融合颜色和增量形状先验的目标轮廓跟踪

周 雪[1] 胡卫明[2]

摘 要 基于主动轮廓的跟踪方法被广泛应用于移动摄像机下运动物体轮廓的跟踪. 针对传统方法容易受噪音、部分遮挡、背景干扰等因素影响的缺点, 提出了一个分层的基于水平集 (Level sets) 的跟踪框架. 该框架将颜色信息和形状先验有效地结合起来. 在框架的第一层, 初始轮廓首先根据颜色信息进化, 通过引入一个反映邻域像素之间关系的惩罚因子来改进传统的速度模型. 然后, 基于 Mahalanobis 距离的判别式被用来决定是否需要引入形状先验, 如果不需要, 则第一层基于颜色进化的结果就作为最终的跟踪结果; 否则, 第一层得到的轮廓需要在第二层中在形状先验的约束下继续进化. 在第二层轮廓进化中, 本文提出了一个权重形状距离因子 (Weighted shape distance term, WSDT), 用来融合全局的形状信息和局部的颜色信息. 形状先验模型建立在主成分分析 (Principal component analysis, PCA) 子空间并通过增量学习算法在线更新. 实验结果证明了方法的有效性和鲁棒性.

关键词 跟踪, 基于水平集, 形状先验, 移动摄像机
中图分类号 TP391

Object Contour Tracking with Fusion of Color and Incremental Shape Priors

ZHOU Xue[1] HU Wei-Ming[2]

Abstract Active contour-based tracking has been widely used for tracking contours of objects with moving cameras. However, traditional methods are sensitive to noise, partial occlusions, background disturbance and some other factors. With respect to this point, we propose a two-layer hierarchical level set-based tracking framework in which color information and shape are fused sequentially. In the first layer, the initial contour is evolved only with the color feature. A penalty term measuring the correlations between neighboring pixels is added to improve the general region-based level set speed model. Then, the Mahalanobis distance-based discriminant criterion is adopted to determine whether the shape model is needed. If the shape model is not needed, the contours obtained in the first layer are considered as the final tracking results; otherwise, the obtained contours are evolved with the shape constraint continuously in the second layer. For the second layer, a weighted shape distance term (WSDT) is introduced to fuse the global shape information and the local color information. Principal component analysis (PCA) subspace of shape samples is trained off-line and updated using an incremental learning algorithm. The experimental results have demonstrated the effectiveness and robustness of our method.

Key words Tracking, level set-based, shape priors, moving cameras

摄像机移动条件下基于主动轮廓的跟踪方法由于能获取跟踪物体的形状信息, 已经成为计算机视觉领域的一个研究热点. 与轮廓的显式参数化[1] 描述方法相比, 隐式的水平集 (Level sets) 方法[2] 具有数值解稳定、能处理拓扑变化等优点而受到研究人员的广泛关注. 基于主动轮廓的方法的思想是根据实际问题建立一个关于轮廓的能量函数、采用变分方法最小化该能量函数、最终得到轮廓的进化方程. 实际问题的最优解往往对应于能量函数的最小值. 能量函数通常由三部分组成: 内部能量、外部能量以及目标物体的先验知识. 内部能量反映的是轮廓自身的约束, 比如轮廓的曲率等; 外部能量通常表现为轮廓与图像数据的耦合, 和轮廓自身没有关系; 而当图像数据被破坏 (比如遭遇噪音、遮挡、边缘模糊等), 就需要引入第三部分, 即目标物体的先验知识.

根据外部能量函数中使用的描述子的不同, 基于主动轮廓的方法可以分为基于边缘的和基于区域的两大类. Snake[1] 是一个典型的基于边缘的主动轮廓模型, 它采用参数化的轮廓表达方式, 考虑了轮廓边缘处图像的梯度. 对比 Snake 模型, Caselles 等[3] 提出了一个更为本质的反映图像几何测度的 Geodesic 模型. 2000 年, Paragios 和 Deriche[4] 改进了 Geodesic 模型, 用 Level set 描述轮廓并采用梯度下降方法建立轮廓的进化方程. 基于边缘的主动轮廓方法存在如下一些缺点: 1) 只考虑了轮廓周围的局部信息, 因此初始化要求靠近目标; 2) 对

收稿日期 2008-07-01 收修改稿日期 2009-01-04
Received July 1, 2008; in revised form January 4, 2009
国家自然科学基金 (60905015, 60672040) 资助
Supported by National Natural Science Foundation of China (60905015, 60672040)
1. 电子科技大学自动化工程学院 成都 610054 2. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190
1. College of Automation, University of Electronic Science and Technology of China, Chengdu 610054 2. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190
DOI: 10.3724/SP.J.1004.2009.01394

噪音比较敏感. 基于上述缺点, 一种有效的解决方案就是采用基于区域的主动轮廓方法, 该方法考虑了图像的全局信息. 常用的区域描述子是一些统计特征, 比如: 均值、方差、纹理或所考虑区域的直方图等. 早在 1985 年, Mumford 和 Shah[5] 就提出了一种基于区域的方法用于图像分割, 对于一幅给定图像 I, 找到一个理想的分段相似 (Piecewise homogeneous) 的图像 U, 使得 U 是 I 的一个近似. 图像 U 中不同区域的交界处即分割物体的边缘处. Chan 和 Vese[6] 改进了 Mumford-Shah 函数[5], 把图像 U 简化成区域的均值图像. 另一种流行的基于区域的方法是基于 Bayesian 理论的主动轮廓方法. 该方法基于分割的思想, 通过建立一个最大化后验概率 (Maximum a posterior, MAP) 问题将图像分成目标区域和背景区域. Zhu 和 Yuille[7] 在一个统计变分的框架下提出了基于 Bayesian 的区域竞争算法用于图像的分割.

无论是基于边缘的还是基于区域的模型, 它们都是图像像素灰度的函数. 当图像存在噪音干扰、目标被遮挡、目标边缘模糊等情况时, 仅靠这些模型很难得到准确的轮廓. 因此需要加入关于目标的颜色[8]、纹理[9]、形状[10−12] 等方面的先验知识, 使得轮廓在这些先验知识的约束下准确进化到目标物体的边缘处. Yilmaz 等[8] 通过对颜色特征和纹理特征分别用核密度估计 (Kernel density estimation) 和 Gabor 小波建模来引导轮廓进化. 相对于颜色和纹理先验, 形状先验在出现背景干扰、目标被遮挡等情况时更为有效. 现有的建立形状先验的方法大都是建立统计形状模型[10−12]. Leventon 等[10] 通过主成分分析 (Principal component analysis, PCA) 把形状训练样本投影到低维子空间, 然后建立降维系数的高斯模型来辅助医学图像的分割. Charpiat 等[12] 则提出了一个非线性、更为本质的形状统计先验模型, 用来解决待分割物体在细节上的变化. Rousson 和 Paragios[11] 建立了一个像素级的形状模型, 每个像素位置处用一个单高斯模型来描述形状的局部变化. 由于上述两种静态形状模型不能在线更新, 无法处理跟踪过程中形状不断变化的情况, 因此需要一种反映连续形状变化的动态形状模型 (Active shape models, ASMs). 最早的 ASMs 是 Cootes 等[13] 于 1995 年提出来的, 一开始也是用于图像分割, 即在训练阶段仅考虑有限的样本, 每一次迭代后的结果被用来在线更新形状模型. Fussenegger 等[14] 把训练样本嵌入到一个正交子空间, 并用一个增量 PCA (Incremental PCA, IPCA) 算法在线更新形状模型. 但该 IPCA 算法一次更新只能处理一个样本. Cremers[15] 将马尔科夫模型和自回归模型融合起来, 提出了一个线性动态形状模型 (Linear dynamical shape model) 用于目标的跟踪, 当前帧目标的形状只与前几帧的形状有关并能由前几帧重构出来. 该模型只是数据的拟合, 没有在线更新的过程.

综合上述方法的优缺点, 本文提出了一个基于 Level sets 的分层的框架, 用来跟踪移动摄像机下的目标轮廓. 该框架有效地将局部颜色信息和全局形状先验相结合, 在噪音、部分遮挡、背景干扰等情况下也能鲁棒跟踪. 本文方法的特点如下:

1) 第一层初始轮廓首先根据颜色信息进化. 采用基于区域的主动轮廓方法, 对目标区域和背景区域分别建立多高斯模型 (Gaussian mixture model, GMM). 与传统方法 (假定区域内各个像素之间独立) 不同的是, 我们考虑了邻域像素之间的关系, 根据马尔科夫随机场 (Markov random field, MRF) 理论引入一个惩罚因子 (Penalty term) 来度量邻域内像素之间的关系.

2) 建立了一个基于 Mahalanobis 距离的决策函数, 用来判断是否需要引入形状先验.

3) 如果需要引入形状先验, 在第二层进化过程中, 我们提出了一个权重形状距离因子 (Weighted shape distance term, WSDT) 来融合在形状子空间的进化结果和在颜色空间的进化结果.

4) 形状模型建立在正交的子空间上, 并采用一个鲁棒的增量学习算法在线更新.

本文内容安排如下: 第 1 节介绍了系统的整体框架; 第 2 节描述了颜色和形状模型的离线训练; 第 3 节给出了目标在线跟踪过程以及形状模型的在线更新; 第 4 节列举了部分实验结果; 第 5 节对本文内容进行了总结.

1 方法概述

本文提出的方法包括离线训练先验模型和在线跟踪两部分. 在离线训练阶段, 通过在目标区域和背景区域分别建立 GMM 模型对颜色建模; 通过对形状训练样本做奇异值分解 (Singular value decomposition, SVD), 建立形状子空间. 在在线跟踪阶段, 对于每一帧, 初始轮廓先根据颜色信息进化; 然后把基于颜色得到的结果与形状模型中的均值形状进行比较, 并根据基于 Mahalanobis 距离的决策准则判断是否引入形状模型, 如果不需要引入, 那么基于颜色进化得到的结果就作为这一帧最终的结果, 否则, 需要引入形状模型使得轮廓在形状约束下继续进化. 返回每一帧得到的最终结果并用一个扩展 R-SVD 算法在线更新形状模型. 图 1 给出了整个算法的框架图.

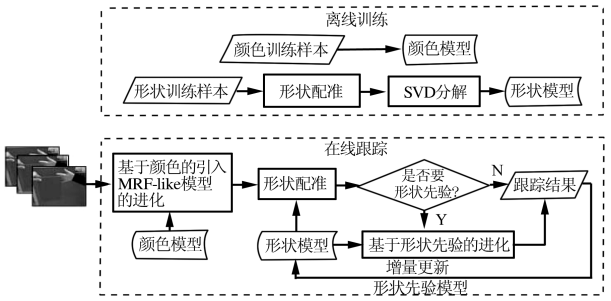


图 1 算法框架图

Fig. 1 The framework of our method

2 离线训练先验模型

先验模型的离线训练阶段包括建立颜色模型和形状模型两部分.

2.1 颜色先验模型

分别对目标区域和背景区域建立颜色 GMM 模型, 选用的颜色空间为 HSV 空间. 通过最大期望 (Expectation maximization, EM) 算法获得模型参数 (权重 ω, 均值 μ^c 和协方差矩阵 Σ^c). 对于像素 x_i, 它在颜色空间的概率密度函数 (Probabity density function, PDF) 为

$$p(x_i|\mu,\Sigma)=\sum_{j=1}^{k}\omega_j\eta(\boldsymbol{x}_i^c,\mu_j^c,\Sigma_j^c) \tag{1}$$

其中, $\boldsymbol{x}_i^c$ 是像素 x_i 的颜色特征, k 为高斯模式的个数, η 是标准的单高斯 PDF, 即

$$\eta\propto\frac{1}{|\Sigma_j^c|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(\boldsymbol{x}^c-\mu_j^c)^{\mathrm{T}}(\Sigma_j^c)^{-1}(\boldsymbol{x}^c-\mu_j^c)\} \tag{2}$$

2.2 形状先验模型

在本文中, 采用 Level sets 方法来描述轮廓[2], 即用 3 维的 Level sets 函数的零值来表达一个 2 维的轮廓. 常用的 Level sets 函数为带符号的距离函数. 我们定义轮廓内的符号为负, 轮廓外为正. 通过这种描述方式, 形状信息被完全嵌入到带符号的距离图像中 (用 ϕ 表示), 即

$$\phi(x,y)=\begin{cases}0, & (x,y)\in C\\ d(x,y,C), & (x,y)\in R_{\mathrm{out}}\\ -d(x,y,C), & (x,y)\in R_{\mathrm{in}}\end{cases} \tag{3}$$

其中, R_{in} 和 R_{out} 分别代表轮廓 C 的内部区域和外部区域, $d(x,y,C)$ 是从像素 (x,y) 到轮廓 C 的最短 Euclidean 距离, 即

$$d(x,y,C)=\min_{x_c,y_c\in C}\sqrt{(x-x_c)^2+(y-y_c)^2} \tag{4}$$

基于以上描述, 每个形状样本都可以用一个带符号的距离图像 ϕ 来表示. 在建立形状模型之前, 所有的形状训练样本都要求配准. 配准参数 (即仿射变换参数, 包含尺度因子、旋转角度和平移向量) 由 Paragios[16] 提出的变分的方法来求解. 该方法通过建立以仿射变化参数为变量的目标函数, 并用梯度下降法求解得到这些参数的迭代方程来完成不同形状之间的配准. 把每个配准后的形状样本 (矩阵形式的带符号距离图像) 拉成一个列向量 (维数为 $M\times 1$), 计算样本的均值 μ; 然后构建一个 $M\times N$ 的矩阵 X, 该矩阵的每一列为每个样本减去均值 μ 得到的向量, N 为样本的个数; 计算矩阵 X 的 SVD 分解

$$X=U\Sigma D^{\mathrm{T}} \tag{5}$$

其中, U 是一个 $M\times N$ 的特征向量矩阵, Σ 为一个 $N\times N$ 的对角矩阵, 对角元素为对应的奇异值. 我们选择前 k ($k\le N$) 个最大的奇异值以及对应的 k 个特征向量和均值 μ 构成形状模型 $\{\mu,\Sigma_k,U_k\}$.

3 在线跟踪

在线跟踪阶段包括三个步骤: 首先初始轮廓基于颜色信息进化; 然后采用基于 Mahalanobis 距离的决策准则判断是否引入形状先验; 最后是在形状约束下的轮廓进化.

3.1 基于颜色的轮廓进化

在基于颜色的进化这一阶段, 本文采用类似分割的思想[7−8] 建立能量函数, 即在当前帧找到一个最优的分割算子把图像中的目标区域和背景区域分开, 这个最优算子就是要跟踪目标的轮廓. 当前帧的结果是下一帧的初始值.

当前帧的分割问题可以建模为一个 MAP 问题. 定义对于一给定图像 I, 得到一种分割 $\varphi(R)$ 的后验概率为 $P(\varphi(R)|I)$. 根据 Bayesian 公式以及考虑到 $P(I)$ 为常量, 后验概率可以写成如下形式

$$P(\varphi(R)|I)\propto P(I|\varphi(R))P(\varphi(R)) \tag{6}$$

通常先验概率 $P(\varphi(R))$ 被建模为与轮廓长度有关的平滑因子项. 在本文中忽略这一项, 因为接下来提出的惩罚因子也有平滑轮廓的作用. 假定不同区域之间是相互独立的, 似然函数 $P(I|\varphi(R))$ 可以写为下面两式的乘积

$$P(I|\varphi(R))=P(I|\varphi(R_{\mathrm{in}}))P(I|\varphi(R_{\mathrm{out}})) \tag{7}$$

其中, R_{in} 和 R_{out} 的定义同式 (3), $P(I|\varphi(R_{\mathrm{in}}))$ 和 $P(I|\varphi(R_{\mathrm{out}}))$ 分别代表目标和背景的区域似然函数.

现有的方法通常假定每个区域内的像素也是独立的[7−8], 因此式 (7) 可以写为

$$P(I|\varphi(R)) = \prod_{x\in R_{\text{in}}} P(I(x)|\varphi(R_{\text{in}}))\times \prod_{x\in R_{\text{out}}} P(I(x)|\varphi(R_{\text{out}})) \tag{8}$$

这种假设有很大的不合理性, 特别是对于纹理区域或者具有重复模式的区域, 相邻像素之间存在很大的关联性.

为了避免以上问题, 本文考虑了邻域内像素之间的关系, 重新推导区域似然函数. 以目标区域似然函数为例, 一维的情况, 像素有序地排列. 根据一阶 One-sided 马尔科夫性, 目标区域似然函数可以有如下的推导

$$\begin{aligned} &P(I|\varphi(R_{\text{in}})) = \\ &\quad P(x_1,x_2,x_3,\cdots,x_i,x_{i+1},\cdots,x_n|\theta_{in}) = \\ &\quad P(x_1|\theta_{\text{in}})P(x_2|x_1,\theta_{\text{in}})\cdots P(x_n|x_1,x_2,\cdots, \\ &\quad\quad x_{n-1},\theta_{\text{in}}) \approx \\ &\quad P(x_1|\theta_{\text{in}})P(x_2|x_1,\theta_{in})\cdots P(x_n|x_{n-1},\theta_{\text{in}}) = \\ &\quad \prod_{x_i\in R_{\text{in}}} P(x_i|N_{x_i},\theta_{\text{in}}) \end{aligned} \tag{9}$$

其中, θ_{in} 表示目标 GMM 模型的参数 $\{\omega_{\text{in}},\mu^c_{\text{in}},\Sigma^c_{\text{in}}\}$, N_{x_i} 代表像素 x_i 的一维邻域像素. 推广到二维图像情况, 由马尔科夫随机场 (MRF) 理论, 二维的目标区域似然函数也有如下近似

$$P(I|\varphi(R_{\text{in}})) \approx \prod_{x_i\in R_{\text{in}}} P(x_i|N_{x_i},\theta_{\text{in}}) \tag{10}$$

其中, N_{x_i} 代表像素 x_i 在二维图像网格中的邻域.

本文引入一个惩罚因子来度量邻域内像素对中心像素的影响. 之所以称为惩罚因子, 是因为即将引入的这个因子体现了惩罚区域不平滑不连续的特性, 类似于 MRF 中的先验概率项的作用. 在本文中该惩罚因子相当于一个权重项. 由于我们是在 Label 层上用 MRF-like 模型建模的, 在解释惩罚因子之前, 首先要定义一个 Label 集. 像素的 Label 指明该像素属于目标还是背景. 如果像素属于目标的后验概率大于属于背景的后验概率, 则像素的 Label 为 1, 反之为 0. 此外还需要定义像素 x_i 周围一个 $(2w+1)\times(2w+1)$ 大小的方形邻域. 考虑邻域内像素之间的关系, 单个像素的目标似然函数正比于不考虑邻域关系的似然函数与惩罚因子的乘积

$$P(x_i|N_{x_i},\theta_{\text{in}}) \propto P(x_i|\theta_{\text{in}})\times \exp\left\{-\frac{\left(\frac{\max(N_1,N_0)}{N_1+N_0}\right)^2}{\sigma^2}\times\text{sign}\{(L-\frac{1}{2})(N_1-N_0)\}\right\} \tag{11}$$

其中, 不考虑邻域关系的似然函数 $P(x_i|\theta_{\text{in}})$ 由式 (1) 计算, 它只考虑了该像素自身在 GMM 模型中的似然概率. 式 (11) 中的指数项就是本文提出的惩罚因子. 惩罚因子中的 N_1 和 N_0 分别是邻域内 Label 为 1 和 0 的像素的个数. σ 是指数函数中控制衰减速度的参数. sign 是一个分段函数, 其定义为

$$\text{sign}\{(L-\frac{1}{2})(N_1-N_0)\} = \begin{cases} -1, & \text{若 } (L-\frac{1}{2})(N_1-N_0)>0 \\ 0, & \text{若 } (L-\frac{1}{2})(N_1-N_0)=0 \\ 1, & \text{若 } (L-\frac{1}{2})(N_1-N_0)<0 \end{cases} \tag{12}$$

其中的 L 具有如下形式

$$L = \begin{cases} 1, & \text{当计算 } P(x_i|N_{x_i},\theta_{\text{in}}) \\ 0, & \text{当计算 } P(x_i|N_{x_i},\theta_{\text{out}}) \end{cases} \tag{13}$$

单个像素的背景似然函数 $P(x_i|N_{x_i},\theta_{\text{out}})$ 可以通过类似的方法定义, 只需把式 (11) 中的 θ_{in} 换成 θ_{out} 即可. 惩罚因子具有如下特点:

1) 如果中心像素的 Label 与邻域内大多数像素的 Label 相同, 则惩罚因子有增加中心像素似然函数的作用. 增加的程度取决于 N_1 和 N_0 之间的差异, 差异越大, 增加的越多.

2) 如果中心像素的 Label 与邻域内大多数像素的 Label 都不一致, 则惩罚因子有减小中心像素似然函数的作用. 减小的程度取决于 N_1 和 N_0 之间的差异, 差异越大, 减小的越多.

3) 如果 N_1 和 N_0 相同, 则惩罚因子等于 1, 对中心像素的似然函数不起作用.

综上所述, 后验概率 $P(\varphi(R)|I)$ 可以展开为

$$P(\varphi(R)|I) \propto \prod_{x_i\in R_{\text{in}}} P(x_i|N_{x_i},\theta_{\text{in}}) \prod_{x_j\in R_{\text{out}}} P(x_j|N_{x_j},\theta_{\text{out}}) \tag{14}$$

将 MAP 问题转化为能量最小化问题, 可以得到

$$E=-\ln P(\varphi(R)|I)\approx-\iint_{\boldsymbol{x}_i\in R_{\rm in}}\ln P(\boldsymbol{x}_i|N_{\boldsymbol{x}_i},\theta_{\rm in})\mathrm{d}\boldsymbol{x}_i-\iint_{\boldsymbol{x}_j\in R_{\rm out}}\ln P(\boldsymbol{x}_j|N_{\boldsymbol{x}_j},\theta_{\rm out})\mathrm{d}\boldsymbol{x}_j \tag{15}$$

类似于文献 [8] 中变分的方法, 通过求解上述方程相关的 Euler-Lagrange 方程最小化能量函数, 可以获得 Level sets 的速度进化方程为

$$F_{x,y}=-\sum_{i=-l}^{l}\sum_{j=-l}^{l}\ln P_{R_{\rm in}}(I_{\tilde{x}})H_a(\phi(\tilde{x},t))+\sum_{i=-l}^{l}\sum_{j=-l}^{l}\ln P_{R_{\rm out}}(I_{\tilde{x}})(1-H_a(\phi(\tilde{x},t))) \tag{16}$$

其中, $\tilde{x}$ 是像素 (x,y) 的邻域像素: $\tilde{x}=(x+i,y+j)$, 邻域大小为 $(2l+1)\times(2l+1)$, $H_a(\phi(\tilde{x},t))$ 是一个 Heaviside 函数, 具有如下形式

$$H_a(\phi(\tilde{x},t))=\begin{cases}0, & \phi(\tilde{x},t)\geq 0\\ 1, & \phi(\tilde{x},t)<0\end{cases} \tag{17}$$

$P_{R_{\rm in}}(I_{\tilde{x}})$ 和 $P_{R_{\rm out}}(I_{\tilde{x}})$ 分别表示 $\tilde{x}$ 属于目标和背景的归一化后的似然函数

$$P_{R_{\rm in}}(I_{\tilde{x}})=\frac{P(\tilde{x}|N_{\tilde{x}},\theta_{\rm in})}{P(\tilde{x}|N_{\tilde{x}},\theta_{\rm in})+P(\tilde{x}|N_{\tilde{x}},\theta_{\rm out})} \tag{18}$$

$$P_{R_{\rm out}}(I_{\tilde{x}})=\frac{P(\tilde{x}|N_{\tilde{x}},\theta_{\rm out})}{P(\tilde{x}|N_{\tilde{x}},\theta_{\rm in})+P(\tilde{x}|N_{\tilde{x}},\theta_{\rm out})} \tag{19}$$

在法线方向迭代地改变 ϕ 值, 初始轮廓在速度模型的作用下最终将进化到目标的边缘处.

$$\frac{\phi^{n+1}-\phi^n}{\Delta t}+(F_{x,y}+F_{\rm curv})|\nabla\phi|=0 \tag{20}$$

$F_{x,y}$ 就是式 (16) 中基于颜色信息得到的速度, $F_{\rm curv}=-\alpha\kappa$ 是与轮廓曲率 κ 相关的速度项, 有平滑轮廓并防止尖点振荡的作用[17]. 这一步得到的最终进化结果用 ϕ_c 表示.

3.2 基于 Mahalanobis 距离的判别决策

将基于颜色的进化结果 ϕ_c 与形状模型中的均值形状 μ 进行配准 (配准参数的求解见文献 [16]), 然后将配准后的结果转化为标准化向量 $\boldsymbol{x}_c$, 并投影到建立好的形状子空间中. 最终, ϕ_c 可以由一个 k 维的向量 $\boldsymbol{\alpha}_c$ 表示

$$\boldsymbol{\alpha}_c=U_k^{\rm T}(\boldsymbol{x}_c-\mu) \tag{21}$$

其中, U_k 和 μ 是训练好的形状模型的参数 (见第 2.2 节). $\boldsymbol{x}_c$ 和 μ 之间的 Mahalanobis 距离表示为

$$\gamma^2=(\boldsymbol{x}_c-\mu)^{\rm T}C^{-1}(\boldsymbol{x}_c-\mu) \tag{22}$$

其中, C 是协方差矩阵: $C\approx U_k\Sigma_k^2U_k^{\rm T}$, 代入式 (22) 中, 可以得到

$$\gamma^2\approx(\boldsymbol{x}_c-\mu)^{\rm T}U_k\Sigma_k^{-2}U_k^{\rm T}(\boldsymbol{x}_c-\mu)=\boldsymbol{\alpha}_c^{\rm T}\Sigma_k^{-2}\boldsymbol{\alpha}_c \tag{23}$$

如果 γ^2 比预先定义的阈值 T 大, 则说明可能由于噪音或遮挡等原因造成根据颜色得到的进化结果 ϕ_c 离形状子空间的距离比较远, 单靠颜色信息不足以得到准确的跟踪结果, 因此需要引入形状先验, ϕ_c 需要在形状先验的约束下继续进化; 否则, 如果 γ^2 小于 T, 则 ϕ_c 被认为是最终的结果, 直接返回用于更新形状模型.

3.3 基于形状先验的轮廓进化

如果根据上述判别准则需要引入形状先验, 则在这一层的进化过程中, ϕ_c 会基于形状先验继续进化. 本文提出了一个权重形状距离因子 (WSDT) 来融合全局形状信息和局部颜色信息.

1) 让 ϕ_c 在形状子空间进化. 根据式 (23) 中描述的 Mahalanobis 距离, 建立一个全局形状能量函数 $E_{G\rm shape}$:

$$E_{G\rm shape}=\boldsymbol{\alpha}^{\rm T}\Sigma_k^{-2}\boldsymbol{\alpha} \tag{24}$$

其中, $\boldsymbol{\alpha}$ 是一个 k 维的向量, 代表形状子空间中的一个轮廓. 采用梯度下降法获得 $\boldsymbol{\alpha}$ 的迭代方程为

$$\frac{\partial\boldsymbol{\alpha}}{\partial t}=-\frac{\partial E_{G\rm shape}}{\partial\boldsymbol{\alpha}}=-2\Sigma_k^{-2}\boldsymbol{\alpha} \tag{25}$$

设定 $\boldsymbol{\alpha}$ 的初始值为将 ϕ_c 投影到子空间得到的向量, 即 $\boldsymbol{\alpha}_{\rm initial}=\boldsymbol{\alpha}_c$. 经过有限步迭代, 得到 $\boldsymbol{\alpha}_{\rm final}$.

2) 由 k 维向量 $\boldsymbol{\alpha}_{\rm final}$ 重构出轮廓形状 $\tilde{\boldsymbol{x}}$ 为

$$\tilde{\boldsymbol{x}}=U_k\boldsymbol{\alpha}_{\rm final}+\mu \tag{26}$$

重构出的形状 $\tilde{\boldsymbol{x}}$ 通过第 3.2 节中求出的配准参数逆变换到图像平面, 并重新计算包含该轮廓的 Level sets 函数 (带符号的距离图像), 我们用 ϕ_0 表示.

3) 现有的方法一般都是通过构建一个形状距离因子来建立形状约束下的像素级别的轮廓进化方程

$$\frac{\partial\phi}{\partial t}=-2(\phi-\phi_m) \tag{27}$$

其中, ϕ_m 是训练样本均值的 Level sets 函数. 在本文中, 用 ϕ_0 代替 ϕ_m, 并在轮廓进化方程中引入一个 WSDT 因子来融合基于颜色的进化结果 (用 ϕ_c 表示) 和在形状子空间中进化得到的结果 (用 ϕ_0 表示). 由于样本的形状空间是由 k 个有限的状态向量

表征, 因此基于形状先验重构出的轮廓 ϕ_0 更侧重于整体性, 在细节部分不准确; 相反, 基于颜色得到的 ϕ_c 在细节上更为准确. 引入 WSDT 因子后, 改进的方程具有如下表现形式

$$\frac{\partial \phi}{\partial t} = -2(\phi - \phi_0)[1 - \mathrm{e}^{-(\frac{\phi-\phi_0}{\tau})^2}] \tag{28}$$

τ 是控制指数函数衰减的参数. 在这个像素级的轮廓进化方程中, ϕ 的初始值为第 3.1 节中基于颜色进化获得的 ϕ_c. 式 (28) 的中括号里面的因子则为 WSDT 因子, 它是 ϕ 和 ϕ_0 之间距离 ($d(\phi, \phi_0) = \phi - \phi_0$) 的函数. 对于该因子考虑如下两种极限情况:

a) 在进化过程中, 对于某个像素位置, 如果 $d(\phi, \phi_0)$ 非常小, 则权重因子趋于 0. 因此, ϕ 的值不会改变太多, 即经过迭代以后它的值仍然接近初始值 ϕ_c. d 非常小意味着在这个像素位置, 基于颜色进化得到的结果 ϕ_c 和基于形状约束得到的结果 ϕ_0 很相近, 由于 ϕ_c 在细节部分更准确, 因此优先考虑.

b) 在进化过程中, 对于某个像素位置, 如果 $d(\phi, \phi_0)$ 非常大, 则权重因子趋于 1. 因此, ϕ 不断地进化, 直到它的值接近 ϕ_0. d 非常大意味着在这个像素位置, 可能出现遮挡或噪音的干扰, 使得基于颜色的结果 ϕ_c 不可靠, 只能相信基于形状先验得到的结果 ϕ_0.

这样综合考虑颜色和形状, 使得所得到的结果更加准确, 整体和局部兼顾. 随着 ϕ 不断地迭代, $d(\phi, \phi_0)$ 会变得越来越小. 由于权重因子的作用, ϕ 很快就都收敛了. 但是对于那些一开始 $d(\phi, \phi_0)$ 比较大的像素位置, ϕ 收敛得太快导致结果不准确. 因此为了解决上述问题, 我们需要调节参数 τ 来控制不同像素位置的进化速度. τ 由 ϕ_c 和 ϕ_0 之间的距离 (用 $d(\phi_c, \phi_0)$ 表示) 决定:

$$\tau = \beta \mathrm{e}^{-(\phi_c - \phi_0)^2} \tag{29}$$

其中 β 是一个常数. 通过这样的设定, 在 $d(\phi_c, \phi_0)$ 大的像素位置处, 即使随着 $d(\phi, \phi_0)$ 变得越来越小, ϕ 仍然能继续进化; 在 $d(\phi_c, \phi_0)$ 小的像素位置处, ϕ 很快停止进化. 最终, 基于颜色的进化结果 ϕ_c 和在子空间进化得到的结果 ϕ_0 很好地融合在一起.

3.4　增量学习算法

获得了最终的进化结果后, 将它们与模型中的均值形状配准, 转化为标准格式后返回用于更新形状模型. 在这一步, 我们采用了一个有效的在基于表观模型的跟踪领域被广泛应用的在线学习算法[18]. 该算法优于其他一些子空间更新算法[19−20]: 首先, 它能正确地更新样本均值和正交基; 其次, 它一次能处理一批样本而不仅限于一个样本. 通过这个增量学习算法, 跟踪目标的形状变化能够被在线学习.

已知时刻 $i-1$ 的形状模型 $\{\mu_{i-1}, U_{i-1}, \Sigma_{i-1}\}$ 和新的数据集 X, 则时刻 i 的新的形状模型 $\{\mu_i, U_i, \Sigma_i\}$ 按照算法 1 进行增量更新.

算法 1 (子空间的增量学习算法).

输入. 时刻 $i-1$ 的均值向量 $\boldsymbol{\mu}_{i-1}$, 正交基 U_{i-1} 以及奇异值 Σ_{i-1}, 新的样本集 X 和均值 $\boldsymbol{\mu}_{\mathrm{new}}$, 旧样本个数 n_0, 新样本个数 n_{new}, "遗忘因子" 系数 ff.

输出. 时刻 i 的均值向量 $\boldsymbol{\mu}_i$, 正交基 U_i 以及奇异值 Σ_i.

步骤 1. 更新均值向量

$$\boldsymbol{\mu}_i = \frac{ff \times n_0}{(n_{\mathrm{new}} + ff \times n_0)}\ \boldsymbol{\mu}_{i-1} + \frac{n_{\mathrm{new}}}{(n_{\mathrm{new}} + ff \times n_0)}\ \boldsymbol{\mu}_{\mathrm{new}};$$

步骤 2. X 为 0 均值:

$$X = X - repmat(\boldsymbol{\mu}_{\mathrm{new}}, [1, n_{\mathrm{new}}]);$$

步骤 3. 构建复合矩阵 X'

$$X' = (ff \times U_{i-1}\Sigma_{i-1} \mid X \mid \sqrt{\frac{n_0 n_{\mathrm{new}}}{n_0 + n_{\mathrm{new}}}}\ (\boldsymbol{\mu}_{i-1} - \boldsymbol{\mu}_{\mathrm{new}}));$$

步骤 4. 计算矩阵 X' 的 QR 分解 $X' = QR$;

步骤 5. 计算矩阵 R 的 SVD 分解 $R = U\Sigma D^{\mathrm{T}}$;

步骤 6. 计算特征值和奇异值

$$U_i' = QU,\quad \Sigma_i' = \Sigma \times \sqrt{\frac{n_0}{(n_{\mathrm{new}} + ff \times n_0)}}\ ;$$

步骤 7. 取 Σ_i' 的前 k 个最大的奇异值构成对角阵 Σ_i, U_i' 中对应的 k 列构成最终的正交基 U_i.

4　实验

为了验证本文提出方法的有效性, 我们在各种视频序列上进行了一系列的实验, 其中还包括一些比较实验.

在实验中, 所有的视频序列都由移动摄像机拍摄. 跟踪目标用一个轮廓来描述. 采用 Narrow Band 方法[21] 实现 Level sets 的快速进化. 颜色模型训练阶段, 目标和背景区域的特征都用 GMM 建模. 式 (11) 中的参数 σ 对于所有的序列都设为 0.21, 式 (16) 中的 l 也和序列无关, 都设为 2. 形状模型训练阶段, 训练样本统一为 3 000 维的向量. 选取前 20 个特征向量构成子空间的正交基. 基于 Mahalanobis 距离的判别准则里的阈值 T 为 16. 每隔 3 帧, 跟踪结果返回用于更新形状模型. 算法 1 中的 "遗忘因子" 系数 ff 为 0.95. 以上参数中, 最重要的是惩罚因子中控制方形邻域大小的参数 w. 很明显, w 越大, 得到的轮廓就越平滑. 当 w 为 0 时, 邻域只有像素自身, 这时惩罚因子不起作用. 图 2 显示了不同 w 下的不同跟踪结果. 这个实验只考虑了颜色信息. 从实验结果可以看出, 随着 w 的不断

增加, 轮廓越来越平滑, 但随之也丢失了更多的细节. 因此, 选择一个合适的参数 w 可以平衡跟踪结果的准确度和平滑度. 在我们的实验中, w 设为 6.

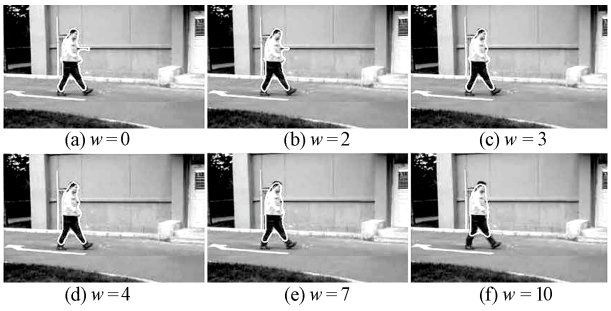


图 2 不同参数 w 下的跟踪结果
(从 (a) 到 (f), w 依次为 0, 2, 3, 4, 7 和 10)

Fig. 2 Tracking results with different values of parameter w (From (a) to (f), w is 0, 2, 3, 4, 7 and 10, respectively.)

4.1 真实视频序列上的实验

我们分别在三个不同的视频序列上做下述实验.

第一个序列是在移动摄像机下跟踪两张人脸. 当人不断改变脸部姿态的时候, 摄像机也在不停地移动和变焦. 在这个序列里只用到了颜色信息. 根据 Mahalanobis 距离判别准则, 没有考虑形状先验. 不同的人对应于不同灰度的轮廓. 实验结果如图 3 (a) 所示. 即使脸部姿态不停地改变, 本文提出的方法仍然能够准确地跟踪这两个轮廓.

第二个序列是跟踪一个快速运动的足球. 与第一个序列相同, 没有用到形状先验. 跟踪结果如图 3 (b) 所示, 每幅图像的右上角为放大的跟踪轮廓. 从实验结果可以看出, 当足球 (箭头所指) 快速运动时, 我们也能很好地进行跟踪. 位于中间位置的那一帧虽然存在严重的运动模糊, 但基于本文提出的带有惩罚因子的速度进化模型仍能得到准确、平滑的轮廓.

第三个序列是跟踪一张运动的人脸. 伴随着脸部姿态变化, 摄像机也在移动和变焦. 在跟踪过程中, 人脸被具有相同肤色的手部分遮挡. 根据 Mahalanobis 距离判别准则, 需要引入形状模型来恢复被遮挡的脸部轮廓. 针对这个序列, 式 (29) 中的参数 β 为 200. 从跟踪结果 (图 3 (c)) 可以发现, 即使存在部分遮挡, 我们的方法还是能比较好地恢复出被遮挡部分的轮廓. 在整个过程中, 人脸的视角由朝前看变为向下看, 导致轮廓形状发生了变化. 由于形状模型具有在线学习的能力, 连续的形状变化被很好地学习到模型中. 因此形状发生变化以后, 轮廓仍能恢复得很准确.

4.2 比较实验

为了能更好地体现本文方法的优势, 本节做了以下两个比较实验.

第一个比较实验是在一个 Mickey head 跟踪序列上做的. 在这个场景中, 背景中的网格颜色和前景相同, 起着干扰跟踪目标的作用. 随着目标的移动, 摄像机也在不断地变焦和倾斜. 跟踪结果如图 4 所示. 我们考虑了三种情况, 分别是: 1) 基于颜色信息没有考虑惩罚因子得到的跟踪结果图 4 (a); 2) 基于颜色信息考虑了惩罚因子得到的结果图 4 (b); 3) 在图 4 (b) 的基础之上又引入了形状先验得到的结果

(a) 两张移动的人脸序列

(a) Two moving faces sequence

(b) 足球运动序列

(b) Football sequence

(c) 脸部遮挡序列

(c) Occluded face sequence

图 3 不同序列的跟踪结果

Fig. 3 Tracking results for different sequences

(a) 基于颜色没有考虑惩罚因子

(a) Only with color feature without considering the penalty term

(b) 基于颜色考虑了惩罚因子

(b) Only with color feature considering the penalty term

(c) 结合颜色和形状先验

(c) Fusion of color feature and shape priors

图 4 三种情况下的跟踪结果

Fig. 4 Tracking results for three different cases

图 4 (c). 在图 4 中, 相同列对应着相同帧. 从这些比较实验结果可以很容易发现本文的方法 (考虑了惩罚因子) 在很大程度上减弱了背景干扰的影响, 比没有考虑惩罚因子的方法 (如 Yilmaz′s) 得到的轮廓更为准确和平滑. 在基于颜色得到比较准确的轮廓基础之上, 引入形状先验 (式 (29) 中的参数 β 为 100), 将局部的颜色信息和全局的形状信息结合起来, 得到的结果比只考虑颜色信息得到的更为规范和准确. 特别是从结果图像的第 4 列 (存在严重的运动模糊) 可以看出, 引入形状先验得到的结果在这三种情况中是最优的.

为了突出形状模型的学习能力, 我们在一个具有明显形状变化的序列上做了如下比较实验. 该序列是跟踪一只被部分遮挡的手. 手的形状在不断地变化, 经历了由最初的五指分开到最终的五指并拢这个过程. 我们进行了三种情况的比较实验. 首先在跟踪过程中只考虑了颜色信息, 结果见图 5 (a). 由于遮挡的存在, 仅考虑颜色信息是不够的. 结果只能得到未被遮挡的手的部分轮廓. 因此需要引入形状先验. 在形状模型的离线训练阶段, 采用没有遮挡的完整轮廓 (五指分开) 作为训练样本. 然后, 训练好的形状模型被用来引导轮廓继续进化. 式 (29) 中的参数 β 为 150. 引入形状约束得到的跟踪结果见图 5 (b) 和图 5 (c). 其中图 5 (b) 是形状模型没有在线更新的情况; 图 5 (c) 是形状模型在线增量更新的情况. 由比较结果很容易发现, 一开始物体的形状没有发生很大的变化, 静态的形状模型 (缺乏在线更新能力) 能够恢复被遮挡部分的轮廓, 但是当跟踪过程中出现新的形状时, 它就不能提供正确的形状先验, 导致跟踪结果有误. 相反我们的形状模型具有自适应能力, 反映了目标形状的变化, 因此能够很好地跟踪具有连续形状变化的目标.

(a) 只考虑颜色信息

(a) Tracking results of using color feature only

(b) 引入形状先验但没有更新形状模型

(b) Tracking results without the shape subspace update

(c) 引入形状先验同时在线更新形状模型

(c) Tracking results of our proposed method which updates the shape subspace model on-line

图 5 三种情况下的跟踪结果

Fig. 5 Tracking results for three different cases

5 结论

本文提出了一个移动摄像机下基于 Level sets 的分层跟踪框架. 该框架将颜色信息和形状先验融合起来, 用于提高在噪音或遮挡情况下跟踪结果的准确度. 主要贡献如下: 首先在第一层基于颜色进化过程中, 利用马尔科夫随机场模型度量邻域像素之间的关系, 并将这种关系加入到速度进化模型中以平衡轮廓的准确度和平滑度. 其次提出一个基于 Mahalanobis 距离的判别准则, 将基于颜色得到的结果和形状模型中的均值形状进行比较, 判断是否需要引入形状先验. 如果不需要, 则根据颜色进化得到的结果就作为最终结果; 否则需要引入形状先验使得轮廓在第二层继续进化. 在基于形状的进化过程中, 提出一个 WSDT 因子融合颜色和形状信息, 同时兼顾了整体的形状和局部的颜色细节. 最终的跟踪结果返回到 ASMs, 并采用增量学习算法在线更新. 我们在大量视频序列上做了实验, 此外还和其他方法进行了比较. 实验结果证明即使存在背景噪音或部分遮挡的情况, 本文提出的方法仍然能够得到准确并且平滑的轮廓.

References

1 Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *International Journal of Computer Vision*, 1988, **1**(4): 321−331

2 Osher S, Sethian J A. Fronts propagation with curvature-dependent speed: algorithms based on Hamilton-Jacobi for mulations. *Journal of Computational Physics*, 1988, **79**(1): 12−49

3 Caselles V, Kimmel R, Sapiro G. Geodesic active contours. *International Journal of Computer Vision*, 1997, **22**(1): 61−79

4 Paragios N, Deriche R. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(3): 266−280

5 Mumford D, Shah J. Boundary detection by minimizing functionals. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 1985. 22−26

6 Chan T F, Vese L A. Active contours without edges. *IEEE Transactions on Image Processing*, 2001, **10**(2): 266−277

7 Zhu S C, Yuille A. Region competition: unifying snakes, region growing and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18**(9): 884−900

8 Yilmaz A, Li X, Shah M. Object contour tracking using level sets. In: Proceedings of Asian Conference on Computer Vision. Jeju, Korea: Springer, 2004. 1−7

9 Houhou N, Thiran J P, Bresson X. Fast texture segmentation model based on the shape operator and active contour. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska: IEEE, 2008. 1−8

10 Leventon M E, Grimson W E L, Faugeras O. Statistical shape influence in geodesic active contours. In: Proceedings of the 5th IEEE EMBS International Summer School on Biomedical Imaging. Hilton Head Island, South Carolina: IEEE, 2000. 316−323

11 Rousson M, Paragios N. Shape priors for level set representation. In: Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2002. 78−92

12 Charpiat G, Faugeras O, Keriven R. Shape statistics for image segmentation with prior. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, Minnesota: IEEE, 2007. 1−6

13 Cootes T F, Taylor C J, Cooper D H, Graham J. Active shape models — their training and application. *Computer Vision and Image Understanding*, 1995, **61**(1): 38−59

14 Fussenegger M, Roth P M, Bischof H, Pinz A. Online, incremental learning of a robust active shape model. In: Proceedings of Symposium of the German Association for Pattern Recognition. Berlin, Germany: Springer, 2006. 122−131

15 Cremers D. Dynamical statistical shape priors for level set based tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(8): 1262−1273

16 Paragios N, Rousson M, Ramesh V. Matching distance functions: a shape-to-area variational approach for global-to-local registration. In: Proceedings of European Conference on Computer Vision. Copenhagen, Denmark: Springer, 2002. 813−815

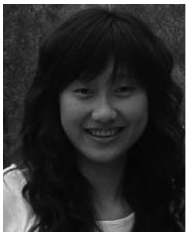
17 Sethian J A. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge: Cambridge University Press, 1999.

18 Lim J, Ross D, Lin R S, Yang M H. Incremental learning for visual tracking. In: Proceedings of Conference on Neural Information Processing Systems. Massachusetts, USA: The MIT Press, 2004. 793−800

19 Skocaj D, Leonardis A. Weighted and robust incremental method for subspace learning. In: Proceedings of International Conference on Computer Vision. Nice, France: IEEE, 2003. 1494−1501

20 Li Y M. On incremental and robust subspace learning. *Pattern Recognition*, 2004, **37**(7): 1509−1518

21 Adalsteinsson D, Sethian J A. A fast level set method for propagating interfaces. *Journal of Computational Physics*, 1995, **118** (2): 269−277



周 雪 电子科技大学自动化工程学院讲师. 主要研究方向为视觉监控, 图像与视频处理. 本文通信作者.
E-mail: zhouxue420@gmail.com
(**ZHOU Xue** Lecturer at the College of Automation, University of Electronic Science and Technology of China. Her research interest covers visual surveillance, image and video processing. Corresponding author of this paper.)

胡卫明 中国科学院自动化研究所研究员. 主要研究方向为视频信息处理和网络信息安全识别.
E-mail: wmhu@nlpr.ia.ac.cn
(**HU Wei-Ming** Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers video information processing and network information security recognition.)