

基于 $l_{1/2}$ 正则化的稀疏表示目标跟踪算法的研究

贵敏, 邓萍, 王保云

(南京邮电大学自动化学院, 江苏南京 210023)

摘要:近年来目标的稀疏表示已经广泛应用到视频跟踪中。文中提出了一种基于局部稀疏表示的鲁棒目标跟踪算法, 目标的表示将局部信息考虑在内, 并且做出了遮挡处理。为了在新的帧中跟踪到目标, 每一个候选目标通过在线构建的过完备字典以及模板解 $l_{1/2}$ 最小化问题稀疏表示。文中用 $l_{1/2}$ 规范最小化来代替 l_0 , 而不是用 l_1 规范最小化, 通过解 $l_{1/2}$ 最小化问题, 可以找到比解 l_1 最小化更稀疏、更精确的解。此外, $l_{1/2}$ 比 l_0 更容易求解。目标稀疏表示后, 通过在线学习的分类器将目标区分出来。实验结果表明, 与现有的一些算法相比, 该算法稳定性好, 性能更优越。

关键词:视频跟踪; 稀疏表示; 过完备字典; $l_{1/2}$ 最小化; 分类器

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2015)01-0082-05

doi: 10.3969/j.issn.1673-629X.2015.01.019

Research on Object Tracking Algorithm of Sparse Representation Based on $l_{1/2}$ Normalization

BEN Min, DENG Ping, WANG Bao-yun

(College of Automation, Nanjing University of Posts and Telecommunications,
Nanjing 210023, China)

Abstract: Recently sparse representation has been widely used in video tracking. In this paper, propose a robust target tracking method based on local sparse representation, considering the local information for object representation and take occlusion into account. In order to track the target in a new frame, each target candidate is sparsely represented by over-complete dictionary online constructed and target templates solving a $l_{1/2}$ -norm minimization problem. In this algorithm, use $l_{1/2}$ -norm minimization to replace l_0 -norm minimization instead of l_1 -norm minimization. By solving $l_{1/2}$ -norm minimization, can find a sparser and more accurate solution than l_1 -norm minimization, moreover, it is much easier to be solved than l_0 -norm minimization. After that, a classifier is learned to distinguish the target from the background. Experimental results show that this method has good stability and the performance is superior to the current algorithms.

Key words: video tracking; sparse representation; over-complete dictionary; $l_{1/2}$ -norm minimization; classifier

0 引言

目标跟踪是计算机视觉领域中一个重要的研究课题, 并且已经广泛应用到很多实际问题中, 如视频监控、人机交互以及车辆检测等。近年来, 目标跟踪算法已经取得了一定的进步, 由于在视频跟踪中, 光照、外观形状和摄像机会发生变化, 有时会出现遮挡等情况^[1]; 因此, 建立一个鲁棒的目标跟踪算法仍具有很强的挑战性, 许多跟踪算法被提出, 用来解决上述问题。

文中提出了一种鲁棒的跟踪算法, 即将跟踪问题视为在模板子空间寻找稀疏近似解。受文献[2]的启发, 用琐碎模板处理遮挡, 每一个琐碎模板只有一个非

零元素, 在跟踪中, 每一个候选目标则是目标模板和琐碎模板的线性结合。一个好的候选目标可以由目标模板有效的表示, 这就使得系数矩阵稀疏, 也就是说对应于琐碎模板的系数即琐碎系数趋于零。当环境中存在遮挡、噪声等不利的因素时, 某些琐碎系数将不再为零, 但整体的系数仍然保持其稀疏性, 系数的稀疏表示是由解 $l_{1/2}$ 规范最小化问题得到的。为了稀疏表示目标, 首先手动确定第一帧目标的位置, 将目标划分成几个小块, 对这些小块进行表示, 每一个小块的系数的稀疏解通过过完备字典解 $l_{1/2}$ 最小化问题得到, 然后直接连接这些小块的稀疏解来表示目标。而后在目标位置

收稿日期: 2014-02-17

修回日期: 2014-05-23

网络出版时间: 2014-11-17

基金项目: 国家自然科学基金资助项目(61271232)

作者简介: 贵敏(1989-), 女, 硕士, 研究方向为图像处理与模式分类; 王保云, 教授, 研究方向为信号信息处理、模式识别与智能系统。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20141117.2208.036.html>

周围提取正、负样本,根据上述的目标表示方法,将这些样本也进行稀疏表示,利用这些数据来学习一个线性分类器将目标从背景中分离出来,这样就可以通过训练过的分类器选出最合适的候选目标。在更新过完备字典和分类器的时候,为了不使跟踪漂移,文中提出一种双重保险跟踪算法,即在判断最好的候选目标时,既使用第一帧所确定的过完备字典和分类器,又使用最近一帧的字典和分类器。文中在很多有挑战性的视频中(严重遮挡、光照变化、姿态变化等)测试上述所提算法的性能,并且与前人提出的算法(增量视频跟踪^[3]、L1跟踪^[4]、多实例学习^[5])性能相比较,结果表明文中所提的算法的性能更好。

1 相关工作

近年来,稀疏表示已经应用于各个视觉领域,比如图像增强^[6]、目标识别、视频跟踪^[7]等。文献[4]中,作者将稀疏表示运用到视频跟踪中,并且利用琐碎模板来处理遮挡问题,尽管已取得一定的成功,但还有一些问题需要解决。琐碎模板可以用来为图像的任何区域建模,包括背景区域,因此图像的遮挡区域和背景的重构误差都会比较小,那么,如果将具有最小重构误差的样本视为跟踪结果,会导致误差的积累,最后使得跟踪失败。由于该算法并未将背景考虑在内,所以不能很好地在复杂的环境中进行目标跟踪。Wang等人^[8]利用前景和背景的分差比提取有效的特征来进行目标追踪。文献[9]提出了一个整体跟踪方法,把跟踪看作二分类问题,在线训练作为整体的多个弱分类器用来区分目标和背景,使用AdaBoost把作为整体的多个弱分类器合并为一个强的分类器,该分类器用于下一帧的分类,区分像素属于目标还是背景,并得出置信图,利用mean shift算法找出目标的位置。在跟踪过程中通过在线训练新的弱分类器从而在连贯时间上保持更新弱分类器这一整体。在线boosting算法也被用来提取有效特征,并应用到各个领域^[10-11]。

前人提出了很多方法来解决视频跟踪中的漂移问题。Grabner等人将视频跟踪结果中的信息视为未标记的数据,并且在半监督学习的框架内学习一个分类器。Babenko等人使用多实例学习来处理模糊标记的正负样本,减缓视频漂移。Kalal等人^[12]也将分类器的分类结果视为未标记的,并且利用底层结构选择正负样本进行更新。

2 基于 $l_{1/2}$ 正则化稀疏表示的目标跟踪

2.1 基于 $l_{1/2}$ 的目标稀疏表示

稀疏表示已经被广泛应用在各个领域,为了得到一个有效的目标表示,利用一个过完备字典对目标区

域的局部块进行稀疏求解,目标图像块相对应的稀疏编码表示目标,文中直接将这稀疏编码集合起来。尽管已经有很多有效的学习过完备字典的方法被提出,但仅仅从第一帧中获得充足的训练数据是很困难的,如文献[4],在建立过完备字典时受到目标周围干扰信息的影响,这必然将误差引入到训练数据中,最终对字典产生影响。文中将使用不同的方法来构建过完备字典。在第一个视频框架中给定一幅图像,在目标区域内通过固定一个滑动的窗口,获得一个图像块集合 $T = [t_1, t_2, \dots, t_n] \in R^{w \times n}$ 。其中, $t_i \in R^w$ 表示的是第*i*个标准向量化的图像块, w 表示图像块特征维数, n 表示图像块的个数。因此,定义过完备字典如下:

$$D = [T, E] \quad (1)$$

其中, $E = [I, -I] \in R^{w \times 2w}$, $I \in R^{w \times w}$ 是单位矩阵。跟文献[4]相似,矩阵 E 为琐碎模板,它的引入是为了处理噪声,并且保证了目标重构稀疏系数的非负性。利用此方法,在构造过完备字典时就不会引入误差或者噪声。

假设 $X = [X_1, X_2, \dots, X_N]$ 表示从目标图像周围提取的向量化的正负样本图像块, N 表示图像块数,通过过完备字典 D ,每一个图像块 X_i 都有一个稀疏系数 $\alpha_i \in R^{n+2w}$ 与之相对应。文中稀疏系数的求解与以往的稀疏求解不同,受文献[13]的启发,用 $l_{1/2}$ 规范最小化来代替 l_0 ,而不是用 l_1 规范最小化,通过解 $l_{1/2}$ 最小化问题,可以找到比解 l_1 最小化更稀疏、更精确的解。此外, $l_{1/2}$ 比 l_0 更容易求解^[14]。计算公式定义如下:

$$\min_{\alpha} \|X_i - D\alpha_i\|^2 + \lambda \|\alpha_i\|_{l_{1/2}} \quad (2)$$

当将所有正负样本图像块所对应的稀疏编码计算出来以后,直接把这些编码集合起来,即 $O = [\alpha_1, \alpha_2, \dots, \alpha_N]$,该稀疏系数矩阵将作为训练数据来训练分类器。

2.2 基于稀疏表示的分类器学习

文中将视频跟踪视为一个二元分类问题,通过线性分类器将前景目标从背景中分离出来。基于上述的目标稀疏表示,前景目标和背景的图像块由完备字典中不同的集合组成,前景目标图像块可以通过上述的目标基 T 实现重构,而背景图像块的重构则需要用到琐碎模板,因此利用提出的稀疏目标表示可以很容易地将前景目标从背景中分离出来。文中利用一个线性分类器来进行目标跟踪,并且能够获得较好的跟踪效果。下面介绍一下文中分类器的学习过程。

初始帧目标的位置由手动确定,在初始目标位置的周围提取正负样本,并且将正负样本都归一化为与手动确定的目标大小相同。样本的选取规则为高斯选取,即靠近目标的周围多选取一些,远离目标的少选或

者不选,并且正样本的选取是在一个圆形内,负样本的选取在一个圆弧内进行。具体描述如下:假设第一帧目标的位置确定为 $L_1 = (x_1, y_1)$,那么正样本的选取范围为 $\|L_p - L_1\| < \theta$,负样本的选取范围为 $\theta < \|L_n - L_1\| < \beta$ 。其中, L_p, L_n 分别代表正负样本的位置, θ, β 则确定圆和圆弧的区域。在提取正负样本后,对样本提取图像块,然后计算样本所对应的稀疏编码,如式(2)所示,计算出的稀疏编码则作为分类器的训练数据 O ,另外还加入了样本标签,来区分正负样本,即 $\{o_i, b_i\}_{i=1}^M$ 。其中, $o_i \in R^{n+2w}$, $b_i \in \{+1, -1\}$ 是样本标签, M 表示训练样本的数量。当有了这些训练数据后,就可以通过学习得到一个分类器参数,定义如下:

$$J(W) = \frac{1}{M} \sum_{i=1}^M s(b_i, W, o_i) + \frac{\lambda}{2} \|W\|_2^2 \quad (3)$$

其中, W 代表分类器参数; $s(\cdot)$ 表示损失函数,定义如下:

$$s(\cdot) = \log(1 + e^{-bW'o}) \quad (4)$$

其中, $o' = [o^T, 1]^T$ 为增广向量。根据上述的分类器参数,得到分类器

$$h(o) = \frac{1}{1 + e^{-W'o}} \quad (5)$$

当损失函数(4)越小,那么 $1 + e^{-W'o}$ 也就越小,那么式(5)的值也就越大。一个样本具有更大分类值,说明该样本源于目标的可能性就越大,把带有最大分类值的样本作为当前帧中的跟踪结果。

2.3 基于双重保险的目标跟踪

基于文献[15],在进行字典和分类器的更新时,提出了一种双重保险的跟踪算法。为了获得第 t 帧的跟踪结果,首先利用由 $t-1$ 帧的跟踪结果获得的过完备字典以及分类器来初步确定第 t 帧的跟踪结果,再利用第一帧的字典和分类器来对初步确定的结果进行修正,最终确定跟踪结果。具体算法如下:

输入:手动确定初始帧位置,视频帧数 F_1, F_2, \dots, F_T 。

输出:第 t 帧目标的状态 e_t 。

(1) for $t = 1, 2, \dots, T$, do

(2) if $t = 1$

对初始帧构建初始过完备字典 D_1 以及初始分类器参数 W_1 。

(3) else

a. 利用上一帧的跟踪结果 e_{t-1} , 以及 D_{t-1} 和 W_{t-1} 初步估算出此时的目标状态 e_t 。

b. 利用第一帧的信息 D_1 以及 W_1 对 e_t 进行修正,得到准确的跟踪结果 e_t , 并且在当前帧中框出目标的

位置。

c. 根据得到的目标状态 e_t , 重新学习过完备字典 D_t 以及分类器参数 W_t , 作为下一帧的训练数据。

(4) end if

(5) end for

通过文中的跟踪算法,可以对感兴趣的目标进行有效的跟踪。

3 实验结果

将所提出的算法在一些有挑战性的视频序列中进行实验,并且将文中的算法与前人所提出的算法(增量视频跟踪 (IVT), 多实例学习跟踪 (MILT), 基于 LI 的跟踪 (LIT)) 进行性能比较。文中用到的视频序列见表 1。

表 1 实验中用到的视频序列

视频序列	主要的挑战
David	光照变化,姿态变化,部分遮挡
Car4	光照变化,其余物体的影响
Jumping	图像模糊,快速移动
Faceocc	部分遮挡,姿态变化
Girl	严重遮挡,快速移动

3.1 定性分析

在 David 视频序列中,刚开始的环境比较昏暗,刚开始几十帧所有的算法都性能相当,随后 LIT 和 MIL 相继出现漂移,IVT 在其中某些帧也出现漂移,但后来又恢复跟踪,文中算法性能良好。Jumping 视频中,由于目标的快速移动产生图像模糊,IVT、MIL 以及文中提出算法的跟踪效果理想,LIT 方法则有较大的误差。Girl 视频中存在严重遮挡和快速移动等因素,IVT 跟踪失败,LI 误差较大,MIL 误差较小,文中的算法一直持续有效的跟踪。在 Car4 视频序列中,MIL 算法在第一次光照变化时就开始漂移,LIT 算法稍有漂移,误差不大,其余两个算法性能较好。Faceocc 中,文中算法和 IVT 跟踪效果最好,LIT 和 MIL 算法稍有漂移,但能够大致跟踪出目标位置。四种算法的跟踪结果对比如图 1 所示。图中,文中的算法用实线框表示,IVT 用点横线表示,LIT 用虚线表示,MILT 用点线进行表示。由图 1 可以看到,文中所提算法一直能保持准确的跟踪,其余算法都会产生漂移或者跟踪失败,因此文中算法性能优于其他三种算法。

3.2 定量分析

定量分析时,计算以上四种算法的精确度,将中心位置误差作为精确度准则,中心位置误差定义为跟踪目标的中心位置与手动标记的初始目标中心位置之间的误差,利用欧氏距离进行对比,距离越大说明离中心

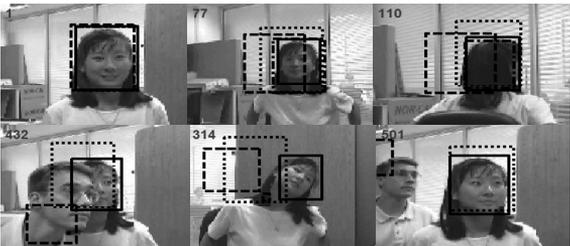
目标位置越远,即越偏离目标。四种算法在上述五个视频序列中的中心位置误差比较结果如图 2 所示。由图 2 可以直观地看出,文中算法与其他三种算法相比中心位置误差小,可见文中算法在稳定性和健壮性方面优于其他三种算法。



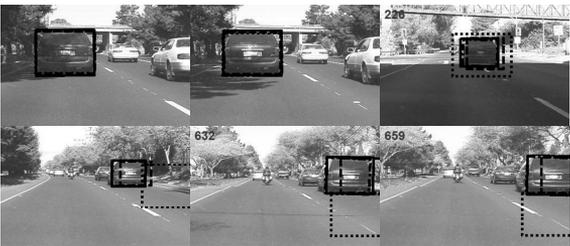
(a)David 跟踪对比图



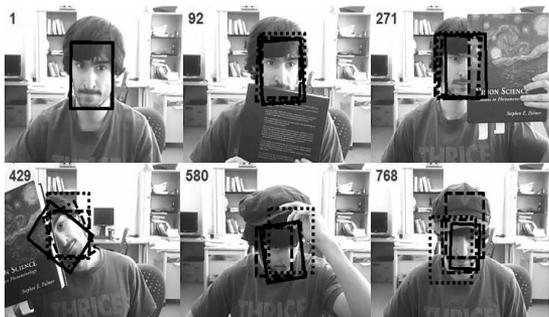
(b)Jumping 跟踪对比图



(c)Girl 跟踪对比图

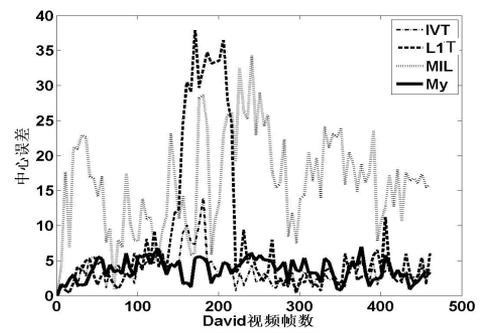


(d)Car4 跟踪对比图

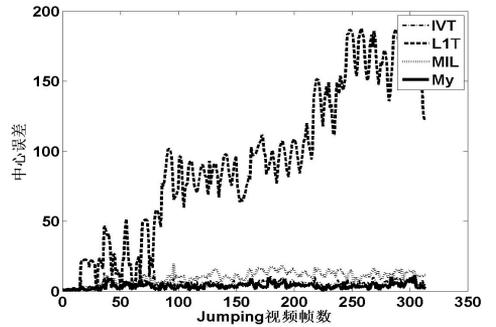


(e)Faceocc 跟踪对比图

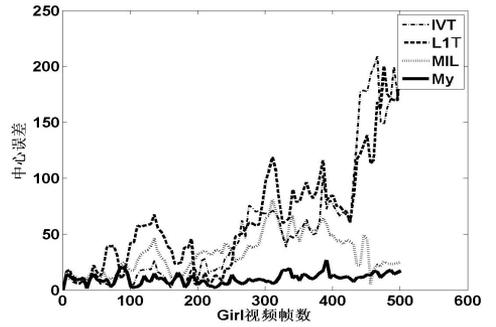
图 1 跟踪结果对比图



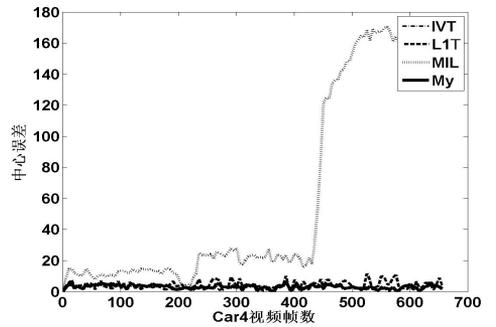
(a)David 中心位置误差图



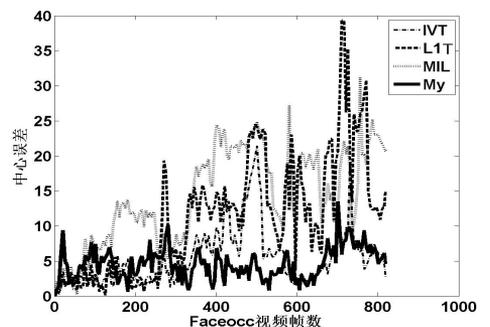
(b)Jumping 中心位置误差图



(c)Girl 中心位置误差图



(d)Car4 中心位置误差图



(e)Faceocc 中心位置误差图

图 2 中心位置误差图

4 结束语

文中提出了一种基于稀疏表示和分类器学习的在线跟踪算法,利用过完备字典对目标局部块进行稀疏编码,然后利用稀疏系数学习一个线性分类器,将目标从背景中分离出来。在进行字典和分类器的更新时,提出了一种双重保险的跟踪算法从而减缓跟踪漂移。上述实验结果证明了文中算法的健壮性和有效性。未来的研究重点将在稀疏表示模型的基础上,解决更为复杂场景中的目标跟踪问题。

参考文献:

- [1] Yilmaz A, Javed O, Shah M. Object tracking: a survey[J]. ACM Computing Surveys, 2006, 38(4): 1-45.
- [2] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.
- [3] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1-3): 125-141.
- [4] Xue Mei, Ling Haibin. Robust visual tracking using l_1 minimization[C]//Proc of international conference on computer vision. Kyoto, Japan; IEEE, 2009: 1436-1443.
- [5] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning[C]//Proc of computer vision and pattern recognition. [s. l.]: [s. n.], 2009: 983-990.
- [6] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
- [7] Liu Baiyang, Huang Junzhou, Yang Lin, et al. Robust tracking using local sparse appearance model and k-selection[C]//Proc of 24th IEEE conference on computer vision and pattern recognition. Colorado; IEEE, 2011: 1313-1320.
- [8] Wang Xiaoyu, Hua Gang, Han T X. Discriminative tracking by metric learning[C]//Proc of ECCV. Heraklion: [s. n.], 2010: 200-214.
- [9] Avidan S. Ensemble tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261-271.
- [10] Qi Z, Xu Y, Wang L. Online multiple instance boosting for object detection[J]. Neurocomputing, 2011, 74(10): 1769-1775.
- [11] Chang W C, Cho C W. Online boosting for vehicle detection[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(3): 892-902.
- [12] Kalal Z, Matas J, Mikolajczyk K. P-N learning: bootstrapping binary classifiers by structural constraints[C]//Proc of computer vision and pattern recognition. [s. l.]: [s. n.], 2010: 49-56.
- [13] Guo S, Ruan Q, Gao Y, et al. Sparse representation via $l_1/2$ -norm minimization for facial expression recognition[C]//Proc of international conference on signal processing. [s. l.]: [s. n.], 2012: 1243-1246.
- [14] Xu Zongben, Zhang Hai, Wang Yao, et al. $l_1/2$ regularization[J]. Information Sciences, 2010, 53(6): 1159-1169.
- [15] Matthews L, Ishikawa T, Baker S. The template update problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6): 810-815.
-
- (上接第 81 页)
- 出版社, 2003.
- [6] 邹恩, 李祥飞, 张泰山. 混沌控制综述[J]. 株洲工学院学报, 2002, 16(4): 16-18.
- [7] 李卫东, 王秀岩. 混沌控制综述[J]. 自动化技术与应用, 2009, 28(1): 1-5.
- [8] Ott E, Grebogi C, Yorke J A. Controlling chaos[J]. Physical Review Letters, 1990, 64(11): 1196-1199.
- [9] Grebogi C, Lai Y C. Controlling chaotic dynamical systems[J]. Systems & Control Letters, 1997, 31(5): 307-312.
- [10] Ditto W L, Spano M L, Lindner J F. Techniques for the control of chaos[J]. Physica D: Nonlinear Phenomena, 1995, 86(1): 198-211.
- [11] Tian Yuping. Controlling chaos using invariant manifolds[J]. International Journal of Control, 1999, 72(3): 258-266.
- [12] Pyragas K. Continuous control of chaos by self-controlling feedback[J]. Physics Letters A, 1992, 170(6): 421-428.
- [13] Tian Yuping, Chen Guanrong. A separation principle for dynamical delayed output feedback control of chaos[J]. Physics Letters A, 2001, 284(1): 31-42.
- [14] Jiang Guoping, Chen Guanrong, Tang W K S. Stabilizing unstable equilibria of chaotic systems from a state observer approach[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2004, 51(6): 281-288.

基于 $l_{1/2}$ 正则化的稀疏表示目标跟踪算法的研究

作者: [贲敏](#), [邓萍](#), [王保云](#), [BEN Min](#), [DENG Ping](#), [WANG Bao-yun](#)
作者单位: [南京邮电大学 自动化学院, 江苏 南京, 210023](#)
刊名: [计算机技术与发展](#) 
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2015(1)

引用本文格式: [贲敏](#), [邓萍](#), [王保云](#), [BEN Min](#), [DENG Ping](#), [WANG Bao-yun](#) [基于 \$l_{1/2}\$ 正则化的稀疏表示目标跟踪算法的研究](#)[期刊论文]-[计算机技术与发展](#) 2015(1)