(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108108652 B
(45)授权公告日 2021.11.26

(21)申请号 201710197426.4
(22)申请日 2017.03.29
(65)同一申请的已公布的文献号
申请公布号 CN 108108652 A
(43)申请公布日 2018.06.01
(73)专利权人 广东工业大学
地址 510062 广东省广州市越秀区东风东路729号大院
(72)发明人 陆光辉 刘波 肖燕珊 聂欢 李子彬
(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 罗满
(51)Int.Cl.
G06K 9/00(2006.01)
G06K 9/62(2006.01)
(56)对比文件
CN 104091169 A,2014.10.08
CN 106056082 A,2016.10.26
CN 106056135 A,2016.10.26
US 2010250241 A1,2010.09.30
US 2016171096 A1,2016.06.16
审查员 刘志军

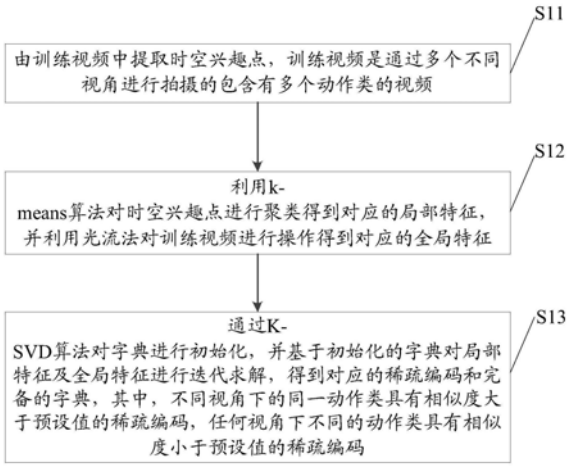
权利要求书2页 说明书14页 附图1页

(54)发明名称

一种基于字典学习的跨视角人体行为识别方法及装置

(57)摘要

本发明公开了一种基于字典学习的跨视角人体行为识别方法及装置，该方法包括：由训练视频中提取时空兴趣点，训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频；利用k-means算法对时空兴趣点进行聚类得到对应的局部特征，并利用光流法对训练视频进行操作得到对应的全局特征；通过K-SVD算法对字典进行初始化，并基于初始化的字典对所述局部特征及所述全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，其中，不同视角下的同一动作类具有相似度大于预设值的稀疏编码，任何视角下不同的动作类具有相似度小于预设值的稀疏编码。能够适应于拍摄视频的视角发生变化时对其中的人类行为进行识别的情况，识别性能较高。



CN 108108652 B

1.一种基于字典学习的跨视角人体行为识别方法,其特征在于,包括:

由训练视频中提取时空兴趣点,所述训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频;

利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征,并利用光流法对所述训练视频进行操作得到对应的全局特征;

通过K-SVD算法对字典进行初始化,并基于初始化的字典对所述局部特征及所述全局特征进行迭代求解,得到对应的稀疏编码和完备的字典,其中,不同视角下的同一动作类具有相似度大于预设值的稀疏编码,任何视角下不同的动作类具有相似度小于预设值的稀疏编码;

通过K-SVD算法对字典进行初始化,并基于初始化的字典对所述局部特征及所述全局特征进行迭代求解,得到对应的稀疏编码和完备的字典,包括:

基于K-SVD算法对所述局部特征及全局特征进行训练,得到初始化的字典;

基于所述初始化的字典通过下列方程进行优化求解,得到优化后的稀疏编码X和完备的字典D:

$$f=\sum_{c=1}^{C}\left\{\begin{array}{l}\|Y_c-DX_c\|_2^2+\|Y_c-D\tilde{Q}_c\tilde{Q}_c^{T}X_c\|_2^2+\|\tilde{Q}_{/c}^{T}X_c\|_2^2\\+\lambda_1\phi(X_c)+\lambda_2\|X_c\|_2^2+\beta\|p_c-AX_c\|_2^2\end{array}\right\}+\alpha\sum_{c=1}^{C+1}\sum_{\substack{j=1\\j\neq c}}^{C+1}Q(D_c,D_j)$$

其中,C表示所述训练视频中包含的动作类的个数,Y_c表示所述训练视频中包含的第c个动作类,D表示字典,X_c表示所述训练视频中包含的第c个动作类所对应的稀疏编码,p_c表示所述训练视频中包含的第c个动作类的理想的稀疏编码,D_c和D_j都表示特性字典,c和j分别表示为第c个特性字典和第j个特性字典的序号,$Q_c=\left[q_1^c,\ldots q_j^c\ldots q_{k_c}^c\right]\in R^{k\times k_c}$,其中$k\times k_c$表示$Q_c$的维度,$q_j^c$表示一个维度与$Q_c$维度一样的矩阵,且只有第$k_c$行第j列的值为1,其他值都为零,$\tilde{Q}_c=[Q_c,Q_{C+1}]$, $\tilde{Q}_c^T$为$\tilde{Q}_c$的转置,$\tilde{Q}_{/c}=[Q_1,\ldots,Q_{c-1},Q_{c+1},\ldots,Q_C]$,$\lambda_1$、$\lambda_2$、α及β为预先设定的系数,A为对应线性转化矩阵,$Q(D_c,D_j)=\|D_i^TD_j\|^2$, $\phi(X_c)=\sum_{i=1}^{N_c}\|x_i^c\|_1$, x_i^c表示第i个视角对应的特性字典D_c的稀疏表示,Nc表示所述训练视频对应的视角个数。

2.根据权利要求1所述的方法,其特征在于,在利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征之前,还包括:

使用PCA技术对提取得到的所述时空兴趣点进行降维操作。

3.根据权利要求2所述的方法,其特征在于,由训练视频中提取时空兴趣点,包括:

利用高斯平滑函数和Gabor滤波器由训练视频中提取时空兴趣点。

4.一种基于字典学习的跨视角人体行为识别装置,其特征在于,包括:

提取模块,用于:由训练视频中提取时空兴趣点,所述训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频;

处理模块,用于:利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征,并利用光流法对所述训练视频进行操作得到对应的全局特征;

训练模块,用于:通过K-SVD算法对字典进行初始化,并基于初始化的字典对所述局部

特征及所述全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，其中，不同视角下的同一动作类具有相似度大于预设值的稀疏编码，任何视角下不同的动作类具有相似度小于预设值的稀疏编码具有相似度不小于预设值的稀疏编码；

所述训练模块包括：

训练单元，用于：基于K-SVD算法对所述局部特征及全局特征进行训练，得到初始化的字典；

基于所述初始化的字典通过下列方程进行优化求解，得到优化后的稀疏编码X和完备的字典D：

$$f=\sum_{c=1}^{C}\left\{\begin{matrix}\|Y_c-DX_c\|_2^2+\|Y_c-D\tilde{Q}_c\tilde{Q}_c^{T}X_c\|_2^2+\|\tilde{Q}_{/c}^{T}X_c\|_2^2\\+\lambda_1\phi(X_c)+\lambda_2\|X_c\|_2^2+\beta\|p_c-AX_c\|_2^2\end{matrix}\right\}+\alpha\sum_{c=1}^{C+1}\sum_{\substack{j=1\\j\neq c}}^{C+1}Q(D_c,D_j)$$

其中，C表示所述训练视频中包含的动作类的个数，Y_c表示所述训练视频中包含的第c个动作类，D表示字典，X_c表示所述训练视频中包含的第c个动作类所对应的稀疏编码，p_c表示所述训练视频中包含的第c个动作类的理想的稀疏编码，D_c和D_j都表示特性字典，c和j分别表示为第c个特性字典和第j个特性字典的序号，$Q_c=\left[q_1^c,\ldots q_j^c\ldots q_{k_c}^c\right]\in R^{k\times k_c}$，其中$k\times k_c$表示$Q_c$的维度，$q_j^c$表示一个维度与$Q_c$维度一样的矩阵，且只有第$k_c$行第j列的值为1，其他值都为零，$\tilde{Q}_c=[Q_c,Q_{C+1}]$，$\tilde{Q}_c^T$为$\tilde{Q}_c$的转置，$\tilde{Q}_{/c}=[Q_1,\ldots,Q_{c-1},Q_{c+1},\ldots,Q_C]$，$\lambda_1$、$\lambda_2$、α及β为预先设定的系数，A为对应线性转化矩阵，$Q(D_c,D_j)=\|D_i^TD_j\|^2$，$\phi(X_c)=\sum_{i=1}^{N_c}\|x_i^c\|_1$，$x_i^c$表示第i个视角对应的特性字典$D_c$的稀疏表示，Nc表示所述训练视频对应的视角个数。

5.根据权利要求4所述的装置，其特征在于，还包括：

降维模块，用于：在利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征之前，使用PCA技术对提取得到的所述时空兴趣点进行降维操作。

6.根据权利要求5所述的装置，其特征在于，所述提取模块包括：

提取单元，用于：利用高斯平滑函数和Gabor滤波器由训练视频中提取时空兴趣点。

一种基于字典学习的跨视角人体行为识别方法及装置

技术领域

[0001] 本发明涉及人体行为识别技术领域，更具体地说，涉及一种基于字典学习的跨视角人体行为识别方法及装置。

背景技术

[0002] 随着现代信息技术的发展，人们相互交流不再仅仅局限于文字、语音和图像等传统媒介，大量的视频甚至是高质量的视频信号日益充满在人类社会中。大量的视频数据存在于的生活中，并且仍然以超出想象的速度迅猛膨胀，如何快速有效的理解和处理好这些视频信息就成了一个十分重大的课题。而人体运动作为视频中的核心信息，对于视频中的人体行为识别的研究就成为了计算机理解视频含义的关键钥匙。

[0003] 目前用于实现视频中人体行为识别的技术方法通常是对预先获取的视频提取特征并进行相关建模，进而通过建出的模型对其他视频中的人体行为进行识别。但是用于训练模型的视频通常都是通过一个固定的视角拍摄的，也即提取特征及相关建模均是基于一个固定的视角实现的，由此建出的模型对于在该固定的视角拍摄的视频中的人体行为能够很好的识别，但是当视角发生变化，人体的形态和运动轨迹都会随之发生改变，对应的特征也会变得不一样，这就会导致建出的模型对于视频中人体行为的识别性能大大降低。

[0004] 综上所述，现有技术中用于识别视频中人体行为的模型存在识别性能较低的问题。

发明内容

[0005] 本发明的目的是提供一种基于字典学习的跨视角人体行为识别方法及装置，以解决现有技术中用于识别视频中人体行为的模型存在的识别性能较低的问题。

[0006] 为了实现上述目的，本发明提供如下技术方案：

[0007] 一种基于字典学习的跨视角人体行为识别方法，包括：

[0008] 由训练视频中提取时空兴趣点，所述训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频；

[0009] 利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征，并利用光流法对所述训练视频进行操作得到对应的全局特征；

[0010] 通过K-SVD算法对字典进行初始化，并基于初始化的字典对所述局部特征及所述全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，其中，不同视角下的同一动作类具有相似度大于预设值的稀疏编码，任何视角下不同的动作类具有相似度小于预设值的稀疏编码。

[0011] 优选的，在利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征之前，还包括：

[0012] 使用PCA技术对提取得到的所述时空兴趣点进行降维操作。

[0013] 优选的，由训练视频中提取时空兴趣点，包括：

[0014] 利用高斯平滑函数和Gabor滤波器由训练视频中提取时空兴趣点。

[0015] 优选的，通过K-SVD算法对字典进行初始化，并基于初始化的字典对所述局部特征及所述全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，包括：

[0016] 基于K-SVD算法对所述局部特征及全局特征进行训练，得到初始化的字典；

[0017] 基于所述初始化的字典通过下列方程进行优化求解，得到优化后的稀疏编码X和完备的字典D：

[0018]
$$f=\sum_{c=1}^{C}\{\| Y_c-DX_c\|_2^2+\| Y_c-D\tilde{Q}_c\tilde{Q}_c^T X_c\|_2^2+\| \tilde{Q}_{/c}^T X_c\|_2^2+\lambda_1\phi(X_c)+\lambda_2\| X_c\|_2^2 +\beta\| p_c-AX_c\|_2^2\}+\alpha\sum_{c=1}^{C+1}\sum_{\substack{j=1\\ j\neq c}}^{C+1}Q(D_c,D_j)$$

[0019] 其中，C表示所述训练视频中包含的动作类的个数，Y_c表示所述训练视频中包含的第c个动作类，D表示字典，X_c表示所述训练视频中包含的第c个动作类所对应的稀疏编码，p_c表示所述训练视频中包含的第c个动作类的理想的稀疏编码，D_c和D_j都表示特性字典，c和j分别表示为第c个特性字典和第j个特性字典的序号，$Q_c=\left[q_1^c,\dots q_j^c\dots q_{k_c}^c\right]\in R^{k\times k_c}$，其中$k\times k_c$表示$Q_c$的维度，$q_j^c$表示一个维度与$Q_c$维度一样的矩阵，且只有第$k_c$行第j列的值为1，其他值都为零，$\tilde{Q}_c=[Q_c,Q_{C+1}]$，$\tilde{Q}_c^T$为$\tilde{Q}_c$的转置，$\tilde{Q}_{/c}=[Q_1,\dots,Q_{c-1},Q_{c+1},\dots,Q_C]$，$\lambda_1$、$\lambda_2$、α及β为预先设定的系数，A为对应线性转化矩阵，$\phi(X_c)=\sum_{i=1}^{N_c}\| x_i^c\|_1$，$x_i^c$表示第i个视角对应的特性字典$D_c$的稀疏表示，$Q(D_c,D_j)=\| D_i^T D_j\|^2$，Nc表示所述训练视频对应的视角个数。

[0020] 一种基于字典学习的跨视角人体行为识别装置，包括：

[0021] 提取模块，用于：由训练视频中提取时空兴趣点，所述训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频；

[0022] 处理模块，用于：利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征，并利用光流法对所述训练视频进行操作得到对应的全局特征；

[0023] 训练模块，用于：通过K-SVD算法对字典进行初始化，并基于初始化的字典对所述局部特征及所述全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，其中，不同视角下的同一动作类具有相似度大于预设值的稀疏编码，任何视角下不同的动作类具有相似度小于预设值的稀疏编码具有相似度不小于预设值的稀疏编码。

[0024] 优选的，还包括：

[0025] 降维模块，用于：在利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征之前，使用PCA技术对提取得到的所述时空兴趣点进行降维操作。

[0026] 优选的，所述提取模块包括：

[0027] 提取单元，用于：利用高斯平滑函数和Gabor滤波器由训练视频中提取时空兴趣点。

[0028] 优选的，所述训练模块包括：

[0029] 训练单元，用于：基于K-SVD算法对所述局部特征及全局特征进行训练，得到初始化的字典；

[0030] 基于所述初始化的字典通过下列方程进行优化求解，得到优化后的稀疏编码X和完备的字典D：

$$f=\sum_{c=1}^{C}\{\| Y_c-DX_c\|_2^2+\| Y_c-D\tilde{Q}_c\tilde{Q}_c^TX_c\|_2^2+\| \tilde{Q}_{/c}^TX_c\|_2^2+\lambda_1\phi(X_c)+\lambda_2\| X_c\|_2^2$$

[0031]

$$+\beta\| p_c-AX_c\|_2^2\}+\alpha\sum_{c=1}^{C+1}\sum_{\substack{j=1\\ j\neq c}}^{C+1}Q(D_c,D_j)$$

[0032] 其中，C表示所述训练视频中包含的动作类的个数，Y_c表示所述训练视频中包含的第c个动作类，D表示字典，X_c表示所述训练视频中包含的第c个动作类所对应的稀疏编码，p_c表示所述训练视频中包含的第c个动作类的理想的稀疏编码，D_c和D_j都表示特性字典，c和j分别表示为第c个特性字典和第j个特性字典的序号，$Q_c=\left[q_1^c,\cdots q_j^c\cdots q_{k_c}^c\right]\in R^{k\times k_c}$，其中$k\times k_c$表示$Q_c$的维度，$q_j^c$表示一个维度与$Q_c$维度一样的矩阵，且只有第$k_c$行第j列的值为1，其他值都为零，$\tilde{Q}_c=[Q_c,Q_{C+1}]$，$\tilde{Q}_c^T$为$\tilde{Q}_c$的转置，$\tilde{Q}_{/c}=[Q_1,\ldots,Q_{c-1},Q_{c+1},\ldots,Q_C]$，$\lambda_1$、$\lambda_2$、α及β为预先设定的系数，A为对应线性转化矩阵，$\phi(X_c)=\sum_{i=1}^{N_c}\| x_i^c\|_1$ x_i^c表示第i个视角对应的特性字典D_c的稀疏表示，$Q(D_c,D_j)=\| D_i^TD_j\|^2$，Nc表示所述训练视频对应的视角个数。

[0033] 本发明提供了一种基于字典学习的跨视角人体行为识别方法及装置，其中该方法包括：由训练视频中提取时空兴趣点，所述训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频；利用k-means算法对所述时空兴趣点进行聚类得到对应的局部特征，并利用光流法对所述训练视频进行操作得到对应的全局特征；通过K-SVD算法对字典进行初始化，并基于初始化的字典对所述局部特征及所述全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，其中，不同视角下的同一动作类具有相似度大于预设值的稀疏编码，任何视角下不同的动作类具有相似度小于预设值的稀疏编码。本申请公开的技术特征中，利用通过不同视角拍摄的视频进行训练，且训练得到的完备的字典中对应于不同的动作类具有相似度小于预设值的稀疏编码，由此，能够适应于拍摄视频的视角发生变化时对其中的人类行为进行识别的情况，识别性能较高。

附图说明

[0034] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据提供的附图获得其他的附图。

[0035] 图1为本发明实施例提供的一种基于字典学习的跨视角人体行为识别方法的流程图；

[0036] 图2为本发明实施例提供的一种基于字典学习的跨视角人体行为识别装置的结构示意图。

具体实施方式

[0037] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0038] 请参阅图1，其示出了本发明实施例提供的一种基于字典学习的跨视角人体行为识别方法的流程图，可以包括以下步骤：

[0039] S11：由训练视频中提取时空兴趣点，训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频。

[0040] 其中时空兴趣点就是在空间和时间轴上变化较为显著的点，时空兴趣点的检测就是通过对视频的图像中每个像素点或者位置进行强度函数的极大值滤波，得到对应的兴趣点。而训练视频可以是对预先设定的环境范围内通过不同视角进行拍摄的包含有多个动作类的视频，也可以是对任意环境内通过不同视角进行拍摄的包含有多个动作类的视频，具体可以根据实际需要进行设定；而多个动作类可以对应不同的人，从而使得得到的训练视频更具有训练价值。

[0041] S12：利用k-means算法对时空兴趣点进行聚类得到对应的局部特征，并利用光流法对训练视频进行操作得到对应的全局特征。

[0042] 需要说明的是，得到训练视频对应的上述局部特征和全局特征后可以将这两种特征存储至一文件中作为待处理文件，以在后续需要使用上述两种特征时直接利用待处理文件进行对应操作即可。其中，利用k-means算法对时空兴趣点进行聚类得到对应的局部特征具体可以包括：由时空兴趣点中随机选取k个聚类质心点(cluster centroids)作为当前选取的聚类质心点，重复下列过程直到收敛：1、对于每一个时空兴趣点i，基于当前选取的聚类质心点计算其应该属于的聚类。2、对于每一个聚类j，重新计算该聚类的质心点，得到当前选取的聚类质心点，返回执行1，直至计算得出的聚类质心点不再发生变化为止。简单来说就是计算其他每个时空兴趣点到当前选取的每个聚类质心点的距离(欧氏距离)，选取某个时空兴趣点到某一个聚类质心点的距离最小的将该时空兴趣点与该聚类质心点归为一类，得到的聚类质心点作为当前选取的聚类质心点，然后重新计算其他每个兴趣点到当前选取的每个聚类质心点的距离，如此循环直至聚类质心点不再发生变化为止。通过上述方式可以得到视频的局部特征，进而通过光流法得到视频的全局特征。

[0043] 具体来说，利用光流法得到全局特征的原理为：给视频图像中的每个像素点赋予一个速度矢量，这样就形成了一个运动矢量场；在某一特定时刻，图像上的点与三维物体上的点一一对应，这种对应关系可以通过投影来计算得到；根据各个像素点的速度矢量特征，可以对图像进行动态分析；如果图像中没有运动物体，则光流矢量在整个图像区域是连续变化的；当图像中有运动物体时，目标和背景存在着相对运动；运动物体所形成的速度矢量必然和背景的速度矢量有所不同，如此便可以计算出运动物体的位置。简单来说，光流是空间运动物体在观测成像平面上的像素运动的“瞬时速度”，而通过光流法获取到的全局特征

即为视频的图像序列中像素强度数据的时域变化和相关性来确定出各像素位置的动态变化。

[0044] S13:通过K-SVD算法对字典进行初始化,并基于初始化的字典对局部特征及全局特征进行迭代求解,得到对应的稀疏编码和完备的字典,其中,不同视角下的同一动作类具有相似度大于预设值的稀疏编码,任何视角下不同的动作类具有相似度小于预设值的稀疏编码。

[0045] 预设值可以根据实际需要进行设定,相似度大于预设值则说明相似度较高,否则则说明相似度较低,因此上述步骤中不同视角下的同一动作类具有相似的稀疏编码,任何视角下不同的动作类则不具有相似的稀疏编码,另外相似度等于预设值的情况也可以归为相似度较高的情况,也即不同视角下的同一动作类可以具有相似度大于或者等于预设值的稀疏编码。具体来说,通过K-SVD算法对字典进行初始化,然后对上一步获得的局部特征和全局特征进行训练,从而可以获得动作类的稀疏编码,再根据得到的稀疏编码训练字典,如此循环便可获得完备的字典和稀疏编码。具体来说,使用K-SVD算法进行字典初始化时,首先用Y表示所要输入的特征(即上述局部特征和全局特征),K-SVD算法下列方程求解得到对应字典D:

[0046] $$D = \arg\min_{D,X} \| Y - DX \|_2^2 \, s.t. \forall i, \| xi \|_0 \leq s$$

[0047] 其中,Y是输入特征,X是稀疏编码,可以通过这个算法获得初始化的字典D,在初始化过程中是一个一个子字典进行初始化的。然后通过初始化的字典求得第一次的稀疏编码,再通过稀疏编码反过来求字典,如此循环,直至收敛即可求得最终的完备的字典。

[0048] 本申请公开的技术特征中,利用通过不同视角拍摄的视频进行训练,且训练得到的完备的字典中对应于不同的动作类具有相似度小于预设值的稀疏编码,由此,能够适应于拍摄视频的视角发生变化时对其中的人类行为进行识别的情况,识别性能较高。

[0049] 具体来说,本发明是一种基于字典学习的跨视角的动作识别的方法,使得不同的视角下由不同的特定字典和共性字典和稀疏编码进行表示,这样不仅保证了不同视角下相同动作具有相似的稀疏表示,而且使得不同视角下相同的动作具有不同的稀疏表示,这样保证了同一个动作有相同的特征,同时保证了同一个动作具有区别性。通过相同动作在不同的视角下的同一时间具有相同的类标签和有相似稀疏编码表示,学习获得完备的字典和稀疏编码表示。对于视角的转换,可以根据字典转移数据而不影响识别的效果,这样保证了对于视角的推广。

[0050] 本发明实施例提供的一种基于字典学习的跨视角人体行为识别方法,在利用k-means算法对时空兴趣点进行聚类得到对应的局部特征之前,还可以包括:

[0051] 使用PCA(Principal Component Analysis)技术对提取得到的时空兴趣点进行降维操作。

[0052] 具体来说,PCA是一种常用的数据分析方法,PCA通过线性变换将原始数据变换为一组各维度线性无关的表示,可用于提取数据的主要特征分量,常用于高维数据的降维。设有m条n维数据特征,则利用PCA进行降维操作的原理可以如下:

[0053] 1、均值归一化:计算出所有特征的均值,然后令$x_j = x_j - \mu_j$,其中μ_j表示均值,x_j表示被计算的数据特征点。如果特征是在不同的数量级上,还需要将其除以标准差σ^2。

[0054] 2、求出协方差矩阵$\Sigma=\frac{1}{m}\sum_{i=1}^{n}(x^i)(x^i)^T$。

[0055] 3、计算协方差矩阵Σ的特征向量。

[0056] 4、将特征向量按对应特征的值大小从上到下按行排列成矩阵，取前k行组成矩阵P’。

[0057] 5、Y’＝P’X’即为降维到k维后的数据特征，其中X’表示被降维的数据特征点。

[0058] 通过将时空兴趣点进行上述降维操作，能够去掉不重要的影响因子，进而有利于对于时空兴趣点的后续处理。

[0059] 本发明实施例提供的一种基于字典学习的跨视角人体行为识别方法，由训练视频中提取时空兴趣点，可以包括：

[0060] 利用高斯平滑函数和Gabor滤波器由训练视频中提取时空兴趣点。

[0061] 具体来说，利用高斯平滑函数和Gabor滤波器提取时空兴趣点的步骤可以包括：首先将视频的视频序列f中每一帧二维坐标到像素点亮度的映射$f^{sp}:R^2\rightarrow R$表示；然后使用空间域中的高斯核函数g^{sp}将f中的每一帧变换到高斯空间，得到$L^{sp}:R^2\times R_+\rightarrow R$，其中R指的是像素点的亮度，具体地有$L^{sp}(x,y,\sigma^2)=g^{sp}(x,y,\sigma^2)*f^{sp}(x,y)$，其中$L^{sp}$是使用空间域中的高斯核函数$g^{sp}$将f中的每一帧变换到高斯空间的表示，$R_+$指的是高斯核函数在将f中的每一帧变换到高斯空间的一个指代(可以理解为高斯核函数即R_+)，

$g^{sp}(x,y,\sigma_l^2)=\frac{1}{2\pi\sigma_l^2}\exp(-(x^2+y^2)/2\sigma_l^2)$;接下来对经过高斯平滑的视频序列$f:^{R2}\times R_+\rightarrow$R沿着f的时间轴方向，对每一列元素进行选定窗口的一维Gabor滤波，其具体运算过程式为：$I=(f*h_{ev})^2+(f*(h_{od})^2$，其中$h_{ev}$和$h_{od}$如下：

[0062] $h_{ev}(t;\tau,\omega)=-\cos(2\pi t\omega)e^{-t^2/\tau^2}$

[0063] $h_{od}(t;\tau,\omega)=-\sin(2\pi t\omega)e^{-t^2/\tau^2}$

[0064] 其中，τ^2表示滤波器在时域上的尺度，ω为Gabor窗口大小的1/8，I为像素点的强度，t为时间，x和y表示像素点的坐标，σ表示高斯函数的函数参数。计算视频序列f中每一点的R值(R为像素点的亮度，彩图中像素点的R值为该像素点的RGB三色像素的值加权求和得到的，灰度图中像素点的R值为该像素点的亮度)，然后选定观测窗口的大小对I进行极大值滤波，就可以得到时空兴趣点的位置。通过上述方式能够快速准确的提取到训练视频中的时空兴趣点，供后续步骤使用。

[0065] 本发明实施例提供的一种基于字典学习的跨视角的人体行为识别方法，通过K-SVD算法对字典进行初始化，并基于初始化的字典对局部特征及全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，包括：

[0066] 基于K-SVD算法对局部特征及全局特征进行训练，得到初始化的字典；

[0067] 基于初始化的字典通过下列方程进行优化求解，得到优化后的稀疏编码X和完备的字典D：

[0068]
$$f=\sum_{c=1}^{C}\{\| Y_c-DX_c \|_2^2+\| Y_c-D\tilde{Q}_c\tilde{Q}_c^T X_c \|_2^2+\| \tilde{Q}_{/c}^T X_c \|_2^2+\lambda_1\phi(X_c)+\lambda_2 \| X_c \|_2^2 + \beta \| p_c-AX_c \|_2^2\}+\alpha\sum_{c=1}^{C+1}\sum_{\substack{j=1\\ j\neq c}}^{C+1}Q(D_c,D_j)$$

[0069] 其中,C表示训练视频中包含的动作类的个数,Y_c表示训练视频中包含的第c个动作类,D表示字典,X_c表示训练视频中包含的第c个动作类所对应的稀疏编码,p_c表示训练视频中包含的第c个动作类的理想的稀疏编码,具体来说X_c和p_c分别为稀疏编码和理想的稀疏编码,用这两个值构造均方误差,使得构造误差最小,就可以使得X_c向理想的稀疏编码靠近,从而提高稀疏编码的精确度。D_c和D_j都表示特性字典,c和j分别表示为第c个特性字典和第j个特性字典的序号,$Q_c=\left[q_1^c,\ldots q_j^c\ldots q_{k_c}^c\right]\in R^{k\times k_c}$,其中$k\times k_c$表示$Q_c$的维度(分别表示$Q_c$的行数和列数),$q_j^c$表示一个维度与$Q_c$维度一样的矩阵,且只有第$k_c$行第j列的值为1,其他值都为零,以使得$D_c=DQ_c$,$\tilde{Q}_c=[Q_c,Q_{C+1}]$,$\tilde{Q}_c^T$为$\tilde{Q}_c$的转置,$\tilde{Q}_{/c}=[Q_1,\ldots,Q_{c-1},Q_{c+1},\ldots,Q_C]$,$\lambda_1$、$\lambda_2$、α及β为预先设定的系数,A为对应线性转化矩阵,$\phi(X_c)=\sum_{i=1}^{N_c}\| x_i^c \|_1$ x_i^c表示第i个视角对应的特性字典D_c的稀疏表示,$Q(D_c,D_j)=\| D_i^T D_j \|^2$,Nc表示训练视频对应的视角个数。

[0070] 令$Y=[Y_1,\ldots,Y_N]\in R^{n\times N}$是一组n维的N个输入信号的特征空间表示,假设字典D的大小为K并且已知,对于Y的稀疏表示$X=[X_1,\ldots,X_N]\in R^{K\times N}$可以有下方程解决:

[0071] $$X=\arg\min_X \| Y-DX \|_F^2 \quad s.t.\forall i,\| xi \|_0\leq s$$

[0072] 其中$\| Y-DX \|_F^2$表示构造误差,“F”表示Frobenius范数,$||xi||_0\leq s$要求少于或等于s个的分解元素x。

[0073] 而字典学习的过程可以包括:

[0074] $$(D,X)=\arg\min_{D,X} \| Y-DX \|_2^2 \quad s.t.\forall i,\| xi \|_0\leq s$$

[0075] 其中$D=[D_1,\ldots,D_C]\in R^{C\times k}$是要经过学习获得的,Y的稀疏表示为$X=[X_1,\ldots,X_N]$,可以通过K-SVD方法学习获得完备的字典。

[0076] 假定在数据源有C个类的动作类$Y=[Y_1,\ldots Y_c\ldots,Y_C]\in R^{d\times N}$,其中$Y_c\in R^{d\times Nc}$($N=\sum_{c=1}^{C}N_c$)表示数据集中的第c个动作类,$y_i^c\in N_c$表示第i个视角下的第c个类的信号。用$D_{C+1}$表示共性字典,则可以得到一个完备的字典$D=[D_1,\ldots D_c\ldots D_C,D_{C+1}]\in R^{d\times K}$,其中$K=\sum_{c=1}^{C+1}K_c$,$D_{C+1}\in R^{d\times k_{C+1}}$,$Dc\in R^{d\times k_c}$表示第c个动作类的特性字典。现在假设有稀疏编码$X=[X_1,\ldots,X_N]\in R^{K\times N}$使得$Y_i\approx DX_i$,$X_i=[x_i^1,\ldots x_i^c\ldots x_i^C,x_i^{C+1}]\in R^k$,$x_i^c\in R^{Kc}$是第i个视角所对应的子字典$D_c$的稀疏表示。I表示相对应的单位矩阵。定义目标方程f:

[0077] $$f=\sum_{c=1}^{C}\sum_{i=1}^{N_c}\{\| Y_i-DX_i \|_2^2+\lambda_1 \| X_i \|_1+\lambda_2 \| X_i \|_2^2+\| Y_i-D_c x_i^c-D_{C+1}x_i^{C+1} \|_2^2\}$$

[0078] 定义一种选择操作：$Q_c=[q_1^c,\dots q_j^c \dots q_{k_c}^c]\in R^{k\times k_c}$

[0079] 其中：

[0080] $$q_j^c=[\underbrace{0,\dots,0}_{\sum_{m=1}^{c-1}k_m},\overbrace{\underbrace{0,\dots,0}^{j-1}\,1,0,\dots,0}^{k_c},\underbrace{0,\dots,0}_{\sum_{m=c+1}^{c+1}k_m},]^T$$

[0081] 所以有：

[0082] $$Q_c^T Q_c=I$$

[0083] $D_c=DQ_c$

[0084] $$X_i^c=Q_c^T X_i\in R^{k_c}$$

[0085] 定义$\tilde{Q}_c=[Q_c,Q_{C+1}]$，$\begin{pmatrix}x_i^c\\x_i^{C+1}\end{pmatrix}=\tilde{Q}_c^T x_i\in R^{k_c+k_{C+1}}$，对于：$X_c=[x_1^c,\dots,x_{N_c}^c]$

[0086] 令：

[0087] $$\phi(X_c)=\sum_{i=1}^{N_c}\| x_i^c \|_1$$

[0088] 因此更新目标方程f为：

[0089] $$f=\sum_{c=1}^{C}\{\| Y_c-DX_c \|_2^2+\| Y_c-D\tilde{Q}_c\tilde{Q}_c^T X_c \|_2^2+\lambda_1\phi(X_c)+\lambda_2 \| X_c \|_2^2\}$$

[0090] 然而，仅仅做到这些去学习有区分的字典是不够的，因为其他特定的字典可能跟第c类的字典共享一些基，例如，来自不同特定字典的元素仍然可能是一致的，因此可以互相交换表示查询数据。为了避免这个问题，使得除了那些对应于特定字典和共性字典除外的系数全为零。令：

[0091] $Q/c=[Q_1,\dots,Q_{c-1},Q_{c+1},\dots,Q_C,Q_{C+1}]$

[0092] $$\tilde{Q}_{/c}=[Q_1,\dots,Q_{c-1},Q_{c+1},\dots,Q_C]$$

[0093] 然后令：

[0094] $$\| \tilde{Q}_{/c}^T X_c \|_F^2=0$$

[0095] 就可以得到如下目标方程式：

[0096] $$f=\sum_{c=1}^{C}\{\| Y_c-DX_c \|_2^2+\| Y_c-D\tilde{Q}_c\tilde{Q}_c^T X_c \|_2^2+\| \tilde{Q}_{/c}^T X_c \|_F^2+\lambda_1\phi(X_c)+\lambda_2 \| X_c \|_2^2\}$$

[0097] 该方程式可能无法获取字典的共性模式，例如，真实共性模式的基础可以出现几

个特性，这样使得学习特性冗余和有较少的区别性，所以加入 $Q(D_c,D_j)=\| D_i^T D_j \|_F^2$ 到上述目标方程式，同时将字典分割成不相交的子集，使得每一个子集负责一个视频动作类，也就是说用相同的子集代表同一动作，用不同的子集代表不同的动作，所以在目标方程式中加入 $\| p_c - AX_c \|_2^2$，其中 $p_i=[p_{i1}^T,\dots,p_{ik}^T,\dots,p_{iK}^T]^T \in R^{j\times 1}$，$p_{ik} \in R^{j_k\times 1}$ 是一个基于有标签的 y_i^c 的理想的区分稀疏编码，如果 y_i^c 来自于第k个类，就令$p_{ik}=1$，而其他的$p_i=0$，其中$A\in R^{J\times J}$是一个线性转化矩阵，把原始的稀疏编码x_i转换到相似的p_i。所以可以得到如下目标方程。

[0098]
$$f=\sum_{c=1}^{C}\{\| Y_c - DX_c \|_2^2+\| Y_c - D\tilde{Q}_c\tilde{Q}_c^T X_c \|_2^2+\| \tilde{Q}_{/c}^T X_c \|_2^2+\lambda_1\phi(X_c)+\lambda_2\| X_c \|_2^2 + \beta\| p_c - AX_c \|_2^2\}+\alpha\sum_{c=1}^{C+1}\sum_{\substack{j=1\\ j\neq c}}^{C+1} Q(D_c,D_j)$$

[0099] 其中特性字典为具有特殊属性，区别于其他字典的字典，例如视频里的人有一个动作，从不同的角度去观看所产生的效果是不一样的，所以每个视角就会存在差异，有自己特殊的性质；而共性字典就是每个字典里面的相同的部分，例如从不同的角度去观测一个人的动作，虽然角度发生了变化，但是终究只是一个人的行为动作，不管从哪个角度观察，本质上还是同一个动作，所以每个视角所对应的字典是存在共同的属性，简称共性。

[0100] 三：对目标方程的优化：

[0101] 对此目标方程的优化分为如下步骤：

[0102] 1、固定字典D和A，计算稀疏编码X；

[0103] 2、固定稀疏编码X和A，计算字典D；

[0104] 3、固定字典D和系数编码X，计算矩阵A。

[0105] 具体步骤：

[0106] 1、计算稀疏编码X：

[0107] 可以把目标方程写成如下方程式：

[0108]
$$f=\sum_{c=1}^{C}\{\| Y_c - DX_c \|_2^2+\| Y_c - D\tilde{Q}_c\tilde{Q}_c^T X_c \|_2^2+\| \tilde{Q}_{/c}^T X_c \|_2^2+\lambda_1\phi(X_c)+\lambda_2\| X_c \|_2^2 + \beta\| p_c - AX_c \|_2^2\}$$

[0109] 可以把上式用如下方程式表示：

[0110]
$$f=\sum_{c=1}^{C}\| \tilde{Y}_c - \tilde{D}X_c \|_2^2+\lambda\phi(X_c)$$

[0111] 其中：

[0112] $$\tilde{Y}_c = \begin{pmatrix} Y_c \\ Y_c \\ 0 \\ 0 \\ \sqrt{\beta} p_c \end{pmatrix}$$

[0113] $$\tilde{D} = \begin{pmatrix} D \\ D\tilde{Q}_c\tilde{Q}_c^T \\ \tilde{Q}_{/c}^T \\ I \\ \sqrt{\beta} A \end{pmatrix}$$

[0114] I为单位矩阵。

[0115] 优化上式是一个多任务组的套索问题，把每一个视角看成一个任务，使用SLEP(Sparse Learning With Efficient Projections)计算出稀疏编码X。

[0116] 2、计算字典D：

[0117] 可以把目标方程写成如下：

[0118] $$f = \sum_{c=1}^{C} \left\{ \| Y_c - DX_c \|_2^2 + \| Y_c - D\tilde{Q}_c\tilde{Q}_c^T X_c \|_2^2 \right\} + \alpha \sum_{c=1}^{C+1} \sum_{\substack{j=1 \\ j \neq c}}^{C+1} Q(D_c, D_j)$$

[0119] 为了更新字典$D=[D_1,...D_c...D_C,D_{C+1}]$，使用逐步迭代方法，比如更新$D_c$，先固定其他的子字典$D_i(i \neq c)$，由于共性字典$D_{c+1}$也有助第c个类的拟合，所以对$D_c$和$D_{c+1}$采用不同的优化，优化步骤如下：

[0120] 第一步：更新D_c：为了不失一般性，在更新D_c的时候固定其他的字典$D_i(i \neq c)$。对于$i=1,...,C+1$指定$X_c^i = Q_i^T X_c$，所以用如下方程跟新第c个类的D_c：

[0121] $$D_c = \underset{D_c}{\operatorname{argmin}} \left\{ \| Y_c - \sum_{\substack{i=1 \\ i \neq c}}^{C+1} D_i A_c^i - D_c A_c^c \|_2^2 + \| Y_c - D_{C+1} A_c^{C+1} - D_c A_c^c \|_2^2 + \sum_{\substack{j=1 \\ j \neq c}}^{C+1} Q(D_c, D_i) \right\}$$

[0122] 定义：$\bar{Y}_c = Y_c - D_{C+1} A_c^{C+1}$

[0123] $$\bar{X}_c = Y_c - \sum_{\substack{i=1 \\ i \neq c}}^{C+1} D_i A_c^i$$

[0124] $B=DQ_{/c}$

[0125] 所以就有如下方程式：

[0126] $$D_c = \underset{D_c}{\operatorname{argmin}} \{ \| \bar{Y}_c - D_c A_c^c \|_2^2 + \| \bar{X}_c - D_c A_c^c \|_2^2 + \alpha \| D_c^T B \|_2^2 \}$$

[0127] 接下来对 $D_c = [d_c^1, ..., d_c^{Kc}]$ 的元素逐个的进行更新，例如更新 d_c^k 时固定其他的元素，令$X_c=[x_1,...,x_{Kc}]$，其中 $x_k \in R^{1 \times N_c}$ 是 A_c^c 的第k行，令：

[0128] $$\hat{Y}_c = \bar{Y}_c - \sum_{j \neq k} d_c^j x_j$$

[0129] $$\hat{X}_c = \bar{X}_c - \sum_{j \neq k} d_c^j$$

[0130] 可以得出：

[0131] $$d_c^k = arg\min_{d_c^k} \left\{ g(d_c^k) \equiv \| \hat{Y}_c - d_c^k x_k \|_2^2 + \| \hat{X}_c - d_c^k x_k \|_2^2 + \alpha \| d_c^{kT} B \|_2^2 \right\}$$ 用 $g(d_c^k)$ 对 d_c^k 进行求导并令其等于0，然后可以得到：

[0132] $$d_c^k = \frac{1}{2} (\| x_k \|_2^2 I + \alpha B B^T)^{-1} (\hat{Y}_c + \hat{X}_c) x_k^T$$

[0133] 作为字典的原子，应当被单位化，所以有：

[0134] $$\hat{d}_c^k = d_c^k / \| d_c^k \|_2$$

[0135] 所以对应的系数应该乘以 $\| d_c^k \|_2$，即 $\hat{x}_k = \| d_c^k \|_2 x_k$。

[0136] 第二步、更新D_{c+1}：

[0137] 令：

[0138] $B = DQ_{/C+1}$

[0139] 得到如下的方程式：

[0140] $$D_{C+1} = \underset{D_{C+1}}{\operatorname{argmin}} \sum_{c=1}^{C} \left\{ \| Y_c - \sum_{i=1}^{C+1} D_i X_c^i - D_{C+1} X_{C+1}^{C+1} \|_2^2 + \| Y_c - D_c X_c^c - D_{C+1} X_{C+1}^{C+1} \|_2^2 \right\} + \alpha \| D_{C+1}^T B \|_2^2$$

[0141] 令

[0142] $$\bar{Y}_c = Y_c - \sum_{i=1}^{C} D_i X_c^i$$

[0143] $$\bar{X}_c = Y_c - D_c X_c^c$$

[0144] 可以得到如下方程：

[0145] $$D_{C+1} = \underset{D_{C+1}}{\operatorname{argmin}} \sum_{c=1}^{C} \{ \| \bar{Y} - D_{C+1} X^{C+1} \|_2^2 + \| \bar{X} - D_{C+1} X^{C+1} \|_2^2 \} + \alpha \| D_{C+1}^T B \|_2^2$$

[0146] 其中

[0147] $$X^{C+1} = [X_1^{C+1}, \ldots, X_C^{C+1}]$$

[0148] $$\bar{Y} = [\bar{X}_1, \ldots \bar{X}_C]$$

[0149] $$\bar{X} = [\bar{Y}_1, \ldots, \bar{Y}_C]$$

[0150] 可以对D_{c+1}的元素进行逐个更新：

[0151] $$d_{C+1}^{k}=\frac{1}{2}(\| x_k \|_2^2 I+\alpha BB^T)^{-1}(\hat{Y}_c+\hat{X}_c)x_k^T$$

[0152] 其中：

[0153] $$X^{C+1}=[x_1,\ldots,x_{k_{C+1}}]\in R^{k_{C+1}\times N}$$

[0154] $$\hat{X}_c=\bar{X}_c-\sum_{j\neq k}d_{C+1}^{j}x_j$$

[0155] $$\hat{Y}_c=\bar{Y}_c-\sum_{j\neq k}d_{C+1}^{j}x_j$$

[0156] 同理所以有：

[0157] $$\hat{d}_{C+1}^{k}=d_{C+1}^{k}/\| d_{C+1}^{k} \|_2$$

[0158] 所以对应的系数应该乘以 $\| d_{C+1}^{k} \|_2$，即 $\hat{x}_k=\| d_{C+1}^{k} \|_2 x_k$。

[0159] 3、计算A：

[0160] 用如下方程式计算A：

[0161] $$\min_A \sum_{c=1}^{C+1}\sum_{i=1}^{N_c}\| p_i-Ax_i^c \|_2^2+\lambda_1\| A \|_2^2$$

[0162] $$A^*=P\sum_{i=1}^{N_c}X^{iT}(\sum_{i=1}^{N_c}X^iX^{iT}+\lambda_1 I)^{-1}$$

[0163] $P=[p_1,\ldots,p_{C+1}]$

[0164] $$X=[x_1,\ldots,x_{N_c}]$$

[0165] 上述公式中，Y表示特征空间表示，X表示稀疏编码，D表示字典，N表示输入信号的个数，Nc表示视角的个数，$Y_c\in R^{d\times Nc}$（$N=\sum_{c=1}^{C}N_c$）表示数据集中的第c个动作类，$y_i^c\in N_c$表示第i个视角下的第c个类的信号，D_{C+1}表示共性字典，$D=[D_1,\ldots D_c\ldots D_C,D_{C+1}]\in R^{d\times K}$表示完备的字典，$Dc\in R^{d\times k_c}$表示第c个动作类的特性字典，$x_i^c\in R^{Kc}$表示第i个视角所对应的子字典Dc的稀疏表示。

[0166] 简单来说，上述算法的实现过程可以表示为：

[0167] 1:Input:$Y=[Y_1,\ldots Y_c\ldots,Y_C]$,$\lambda_1$,$\lambda_2$,α,β,P

[0168] 2:Initialize $D=[D_1,\ldots D_c\ldots D_C,D_{C+1}]$by K-SVD

[0169] 3:Repeat

[0170] 4:Compute spare codes X by(1)

[0171] 5:Updating D using(2)and(3)

[0172] 6:Updating A using(4)

[0173] 7:until convergence of certain rounds

[0174] 8:Output:D=[D_1,...D_c...D_C,D_{C+1}]

[0175] 其中算法中的各公式即为上文中包含的：

[0176] $$f=\sum_{c=1}^{C}\|Y_c-\tilde{D}X_c\|_2^2+\lambda\phi(X_c) \quad (1)$$

[0177] $$d_c^k=\frac{1}{2}\left(\|x_k\|_2^2 I+\alpha BB^T\right)^{-1}\left(\hat{Y}_c+\hat{X}_c\right)x_k^T \quad (2)$$

[0178] $$d_{C+1}^k=\frac{1}{2}\left(\|x_k\|_2^2 I+\alpha BB^T\right)^{-1}\left(\hat{Y}_c+\hat{X}_c\right)x_k^T \quad (3)$$

[0179] $$A^*=P\sum_{i=1}^{N_c}X^{iT}\left(\sum_{i=1}^{N_c}X^iX^{iT}+\lambda_1 I\right)^{-1} \quad (4)$$

[0180] $$P=[p_1,\ldots,p_{C+1}],\ X=\left[x_1,\ldots,x_{N_c}\right]$$

[0181] 另外需要说明的是，本发明公开的上述技术方案中使用到的算法或者执行步骤未完全阐述清楚的部分均与现有技术中的对应算法或者执行步骤的实现原理一致，在此不做过多赘述。

[0182] 本发明实施例还提供了一种基于字典学习的跨视角人体行为识别装置，如图2所示，可以包括：

[0183] 提取模块11，用于：由训练视频中提取时空兴趣点，训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频；

[0184] 处理模块12，用于：利用k-means算法对时空兴趣点进行聚类得到对应的局部特征，并利用光流法对训练视频进行操作得到对应的全局特征；

[0185] 训练模块13，用于：通过K-SVD算法对字典进行初始化，并基于初始化的字典对局部特征及全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，其中，不同视角下的同一动作类具有相似度大于预设值的稀疏编码，任何视角下不同的动作类具有相似度小于预设值的稀疏编码。

[0186] 本发明实施例提供的一种基于字典学习的跨视角人体行为识别装置，还可以包括：

[0187] 降维模块，用于：在利用k-means算法对时空兴趣点进行聚类得到对应的局部特征之前，使用PCA技术对提取得到的时空兴趣点进行降维操作。

[0188] 本发明实施例提供的一种基于字典学习的跨视角人体行为识别装置，提取模块可以包括：

[0189] 提取单元，用于：利用高斯平滑函数和Gabor滤波器由训练视频中提取时空兴趣点。

[0190] 本发明实施例提供的一种基于字典学习的跨视角的人体行为识别装置，训练模块可以包括：

[0191] 训练单元，用于：基于K-SVD算法对局部特征及全局特征进行训练，得到初始化的字典；

[0192] 基于初始化的字典通过下列方程进行优化求解，得到优化后的稀疏编码X和完备的字典D：

[0193]
$$f=\sum_{c=1}^{C}\{\| Y_c-DX_c\|_2^2+\| Y_c-D\tilde{Q}_c\tilde{Q}_c^TX_c\|_2^2+\| \tilde{Q}_{/c}^TX_c\|_2^2+\lambda_1\phi(X_c)+\lambda_2\| X_c\|_2^2 + \beta\| p_c-AX_c\|_2^2\}+\alpha\sum_{c=1}^{C+1}\sum_{\substack{j=1\\ j\neq c}}^{C+1}Q(D_c,D_j)$$

[0194] 其中，C表示训练视频中包含的动作类的个数，Y_c表示训练视频中包含的第c个动作类，D表示字典，X_c表示训练视频中包含的第c个动作类所对应的稀疏编码，p_c表示训练视频中包含的第c个动作类的理想的稀疏编码，D_c和D_j都表示特性字典，c和j分别表示为第c个特性字典和第j个特性字典的序号，$Q_c=\left[q_1^c,\dots q_j^c\dots q_{k_c}^c\right]\in R^{k\times k_c}$，其中$k\times k_c$表示$Q_c$的维度，$q_j^c$表示一个维度与$Q_c$维度一样的矩阵，且只有第$k_c$行第j列的值为1，其他值都为零，$\tilde{Q}_c=[Q_c,Q_{C+1}]$，$\tilde{Q}_c^T$为$\tilde{Q}_c$的转置，$\tilde{Q}_{/c}=[Q_1,\dots,Q_{c-1},Q_{c+1},\dots,Q_C]$，$\lambda_1$、$\lambda_2$、α及β为预先设定的系数，A为对应线性转化矩阵，$\phi(X_c)=\sum_{i=1}^{N_c}\| x_i^c\|_1$ x_i^c表示第i个视角对应的特性字典D_c的稀疏表示，$Q(D_c,D_j)=\| D_i^TD_j\|^2$，Nc表示训练视频对应的视角个数。

[0195] 本发明实施例提供的一种基于字典学习的跨视角人体行为识别装置中相关部分的说明请参见本发明实施例提供的一种基于字典学习的跨视角人体行为识别方法中对应部分的详细说明，在此不再赘述。

[0196] 对所公开的实施例的上述说明，使本领域技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下，在其它实施例中实现。因此，本发明将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

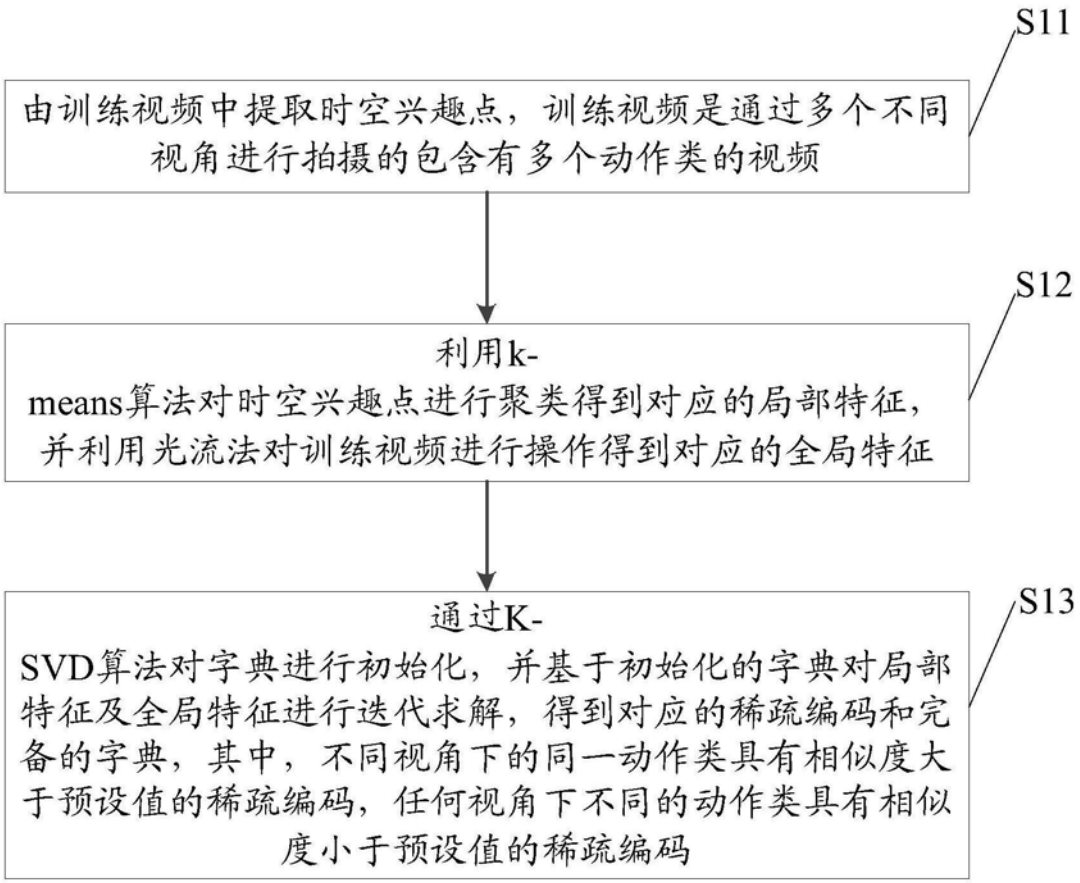

S11
由训练视频中提取时空兴趣点，训练视频是通过多个不同视角进行拍摄的包含有多个动作类的视频
S12
利用k-means算法对时空兴趣点进行聚类得到对应的局部特征，并利用光流法对训练视频进行操作得到对应的全局特征
S13
通过K-SVD算法对字典进行初始化，并基于初始化的字典对局部特征及全局特征进行迭代求解，得到对应的稀疏编码和完备的字典，其中，不同视角下的同一动作类具有相似度大于预设值的稀疏编码，任何视角下不同的动作类具有相似度小于预设值的稀疏编码

图1

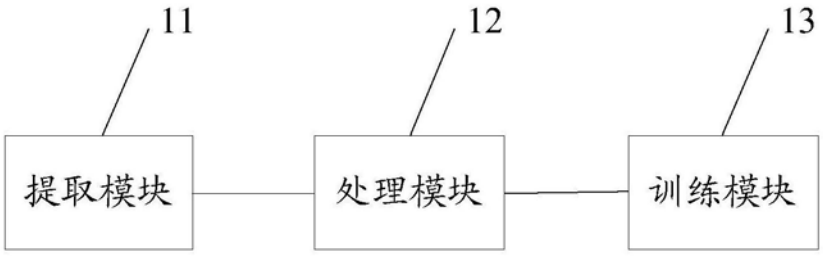

11
12
13
提取模块
处理模块
训练模块

图2