都可以由  $I(x, y, t)$  中相应窗口的像素点平移  $d = (x, y)$  得到。KL T 算法的目的就是求出  $d$ 。

假设  $t +$  时刻的特征窗口为  $B(X) = B(x, y, t + )$ , 其中  $X = (x, y)$  为视口坐标。  $t$  时刻的特征窗口为  $A(X - d) = A(x - x, y - y, t)$ 。

考虑一般情况,有

$$B(X) = A(X - d) + n(X) \quad (2)$$

其中  $n(X)$  为在时间 内由于光照条件变化产生的噪声。

将  $n(X)$  平方并在整个窗口上积分,就得到了窗口图像的 SSD

$$\begin{aligned} &= \int_V n(X)^2 (X) dX = \\ &\int_V [A(X - d) - B(X)]^2 (X) dX \quad (3) \end{aligned}$$

其中,  $X = [x, y]^T$ ,  $d = [d_x, d_y]^T$ ,  $(X)$  通常可取为 1。如果强调中心部分纹理的作用,  $(X)$  可以采用高斯分布函数。

当  $d$  为一和  $X$  比起来可忽略的小量时,将  $A(X - d)$  泰勒展开,去除高次项,得到

$$A(X - d) = A(X) - g \cdot d \quad (4)$$

将式(4)代入式(3),并对式(3)的两边同时对  $d$  求导之后取 0,可以得到

$$\begin{aligned} \frac{\partial}{\partial d} &= \int_V [A(X) - B(X) - g^T d] \times \\ &g(X) dX = 0 \quad (5) \end{aligned}$$

此时 取得极小值。式(5)可变换为

$$\left( \int_V g g^T dX \right) d = \int_V [A(X) - B(X)] g dX \quad (6)$$

若令

$$\begin{aligned} Z &= \int_V g(X) g^T(X) dX = \\ &\int_V \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix} (X) dX = \begin{bmatrix} G_{xx} & G_{xy} \\ G_{xy} & G_{yy} \end{bmatrix} \quad (7) \end{aligned}$$

$$e = \int_V [A(X) - B(X)] g(X) (X) dX \quad (8)$$

则式(6)可以表示为

$$Zd = e \quad (9)$$

对于每两帧图像,解方程(9),可以求出特征窗口的位移  $d = (x, y)$ 。

在 KL T 跟踪算法中,并不是所有包含纹理信息的特征窗口都是适合跟踪的,对于某特征窗口,当其矩阵  $Z$  的两个特征值  $\lambda_1$  和  $\lambda_2$  满足条件  $\lambda_2 > \lambda_1 > \lambda_{\max}$  时,此特征窗口具有较好的跟踪效果。阈值  $\lambda_{\max}$  与拍摄条件有关,可以通过实验来确定。

## 3 基于经验约束的 KL T 跟踪算法

### 3.1 KL T 算法用于人脸特征点跟踪

人面部的一些特征部位,如嘴角,鼻孔,眉毛边缘等都可以借助 KL T 跟踪算法进行跟踪,如图 1 所示。



图 1 面部特征点

但是,直接采用 KL T 进行人面部特征点跟踪带来的问题是:跟踪过程可能常会出现跟踪漂移和跟踪丢失的情况。如图 2 中圆圈部分所示,由于漂移,跟踪点由“+”位置跳到了点的位置。跟踪丢失情况如图 3 所示,“+”部分为应该出现的跟踪点,但跟踪丢失了。跟踪丢失是由于人面部的运动改变了特征窗口内的纹理而造成的。



图 2 跟踪漂移



图 3 跟踪丢失

3.2 加入经验约束的 KL T 算法

KL T 算法在跟踪特征窗口时,仅仅使用了窗口本身的纹理信息  $I(x, y)$  作为跟踪特征. 而人在辨识人像视频时,除了注意特征部位外,还充分地利用

了人脸各特征部位之间的统计信息如表 1 所示<sup>[7]</sup>,用于协助判断. 正是这种基于统计信息的约束,使得人眼在跟踪视频中的人像时,很少发生跟踪漂移和丢失的情况.

表 1 面部特征构造关系

与双眼间距离的比值	眼睛宽度	眼睛到鼻间	眼睛到嘴唇	眼睛到下巴尖	鼻子宽度	嘴宽度	嘴高度	眼睛到眉毛
平均值	0.42	0.51	0.88	1.59	0.57	0.70	0.27	0.35
变动范围	0.03	0.05	0.04	0.09	0.05	0.06	0.02	0.05

3.2.1 基准点的选择

可以把这种统计信息看成是两种约束的合成:一种是以某个点(如鼻孔连线的中点)为主基准点,面部的各个器官相对于这个点的角度和距离约束;另一种是各个器官内部的形状约束,同样可以用和基准点之间的距离角度约束给出. 日常生活中所说的“樱桃嘴”“丹凤眼”即是形状约束.

主基准点的选择. 鼻孔是人共有的特征器官,在各种表情动作中,鼻孔的纹理变化是最小的. 在正常光照条件下,采用 KL T 跟踪方法可以保证鼻孔有较高的跟踪正确率. 综合两个鼻孔的情况,鼻孔连线的中点就是最佳的主基准点.

各个器官中次基准点的选择. 为了保证跟踪的可靠性,一般均选用器官中纹理特性比较明显的部分,如左右嘴角点、眉毛靠近鼻孔端点、眼睛靠近鼻孔的眼角点等.

鉴于以上分析,以鼻孔连线的中点为原点,从左鼻孔到右鼻孔的连线作为  $x$  轴建立坐标系,这样设立的坐标系可以防止由于面部的小角度旋转造成跟踪偏移. 在后面的实验中,均在此坐标系中讨论问题.

3.2.2 特征窗口的选择

在第 1 节中曾提到,并不是所有包含纹理信息的特征窗口  $W$  都是适合跟踪的. 因此对待跟踪的以特征点为中心的特征窗口  $W$  进行如下选择:

Step1. 在第  $k$  帧中,以面部特征点为中心,定出一定大小(实验中取 20 个像素大小)的矩形区域.

Step2. 在此矩形区域中,依次计算出以每个像素点为中心,固定大小的特征窗口的特征值  $1, 2$ .

$$1 = (G_{xx} + G_{yy} + \sqrt{G_{xx}^2 - 2G_{xx}G_{yy} + G_{yy}^2 + 4G_{xy}^2})/2$$
(10)

$$2 = (G_{xx} + G_{yy} - \sqrt{G_{xx}^2 - 2G_{xx}G_{yy} + G_{yy}^2 + 4G_{xy}^2})/2$$
(11)

其中,  $G_{xx}, G_{yy}, G_{xy}$  同式(7).

Step3. 按  $\min(1, 2)$  进行排序,选取此值最大的像素点作为跟踪特征窗口的中心点.

Step4. 记录面部特征点和此像素点之间的相对位置,作为  $k+1$  帧找回面部特征点的依据.

Step5. 根据 KL T 算法,找出  $k+1$  帧中特征窗口的位置. 根据 Step4 中给出的相对位置,得到  $k+1$  帧中此面部特征点的位置.

3.2.3 改进的 KL T 跟踪算法

对待跟踪的特征点  $K_i$ ,加入基于经验的约束,将用于跟踪的特征空间扩充为  $\{I(x, y), R_{ji}, \theta_{ji}\}$ . 其中,  $R_{ji}$  为基点对该点的距离约束,  $\theta_{ji}$  为基点对它的角度约束. 如图 4 所示.

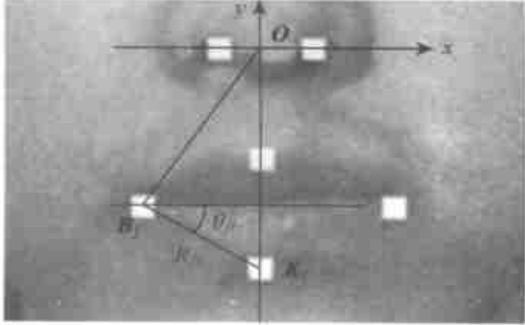


图 4 距离和角度约束

在式(8)的推导过程中强调了  $d$  必须为一与图像大小比起来可忽略的小量. KL T 算法的跟踪丢失和漂移在一定程度上是由于  $d$  太大造成的. 因此必须借助于经验加入对特征点帧间位移的约束.

设  $t$  时刻特征点  $K_i$  的坐标为  $K_i = (x_i, y_i)$ , 一个跟踪较稳定的特征点  $B_j$  坐标为  $B_j = (x_j, y_j)$ . 此时可以求出以  $B_j$  作为基点时对  $K_i$  点的约束为

$$R_{ji} = |B_j K_i|$$
(12)

$$\theta_{ji} = -\arccos\left(\frac{|OB_j|^2 + |B_j K_i|^2 - |OK_i|^2}{2|OB_j||B_j K_i|}\right) + \arccos\left(\frac{x_j}{|OB_j|}\right)$$
(13)

其中,  $\theta_{ji}$  为  $x$  轴和向量  $B_j K_i$  之间的夹角.

这一约束条件可以应用于  $t+$  时刻的跟踪中. 对于任何一个特征点  $K_i$ ,加入这一约束后,其跟踪过程变成了一个有约束的迭代过程.

设  $t +$  时刻  $B_j$  点坐标为  $B_j = (x_j, y_j)$ , 则有

- Step1. 在  $t$  时刻, 求出  $B_j$  对  $K_i$  的距离和角度约束  $R_{ji}$ ,  $j$ ;
- Step2. 求出  $t +$  时刻  $B_j$  点坐标为  $B_j = (x_j, y_j)$ ;
- Step3. 使用在  $t$  和  $t +$  时刻图像, 分别求  $K_i$  点所在特征窗口的  $Z, e$ ;
- Step4. 解  $Zd = e$  求出  $d = (x_i, y_i)$ ,  $K_i = (x_i + x_i, y_i + y_i)$ ;
- Step5. 根据  $t$  时刻的距离和角度约束  $R_{ji}$  与  $j$  计算出  $K_i$  的约束位置  $K_i$ ;
- Step6. if ( $|K_i K_i| > l_i$ )
- {
- 将  $d_i$  修改为  $d_i/n$ , 将修正后的  $d_i$  值赋给特征窗口  $w$  的新位置.
- 转 Step2;
- }
- Step7. 将  $t +$  时刻图像赋给  $t$  时刻图像. 进入下一帧, 转 Step1.

其中步长阈值  $l_i$  由实验确定, 通过这个值进行经验约束. 值过大时约束不明显, 容易造成跟踪漂移; 值过小时约束太强, 则容易造成跟踪丢失.  $1/n$  是一个缩放系数, 用于缩小帧间特征点位移.

对点  $K_i$  进行经验约束的原则是: 作为约束基准点的点必须具有较好的跟踪稳定性, 同时又要与  $K_i$  点相关. 实验中, 采用了两个鼻孔中点对左右嘴角点、靠近鼻孔的嘴角点进行约束; 然后再采用左右嘴角点对上下唇中点进行约束, 采用眼角点对眼睛的其他特征点进行约束.

## 4 实验及结论

### 4.1 实验方法

我们分别采用 KL T 算法, FC- KL T 算法和本文的加入了经验约束的 KL T 算法, 对同一视频系列进行了跟踪实验比较.

其中加入了经验约束的本文 KL T 算法流程如下:

- Setp1. 在首帧中采用文献[7]的方法, 确定需要跟踪的 12 个人脸特征点, 同时记录统计约束  $\{R_{ji}, j\}, 1 \leq i, j \leq 12$ ,  $i \neq j$ .
- Setp2. 对于稳定性最好的特征点 (如鼻孔), 采用第 3.2.2 节的方法选择特征窗口.
- Setp3. 按照第 3.2.2 节的流程求出此特征点的位移.
- Setp4. 以此特征点为约束基准, 采用基于经验约束的本文 KL T 算法, 求出与之相关的所有特征点的位移, 对部分跟踪效果较差的特征点可以采用本文方法.
- Setp5. 重复过程 Step2 ~ Step4, 直到求出所有特征点的位移.
- Setp6. 根据本帧跟踪结果, 计算新的统计约束量  $\{R_{ji}, j\}, 1 \leq i, j \leq 12, i \neq j$ . 进入下一帧.

### 4.2 实验结果及结论

我们在实验室的普通光照条件下, 采用分辨率为  $240 \times 320$  的 logitech 摄像机摄取序列图像, 采用 KL T 算法和基于经验约束的 KL T 算法分别进行跟踪处理, 结果如表 2 所示.

表 2 几种方法跟踪视频系列 1 (150 帧) 结果统计

实验方法	特征点数	漂移数	丢失数	漂移首个特征点的帧位置	丢失首个特征点的帧位置
KL T 方法	12	5	2	38	43
FC- KL T 方法	12	2	0	41	无
约束 KL T	12	0	0	无	无

图 5 所示为采用约束 KL T 跟踪算法对实验室条件下的 150 帧视频的跟踪结果.

图 6 所示为采用约束 KL T 跟踪算法对一个 40 帧的表情库系列图的跟踪结果.

对视频系列 2 的跟踪结果和理想位置之间的像素误差情况如图 7 所示.

可见, 使用单纯的 KL T 算法进行人脸特征跟踪容易产生漂移和丢失, 而使用加入了约束的 KL T 算法没有产生丢失和漂移, 跟踪的平均误差仅为  $2 \sim 3$



图 5 约束 KL T 方法对视频系列 1 (150 帧) 的跟踪结果

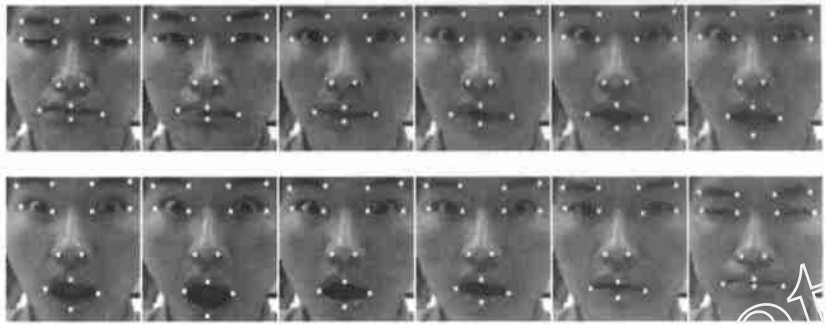


图 6 约束 KL-T 方法对视频系列 2(40 帧)的跟踪结果

像素,可以适应一定范围的光照变化,具有较好的结果. 在人脸小角度转动(左右  $\pm 5^\circ$ ,上下  $\pm 5^\circ$ )的范围,分辨率变化不大的实验条件下可以达到 95 %的跟踪正确率.

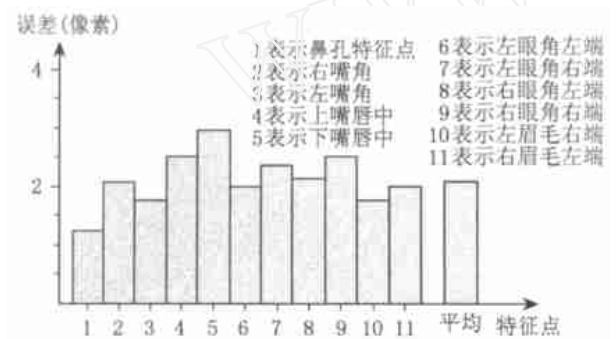


图 7 实际跟踪结果和理想位置之间的误差统计

该跟踪方法获取的特征点位移数据可用于制作虚拟主持人、网络课堂中虚拟教师的表情口型动画,以及虚拟人的表情库等.

参 考 文 献

[1] Duan Hong , Wang Yixiao , Cheng Yiming , *et al.* Drive muscle face model using single video [J]. Journal of System Simulation , 2003 ,15(9) : 1268 ~ 1271(in Chinese)  
(段 鸿,王以孝,程义民,等. 人脸肌肉模型得单视频驱动[J]. 系统仿真学报, 2003, 15(9) : 1268 ~ 1271)

[2] Kouadio C, *et al.* Real-time facial animation based upon a bank of 3D facial expressions [A]. In: Proceedings of IEEE Conference Computer Animation98, Philadelphia, 1998. 128 ~ 136

[3] Demetri Terzopoulos, Keith Waters. Analysis and synthesis of facial image sequences using physical and anatomical models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(6) : 569 ~ 579

[4] Fabrice Bourel, Chibelushi Claude C, Low Adrian A. Robust facial feature tracking [A]. In: Proceedings of the 11th British Machine Vision Conference, Bristol, UK, 2000, 1 : 232 ~ 241

[5] Shi J, Tomasi C. Good features to track [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 1994. 593 ~ 600

[6] Carlo Tomasi, Takeo Kanade. Detection and tracking of point features [R]. Pittsburgh, Carnegie Mellon University, CMU-CS-91-132, 1991

[7] Yan Chao, Su Guangda. Location and acquisition of characters of human face [J]. Journal of Image and Graphics, 1998, 3(5) : 375 ~ 379 (in Chinese)  
(严 超,苏光大. 人脸特征的定位与提取[J]. 中国图象图形学报, 1998, 3(5) : 375 ~ 379)