

# 基于模糊神经网络和图模型推理的动作识别

赵一丹,肖秦琨,高 嵩

(西安工业大学 电子信息工程学院,陕西 西安 710021)

**摘 要:**提出了一种基于模糊神经网络和图模型推理的人体动作识别方法。该方法将模糊神经网络和概率图模型推理有效地结合起来,目的在于能够更加准确、容易地对复杂视频动作序列进行识别,获得较高的动作识别精度。该方法由系统学习阶段和动作识别阶段两部分组成。首先,在系统学习阶段,构建了一个动态语义识别的层次图模型结构;第一级是基于模糊神经网络的关键帧动作识别图模型,用于关键帧的动作识别;第二级是关键帧序列分类模型,用于关键帧序列的语义识别。其次,在动作识别阶段,使用模糊神经网络和图模型推理算法进行动作识别,从而得到识别结果。最后,通过对实验结果的分析比较可以看出,相比现有的人体动作识别方法,该方法具有更好的识别性能,识别结果精度更佳。

**关键词:**动作识别;模糊神经网络;图模型;推理;概率

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2018)08-0080-06

doi:10.3969/j.issn.1673-629X.2018.08.017

## Action Recognition Based on Fuzzy Neural Network and Graph Model Inference

ZHAO Yi-dan, XIAO Qin-kun, GAO Song

(School of Electronic Information Engineering, Xi'an University of Technology, Xi'an 710021, China)

**Abstract:** We propose a novel action recognition method based on fuzzy neural network and graph model (FNNGM). It combines the fuzzy neural network and the probability graph model reasoning effectively, in order to be able to identify the complex video action sequence more accurately and easily and obtain a higher accuracy of action recognition. The algorithm consists of system learning and action recognition. In system learning, we firstly build a hierarchical graph model for action semantic recognition; the first level is the FNN-based representative frame (RF) gesture identification graph model which is utilized to recognize RF's gestures; the second level is RF sequence classification model which is used to identify final semantic of RF sequence. Then in action recognition, the action recognition results are computed using the FNNGM inference algorithm. Finally, through the analysis and comparison of the experimental results, it can be concluded that the proposed method has better recognition performance and accuracy than some existing methods.

**Key words:** action recognition; fuzzy neural network; graph model; inference; probability

## 0 引言

近年来,人体动作识别研究已经成为计算机视觉领域的核心问题。由于人体动作识别在空间上的复杂性和时间上的差异性,使其在识别研究领域并没有统一有效的研究框架,所以人体动作识别仍然是一个非常具有挑战性的课题。

许多动作识别方法侧重于设计有效的描述符呈现运动特征或提取运动特征,然后通过特征匹配的方法将待识别动作进行分类<sup>[1-2]</sup>。以往的动作识别主要包含两类:特征表示和动作分类。根据文献[3]了解到,

特征表示始终是动作识别过程中的关键任务。一般而言,运动序列特征的表示通常分为全局特征表示和局部特征表示。全局特征呈现总图像,但是,全局特征经常受到遮挡、视点变化和噪声的干扰。基于全局的特征表示包括基于光流的呈现<sup>[4]</sup>、基于剪影的描述符<sup>[5]</sup>、基于边缘的特征表示<sup>[6]</sup>和运动历史图像(MHI)<sup>[7]</sup>等。局部特征描述子可以更有效地呈现动作视频,尤其是处理噪声图像、部分遮挡等复杂的情况。

有研究人员利用许多常规方法对人体动作进行分类<sup>[1-2,8-9]</sup>。比如,文献[2]提出了一种SVM多类分类

收稿日期:2017-09-13

修回日期:2018-01-17

网络出版时间:2018-04-28

基金项目:国家自然科学基金(61671362,61271362);陕西省自然科学基金(2017JM6041)

作者简介:赵一丹(1993-),女,硕士研究生,研究方向为智能信息处理;肖秦琨,博士,教授,研究方向为图模型理论及智能信息处理;高 嵩,博士,教授,研究方向为智能信息处理。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180427.1640.048.html>

器,使用动态规划将运动序列分段。在文献[1]中,局部描述符被组合成用于动作识别的 SVM。在文献[8]中,使用 K-最近邻分类器来预测动作标签。然而,这些常规的识别方法难以及时捕获动态时空序列信息。

除了常规的分类器方法,还有条件随机场(CRF)<sup>[10]</sup>、隐马尔可夫模型(HMM)<sup>[11]</sup>、动态贝叶斯网络(DBN)<sup>[12]</sup>等基于图模型的多序列标注模型是识别动作的有效工具,适用于建模和分析人类行为。在文献[13]中,也提出利用线性图序列模型来构建语义和动作识别的非线性映射。

在上述研究的基础上,文中提出了一种复杂视频序列语义识别的图模型结构。基于以下三点考虑这一问题:(1)为了减少动作视频维度和识别时间,采用模糊 c-均值(fuzzy c-means,FCM)聚类方法<sup>[14-16]</sup>获得关键帧。所选择的辨别功能比原始功能具有更好的性能。基于 FCM,从原始动作视频获得更好的辨别功能。(2)为了提高鲁棒性,考虑使用层次图模型进行动作语义识别,基本识别级别是基于模糊神经网络的关键帧动作识别,较高的识别级别是基于概率图模型将关键帧动作序列进行分类。(3)同时,还发现概率图模型是挖掘隐藏状态信息的有效途径<sup>[17-19]</sup>。DBN是目前呈现随机时间序列信号很有效的工具,因此,选择模糊神经网络图模型(FNNGM)来呈现人体动作。基于图模型的信号处理可以完成三个任务:一是完全清楚地呈现动作;二是挖掘更多隐藏的语义信息;三是在此基础上,将高级语义分析和图模型结合起来,以获得有效的动作表示。

1 检索算法

1.1 算法概述

所提出的算法在图 1 中给出了图解说明,该算法那分为系统学习和动作识别两个阶段。在系统学习阶段,首先提取运动数据集中动作的帧图像。其次,使用模糊 c-均值聚类算法对帧图像进行聚类获取运动的

关键帧。设计了一个模糊神经网络分类器来识别关键帧的运动姿态,关键帧标记的姿态特征数据集被用作教师信号来训练模糊神经网络。最后,使用模糊神经网络图模型的概率图模型对关键帧序列进行分类。文中设计的图模型是基于隐马尔可夫模型和模糊神经网络的组合。在动作识别阶段,自动提取待识别动作的关键帧特征,然后使用 FNNGM 推理算法得到动作语义分类结果。

1.2 系统学习

通过系统学习来构建和训练 FNNGM 模型,从而进行人体运动序列动作识别。

1.2.1 运动序列预处理

首先,采用 FCM 通过聚类获取关键帧<sup>[20-22]</sup>。假设获取的一个动作其关键帧运动序列为  $\{rf_i\}_{i=1,2,\dots,k}$ ,其中  $rf_i$  表示第  $i$  个运动关键帧特征。可以使用  $s$  维四元数矢量  $rf_j = (x_1, x_2, \dots, x_n)$  来描述身体姿态。文中考虑在骨架中使用 26 个关节状态信息来呈现身体姿态,每个关节采用一个四元数描述符  $(1 \times 4)$  进行描述,其中  $n = 26 \times 4 = 104$ 。

其次,基于所有动作关键帧训练数据集。首先建立动作语义集  $S = \{S_1, S_2, \dots, S_\omega\}$ 。然后关键帧动作通过 K-means 进行姿态聚类以形成相似运动姿态集合  $G = \{G_1, G_2, \dots, G_q\}$ ,其中  $q$  是聚类组的个数。

1.2.2 使用 FNN 对关键帧进行识别

一般而言,动作语义是由不同关键帧组合决定的,这意味着不同的身体姿态组合代表不同的动作语义。文中首先构建一个 T-S 模糊神经网络模型,来识别由每个关键帧确定的身体姿态语义,此外,使用概率图推理来分类动作序列语义。使用数据  $(rf_i, G^j), i = 1, 2, \dots, n, j = 1, 2, \dots, p$  对 FNN 进行训练,具体过程如下:

T-S 模糊系统具有较强的适应性能,并且能够自我更新,所以可以自动地修改模糊子集的隶属函数。T-S 模糊系统的使用遵循“if-then”规则,规则如下:

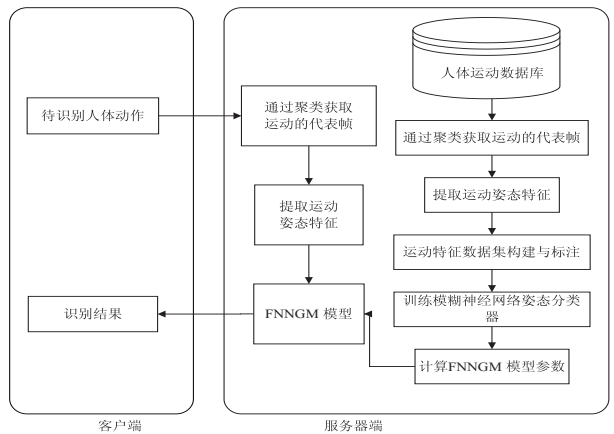
if  
$$x_1 = A_1^i, x_2 = A_2^i, \dots, x_k = A_k^i$$
  
then  
$$y_i = p_0^i + p_1^i x_1 + \dots + p_k^i x_k$$

其中,  $A_j^i$  为模糊集;  $p_j^i$  为系统参数;  $y_i$  为模糊规则下的推理输出。

假定输入  $rf_i = (x_1, x_2, \dots, x_n)$ , 则根据模糊规则可计算  $x_j$  的模糊隶属度为:

$$u_{A_j^i}(x_j) = \exp(-(x_j - c_j^i)^2/b_j^i), j = 1, 2, \dots, k, i = 1, 2, \dots, n \tag{1}$$

其中,  $c_j^i, b_j^i$  分别为隶属函数中心和宽度;  $k$  为输入参数的数量;  $n$  为模糊集的数量。



万方数据 图 1 算法流程

则  $\text{rf}_i$  的模糊隶属度计算公式为:

$$w^i = u_{A_j^i}(x_1) * u_{A_j^i}(x_2) * \cdots * u_{A_j^i}(x_k), j = 1, 2, \cdots, n \quad (2)$$

同时,输出  $y_i$  计算为:

$$y_i = \sum_{i=1}^n \omega^i (p_0^i + p_1^i x_1 + \cdots + p_k^i x_k) / \sum_{i=1}^n \omega^i \quad (3)$$

T-S 模糊神经网络分为四层,包括输入层、模糊层、模糊规则层和输出层。在输入层中没有计算操作,在模糊层使用式 1 计算模糊隶属度,然后根据式 2 得到规则推理结果,最终通过式 3 计算输出。FNN 训练包括以下三个步骤:

(1) 误差计算。

$$e = \frac{1}{2} (y_d - y_c)^2 \quad (4)$$

其中,  $y_d$  是期望输出;  $y_c$  是实际输出;  $e$  是期望输出与实际输出之间的误差。

(2) 系数修正。

$$p_j^i(k) = p_j^i(k-1) - \alpha \frac{\partial e}{\partial p_j^i} \quad (5)$$

$$\frac{\partial \mathcal{E}}{\partial p_j^i} = (y_c - y_d) \omega^i / \sum_{i=1}^m \omega^i x_j$$

其中,  $p_j^i$  是系数;  $\alpha$  是学习效率;  $\omega^i$  是输入量  $x_j$  和  $u_{A_j^i}$  的乘积。

(3) 参数修正。

$$\begin{cases} c_j^i(k) = c_j^i(k-1) - \beta \frac{\partial \mathcal{E}}{\partial c_j^i} \\ b_j^i(k) = b_j^i(k-1) - \beta \frac{\partial \mathcal{E}}{\partial b_j^i} \end{cases} \quad (6)$$

其中,  $c_j^i, b_j^i$  分别是隶属函数中心和宽度。

### 1.3 图模型和推理

#### 1.3.1 图模型构建

FNNGM 模型如图 2 所示。

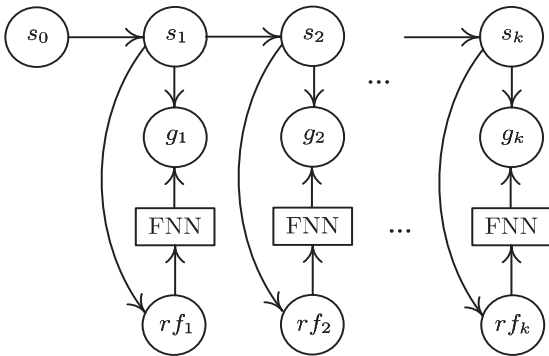


图 2 用于动作识别的 FNNGM 图模型结构

该模型包括三层:最低层是测量信号层,从最低层输入 RF 序列(  $\text{rf}_1, \text{rf}_2, \cdots, \text{rf}_k$  );中间层是运动姿势层,在这一层呈现由 FNN 推断的离散姿势状态信号(  $g_1, g_2, \cdots, g_k$  );最高层是动作语义层,根据中间层

和最低层的信息,在该层中计算最终动作语义(  $s_1, s_2, \cdots, s_k$  )。

FNNGM 的参数定义为  $\lambda = (\Pi_s, A_s, B_G, B_{rf})$ , 其中  $\Pi_s = [\pi_i]_{1 \times \omega}$  是先验概率矩阵,  $\omega$  是动作语义的数量,  $A_s = [a_{ij}]_{\omega \times \omega}$  是语义状态转换矩阵,  $B_G = [b_{ij}^G]_{\omega \times \gamma}$  是姿势观察矩阵,  $B_{rf} = [b_{ij}^{rf}]_{\omega \times 1}$  是 RF 观察矩阵。此外,所有矩阵的元素定义为:

$$\begin{cases} \pi_i = P(s_0) = P(s_0 = S^i), S^i \in S \\ a_{ij} = P(S_{i+1} | S^i) = P(S_{i+1} = S^j | S^i = S^i), S^i, S^j \in S \\ b_{ij}^{rf} = P(\text{rf}_i | s_i) = P(\text{rf}_i | s_i = S^i) = N(\mu_{s_i}, \Sigma_{s_i}) (\text{rf}_i), \\ b_{ij}^G = P(g_i | s_i) = P(g_i = G^j | s_i = S^j), G^i \in G, S^j \in S \end{cases} \quad (7)$$

其中,  $S = \{S^1, S^2, \cdots, S^w\}$  和  $G = \{G^1, G^2, \cdots, G^p\}$  分别是动作语义集和相似姿态集合。

基于  $S$  和  $G$  中的  $\text{rf}_i$  数量,计算图模型参数。首先,先验概率的计算为:

$$P(S_i = S^i) \approx \frac{n_{\text{rf} \in S^i}}{n_{\text{rf}}} \quad (8)$$

其中,  $n_{\text{rf} \in S^i}$  为属于语义  $S^i$  的关键帧数;  $n_{\text{rf}}$  为训练数据集中所有关键帧的数量。

其次,估计语义交易可能性  $a_{ij}$ :

$$a_{ij} = P(s_{i+1} = S^j | s_i \in S^i) \approx \frac{n_{S^j | S^i}}{n_{S | S^i}} \quad (9)$$

其中,  $n_{S^j | S^i}$  表示训练数据中从  $\text{rf}_i$  (  $\text{rf}_i \in S^i$  ) 到  $\text{rf}_{i+1}$  (  $\text{rf}_{i+1} \in S^j$  ) 相邻链接的数目;  $n_{S | S^i}$  表示训练数据中从  $\text{rf}_i$  (  $\text{rf}_i \in S^i$  ) 到  $\text{rf}_{i+1}$  (  $\text{rf}_{i+1} \in S^j$  ,  $S$  所有可能的语义)的邻接链路的数量。

第三,观测可能性  $b_{ij}^{rf}$  是多变量高斯分布,其中参数  $\mu_{s^i}$  和  $\Sigma_{s^i}$  分别是期望和方差,可以通过最大似然 (ML) 方法来学习。

最后,计算观测可能性  $b_{ij}^G$ :

$$b_{ij}^G = P(g_i = G^j | s_i = S^i) \approx \frac{n_{G^j | S^i}}{n_{S^i}} \quad (10)$$

其中,  $n_{G^j | S^i}$  表示训练数据中  $\text{rf}_i$  (  $\text{rf}_i \in S^i \cap G^j$  ) 的数目;  $n_{S^i}$  表示训练数据中  $\text{rf}_i$  (  $\text{rf}_i \in S^i$  ) 的数量。

基于初始参数估计,进一步使用 Baum-Welch 算法优化参数,对数似然值对应于参数学习的全局收敛性能,当对数似然值变得稳定时,意味着参数收敛到最优值。

#### 1.3.2 图模型推理

基于所构建的 FNNGM 图模型结构进行图模型推理,对动作语义序列进行分类。将  $\{\text{rf}_{1:k}\}$  作为一个测量的图模型。系统状态序列为  $S = \{s_{1:k}\}$  和  $G = \{g_{1:k}\}$ , 基于图模型推理的动作识别描述如下:

给定检测量  $\{\text{rf}_{1:k}\}$ , 以更新隐藏状态信号  $S =$

$\{s_{1:k}\}$  和  $g = \{h_{1:k}\}$ 。首先,根据贝叶斯规则,计算  $P(s_1)$  :

$$P(s_1) = \sum_{s_0=1}^{\omega} P(s_1 | s_0) P(s_0) \quad (11)$$

将  $s_0$  和  $P(s_1 | s_0)$  设置为初始系统输入,因此,根据式 8、9、11 可以得到  $P(s_1)$ 。

接下来,预测 C 使用检测量更新为:

$$\begin{cases} P(s_1 | \text{rf}_1) = \frac{P(s_1 | \text{rf}_1) P(s_1)}{P(\text{rf}_1)} = \alpha P(s_1 | \text{rf}_1) P(s_1) \\ P(s_1 | g_1) = \frac{P(g_1 | s_1) P(s_1)}{P(g_1)} = \alpha P(g_1 | s_1) P(s_1) \end{cases} \quad (12)$$

同时,确保  $\alpha$  的输出结果在区间  $[0,1]$  中。此外,通过测量结果,滤波方程可以重新写为:

$$\begin{aligned} P(s_{1+t} | g_{1:1+t}, \text{rf}_{1:1+t}) &= \\ P(s_{1+t} | \text{rf}_{1:t}, \text{rf}_{1:t}) P(s_{1+t} | g_{1:t}, g_{1:t}) &= \\ \alpha P(\text{rf}_{1:t} | s_{1:t}, \text{rf}_{1:t}) & \\ \int P(s_{1+t} | \text{rf}_{1:t}) P(g_{1+t} | s_{1+t}) \sum_{g(t)} P(s_{1+t} | g_{1:t}) &= \\ \alpha P(\text{rf}_{1:t} | s_{1:t}) \int P(s_{1+t} | s_t) P(s_t | \text{rf}_{1:t}) \times & \\ P(g_{1+t} | s_{1+t}) \sum_{g(t)} P(s_{1+t} | s_t) P(s_t | g_{1:t}) & \quad (13) \end{aligned}$$

基于滤波方程(式 13),进一步得到最大可能状态路径:

$$\begin{aligned} \max_{s_1, s_2, \dots, s_t} P(s_{1:1+t} | g_{1:1+t}, \text{rf}_{1:1+t}) &= \\ \alpha P(\text{rf}_{1:t} | s_{1:t}) \max_{s_t} P(s_{1+t} | s_t) \max_{s_1, s_2, \dots, s_{t-1}} P(s_t | \text{rf}_{1:t}) \times & \\ P(g_{1+t} | s_{1+t}) \max_{s_t} P(s_{1+t} | s_t) \max_{s_1, s_2, \dots, s_{t-1}} P(s_t | g_{1:t}) & \quad (14) \end{aligned}$$

最终获得语义序列最大可能性状态值。

## 2 实验

为了评估所提出的动作识别方法的性能<sup>[23]</sup>,构建了一个人体运动数据库,包括 360 个不同的动作剪辑。动作剪辑大致分为 20 个动作语义组,这里选用四元数描述符<sup>[24]</sup>来呈现身体运动姿态。文中采用 26 个骨骼关节<sup>[25]</sup>呈现身体运动姿态,因此得到 104 维 ( $26 \times 4 = 104$ ) 的四元数描述符。

使用 FNNGM 模型来实现动作序列数据分类。利用两个评价标准以评估提出方法的性能。帧精度计算如下:

$$\text{Frame\_accuracy} = \frac{\# \text{correctly\_recognized\_frames}}{\# \text{frames}} \quad (15)$$

由于运动数据集中的视频被分割,每个视频中只有一种动作。每个预测序列中的类标签通过多数

表决计算。基于上述,视频精度可以计算如下:

$$\text{Video\_accuracy} = \frac{\# \text{correctly\_recognized\_videos}}{\# \text{videos}} \quad (16)$$

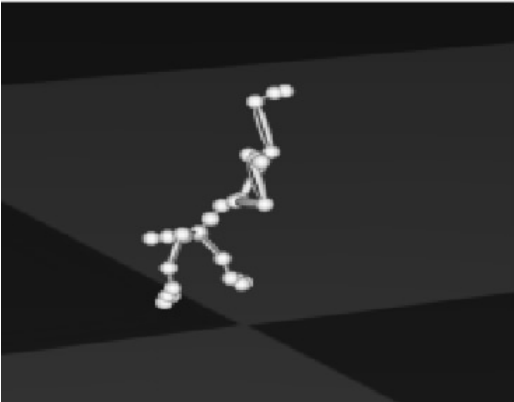
所有的比较方法可以分为两类:(1)包含 CRF 的序列模型、条件随机场 (CRF) 与单隐层神经网络 (NNCRF) 耦合的序列模型;(2)包含支持向量机 (SVM) 的非序列模型、单隐层神经网络的非序列模型。对基于帧的样本进行非序列模型训练。SVM 模型是具有线性内核的多类 SVM 模型。作为非线性模型,神经网络用一个隐层进行训练。对于序列模型,选择标准线性链 CRF 作为基线方法之一。选择非线性链 NNCRF 作为非线性序列模型的基线。使用一个隐藏层,枚举隐藏节点的数量。在线性链 CRF 和非线性 NNCRF 方法中都加入正则化术语,以避免在训练过程中过度拟合。调整正则化参数<sup>[27]</sup>,并通过交叉验证调整正则化的参数。将提出的方法在两种配置下进行测试。

动作识别示例如图 3 所示。其中,图 (a) 是动作的“cartwheelLHandStart1Reps”插图,图 (b) 是 HMM 参数学习的对数似然-时间图 (loglik-time)。基于先验知识估计初始参数,然后使用 Baum-Welch 算法优化参数,对数似然值对应于参数学习的全局收敛性能,当对数似然值变得稳定时,意味着参数收敛到最优值。图 (c) 是 HMM 参数矩阵,参数为  $\lambda = (\pi_s, A_s, B_G, B_{rf})$ ,因为  $B_{rf}$  包含太多参数,只将  $\pi_s, A_s$  和  $B_G$  组合成一个大矩阵在 (c) 中显示。其中,  $\pi_s$  为  $1 \times 20$  的矩阵,对应于 20 个动作语义;  $A_s$  为  $20 \times 20$  的矩阵;  $B_G$  为  $20 \times 25$  的矩阵,对应于 20 个动作语义中的 25 个待识别姿势组。图 (d) 是 FNNGM 的对数概率矩阵,对应于图模型推理对数概率值,在运动数据集中,动作“cartwheelLHandStart1Reps”包含 21 个样本,前 50% 样本为训练数据,其余样本为检测数据。检测数据的识别结果如图 (d) 所示,矩阵中的每一行对应于动作样本与不同 FNNGM 模型之间的匹配概率。在实验中,由于动作“carwheelLHandStart1Reps”在数据集中被标注为数字 #1 动作语义组,从识别结果看出,第一列具有最高匹配概率值。

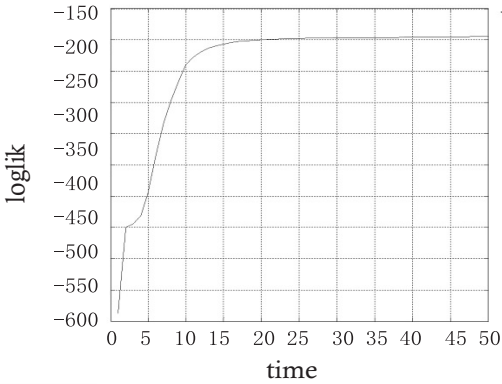
同时,为了评估 FNNGM 动作识别模型的识别性能,将其与基于构建框架的其他方法进行比较,结果如表 1 所示。从中可以得出以下结果:线性序列模型 (CRF) 优于线性支持向量机 (SVM),表明结构依赖于动作识别很重要;(2) 非线性模型比线性模型获得了更好的性能;(3) 通过将 CRF 与隐藏层相结合,NNCRF 模型进一步提高了标准 CRF 的性能;(4) 所有三个特征函数的组合比标准 CRF 特征函数组合有更



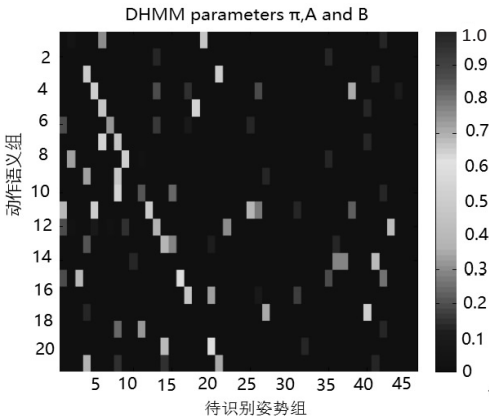
好的性能。因此,可以得出结论,文中方法得到的识别结果精度最佳。



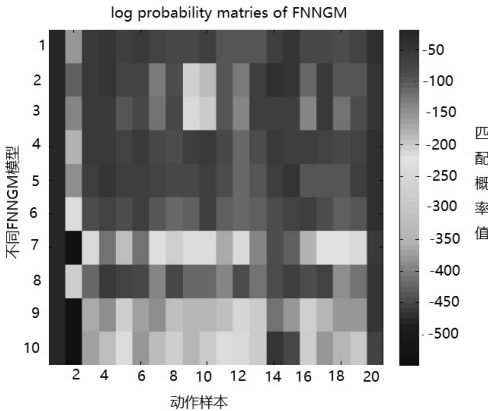
(a)动作说明(#1)



(b)对数概率与时间图



(c) HMM 参数矩阵



(d) 对数概率矩阵

图 3 FNNGM 模型动作识别示例

表 1 与现有方法的比较结果 %

算法	Frame_accuracy	Video_accuracy
SVM	76.3	89.5
NN	83.2	92.4
CRF	86.4	91.5
NNCRF	90.3	95.8
文中方法	92.5	96.2

3 结束语

提出了一种新的基于图模型和模糊神经网络的动作识别方法。构建了一个新的基于图模型的动作识别模型,该模型将模糊神经网络和概率图模型推理有效地结合起来,并采用基于递归的方法识别视频动作数据。基于贝叶斯规则、图模型等理论,提出了基于概率的递归计算结构,以获得较高的动作识别精度。通过与现有模型进行比较,验证了该算法具有良好的精确性和有效性。

参考文献:

[1] JHUANG H, SERRE T, WOLF L, et al. A biologically inspired system for action recognition[C]//IEEE international

conference on computer vision. Rio de Janeiro, Brazil, IEEE, 2007:1–8.

[2] TORRE F D L. Joint segmentation and classification of human actions in video[C]//Computer vision and pattern recognition. [s. l. ]:IEEE,2011:3265–3272.

[3] WEINLAND D, RONFARD R, BOYER E. A survey of vision-based methods for action representation, segmentation and recognition[J]. Computer Vision and Image Understanding,2011,115(2):224–241.

[4] WALKER J, GUPTA A, HEBERT M. Dense optical flow prediction from a static image[C]//International conference on computer vision. [s. l. ]:IEEE,2015:2443–2451.

[5] CHAARAOU I A A, CLIMENT-PÉREZ P, FLÓREZ-REVUELTA F. Silhouette-based human action recognition using sequences of key poses[J]. Pattern Recognition Letters, 2013,34(15):1799–1807.

[6] ZHANG Ziming, HU Yiqun, CHAN S, et al. Motion context: a new representation for human action recognition[M]//European conference on computer vision. Marseille, France; [s. n. ],2008:817–829.

[7] BOBICK A F, DAVIS J W. The recognition of human movement using temporal templates[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2001,23(3):257–267.

[8] MUKHERJEE S,BISWAS S K,MUKHERJEE D P. Recognizing human action at a distance in video by key poses[J]. IEEE Transactions on Circuits & Systems for Video Technology,2011,21(9):1228–1241.

[9] LAERTY J, LIU Y, ZHU X. Kernel conditional random fields;representation, clique selection, and semi-supervised learning[C]//International conference on machine learning. [s. l. ]:[s. n. ],2004.

[10] 姜 维,关 毅,王晓龙. 基于条件随机场的词性标注模型[J]. 计算机工程与应用,2006,42(21):13–16.

[11] SMINCHISESCU C,KANAUJIA A,Li Zhiguo,et al. Conditional models for contextual human motion recognition[J]. Computer Vision & Image Understanding,2005,104(2–3):210–220.

[12] YANG Q,XUE Dingyu,CUI Jianwei. Human action recognition using dynamic Bayesian network[J]. International Journal of Advancements in Computing,2002,4(12):291–298.

[13] LAPTEV I. On space-time interest points[J]. International Journal of Computer Vision,2005,64(2–3):107–123.

[14] WOLF W. Key frame selection by motion analysis[C]//Proceedings of the international conference on acoustics, speech, and signal processing. [s. l. ]:IEEE,1996:1228–1231.

[15] LEE H,GROSSE R,RANGANATH R,et al. Unsupervised learning of hierarchical representations with convolutional deep belief networks[J]. Communications of the ACM,2011,54(10):95–103.

[16] ZHOU Feng,DE LA TORRE F,HODGINS J K. Hierarchical aligned cluster analysis for temporal clustering of human motion[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2013,35(3):582–596.

[17] BLANK M,GORELICK L,SHECHTMAN E,et al. Actions as space-time shapes[C]//Tenth IEEE international conference on computer vision. Beijing:IEEE,2005:1395–1402.

[18] LIU Jingen,LUO Jiebo,SHAH M. Recognizing realistic actions from videos[C]//IEEE conference on computer vision and pattern recognition. Miami,FL,USA:IEEE,2009:1996–2003.

[19] 张宏毅,王立威,陈瑜希. 概率图模型研究进展综述[J]. 软件学报,2013,24(11):2476–2497.

[20] 杨 涛. 运动捕获数据关键帧提取及检索研究[D]. 杭州:浙江大学,2006.

[21] 王 深. 基于聚类算法的多特征融合关键帧提取技术研究[D]. 武汉:华中科技大学,2012.

[22] 瞿 中,高腾飞,张庆庆. 一种改进的视频关键帧提取算法研究[J]. 计算机科学,2012,39(8):300–303.

[23] 邓美玲. 一种基于骨骼关节点的行为识别算法研究[D]. 沈阳:辽宁大学,2016.

[24] 肖秦琨,李俊芳,肖秦汉. 基于四元数描述和 EMD 的人体运动捕获数据检索[J]. 计算机技术与发展,2014,24(3):90–93.

[25] 朱国刚,曹 林. 基于 Kinect 传感器骨骼信息的人体动作识别[J]. 计算机仿真,2014,31(12):329–333.

[26] 胡 琼,秦 磊,黄庆明. 基于视觉的人体动作识别综述[J]. 计算机学报,2013,36(12):2512–2524.

[27] 胡 彬. 基于模型函数方法的正则化参数选取[D]. 抚州:东华理工大学,2012.

(上接第 61 页)

字图像相关测量系统[J]. 光学学报,2013,33(4):97–107.

[2] 常志华,王子立. 一种基于直方图匹配的颜色矫正方法[J]. 科技资讯,2007(13):206–207.

[3] WESTON J. Support vector machine tutorial[R]. Princeton, USA:NEC Labs America Independence Way,2014.

[4] 胡斌斌,姚明海. 基于 SVM 的图像分类[J]. 微计算机信息,2010,26(1):105–106.

[5] ULABY F T,KOUYATE F,BRISCO B,et al. Textural information in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing,1986,24(2):235–245.

[6] 杨 帆. 数字图像处理与分析[M]. 北京:北京航空航天大学出版社,2007.

[7] 李伯宇. 图像纹理分析及分类方法研究[D]. 上海:复旦大学,2007.

[8] 孙兴华. 基于内容的图像检索研究[D]. 南京:南京理工大学,2001.

[9] 薄 华,马缚龙,焦李成. 图像纹理的灰度共生矩阵计算问题的分析[J]. 电子学报,2006,34(1):155–158.

[10] LI Yufeng,ZHOU Zhihua. Towards making unlabeled data never hurt[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(1):175–188.

[11] BELKIN M,MATVEEVA I,NIYOGI P. Regularization and semi-supervised learning on large graphs[C]//IEEE international conference on acoustics, speech, and signal processing. Montreal,Que,Canada:IEEE,2004.

[12] ZHOU Dengyong,SCHÖLKOPF B,HOFMANN T,et al. Semi-supervised learning on directed graphs[C]//Advances in neural information processing systems. Cambridge MA:MIT Press,2004:1633–1640.

[13] HARALICK R M. Texture features for image classification[J]. IEEE Transactions on Systems, Man, and Cybernetics,1973,3(6):610–621.

[14] 贾永红. 计算机图像处理与分析[M]. 武汉:武汉大学出版社,2001.

[15] 宋 辉,王明威. 基于双远心镜头的虹膜图像采集方法[J]. 计算机技术与发展,2016,26(8):65–68.