第47卷 第6期 计 算 机 工 程 2021年6月
Vol.47 No.6 Computer Engineering June 2021

·图形图像处理· 文章编号：1000-3428(2021)06-0210-07 文献标志码：A 中图分类号：TP391

基于视角转换的多视角步态识别方法

瞿斌杰[1,2]，孙韶媛[1,2]，Samah A.F.Manssor[1,2]，赵国顺[1,2]

（1.东华大学 信息科学与技术学院，上海 201620；2.东华大学 数字化纺织服装技术教育部工程研究中心，上海 201620）

摘 要：针对步态识别中步态视角变化、步态数据样本量少及较少利用步态时间信息等问题，提出一种基于视角转换的步态识别方法。通过VTM-GAN网络，将不同视角下的步态能量图及含有步态时间信息的彩色步态能量图，统一映射到保留步态信息最丰富的侧视图视角，以此突破步态识别中多视角的限制，在视角转换的基础上，通过构建侧视图下的步态正负样本对来扩充用于网络训练的数据，并采用基于距离度量的时空双流卷积神经网络作为步态识别网络。在CASIA-B数据集上的实验结果表明，该方法在各状态、各角度下的平均识别准确率达到92.5%，优于3DCNN、SST-MSCI等步态识别方法。

关键词：步态识别；视角转换；VTM-GAN网络；时空双流卷积神经网络；CASIA-B数据集

开放科学（资源服务）标志码（OSID）：

中文引用格式：瞿斌杰，孙韶媛，Samah A.F.Manssor，等.基于视角转换的多视角步态识别方法[J].计算机工程，2021，47(6)：210-216.

英文引用格式：QU Binjie，SUN Shaoyuan，Samah A.F.Manssor，et al.Multi-view gait recognition method based on view transformation[J].Computer Engineering，2021，47(6)：210-216.

Multi-view Gait Recognition Method Based on View Transformation

QU Binjie[1,2]，SUN Shaoyuan[1,2]，Samah A.F.Manssor[1,2]，ZHAO Guoshun[1,2]

（1.College of Information Science and Technology，Donghua University，Shanghai 201620，China；2.Engineering Research Center of Digitized Textile and Fashion Technology of Ministry of Education，Donghua University，Shanghai 201620，China）

【Abstract】The existing gait recognition methods are limited by multiple factors，including the changes of view，small gait sample size，and under-utilization of the temporal information of gaits.To address the problems and improve the gait recognition performance，this paper proposes a gait recognition method based on view transformation.Through VTM-GAN，Gait Energy Images（GEI）and Chrono Gait Images（CGI）with temporal gait information of different views are mapped to the side view that contains the most abundant gait information in order to break the limitation of views in gait recognition.On the basis of view transformation，the positive and negative sample pairs of gait data in the side view are constructed to extend the volume of network training data.The Spatial-temporal double flow convolutional neural network based on distance measurement is taken as the gait recognition network.Experimental results on the CASIA-B dataset show that the average recognition accuracy of this method in all states and views reaches 92.5%，higher than that of 3DCNN，SST-MSCI and other gait recognition methods.

【Key words】gait recognition；view transformation；VTM-GAN network；spatial-temporal double flow Convolutional Neural Network（CNN）；CASIA-B dataset

DOI：10. 19678/j. issn. 1000-3428. 0057899

0 概述

生物识别技术通过计算机与光学、声学、生物传感器及生物统计学原理等高科技手段密切结合，利用人体固有的生理特性和行为特征来进行个人身份的鉴定。当前诸如面部、虹膜、指纹和签名之类的生物识别技术已广泛用于身份认证。这些生物识别技术的局限性在于需要被测者的配合来获取特征信息。步态是一种重要的生物特征，它克服了上述限制，在不受控制的情况下无需对象合作，摄像机可以很容易地在远距离捕获目标并采集信息[1]。目前步态识别方法主要分为基于模型的方法和基于外观的方法。

基金项目：上海市科委基础研究基金（15JC1400600）。

作者简介：瞿斌杰（1996—），男，硕士研究生，主研方向为图像处理、计算机视觉；孙韶媛，教授、博士后；Samah A.F.Manssor，博士；赵国顺，硕士研究生。

收稿日期：2020-03-30 **修回日期：**2020-05-12 **E-mail：**qubinjieit@163.com

基于模型的方法试图建立模型以从视频序列重建人体的基础结构。文献[2]通过构建身体结构模型来完成步态识别，文献[3]从多个相机重建的3D步态数据进行识别。3D数据比2D数据能够传递更多的信息，但采集成本高限制了其应用。

基于外观的方法将步态图像序列作为输入。此类方法首先从视频序列中提取二进制轮廓序列，然后用多种方法将步态轮廓序列处理成单张的步态特征图，如步态能量图像（Gait Energy Image，GEI）[4]即是广泛使用的一种步态特征，其通过在整个步态周期上平均轮廓序列获得，GEI包含了步态周期的空间信息。此外，还有其他步态特征，如运动历史图像（Motion History Image，MHI）[5]。受MHI的启发，文献[6]提出了运动剪影图像来嵌入步态剪影的时空信息，文献[7]提出了用于步态识别的步态流图像。

尽管目前基于深度学习的步态识别算法已经取得了一些成果，但是步态识别面临着诸多挑战，如视角的变换、着装变化和携带物等。在视角的变换方面[8]，由于一般摄像头均为固定状态，当行人由不同的方向进入摄像采集区域时，会造成多视角下目标姿态不同的问题。针对这种视角变化问题，文献[9]将来自不同视角的样本表示为在原对应视图下的线性组合，通过提取特征表示系数进行分类，文献[10]提取一种基于均值的视角不变步态特征方法，并基于形状距离测量步态相似性。文献[11]提出了基于频域特性和视图转换的模型。在此基础上，文献[12]进一步应用线性判别分析简化计算。因此，构造视角转换模型可以在较小代价下实现精度较高的识别效果。在着装变化方面，人的着装在很大程度上会改变人体的轮廓，尤其在着装较厚或者衣物较长的情况下，会对人体姿态形成遮挡，影响步态的识别效果。在携带物方面，在前景分离的过程中，人所携带的物品极有可能被当作人体的一部分而被提取出来，从而影响步态特征的准确性。若该行人在其他时刻未携带物品或携带其他物品，则较难实现精确识别。

本文提出一种基于视角转换的步态识别方法。通过VTM-GAN网络将不同视角下的步态特征统一转换至90°状态下，即目标运动方向与拍摄方向呈垂直状态，采用视角转换后的步态数据构建步态正负样本对，扩充用于网络训练的数据来增加模型的泛化能力，将时空双流卷积神经网络作为步态识别网络，并在CASIA-B数据集上进行了实验验证。

1　步态数据预处理

良好的步态序列预处理可以提升实验效果，因此将原始步态序列中的图像经过混合高斯模型背景差分[13]、形态学处理以及尺寸归一化等操作后能够得到质量较高的步态二值图。

基于混合高斯模型的背景差分法是一种较常用的背景减除算法，该算法认为像素间颜色互不相关，对每一像素点的处理都是独立的。通过对每个背景像素点建立2个~3个高斯模型后分别进行前背景匹配，得到背景图象后进行差分运算得到步态前景图。

通过背景差分后得到的步态二值图往往存在较多的空洞以及噪声，因此，引入图像形态学处理中的开运算[14]，即先腐蚀再膨胀来消除噪声以提高步态轮廓的质量，开运算数学表达式如式（1）所示：

$$\boldsymbol{A}\circ\boldsymbol{B}=(\boldsymbol{A}\ominus\boldsymbol{B})\oplus\boldsymbol{B} \tag{1}$$

其中，$\boldsymbol{A}$表示待处理图像，$\boldsymbol{B}$表示开运算所定义的结构体用以遍历图像。

经过前两步得到的步态二值图中，目标位置及大小往往因为拍摄角度以及目标位置的改变不尽相同，所以采用步态尺寸归一化得到步态二值图序列。将每个归一化后的轮廓图$\boldsymbol{B}$表示为：

$$\boldsymbol{B}=(x,\ y,\ w,\ h) \tag{2}$$

其中，(x,y)表示步态图左上角坐标，w和h分别表示为归一化后二值图的宽度和高度，实验中为了减少因缩放尺度给图像质量带来的影响，将归一化二值图的宽高比例设置为1，即$w=h=240$。归一化二值图的高度为目标的高度，再对每个目标的中心(G_x,G_y)进行计算，其中x的计算如式（3）所示：

$$x=G_x-\frac{h}{2} \tag{3}$$

对步态二值图进行平均化相加操作来获取GEI，其定义如式（4）所示：

$$\boldsymbol{g}(x,y)=\frac{1}{T}\sum_{q=1}^{Q}\sum_{t=1}^{T}S_{q,t}(x,y) \tag{4}$$

其中，$\boldsymbol{g}(x,y)$为步态能量图，$S_{q,t}(x,y)$表示在第q个步态序列中时刻t的步态剪影图中坐标为(x,y)的像素值。

对于CGI的合成，本文参考了文献[15]提出的方法，在GEI基础上，通过该方法将步态二值图序列根据三通道RGB色彩的映射，来保留GEI图像中损失的时间信息。

2　网络架构及参数配置

本文主要以VTM-GAN网络作为视角转换网络，以时空双流卷积神经网络作为步态识别网络，步态视角转换网络VTM-GAN基于Cycle-GAN网络[16]，将不同视角下的GEI转换成90°状态下的GEI，从而构建扩充的正负数据样本对。时空双流卷积神经网络同时保留时间空间信息，通过距离度量判断同时输入的样本是否来源于同一个目标。

2.1　VTM-GAN网络

VTM-GAN网络结构如图1所示。VTM-GAN网络具有2个输入端，本质上是2个镜像对称的GAN网络，共有2个G_1生成器、2个G_2生成器及2个鉴别器，在每一对生成器间共享权值且采用跳跃链接技术。该网

络思想是源于Cycle-GAN网络，在网络中有两个数据域A和B，该网络从域A获取输入图像并传递到第一个生成器G_1，完成从域A至域B图像的转换。然后新生成的图像被传递到另一个生成器G_2，再完成从域B到域A图像的转换，其目的是在原始域A转换回图像A'。输出A'必须与原始输入图像相似，用来定义非配对数据集中原来不存在的映射。

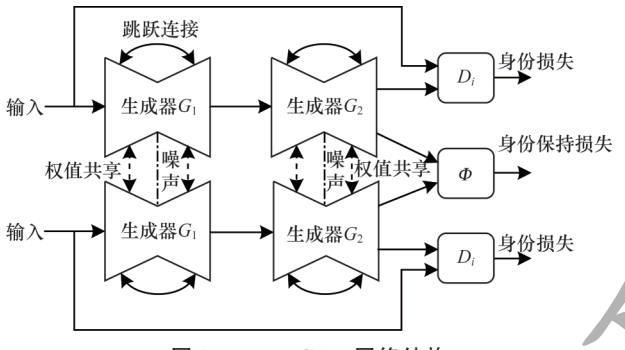

图1 VTM-GAN网络结构

Fig.1 Network structure of VTM-GAN

该网络的损失函数主要由对抗损失、身份损失、身份保持损失、循环一致性损失以及分类损失组成。网络的对抗损失主要由两部分组成，对于映射$A\rightarrow B$和它对应的判别器D_B，损失函数可以定义为：

$$\mathcal{L}\left(G_1, D_B, A, B\right)=E_{b\sim p_{\text{data}}(b)}\left[\log_a D_B(b)\right]+E_{a\sim p_{\text{data}}(a)}\left[1-\log_a D_B\left(G_1(a)\right)\right] \tag{5}$$

同理，对于映射$B\rightarrow A$和它对应的判别器D_A，损失函数可以定义为：

$$\mathcal{L}\left(G_2, D_A, B, A\right)=E_{a\sim p_{\text{data}}(a)}\left[\log_a D_A(a)\right]+E_{b\sim p_{\text{data}}(b)}\left[1-\log_a D_A\left(G_1(b)\right)\right] \tag{6}$$

当网络优化时，对于生成器G_1即去最小化目标函数式(5)，而对于鉴别器即去最大化目标函数式(5)，如式(7)所示：

$$\min_{G_1}\left(\max_{D_B}\mathcal{L}\left(G_1, D_B, A, B\right)\right) \tag{7}$$

对于生成器G_2就是去最小化目标函数式(6)，而对于鉴别器就是去最大化目标函数式(6)，如式(8)所示：

$$\min_{G_2}\left(\max_{D_A}\mathcal{L}\left(G_2, D_A, B, A\right)\right) \tag{8}$$

因为映射G_1完全可将所有域A中的图像转换成域B中同一张图像，从而导致损失无效化，所以引入循环一致性损失同时学习G_1和G_2两个映射，并要求$G_2(G_1(a))\approx a$以及$G_1(G_2(b))\approx b$，该损失如式(9)所示：

$$\mathcal{L}_{\text{cyc}}\left(G_1, G_2\right)=E_{a\sim p_{\text{data}}(a)}\left\|G_2(G_1(a))-a\right\|_1+E_{b\sim p_{\text{data}}(b)}\left\|G_1(G_2(b))-b\right\|_1 \tag{9}$$

综上所述，可得网络总损失如式(10)所示：

$$\mathcal{L}\left(G_1, G_2, D_A, D_B\right)=\mathcal{L}\left(G_1, D_B, A, B\right)+\mathcal{L}\left(G_2, D_A, B, A\right)+\lambda\mathcal{L}_{\text{cyc}}\left(G_1, G_2\right) \tag{10}$$

其中，λ用于调节循环一致性损失在总损失中的比重。

生成器是负责域A到域B的图像的生成，输入一个域B的图片生成域B的图片b'，用于计算b'与输入b的损失称为身份损失，而身份保持损失的作用是不仅希望生成的GEI看上去像同一个人，而且能够保持原有的身份信息，因此引入分类损失，并使分类器与生成器进行竞争，与鉴别器进行协作，将真实图片与生成的图片作为输入，然后预测它们的标签，希望两者间预测得到的是同样的标签，并且使生成的特征尽可能地靠近真实的特征。

基于以上的思想，将步态数据库中不同视角、不同状态风格下的步态特征图统一转换为90°状态下正常行走风格的步态特征图，从而完成对步态数据库样本的扩充并且消除了多视角对于步态识别的影响[17]，如图2所示，其中，NM表示正常状态，BG表示背包状态，CL表示穿外套状态。

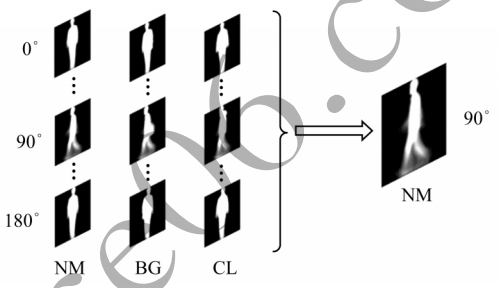

图2 视角转换示意图

Fig.2 Schematic diagram of view transformation

表1所示为网络生成器的具体参数，其中，输入尺寸为240像素×240像素×1通道。采用的Unet网络结构由编码器及解码器两部分构成，并且在第4层编码器之后之间引入[-1,1)范围内遵循均匀分布的噪声，反卷积层的第一层设置Dropout大小为0.5。跳跃连接将d1与e3、d2与e2、d3与e1进行合并。

表1 生成器参数

Table 1 Generator parameters

层名	输出尺寸	卷积核
g_bn_e1	120像素×120像素×64通道	尺寸:4像素×4像素，步长:2
g_bn_e2	60像素×60像素×128通道	尺寸:4像素×4像素，步长:2
g_bn_e3	30像素×30像素×256通道	尺寸:4像素×4像素，步长:2
g_bn_e4	15像素×15像素×512通道	尺寸:4像素×4像素，步长:2
g_bn_e5	15像素×15像素×512通道	尺寸:1像素×1像素，步长:1
g_bn_d1	30像素×30像素×768通道	尺寸:4像素×4像素，步长:2
g_bn_d2	60像素×60像素×384通道	尺寸:4像素×4像素，步长:2
g_bn_d3	120像素×120像素×192通道	尺寸:4像素×4像素，步长:2
g_bn_d4	240像素×240像素×1通道	尺寸:4像素×4像素，步长:2

表2所示为网络鉴别器参数，其中输入尺寸为240像素×240像素×1通道。对输入图像进行特征提

取，然后确定该输入是否属于某特定类别，鉴别器最后一层卷积层通过产生一维输出来完成鉴别。

表2　鉴别器参数

Table 2　Discriminator parameters

层名	输出尺寸	卷积核
d_h0	120像素×120像素×64通道	尺寸：4像素×4像素，步长：2
d_h1	60像素×60像素×128通道	尺寸：4像素×4像素，步长：2
d_h2	30像素×30像素×256通道	尺寸：4像素×4像素，步长：2
d_h3	30像素×30像素×512通道	尺寸：4像素×4像素，步长：1
d_h4_pred	30像素×30像素×1通道	尺寸：4像素×4像素，步长：1

2.2　时空双流卷积神经网络

时空双流卷积神经网络具有左右两侧输入端，网络的左右两侧具有相同结构和参数，两侧输入的样本对可以通过模仿减法运算来得到一对特征的差别，随后可以通过两者的差别进而得到样本间的相似度。其主要思想是通过一个函数将输入映射到目标空间，在目标空间基于距离度量的欧式距离来对比相似度。在分别计算每对样本差值之前，只使用线性投影，这是由卷积核在最底卷积阶段实现的，即在网络底层进行特征的融合，能够在输入端减少数据复杂度，进而能够一定程度地减少网络复杂度，更易于计算。一对卷积核可接受两个输入，可以看作权重比较器。在每个空间位置，首先分别对其两个输入的局部区域重新加权，然后将这些加权后的项相加来模拟减法。在底层融合之后的深层卷积层可以从样本对之间的差异中学习更多复杂信息。

在网络的顶层设置一个Softmax二分类器来判断输入样本对是否为同一目标，利用逻辑回归损失对整个网络进行训练。该预测器可用式(11)表示：

$$S_i=\eta(\varphi(\phi(\boldsymbol{x}),\phi(\boldsymbol{x}_i))) \tag{11}$$

其中，$\boldsymbol{x}$和$\boldsymbol{x}_i$为输入样本对，ϕ将$\boldsymbol{x}$和$\boldsymbol{x}_i$映射到同一空间，与此同时，φ计算两个输入的权值差，η作为预测器来预测最终的相似度，ϕ可由一层或多层卷积层和全连接层组成，φ必须有两个输入并且可由一个卷积层或全连接层构成，预测器η由全连接层和Softmax层构成，S_i为预测器。

该网络在训练阶段最小化来自相同类别的一对样本的损失函数值，最大化来自不同类别的一对样本的损失函数值。给定一组映射函数$G_w(\boldsymbol{X})$，其中参数$\boldsymbol{w}$为共享参数向量，目的是寻找一组参数$\boldsymbol{w}$，使得当$\boldsymbol{X}_1$和$\boldsymbol{X}_2$属于同一类别时，相似性度量较小，且最小化损失函数。当属于不同类别时，相似性度量较大，且最大化损失函数。其中，$\boldsymbol{X}_1$和$\boldsymbol{X}_2$是网络的一组输入图像，Y为输入组的一个0，1标签，如果输入为同一个人，即一组正对，那么$Y=0$，否则$Y=1$，相似性度量公式如式(12)所示：

$$E_w(\boldsymbol{X}_1,\boldsymbol{X}_2)=\left\|G_w(\boldsymbol{X}_1)-G_w(\boldsymbol{X}_2)\right\| \tag{12}$$

综上，该网络可以用式(13)简明地表示为：

$$S_i=\boldsymbol{W}_4 f(\boldsymbol{W}_3 f(\boldsymbol{W}_2(f(\boldsymbol{W}_1\boldsymbol{x})+f(\boldsymbol{W}_1'\boldsymbol{x}_i)))) \tag{13}$$

其中，$\boldsymbol{W}_l$和$\boldsymbol{W}_l'$分别代表了一对样本第l层的权值，f为非线性激活函数，本文采用Relu作为非线性激活函数。

网络结构及各层参数设置如图3和表3所示，在表3中，输入尺寸为240像素×240像素×2通道。因为是对称的网络，所以左右两侧网络参数相同网络中的N为批归一化技术，网络中的D为Dropout技术，通过减少每次训练时的参数量，提高了模型准确率，增强了模型的泛化能力。

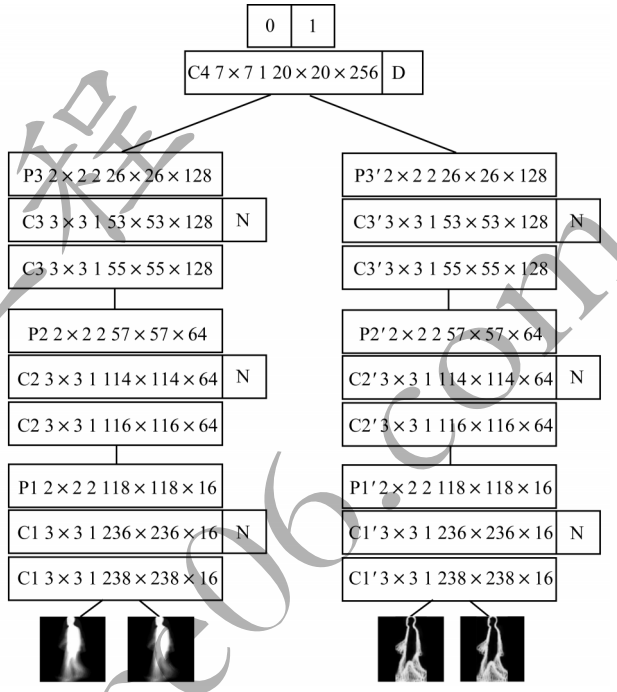

图3　时空双流卷积神经网络结构

Fig.3　Structure of spatial-temporal double flow convolutional neural network

表3　网络参数设置

Table 3　Setting of network parameters

层名	输出尺寸	卷积核
卷积层C1	238像素×238像素×16通道	尺寸：3×3，步长：1
卷积层C1	236像素×236像素×16通道	尺寸：3×3，步长：1
池化层P1	118像素×118像素×16通道	尺寸：2×2，步长：2
卷积层C2	116像素×116像素×64通道	尺寸：3×3，步长：1
卷积层C2	114像素×114像素×64通道	尺寸：3×3，步长：1
池化层P2	57像素×57像素×64通道	尺寸：2×2，步长：2
卷积层C3	55像素×55像素×128通道	尺寸：3×3，步长：1
卷积层C3	53像素×53像素×128通道	尺寸：3×3，步长：1
池化层P3	26像素×26像素×128通道	尺寸：2×2，步长：2
卷积层P4	20像素×20像素×256通道	尺寸：7×7，步长：1

3　实验结果与分析

本文采用CASIA-B[17]数据集作为实验数据，该数据集是一个大规模、多视角的步态库，其中共有124个人，每个人有11个视角(0°，18°，36°，…，180°)，

并在普通条件(NM)、穿大衣(CL)以及携带包裹(BG)这3种行走条件下采集。实验的具体流程依次为步态图像预处理、GEI及CGI的合成、VTM-GAN网络视角转换、步态样本对构建、时空双流卷积神经的网络训练,最后通过测试得到各视角下的准确率。

本文算法使用PyTorch1.4框架,实验的软硬件配置如下:CPU为Intel i5-6600,内存为8 GB,GPU为NVIDIA GTX 1060,操作系统为Windows10,CUDA为9.2。

3.1 数据预处理结果

图4所示为实验得到的各角度不同状态下CASIA-B的GEI,其能很好地表现步态的速度、形态等特征。

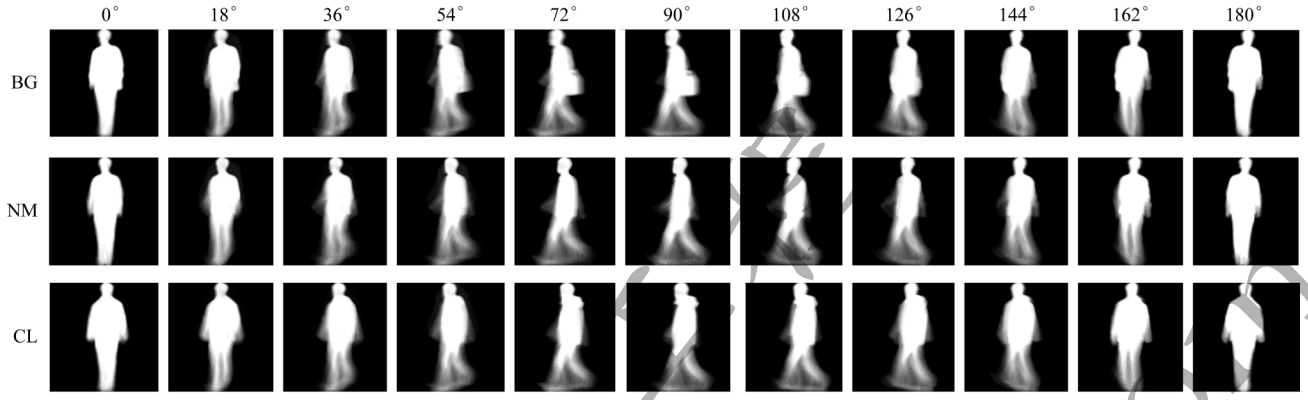

图4 不同角度及状态下的步态能量图

Fig.4 Gait energy diagrams at different angles and states

图5所示为将步态二值图序列通过RGB三通道色彩映射后得到的CGI图像,通过色彩映射保留了步态序列的时间信息。

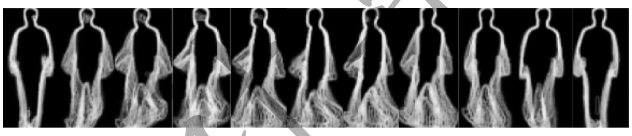

图5 CGI图像示例

Fig.5 Examples of CGI image

3.2 视角转换及样本扩充

图6所示为通过视角转换后的到的计时步态图像(CGI)步态样本。图7所示为将不同视角下、不同行走状态下的步态能量图像(GEI),通过VTM-GAN网络转换成90°状态下的结果示例,图中每3列为一组,每一列分别包含输入步态、实际步态和正常视图中的合成步态。

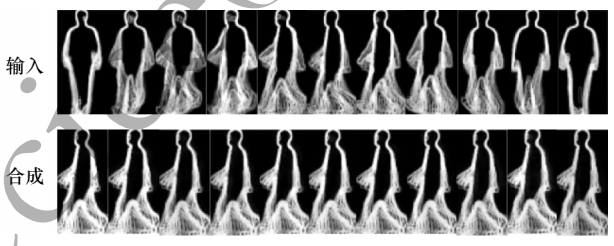

图6 CGI视角转换后结果

Fig.6 Result of CGI after view transformation

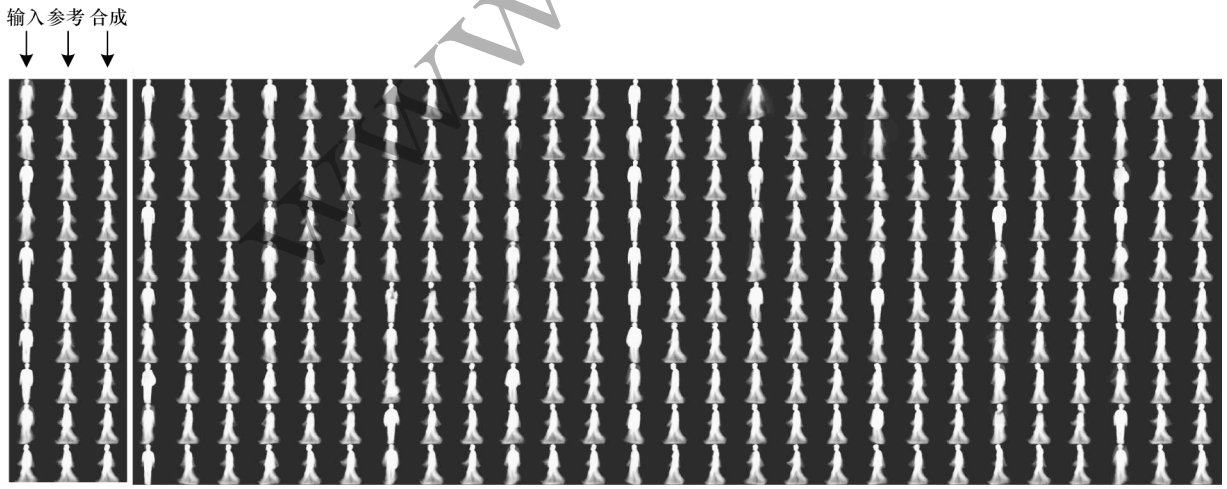

图7 GEI视角转换后结果

Fig.7 Results of GEI after view transformation

为衡量该视角转换网络的准确性，本文采用余弦相似度来判断参考图像与合成图像的相似度，余弦相似度公示如式（14）所示：

$$\cos\theta=\frac{\sum_{i=1}^{n}\boldsymbol{A}_i\times\boldsymbol{B}_i}{\sqrt{\sum_{i=1}^{n}(\boldsymbol{A}_i)^2}\times\sqrt{\sum_{i=1}^{n}(\boldsymbol{B}_i)^2}} \tag{14}$$

其中，$\boldsymbol{A}_i$、$\boldsymbol{B}_i$分别代表参考图像与合成图像。如果合成图像与参考图像通过式（14）计算所得结果越接近1，那么两图就越接近，即合成GEI的信息保留程度就越高。

如图8所示，通过计算参考步态以及合成步态的余弦相似度，发现两者的相似度总体在98.5%以上，因为从0°及180°转换至90°的视角跨度较大，所以相似度有所下降。

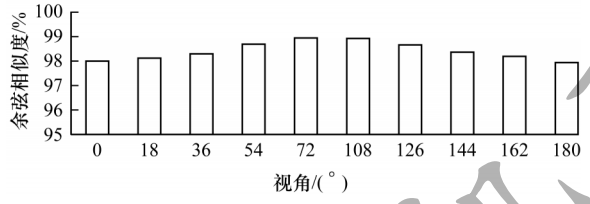

图8　视角转换后合成步态与参考步态的余弦相似度

Fig.8　Cosine similarity between the synthesized gaits and the reference gaits after the view transformation

在视角转换前，一共拥有124×10=1 240个侧视图下的步态样本，而将CASIA-B的124个目标的GEI以及CGI样本全部经过视角转换后，每种共有124×10×11=13 640个侧视图下的样本，视角转换突破了原数据库多视角上的限制，大大扩充了可训练的步态样本数量。本文使用CASIA-B中的NM01-NM04、CL01、BG01序列作为训练集，即124×6×11=8 184个特征来构建正负样本对用于时空双流卷积神经网络的训练，每个目标包含了6×11=66个侧视图下的GEI及CGI，任选同一目标的两个GEI及两个CGI构成一组正样本对，则共可构成C_{66}^{2}对，即2 145对。因此，对于124个目标，总共可以构成265 980对正样本对。同样，随机选取某一目标的一个GEI及一个CGI，再选取其他目标的某一个GEI及某一个CGI来构建负样本对。本文构建的正负样本对都为265 980对，即各占总样本对的一半。通过这种样本构建方法，网络训练过程中的样本量成倍增加，缓解了模型因数据量较少而产生过拟合问题。

3.3　时空双流卷积神经网络训练及测试方式

时空双流卷积神经网络采用Adam进行优化，设置初始学习率为0.000 1，训练过程中采用自适应方式调节各个参数的学习率。用二分类交叉熵损失作为模型的目标函数训练网络，训练集由数量相同的正负样本对构成，Mini-Batch尺寸为128，每训练一个Mini-Batch进行一次网络的权值更新，损失函数如式（15）所示：

$$L(\hat{y},y)=-(y\log_a\hat{y}+(1-y)\log_a(1-\hat{y})) \tag{15}$$

其中，$L(\hat{y},y)$为二分类交叉熵损失函数，y和$\hat{y}$分别为预测值与真实值。

在训练中，每5 000次对模型进行一次测试得到如图9所示的训练准确率及损失变化曲线，时空双流卷积神经网络经过将近1 600K次迭代后，网络损失函数和训练准确率逐渐开始趋于稳定，表明模型已经趋于稳定。

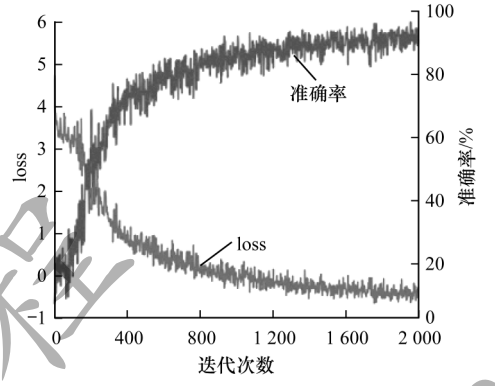

图9　网络损失及训练准确率变化曲线

Fig.9　Network loss and training accuracy curves

对网络进行测试时，分别采用CASIA-B中NM05-NM06、CL02、BG03序列中的每个视角下的GEI作为测试集。首先将待测试样本利用VTM-GAN网络进行视角转换至侧视状态下，此时一共可以得到124×11×4=5 456个测试GEI。然后随机在某类样本中选择一个作为基样本，将待测样本与基样本构成一组样本对输入至训练好的时空双流卷积神经网络中，如图10所示，最终得到NM、BG和CL各状态下各视角的平均准确率分别为94.4%、92.5%和90.5%。

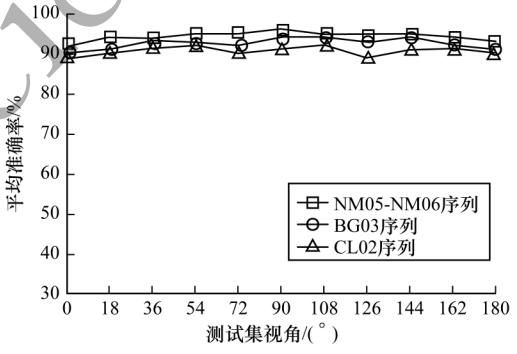

图10　NM、BG及CL状态下平均测试准确率

Fig.10　Average test accuracy in NM，BG and CL states

将本文方法与文献[18]提出的基于3DCNN方法、文献[19]提出的Deterministic Learning方法以及文献[7]提出的SST-MSCI方法进行比较，如图11所示，本文方法在视角为0°时识别准确率略低于SST-MSCI方法，但是在剩余的视角下，识别准确率均高于其他几种算法。此外，将本文方法与文献[20]提出的GaitGAN方法进行了对比，该方法在NM状态下识别准确率达到了98.75%，但在BG以及CL状态下本文方法均高于GaitGAN算法在这两种状态下的识别准确率。

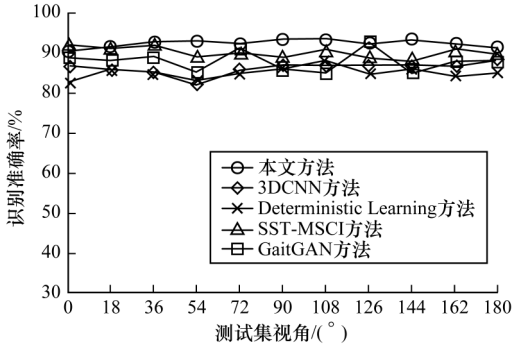

图11 不同方法识别准确率的对比

Fig.11 Comparison of recognition accuracy of different methods

4 结束语

本文采用基于视角转换的方法进行步态识别研究,针对目前步态识别中多视角、缺少对步态时间信息的利用以及数据量较少等问题,结合VTM-GAN网络将不同视角下的步态样本统一转换至保留步态信息最丰富的90°状态下,从而构建扩充的步态样本对训练时空双流卷积神经网络。实验结果表明,与3DCNN、Deterministic Learning等步态识别方法相比,本文方法在各角度下步态识别准确率有所提升,验证了基于视角转换方法的有效性。但是针对多视角状态下的步态识别仍需改进,如研究更精准的行人检测模型来获取精确的步态数据,并结合多种生物特征的优点研究特征融合识别算法。

参 考 文 献

[1] WANG Kejun, DING Xinnan, XING Xianglei, et al. A survey of multi-view gait recognition[J]. Acta Automatica Sinica, 2019, 45(5): 841-852. (in Chinese)
王科俊,丁欣楠,邢向磊,等. 多视角步态识别综述[J]. 自动化学报,2019,45(5):841-852.

[2] JOHNSON A Y, BOBICK A F. A multi-view method for gait recognition using static body parameters [C]// Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication. Halmstad, Sweden: [s. n.], 2001: 301-311.

[3] ARIYANTO G, NIXON M S. Model-based 3D gait biometrics [C]//Proceedings of International Joint Conference on Biometrics. Washington, D. C., USA: IEEE Press, 2011: 1-7.

[4] HAN J, BHANU B. Individual recognition using gait energy image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(2): 316-322.

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

[6] LAM T H W, LEE R S T. A new representation for human gait recognition: motion silhouettes image [C]// Proceedings of International Conference on Biometrics. Hong Kong, China: [s. n.], 2006: 612-618.

[7] LAM T H W, CHEUNG K H, LIU J N K. Gait flow image: a silhouette-based gait representation for human identification [J]. Pattern Recognition, 2011, 44(4): 973-987.

[8] HE Yiwei, ZHANG Junping. Deep learning for gait recognition: a survey[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(5): 442-452. (in Chinese)
何逸炜,张军平. 步态识别的深度学习:综述[J]. 模式识别与人工智能,2018,31(5):442-452.

[9] LIU Nini, LU Jiwen, TAN Yaping. Joint subspace learning for view-invariant gait recognition [J]. IEEE Signal Processing Letters, 2011, 18(7): 431-434.

[10] KUSAKUNNIRAN W, WU Q, ZHANG J, et al. A new view-invariant feature for cross-view gait recognition[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(10): 1642-1653.

[11] MAKIHARA Y, SAGAWA R, MUKAIGAWA Y, et al. Gait recognition using a view transformation model in the frequency domain [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2006: 151-163.

[12] KUSAKUNNIRAN W, WU Q, LI H D, et al. Multiple views gait recognition using view transformation model based on optimized gait energy image[C]//Proceedings of the 12th IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2009: 1058-1064.

[13] GODBEHERE A B, MATSUKAWA A, GOLDBERG K. Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation [C]// Proceedings of the American Control Conference. Montreal, USA: IEEE Press, 2012: 4305-4312.

[14] RAFAEL C G, RICHARD E W. Digital image processing [M]. Englewood Cliffs, USA: Prentice-Hall, Inc., 2007.

[15] WANG Chen, ZHANG Junping, WANG Liang, et al. Human identification using temporal information preserving gait template[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2164-2176.

[16] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[EB/OL]. [2020-02-20]. https://arxiv. org/abs/1703. 10593.

[17] YU Shiqi, TAN Daoling, TAN Tieniu. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition [C]//Proceedings of IEEE International Conference on Pattern Recognition. Washington D. C., USA: IEEE, 2006: 441-444.

[18] WU Zifeng, HUANG Yongzhen, WANG Liang, et al. A comprehensive study on cross-view gait based human identification with deep CNNs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 209-226.

[19] ZENG Wei, WANG Cong. View-invariant gait recognition via deterministic learning [J]. Neurocomputing, 2016, 175(29): 324-335.

[20] YU S Q, CHEN H F, REYES E B G, et al. GaitGAN: invariant gait feature extraction using generative adversarial networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Press, 2017: 532-539.

编辑 索书志