

矩阵的低秩稀疏表达在视频目标分割中的研究

顾 蕊^{1,2}, 马 争¹, 解 梅³

(1. 电子科技大学通信与信息工程学院 成都 611731; 2. 成都航空职业技术学院航空工程学院 成都 610100;
3. 电子科技大学电子工程学院 成都 611731)

【摘要】提出了一种视频目标跟踪与分割的在线算法。该算法将每帧图像中的超级像素作为一个数据点, 根据已知的目标和背景建立模板, 当前帧中待分割的目标可以看成已知模板的稀疏线性表达。根据此线性表达的系数可以建立描述当前帧与模板的相似性矩阵, 即表达子。由于视频图像的连续性, 表达子具有低秩和稀疏的特征。因此通过求解矩阵的低秩稀疏的优化问题可以得到当前帧中所有数据点属于目标的概率分布。为了获得基于像素级的分割结果, 通过能量最小框架, 并利用图分割的方法最终实现目标的分割。实验结果表明该算法具有良好的分割效果。

关键词 能量最小; 图分割; 低秩; 稀疏; 视频目标分割

中图分类号 TP39 文献标志码 A doi:10.3969/j.issn.1001-0548.2017.02.008

Video Object Segmentation Via Low-Rank Sparse Representation

GU Song^{1,2}, MA Zheng¹, and XIE Mei³

(1. School of Communication and Information Engineering, University of Electronic Science and Technology of China Chengdu 611731;
2. Department of Aircraft Maintenance Engineering, Chengdu Aeronautic Polytechnic Chengdu 610100;
3. School of Electronic Engineering, University of Electronic Science and Technology of China Chengdu 611731)

Abstract We present a novel on-line algorithm for target segmentation and tracking in video. Superpixels, which are abstracted in every frame, are treated as data points in this paper. The object in current frame is represented as sparse linear combination of dictionary templates, which are generated based on the segmentation result in the previous frame. Then the algorithm capitalizes on the inherent low-rank structure of representation that are learned jointly. A low-rank sparse matrix optimal solution results in the construction of the trimap. At last, a simple energy minimization solution is adopted in segmented stage, leading to a binary pixel-wise segmentation. Experiments demonstrate that our approach is effective.

Key words energy minimization; graph cut; low rank; sparse; video object segmentation

基于视频的目标分割(video object segmentation)不仅能够跟踪目标的位置, 还能精确地描述目标的形状。它可以看成是目标的精确跟踪。视频目标分割的关键在于时间一致性(temporal coherence)和空间一致性(spatial coherence)的表达。时间一致性描述了在连续帧中目标的相似性, 空间一致性描述了在一帧图像中目标与背景的分辨能力。文献[1-2]提出了基于水平集(level set)的分割算法。这种算法的缺点在于将运动估计与分割过程分别独立, 将运动估计的结果作为分割的输入。这样当运动估计不准确时, 会影响分割的精度。建立能量函数, 利用能量最小化的方法进行目标分割是当前比较流行的方法。文献[3-6]分别利用目标的运动模型、颜色纹理

等信息建立了不同的能量函数。这种方法的优点在于能够将时间与空间信息利用概率模型融合在一起, 最终获得较好的分割效果。文献[3-4]通过对某些关键帧中已知目标的学习, 提出了一种离线算法。文献[5]利用深度摄像机获得了目标的深度信息。但 these 方法都大大限制了其应用范围。文献[6]利用目标的颜色信息建立了3D条件随机场模型(conditional random field)。但由于其信息量较少, 分割精度不高。本文也将利用能量最小的框架结构, 创建新的能量函数, 只需要较少的已知信息就能够进行视频目标的在线分割。

另一方面, 本文将视频目标分割看作一种目标跟踪算法, 将目标的分割转化成矩阵的低秩稀疏表

收稿日期: 2015-11-05; 修回日期: 2016-03-18

基金项目: 国家自然科学基金(61271288); 教育部博士点基金(20130185130001)

作者简介: 顾蕊(1977-), 男, 博士, 主要从事数字图像处理方面的研究。

达。近年来,矩阵的低秩稀疏表达已经被广泛应用在目标跟踪上。文献[7]提出了一种L1范数模型,将目标作为模板集的稀疏表达形式,实现了目标的在线跟踪。文献[8]提出了基于粒子滤波的跟踪算法,将对目标的跟踪转化为矩阵低秩稀疏的优化问题。

本文提出一种基于区域的在线目标视频分割算法。首先将图像过分割(over segmentation)成超级像素(superpixel),将超级像素作为数据点,这不仅滤除了不必要的细节特征,而且可以大大简化计算量。然后根据上一帧的分割结果建立模板集,将当前帧所有数据点看作模板集的稀疏线性组合。将目标分割问题转化为矩阵的低秩稀疏的优化问题。根据计算出的线性组合系数矩阵建立当前帧中每个超级像素的概率分布。最后将此概率分布作为能量函数的一个线索(cue),结合其他基本信息,构建出新的能量函数,利用能量最小化求得最终的结果。此外,由于超级像素的形状不规则,本文将采用L2ECM^[9](local log-euclidean covariance matrix)算法提取其特征。

1 特征提取

基于区域的分割算法已经被广泛应用。本文将超级像素^[10]作为一个数据点进行图像的分割。这种方法不仅能够大大地减少计算的数据量,而且超级像素本身已经将某些相似的像素聚集在一起,滤除了大量的细节噪声,增加了算法的鲁棒性。考虑到超级像素形状的不规则性,本文采用L2ECM算法提取其特征。在一幅图像中,一个像素的基本特征可以表示为:

$$f(x, y) = [I(x, y), |I_x(x, y)|, |I_y(x, y)|, |I_{xx}(x, y)|, |I_{yy}(x, y)|]^T \quad (1)$$

式中, $|\cdot|$ 表示绝对值; $I(x, y)$ 表示图像中像素点 (x, y) 的像素值; $I_x(x, y)$ 和 $I_{xx}(x, y)$ 分别表示 x 方向的一阶和二阶导数; $I_y(x, y)$ 和 $I_{yy}(x, y)$ 分别表示 y 方向的一阶和二阶导数。

在图像中给定一个超级像素 S , 则 $F_S = (f(x_i, y_i) \in R^d)_{i=1}^{N_S}, (x_i, y_i) \in S$ 。显而易见, F_S 是一个 $d \times N_S$ 维的矩阵, 其中 d 为基本特征的维度, N_S 为超级像素中所包含像素的个数。 F_S 的每一列表示 S 中一个像素的基本特征。计算 F_S 的协方差并表示为 C_S , 则 C_S 是一个 $d \times d$ 维的方阵, 它与所包含的像素个数 N_S 无关。由于协方差矩阵所构成空间满足黎曼流形的要求, 无法在欧式空间中进行计算, 因此必须将 C_S 通过 $\log(C_S)$ 转化到欧氏空间中^[9]。取 $\log(C_S)$ 的上对角

并转化成矢量形式, 因此一个超级像素 S_i 的特征是长度为 $\frac{d(d+1)}{2}$ 的向量, 并表示为 f_i 。用L2ECM算法提取超级像素的特征具有较强的抗噪能力, 对旋转、缩放和光照变化都不敏感^[11], 并且提取的特征维度只与基本特征的维度相关, 与超级像素的形状和大小无关。

2 低秩稀疏表达

文献[12-14]均提出了基于区域的图像(视频)分割算法, 它们的基本思想是一幅图像中的元素可以用其他的元素进行线性表达。结合文献[8], 本文将目标分割转化为矩阵的低秩稀疏表达问题进行求解。

2.1 模板的建立

基于上一帧 f^{t-1} 中目标分割的结果, 可以创建一个模板集(dictionary templates), 用 D^{t-1} 表示。 D^{t-1} 是一个 $d_s \times M$ 维的矩阵, 其中 d_s 是一个超级像素的特征维度, M 为模板集中超级像素的个数。为了表达清晰, 可以将 D^{t-1} 的所有列进行重新组合, 使得 $D^{t-1} = [D^O \ D^B]$, 其中 D^O 表示所有的目标特征, D^B 表示所有的背景特征。

2.2 低秩稀疏表达

将当前帧 f^t 中所有的元素构成样本集 $X = [x_1 \ x_2 \ \dots \ x_N]$, 则 X 为一个 $d_s \times N$ 维的矩阵, 其中 N 为 f^t 中元素的个数。在没有噪声的情况下, X 中的任意一列 x_i 都可以用 D^{t-1} 进行线性表达, 将这种关系写成矩阵的形式:

$$X = D^{t-1}Z + E \quad (2)$$

式中, Z 的每一列表示 X 的对应列用 D^{t-1} 进行线性表达的系数, 矩阵 Z 称之为表达子。对应于 D^{t-1} 被分解成 D^O 和 D^B , 表达子 $Z = \begin{bmatrix} Z^O \\ Z^B \end{bmatrix}$ 。 E 为由于噪声引起的误差, 如图1所示。

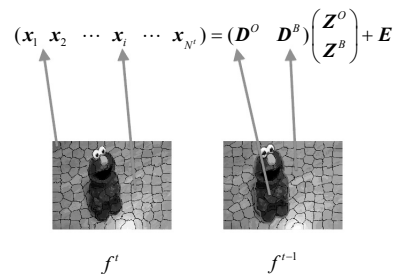


图1 线性表达关系示意图(视频图像参考文献[2])

从图1中可以看出:

1) 由于视频图像的连续性, f^t 中很多元素具有

很大的相似性, 也就是说, 样本集中的很多元素都可以用模板集进行相同的表达, 因此最终的表达子 \mathbf{Z} 应为低秩矩阵。

2) 为了尽可能用最相似的模板来表达 f' 中的各个元素, 即 \mathbf{Z} 的每一列中不为0的元素尽可能少, 因此最终的表达子 \mathbf{Z} 应为稀疏矩阵。

3) 根据文献[7-8]可知, 图像中的噪声可以用稀疏模型进行拟合。因此 \mathbf{E} 也应为稀疏矩阵。

通过以上3点, 可以得出式(2)的最优解为:

$$\begin{aligned} \arg \min_{\mathbf{Z}, \mathbf{E}} (\|\mathbf{Z}\|_* + \beta \|\mathbf{Z}\|_1 + \gamma \|\mathbf{E}\|_1) \\ \text{s.t. } \mathbf{X} = \mathbf{D}^{t-1} \mathbf{Z} + \mathbf{E} \end{aligned} \quad (3)$$

式中, 参数 β 和 γ 分别表示各项的关注因子 (important factor), 它们的值越大, 表示相关项越受关注。在本文中, $\beta = 2$, $\gamma = 2$ 。式(3)中分别用核范数(nuclear norm) $\|\cdot\|_*$ 和L1范数 $\|\cdot\|_1$ 求解低秩和稀疏问题。虽然式(3)在表达方式上与文献[8]中的公式相似, 但由于应用领域的不同, 文献[8]和本文对模板集和样本集的采样方式均不同。文献[8]在目标周围进行密实采样, 而本文是将每个超级像素作为数据点进行采样; 文献[14]是对静态图像进行分割, 无需考虑分割目标的时间相关性; 文献[13]只是文献[14]的一种延拓, 虽然它是对视频图像的分割, 但在进行低秩稀疏表达时, 也没有考虑目标的时间相关性。

2.3 式(3)的求解

根据文献[15], 添加约束条件使变量独立, 将式(3)转化为(为了表达清晰, 以下将 \mathbf{D}^{t-1} 简化成 \mathbf{D}):

$$\begin{aligned} \arg \min_{\mathbf{Z}, \mathbf{E}} (\|\mathbf{Z}\|_* + \beta \|\mathbf{W}\|_1 + \gamma \|\mathbf{E}\|_1) \\ \text{s.t. } \mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{E}, \mathbf{Z} = \mathbf{W} \end{aligned} \quad (4)$$

利用增广拉格朗日乘数法(augmented lagrange multiplier)得到目标函数为:

$$\begin{aligned} L(\mathbf{Z}, \mathbf{W}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) = & \|\mathbf{Z}\|_* + \beta \|\mathbf{W}\|_1 + \gamma \|\mathbf{E}\|_1 + \\ & p(\mathbf{Z}, \mathbf{W}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) - q(\mathbf{Y}_1, \mathbf{Y}_2) \\ p(\mathbf{Z}, \mathbf{W}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) = & \frac{\mu}{2} \left(\left\| \mathbf{X} - \mathbf{D}\mathbf{Z} - \mathbf{E} + \frac{\mathbf{Y}_1}{\mu} \right\|_{\mathbb{F}}^2 + \right. \\ & \left. \left\| \mathbf{Z} - \mathbf{W} + \frac{\mathbf{Y}_2}{\mu} \right\|_{\mathbb{F}}^2 \right) \\ q(\mathbf{Y}_1, \mathbf{Y}_2) = & \frac{1}{2\mu} (\|\mathbf{Y}_1\|_{\mathbb{F}}^2 + \|\mathbf{Y}_2\|_{\mathbb{F}}^2) \end{aligned} \quad (5)$$

分别对各参数进行迭代优化, 则参数的迭代过程为:

$$\begin{aligned} \mathbf{Z}_{k+1} = & J_{(\eta\mu_k)^{-1}} \left(\mathbf{Z}_k + \frac{Q(\mathbf{D}, \mathbf{X}, \mathbf{Z}_k, \mathbf{Y}_{1,k}, \mathbf{Y}_{2,k}, \mu_k)}{\eta} \right) \\ Q(\mathbf{D}, \mathbf{X}, \mathbf{Z}_k, \mathbf{Y}_{1,k}, \mathbf{Y}_{2,k}, \mu_k) = & \mathbf{D}^H \left(\mathbf{X} - \mathbf{D}\mathbf{Z}_k - \mathbf{E}_k + \frac{\mathbf{Y}_{1,k}}{\mu_k} \right) - \\ & \left(\mathbf{Z}_k - \mathbf{W}_k + \frac{\mathbf{Y}_{2,k}}{\mu_k} \right) \\ \mathbf{W}_{k+1} = & \Theta_{\beta\mu_k^{-1}} \left(\mathbf{Z}_{k+1} + \frac{\mathbf{Y}_{2,k}}{\mu_k} \right) \\ \mathbf{E}_{k+1} = & \Theta_{\gamma\mu_k^{-1}} \left(\mathbf{X} - \mathbf{D}\mathbf{Z}_{k+1} + \frac{\mathbf{Y}_{1,k}}{\mu_k} \right) \\ \mathbf{Y}_{1,k+1} = & \mathbf{Y}_{1,k} + \mu_k (\mathbf{X} - \mathbf{D}\mathbf{Z}_{k+1} - \mathbf{E}_{k+1}) \\ \mathbf{Y}_{2,k+1} = & \mathbf{Y}_{2,k} + \mu_k (\mathbf{Z}_{k+1} - \mathbf{W}_{k+1}) \\ \mu_{k+1} = & \rho\mu_k \end{aligned} \quad (6)$$

式中, 函数 J 和 Θ 分别表示奇异值压缩算子^[8] (singular value soft-thresholding operator)和压缩算子^[8] (soft-thresholding operator); $\eta = \|\mathbf{D}\|_2^2$; $\rho = 1.1$; 各参量的初始值为 $\mathbf{Z}_0 = \mathbf{W}_0 = \mathbf{E}_0 = \mathbf{Y}_{1,0} = \mathbf{Y}_{2,0} = \mathbf{0}$ 。

3 目标分割

本文所提出的矩阵低秩稀疏表达是将超级像素看为一个数据点, 这是一种基于区域的计算方法。要想获得基于像素的分割结果, 本文将图像分割作为图分割问题, 利用不同的信息为图像中的每个像素点建立不同的概率分布, 并将此作为能量函数的线索, 通过能量函数最小化的框架, 利用最大流最小割(max flow/min cut)定理进行求解, 达到目标分割的目的, 提高分割精度。

3.1 能量函数框架

给定一幅图像 I , 分别定义 $\{I_i\}$ 和 $\{l_i\}$ 为图像中的像素集和所对应的标签。当任一像素 I_i 属于目标时, 则 $l_i = 1$, 反之 $l_i = 0$ 。最佳的分割结果为最小化能量函数:

$$\begin{aligned} \arg \min_L E(L) = \arg \min_L \left(\sum_{i \in I} \sum_k \lambda_k \Phi_k(l_i, I_i) + \right. \\ \left. \sum_{(i,j) \in \Omega} \Psi(l_i, l_j, I_i, I_j) \right) \end{aligned} \quad (7)$$

式中, $E(L)$ 为能量函数; Φ_k 表示不同的线索函数; $\lambda_k > 0$ 为相对应的关注因子, 并且 $\sum_k \lambda_k = 1$ 。线索函数评估了在不考虑其他像素的情况下, 每个像素 i 属于 l_i 的概率。定义 $\Phi_k(l_i, I) = \log(p(l_i | f_k(I_i)))$, 其中函数 $f_k(\cdot)$ 将输入图像映射到不同的特征空间。本文采用两个特征空间, 分别为目标表观特征(object

appearance cue)和目标显著性特征(object salient cue)。函数 Ψ 为相关能量函数,表示相邻两个元素的相关程度, Ω 选取其4邻域范围。

根据文献[6],相关能量函数定义为:

$$\Psi(l_i, l_j, I_i, I_j) = \delta(l_i \neq l_j) \frac{\varepsilon + e^{-\xi \|l_i - l_j\|^2}}{1 + \varepsilon} \quad (8)$$

式中, $\varepsilon = 1$; $\xi = \frac{1}{2 \langle \|I_i - I_j\|^2 \rangle}$; $\langle \cdot \rangle$ 为期望算子;函数 $\delta(\cdot)$ 为狄拉克 δ 函数。

3.2 目标表现特征

基于目标与背景的直方图为图像中每个像素建立颜色的概率分布。本文根据上一帧分割的结果,在YUV空间中分别建立目标与背景的颜色直方图,并且将此直方图通过高斯滤波器进行平滑。

3.3 目标显著性特征

根据式(3)求得的表达子 \mathbf{Z} 为图像中每个超级像素建立概率模型,使得待分割图像中的目标能够具有较大的概率。

通过式(2)可知,表达子 \mathbf{Z} 表明了当前图像的所有样本集与模板集的相似程度。矩阵 \mathbf{Z} 中的任意元素 z_{ij} 越大,表明模板集中的第 i 个元素(矩阵 \mathbf{D}^{t-1} 中的第 i 列)与样本集中的第 j 个元素(矩阵 \mathbf{X} 中的第 j 列)越相似。从图1可以看出,对于矩阵 \mathbf{Z} 中的任意一列 \mathbf{z}_i ,

可以分解成 $\mathbf{z}_i = \begin{pmatrix} \mathbf{z}_i^O \\ \mathbf{z}_i^B \end{pmatrix}$,其中 \mathbf{z}_i^O 表示矩阵 \mathbf{X} 中第 i 列与

矩阵 \mathbf{D}^O 中各个元素的相似程度, \mathbf{z}_i^B 表示矩阵 \mathbf{X} 中第 i 列与矩阵 \mathbf{D}^B 中各个元素的相似程度。根据以上分析,在当前帧中,定义元素 i 属于目标的概率为:

$$T_i = \exp \left(- \frac{\max(|\mathbf{z}_i^B|)}{\max(|\mathbf{z}_i^O|)} \right) \quad (9)$$

式中, $|\cdot|$ 表示将向量中的所有元素取绝对值,函数 $\max(\cdot)$ 返回向量中最大的一个数。为了获得基于像素的概率分布,本文在进行计算时,超级像素 S_i 中所包含的所有像素值均为 T_i 。

3.4 能量函数最小化求解与图像分割

通过目标表现特征和目标显著性特征可以将图像分别映射到两个不同的特征空间,分别构建两个线索函数,并结合式(8)创建能量函数。根据文献[16],式(7)可以通过图分割的方式进行优化求解,达到目标分割的目的。为了提高分割精度,本文采用了文献[17]的方法,利用形态学的理论,在每次分割的基础上,对分割结果进行膨胀与腐蚀操作,建立新的

粗略划分(trimap),利用GrabCut^[18]的方法对图像进行多次分割,提高图像的分割精度。

4 实验分析

本文共设计了两个实验。第一个实验验证了将矩阵的低秩稀疏表达用于建立目标的显著性特征的可能性,可以将目标显著地表达在待分割图像中。第二个实验将本文算法与现有的一些算法进行比较,并将比较结果作了详细的分析。

4.1 矩阵的低秩稀疏表达

矩阵的低秩稀疏表达是本文算法的关键步骤。本文实验采用GT-SegTrack^[3]数据库中的视频图像作为实验数据,显示了矩阵的低秩稀疏表达的有效性。图2为4个视频(birdfall, girl, monkey, parachute)中的两帧图像。其中左边4个图像分别为上一帧中已经分割好的目标,将其分别作为模板按照前面介绍的方法进行矩阵的低秩稀疏表达。中间的4个图像分别为当前帧所计算出的目标显著性概率图(salient map)。从中间列可以看出,此方法能够较好地突出目标的形状和位置,并且去除了大部分的背景噪声,为后续的图像分割打下了坚实的基础。图2的右列为计算矩阵的低秩稀疏表达时的收敛情况。其横坐标为迭代次数,纵坐标为计算的误差值,表示为 $\text{error} = \frac{\|\mathbf{X} - \mathbf{D}^{t-1}\mathbf{Z} - \mathbf{E}\|_F}{\|\mathbf{X}\|_F}$, $\|\cdot\|_F$ 表示矩阵的Frobenius范数。

从图2的右列可以看出迭代过程是收敛的。在实验中可以发现,矩阵的低秩稀疏表达与模板样本的个数相关。当超级像素的个数较大时,模板样本的个数较多,无法精确地表达目标的局部特征,待分割图像的样本集与模板样本中线性相关的元素较多;当超级像素的个数较少时,目标模板中包含了较多的背景噪声,因此,这两种情况都会使得显著性特征不理想,影响分割的精度。在本文的所有实验中,超级像素的最大个数为200个。

对于整个算法而言,矩阵的低秩稀疏优化的计算占据了绝大部分的计算时间。然而,对于矩阵 \mathbf{Z} 的SVD分解是优化计算的时间瓶颈。由于矩阵 \mathbf{Z} 为低秩矩阵,因此其SVD分解的时间复杂度为 $O(MNr)$,其中, r 为矩阵 \mathbf{Z} 的秩,并且满足 $r \leq \sqrt{\min(M, N)}$,则计算一帧图像显著性特征的时间复杂度为 $O(MNd_s)$,这说明它与采样样本的个数、模板集样本个数和特征维数均保持线性关系。图3为矩阵的低秩稀疏运算时间与采样样本个数 N 的关系,其中在计

算式(3)时的迭代次数为200。

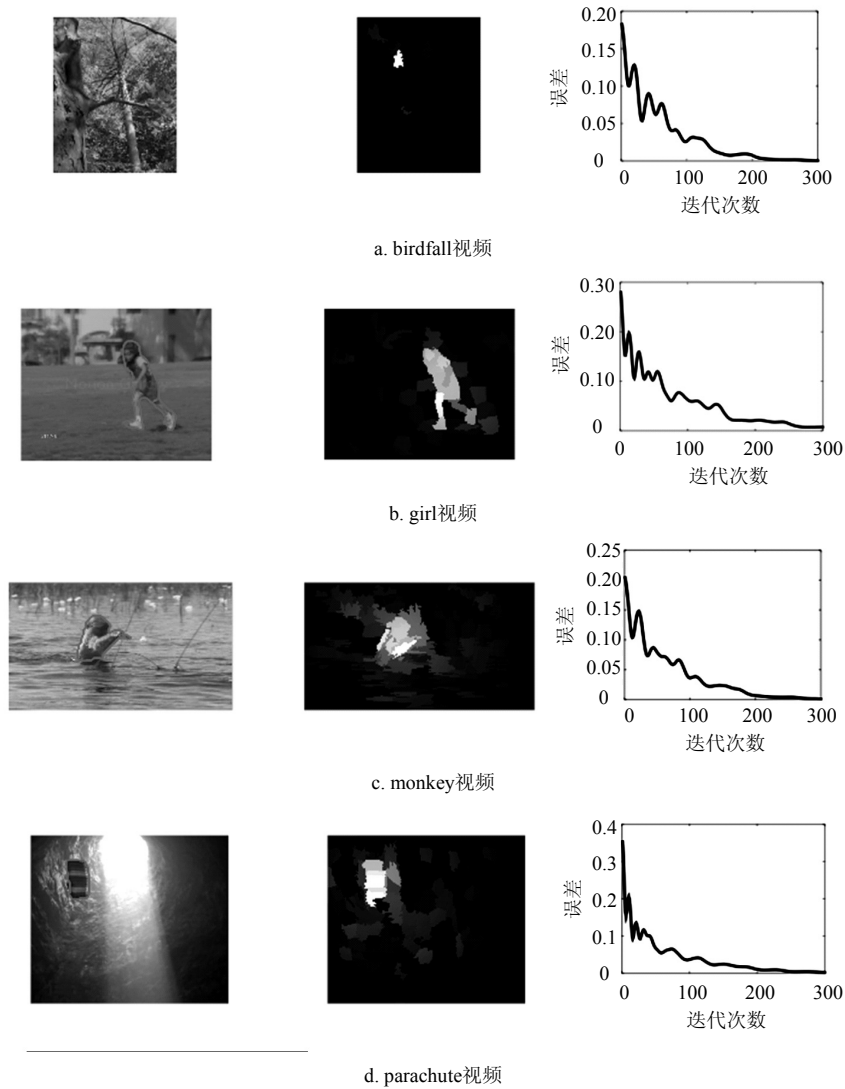


图2 矩阵的低秩稀疏表达和显著性特征

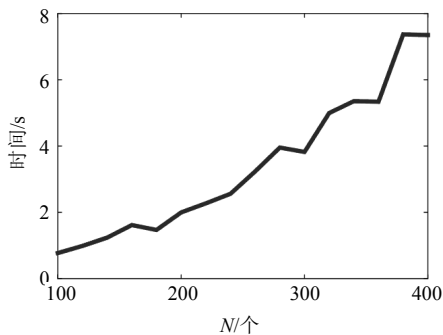


图3 矩阵的低秩稀疏运算时间与采样样本个数的关系

4.2 对比实验

本实验采用GT-SegTrack^[3]数据库中的视频图像作为实验数据, 将文献[2](在线算法)和文献[3](离线算法)所提出的方法与本算法进行比较。比较结果如表1所示。

表1 跟踪精度对比

视频	文献[2]	文献[3]	$\lambda_c = 0.5$	$\lambda_c = 0.2$	目标大小	帧数
parachute	502	235	206	207	3 695	50
girl	1 755	1 304	1 280	1 680	8 201	20
monkey	4 142	2 814	2 790	2 805	6 054	30
birdfall	454	252	—	319	495	29

其中将每个视频中的平均误差作为分割精度的

度量单位, 表示为 $\frac{\sum_{i=1}^L e_i}{L}$, L 为视频图像的个数, e_i 为一帧图像中分割错误的像素个数。表1中的目标大小为每个视频中目标所包含像素个数的平均值。

为了体现目标表观特征和显著性特征在能量最小化框架下不同的关注程度, 本实验分别列举了两个不同关注系数情况下的分割精度, 其中 λ_c 为目

标表观特征系数, 目标显著性特征系数 $\lambda_s = 1 - \lambda_c$ 。可以看出:

1) 由于在视频parachute中目标形状变化不大, 在视频monkey中背景较为单一, 因此不同的关注系数对分割结果影响不大, 并且其分割误差均小于其他两种方法。

2) 对于视频girl, 当 $\lambda_c = 0.5$ 时本算法的分割精度略优于文献[3]中的算法。但由于背景比较复杂, 目标形状变化较大, 矩阵的低秩稀疏表达会造成某些局部信息丢失, 如图2中, 女孩腿部的显著性不高。当过多地关注显著性特征 ($\lambda_c = 0.2$) 时, 会造成部分目标丢失。并且, 由于本算法的模板更新仅仅基于上一帧的分割结果, 虽然在整个视频中, 目标的大部分信息可以被较好地分割(比如女孩的身躯), 但部分信息丢失(如女孩的头、腿和头发), 造成分割精度降低。

3) 在视频birdfall中, 由于目标较小, 背景复杂, 过多的关注表观特征 ($\lambda_c = 0.5$ 时) 会造成目标的漂移(drift)。但目标的显著性特征可以很好地进行目标分割。



图4 分割效果示例

通过以上分析可知, 虽然矩阵的低秩稀疏表达可以在大部分视频中较好地分割目标, 但仅仅使用一种特征无法精确地分割图像, 因此利用能量最小框架将多个特征进行融合, 可以提高目标的分割精

度。图4显示了本文中所涉及视频的分割效果。

5 结 束 语

本文通过矩阵的低秩稀疏表达提出了一种在线的目标分割算法。将已知目标图像进行过分割, 提取每个超级像素的L2ECM特征, 建立模板集。对待分割图像进行相对于模板集的线性表达; 利用视频图像的连续性和噪声的稀疏性, 对线性表达方程添加低秩和稀疏约束, 将目标的显著性问题转化为方程的低秩稀疏优化问题; 进而通过能量最小化框架, 将图像的多种信息融合, 达到像素级目标跟踪与分割的目的。在实验部分本文将该算法与已有的目标分割算法进行比较, 提高了在线算法的分割精度。

但从本文的分析中可知, 超级像素的大小会影响目标的显著性特征; 并且能量最小化框架中不同特征的关注系数对于不同的图像环境影响较大。如何实现这些参数的自适应算法将是未来工作的重点。

参 考 文 献

- [1] NING J, ZHANG L, ZHANG D, et al. Joint registration and active contour segmentation for object tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(9): 1589-1597.
- [2] CHOCKALINGAM P, PRADEEP N, BIRCHFIELD S. Adaptive fragments-based tracking of non-rigid objects using level sets[C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 1530-1537.
- [3] TSAI D, FLAGG M, NAKAZAWA A, et al. Motion coherent tracking using multi-label MRF optimization[J]. International Journal of Computer Vision, 2012, 100(2): 190-202.
- [4] CRIMINISI A, CROSS G, BLAKE A, et al. Bilayer segmentation of live video[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2006, 1: 53-60.
- [5] WANG L, GONG M, ZHANG C, et al. Automatic real-time video matting using time-of-flight camera and multichannel poisson equations[J]. International Journal of Computer Vision, 2012, 97(1): 104-121.
- [6] YIN Z, COLLINS R T. Online figure-ground segmentation with edge pixel classification[C]//BMVC. Leeds: IEEE, 2008: 1-10.
- [7] BAO C, WU Y, LING H, et al. Real time robust l1 tracker using accelerated proximal gradient approach[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence: IEEE, 2012: 1830-1837.
- [8] ZHANG T, GHANEM B, LIU S, et al. Low-rank sparse learning for robust visual tracking[M]//Computer Vision-ECCV 2012. Berlin Heidelberg, Springer, 2012: 470-484.

(下转第406页)