文章编号：1003-501X(2012)09-0065-07

视频监控中基于在线多核学习的目标再现识别

陈 方[1]，许允喜[1,2]

(1. 湖州师范学院 信息与工程学院，浙江 湖州 313000；
2. 浙江大学 信息与电子工程系，杭州 310027)

摘要：在非重叠多摄像机或单摄像机视频监控中，识别跟踪目标的再次出现很重要。针对传统支持向量机方法在特征融合方面的缺陷，本文提出了一种新的基于在线多核学习的人体目标再现识别方法。该方法对跟踪目标视频前景图像序列提取具有互补性的视觉单词树直方图和全局颜色直方图二种特征，再采用多核学习方法在线训练人体目标视觉外观，从而得到多核特征融合模型。实验结果表明，该方法能快速训练人体目标外观模型，满足视频监控的实时要求，多核融合模型获得了比单一特征模型和单核支持向量机方法更高的识别性能。

关键词：视频监控；多核学习；局部描述子；目标再现识别；单词树

中图分类号：TP391.41 **文献标志码**：A doi：10.3969/j.issn.1003-501X.2012.09.011

People Re-identification Based on Online Multiple Kernel Learning in Video Surveillance

CHEN Fang[1]，XU Yun-xi[1,2]

(1. *School of Information & Engineering*, *Huzhou Teachers College*, *Huzhou* 313000, *Zhejiang Province*, *China*;
2. *Department of Information Science & Electronic Engineering*, *Zhejiang University*, *Hangzhou* 310027, *China*)

Abstract: In the non-overlapping multi-camera or single camera video surveillance, re-identification of tracked target is very important. Due to weakness of traditional support vector machine in feature fusion, a new people re-identification method is proposed based on online multiple kernel learning. We extract complementary visual word tree histogram and global color histogram from tracked people foreground image sequence in video, and then multiple kernel learning method is used for online train people visual appearance. Finally, we obtain multiple kernel feature fusion model of people appearance. Experimental results show that our method can train people appearance model rapidly, meet the real-time requirement of video surveillance, and attain higher recognition performance than single feature appearance model and single kernel support vector machine method.

Key words: video surveillance; multiple kernel learning; local descriptor; people re-identification; vocabulary tree

0 引 言

由于智能视频监控在安全领域能发挥重要的作用，近年来受到越来越多研究者的关注。早期的研究主要集中在视频信息的低层处理，如目标检测、跟踪、背影去除等。最近，研究兴趣主要转向高层事件检测上，如行为分析、丢弃目标检测等。高层事件检测的一个主要任务是确定出现在不同时间段的目标对应问题(即目标再现识别)，由此可进一步提取活动场景的语义信息。人体目标再现识别两个最直接的应用分别为单摄像机智能视频监控中可疑目标的逗留徘徊检测；多摄像机视频监控中跨非重叠视域目标再识别(即持

收稿日期：2012-04-04；收到修改稿日期：2012-05-23
基金项目：国家自然科学基金项目(60872057)；浙江省自然科学基金项目(R1090244,Y1101237, Y1110944)；浙江省公益技术应用研究项目(2011C23132)；湖州市自然科学基金项目(2011YZ07)；湖州师范学院校级科研项目成果(KX24056)
作者简介：陈方(1987-)，女(汉族)，浙江诸暨人。讲师，硕士，主要研究工作是图像处理、计算机视觉、人工智能等。E-mail:cf@hutc.zj.cn。

续跟踪问题)。由于光照、视角、背景方面的变化，跨非重叠视域人体目标再现识别很具挑战性。对于没有精确空时限制情况下(如摄像机视角间存在大的间隙)，则只可以利用视觉外观特征[1-3]。

在单摄像机跟踪中，对于行人这样的非刚体来说，颜色是最鲁棒的视觉外观特征。但颜色特征受光照条件、摄像机参数等因素的影响很大，因此跨非重叠视域多摄像机跟踪中仅采用颜色特征不能获得好的目标再识别结果。近年来，另一种视觉外观特征：局部描述符，在图像分类等领域[4-5]获得了广泛应用。由于局部描述符能够适应不精确的目标定位、部分遮挡以及光照变化，文献[6-7]把其应用于人体目标跟踪和再识别。文献[8]采用 SIFT 视觉单词树、Learning++.MT 在线学习和支持向量机方法(SVM)进行跨非重叠摄像机人体目标再识别。实验结果表明：视觉单词树方法比颜色直方图方法的识别率更高；利用 SVM 核的辨别性方法比利用相似度度量的产生式模型方法识别率更高。

采用传统支持向量机无法实现不同特征的高效融合。近年来，针对传统支持向量机方法的缺陷，有研究者提出了多核学习方法(MKL)[9-11]并成为当前机器学习领域一个新的研究热点。多核学习通过将不同的核函数进行组合，增强决策函数的可解释性，得到比单核(或均核)模型更优的性能。颜色和 SIFT 局部描述符这二种特征具有一定的互补性，因此，融合这二种视觉外观特征可以提高目标再现识别率。本文提出了一种新的利用在线多核学习融合颜色和局部特征描述符的人体目标再现识别方法，获得了比单一特征更高的识别性能，并且也获得比单核(或均核)支持向量机更高的识别率。

1 人体目标特征提取

1.1 颜色直方图特征

颜色直方图特征广泛应用于单摄像机或多摄像机人体目标跟踪中。本文使用 RGB 颜色空间。每个颜色通道计算 32 维的直方图。整个人体前景图像全局颜色直方图特征共有 96(3×32)维。

1.2 SIFT 描述符

SIFT 描述符的计算步骤如下[12]：1) 对前景图像进行高斯滤波。2) 以采样点为中心取 16×16 的邻域区域，计算邻域区域的每个像素的梯度大小及方向。3) 将邻域区域均分成 4×4 个子区域，累计每个子区域的梯度方向直方图。梯度方向分为 8 个方向，则 SIFT 描述符向量长度为 4×4×8=128 维。4) 将描述符向量归一化，去除光照影响。

1.3 SIFT 视觉单词树构建

视觉单词树[13]采用了一种分层量化策略。该量化策略由分层 k 均值聚类实现。k 定义了树中每个节点的分支数。首先，对描述符向量进行 k 均值聚类，把描述符向量分为 k 组。每一组由与聚类中心距离最近的描述符向量组成。对每一组描述符递归地执行 k 均值聚类得到下一层的 k 个更细分组，直到预定的最大层数 L，则最后的分支点单词即为该分支点分组的聚类中心。该过程如图 1 所示。

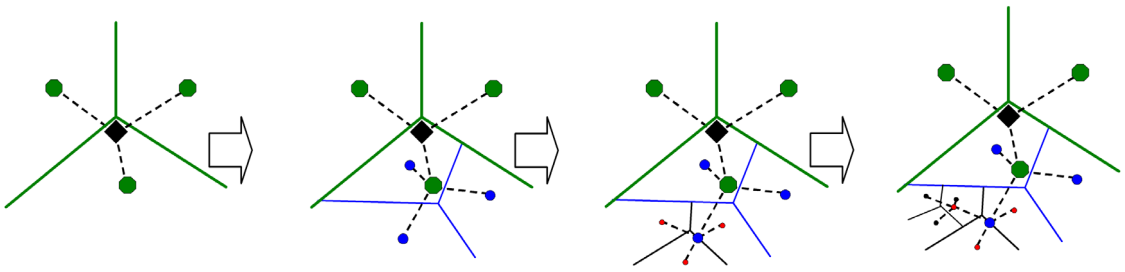

图 1 单词树构建进程图

Fig.1 An illustration of the process of building the vocabulary tree

首先，采用离线训练和一般人体目标数据集构建视觉单词树。在 L 层 k 分支的视觉单词树中表示的单词个数为

$$M=\sum_{i=1}^{L}k^{i}=\frac{k^{L+1}-k}{k-1}\approx k^{L} \tag{1}$$

视觉单词树中不同的分支点具有不同的重要性，本文采用逆向文档频率加权法给每个分支点定义一个权重：

$$w_i=\ln\frac{N}{N_i} \tag{2}$$

其中：N 为训练数据库中总图像数，N_i 为至少有一个 SIFT 描述符通过分支点 i 的图像数。

1.4 SIFT 视觉单词树直方图

在线识别时，对前景图像中每一个描述符向量自上而下逐层与 k 个分支点单词进行比较，选择最近邻的分支点，统计 SIFT 描述符经过的每个分支点，形成 SIFT 视觉单词树直方图。则人体目标 c 在 t 时刻运动前景图像的 SIFT 局部描述符外观特征可表达如下：

$$I_t^c=\{x_1,x_2,...,x_M\} \tag{3}$$

式中：M 为式(1)定义的 SIFT 视觉单词树中单词总数，x_i 为加权的视觉单词直方图，表达式如下：

$$x_i=n_i w_i \tag{4}$$

式中：n_i 为在线识别时人体目标图像包含视觉单词 i 的 SIFT 描述符的个数。w_i 为式(2)定义的视觉单词 i 的权重。文献[9]对 k 和 L 的参数选择进行了详细的实验对比。实验结果得出，k=10 和 L=4 时识别效果最好。所以，本文也采用同样的参数设置。

2　多核学习

多核学习是近几年出现的新的机器学习方法。多核学习对不同特性的核函数进行组合，从而得到比单核模型更优的学习性能。在多核学习框架下，样本在特征空间中的表示可以转化为基本核与权系数的选择问题。目前，多核学习方法已用来进行各种分类任务。相对于传统的单核支持向量机方法，多核方法能够通过学习得到一个最优的核组合方式和相关分类器，是一种有效的特征融合的方式。多核学习方法的合成核可表达如下：

$$K(x_i,x_j)=\sum_{k=1}^{M}d_k K_k(x_i,x_j)\ ,\ \ \sum_{k=1}^{M}d_k=1\ ,\ \ d_k\ge 0 \tag{5}$$

其中：x_i 为数据点，$K_k(x_i,x_j)$ 为第 k 个核，d_k 为每个信息源(核)的权重。M 为合成核的数目，本文中 M=2。

多核学习有许多表示形式和优化方法，本文采用 Rakotomamonjy 等提出的称为简单多核学习(SimpleMKL)的方法[11]。其表示形式使得核联合权重能在标准 SVM 优化框架下学习，与其他多核学习方法相比，简单多核学习收敛速度更快且效率、精确度更高。

简单多核学习的优化方程为

$$\begin{cases}\min_d J(d) & \text{s.t}\ \sum_k d_k=1, \quad (d_k\ge 0\quad \forall k)\\ J(d)=\begin{cases}\min\limits_{w_k,b,\xi}\sum\limits_k\frac{1}{d_k}w_k w_k^T+C\sum\limits_i\xi_i\\ \text{s.t}\ \ y_i\sum\limits_k\varphi_k(x_i)+y_i b\ge 1-\xi_i \quad (\xi_i\ge 0)\end{cases}\end{cases} \tag{6}$$

由式(6)可以得到其对偶形式，则式(6)可以等价为

$$\begin{cases}\max_\alpha & -\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j\sum_k d_k K_k(x_i,x_j)+\sum_i\alpha_i\\ \text{s.t.} & \sum_i\alpha_i y_i=0 \qquad (C\ge\alpha_i\ge 0)\end{cases} \tag{7}$$

使用单核 $\sum_k d_k K_k(x_i,x_j)$，上式就是标准的 SVM 对偶问题。给定 d，上式可由 SVM 标准优化方法得到参数 α^*。给定 d，$J(d)$也是对偶问题的目标值：

$$\begin{cases} J(d)=-\dfrac{1}{2}\sum\limits_{i,j}\alpha_i^*\alpha_j^*\sum\limits_k d_k K_k(x_i,x_j)+\sum\limits_i \alpha_i^* \\ \text{s.t}\sum\limits_k d_k=1, \qquad (d_k\geq 0) \end{cases} \tag{8}$$

因为 SVM 的解唯一，所以 $J(d)$是可微的：

$$\frac{\partial J}{\partial d_k}=-\frac{1}{2}\sum_{i,j}\alpha_i^*\alpha_j^* y_i y_j K_k(x_i,x_j) \qquad \forall k \tag{9}$$

式(6)为线性约束的非线性优化问题，其核矩阵正定，$J(d)$是凸函数，且可微，则式(6)可以用投影梯度法解决。整个 SimpleMKL 多核学习优化问题求解可由 2 步交替迭代优化方法完成，算法流程如下：

1) 令 t=1，$d_k^1=\dfrac{1}{M}$，$k=1,...,M$；

2) 利用 $K=\sum_k d_k^t K_k$ 求解标准的 SVM 问题；

3) 计算 $\dfrac{\partial J}{\partial d_k}$，$k=1,...,M$；

4) 计算函数 J 的下降方向向量 $\boldsymbol{D}_t$ 和最佳步长 γ_t，则 $d_k^{t+1}=d_k^t+\gamma_t \boldsymbol{D}_t$；

5) t=t+1，返回步骤 2)，直至满足一定的收敛判断条件，收敛条件为达到预定的迭代次数。

最后，训练得到的决策函数为

$$f(x)=\sum_j\sum_k d_k\alpha_j y_j K_k(x_i,x)+b \tag{10}$$

3 人体目标多核分类器在线训练

为了计算效率，本文的核函数均采用线性核，其分类函数为

$$f(x)=\sum_{j=1}^{n}\sum_{k=1}^{2} d_k\alpha_j^* y_j(x_j\bullet x)+b^* \tag{11}$$

式中：α_i^* 不为零所对应的样本为支持向量，共有 n 个。在视频监控中，训练数据为跟踪目标前景图像视频流。很显然，一次对所有的数据训练多核分类器不能满足视频监控的实时要求，所以本文对人体目标多核分类器进行在线训练。即不断为已训练的多核分类器增添新的学习样本，提高其分类精度；同时使多核分类器在新的学习中充分利用以前的学习结果，从而减少后继的学习时间。一种精确的 SVM 在线训练方法就是训练新数据时保持以前所有数据的 KKT 条件，但该方法训练过于复杂。

本文把 SVM 增量学习方法用于 SimpleMKL 多核分类器在线学习中，其在很小的时间空间代价下能实现新样本的学习。本文采用的 SimpleMKL 多核学习方法在标准 SVM 优化框架下迭代学习，其多核分类器支持向量可完全描述整个数据样本集的分类特征，而支持向量集只是数据样本集的一小部分。分类仅由在超平面边缘上的少数支持向量决定。所以，在线训练过程中仅保留支持向量可大大减少计算负担。

本文首先利用前几帧训练得到人体目标的初始模型，在随后的视频跟踪中，当新样本积累到一定数目时则新样本联合支持向量重新训练得到新模型和新的支持向量。在后续的增量式学习中，为了加快训练速度，SimpleMKL 迭代步骤中 d_k 的初始值设置为以前的学习结果。由于 SimpleMKL 多核学习方法为二值分类算法，而本文要实现多目标识别，因此采用一对多组合(one against all)的训练方式。

4 实验结果与分析

4.1 实验数据集

本文采用 CAVIAR 视频监控数据集[14]评价本文算法的识别性能。该数据集广泛应用于视频跟踪和跨摄像机人体目标再现识别算法的实验评价[7-8]。如图 2 所示，数据集由安装在购物中心走廊上的 2 个不同拍摄角度的摄像机采集得到，共有 26 段视频片段。本文从该数据集中提取 26 个不同人体目标的跟踪序列。实验使用的 26 个人体目标数据库如图 3。本文中视觉单词树的码书采用 PASCAL 人体目标数据库[15]离线训

练得到。

图 2　CAVIAR 视频采样

Fig.2　Samples of videos in CAVIAR

图 3　26 个跟踪人体目标数据库

Fig.3　26 tracked people datasets

4.2 在线学习性能

为了评价算法的在线学习能力，对本文算法和保留所有样本数据的离线训练方法进行了对比，实验结果如表 1 所示。本文的识别率为跟踪序列图像的识别率，采用投票方式得到。从表 1 可以看出，本文的在线学习方法获得了和离线学习方法相当的识别率。与文献[8]的 Learning++.MT 增量算法相比，本文的在线学习能力明显优越。Learning++.MT 算法只是简单地对 SVM 支持向量机训练数据进行分批集成学习，在线学习的分类器识别性能远低于离线学习。本文的在线 SimpleMKL 多核学习在增量学习过程中每次只保留分类器模型的支持向量，有效地利用了核学习方法的 KKT 特性，损失的历史训练数据对最后的分类器影响非常有限。

表 1　采用不同训练策略的识别性能

Table 1　Recognition performance with different training methods

Learning methods	Offline learning	MKL online learning
Identification rate	88.9%	87.6%

4.3 多核特征融合性能

为了评价本文算法的特征融合性能，对仅采用颜色、仅采用 SIFT 局部描述符、采用单核(均核)学习，采用本文的多核学习这四种方法进行了实验对比，实验结果如表 2 所示。从实验结果看，融合颜色和 SIFT 局部描述符能大大提高识别性能。这二种特征有一定的互补性，局部描述符对光照、局部遮挡较鲁棒，颜色特征对人体目标形变较鲁棒，融合后的人体目标分类器辨别性更强。另外，多核学习获得了比单核学习更高的识别率。图 4 给出了三组单核和多核学习识别结果，图左边为待识别人体目标图像，右边为人体目标分类器取值最大的前 10 个决策结果。从图 4 可以看出，多核分类器三组识别结果都正确，而单核学习仅一组正确。单核学习为把颜色和 SIFT 局部描述符这二种特征接连形成一种特征[x^c x^s]后再利用标准 SVM

训练分类器，均核学习为多核学习中 d_1, d_2 取值都固定为 0.5。由于本文中颜色和 SIFT 局部描述符二种特

表 2 采用单特征、单核和多核方法的识别性能对比

Table 2 Recognition performance with single feature, single kernel and multiple kernel

Method	Color	SIFT local descriptor	Single kernel learning	Multiple kernel learning
Identification rate	71.2%	74.7%	85.3%	87.6%

(a) Multiple kernel learning

(b) Single kernel learning

图 4 多核和单核学习识别结果对比

Fig.4 Recognition results with multiple kernel and single kernel learning

征均采用线性核，所以在这种情况下，单核学习和均核学习是等价的。

4.4 算法实时性能

由于本文采用在线多核学习方法，对于每帧图像，算法的计算时间主要用在 SIFT 局部描述符提取上。整个算法采用 C++语言编写，在主频为 2.8 GHz，内存为 4G 的 PC 上运行。本文算法平均每秒约处理 8 帧，能满足视频跟踪的实时性要求，明显优于 Learning++.MT 分批集成在线学习方法。

5 结 论

本文提出了一种新的用于视频监控的目标再现识别算法。算法采用 SimpleMKL 多核学习算法训练人体目标分类器，在线融合 SIFT 局部描述符和颜色直方图二种特征。实验结果表明，本文方法优于传统方法，获得了比单一特征更高的识别性能，以及比单核学习更高的识别率。下一步研究工作为：1) 融合其他特征，如图像区域协方差特征，进一步提高识别率；2) 视频监控中基于目标再识别算法的跟踪目标徘徊逗留检测。

参考文献：

[1] Matei B C，Sawhney H S，Amarasekera S. Vehicle tracking across nonoverlapping cameras using joint kinematic and

appearance features [C]// **Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)**，Colorado Springs，CO，USA，June 20-25，2011. Piscataway：IEEE Computer Society，2011：3465−3472.

[2] Wei-Shi Z，Shaogang G，Tao X. Person Re-identification by Probabilistic Relative Distance Comparison [C] // **Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)**，Colorado Springs，CO，USA，June 20-25，2011. Piscataway：IEEE Computer Society，2011：649−656.

[3] Aziz K-E，Merad D，Fertil B. People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation [C]// **International Conference on Advanced Video and Signal-based Surveillance (AVSS)**，Klagenfurt，Austria，August 30-September 2，2011. Piscataway：IEEE Computer Society，2011：303−308.

[4] 张朝亮，江汉红，姜春良，等. 基于 SIFT 和加权信息熵的红外小目标检测 [J]. 光电工程，2010，**37**(11)：19−25.
ZHANG Chao-liang，JIANG Han-hong，JIANG Chun-liang，*et al*. Detecting Infrared Small Targets Based on SIFT and Weighted Entropy [J]. **Opto-Electronic Engineering**，2010，**37**(11)：19−25.

[5] Jia Deng，Berg A C，Li Fei-Fei. Hierarchical semantic indexing for large scale image retrieval [C]// **Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)**，Colorado Springs，CO，USA，June 20-25，2011. Piscataway：IEEE Computer Society，2011：785−792.

[6] Farenzena M，Bazzani L，Perina A，*et al*. Person re-identification by symmetry-driven accumulation of local features [C]// **Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)**，San Francisco，CA，United States，June 13-18，2010. Piscataway：IEEE Computer Society，2010：2360−2367.

[7] Bäuml M，Stiefelhagen R. Evaluation of Local Features for Person Re-Identification in Image Sequences [C]// **International Conference on Advanced Video and Signal-based Surveillance (AVSS)**，Klagenfurt，Austria，August 30-September 2，2011. Piscataway：IEEE Computer Society，2011：291−296.

[8] TEIXEIRA L F，CORTE_REAL L. Video object matching across multiple independent views using local descriptors and adaptive learning [J]. **Pattern Recognition Letters**(S0167-8655)，2009，**30**(2)：157−167.

[9] Subrahmanya N，Shin Y C. Sparse Multiple Kernel Learning for Signal Processing Applications [J]. **IEEE Transactions on Pattern Analysis and Machine Intelligence**(S0162-8828)，2010，**32**(5)：788−798.

[10] Sonnenburg S，Ratsch G，Schafer C，*et al*. Large scale multiple kernel learning [J]. **The Journal of Machine Learning Research**(S1533-7928)，2006，**7**(7)：1531−1565.

[11] Rakotomamonjy A，Bach F R，Canu S，*et al*. Simple MKL [J]. **The Journal of Machine Learning Research**(S1533-7928)，2008，**9**(11)：2491−2521.

[12] LOWE D G. Distinctive Image features from scale-invariant keypoints [J]. **International Journal of Computer Vision**(S0920-5691)，2004，**2**(60)：91−110.

[13] NISTER D，STEWENIUS H. Scalable recognition with a vocabulary tree [C]// **2006 Conference on Computer Vision and Pattern Recognition**，New York，NY，United States，June 17-22，2006. Piscataway：Institute of Electrical and Electronics Engineers Computer Society，2006：2161−2168.

[14] EVERINGHAM M，GOOL L V，WILLIAMS C，*et al*. The PASCAL Visual Object Classes Challenge[EB/OL].[2012-02-20]. http：//www.pascal-network.org/challenges/VOC/.

[15] FISHER R，SANTOS-VICTOR J，CROWLEY J. CAVIAR：Context Aware Vision using Image-based Active Recognition[EB/OL]. [2012-02-20] .http://homepages.inf.ed.ac.uk/rbf/CAVIAR/.