



(12) 发明专利申请

(10) 申请公布号 CN 115408564 A

(43) 申请公布日 2022. 11. 29

(21) 申请号 202211063493.4

G06V 10/764 (2022.01)

(22) 申请日 2022.08.31

G06V 40/16 (2022.01)

(71) 申请人 杭州网易云音乐科技有限公司

地址 310052 浙江省杭州市萧山区钱江世纪城奔竞大道353号杭州国际博览中心A座1201室

(72) 发明人 金强 李鹏 蔡苗苗 曹偲

(74) 专利代理机构 北京同达信恒知识产权代理有限公司 11291

专利代理师 金银花

(51) Int. Cl.

G06F 16/783 (2019.01)

G06F 16/75 (2019.01)

G06V 10/74 (2022.01)

G06V 10/762 (2022.01)

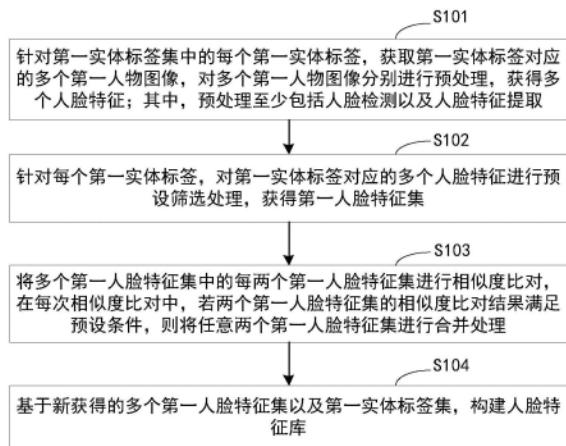
权利要求书2页 说明书23页 附图6页

(54) 发明名称

人脸特征库构建方法、装置、电子设备和存储介质

(57) 摘要

本公开涉及互联网技术领域,尤其涉及一种人脸特征库构建方法、装置、电子设备和存储介质,针对第一实体标签集中的每个第一实体标签,获取第一实体标签对应的多个第一人物图像,对多个第一人物图像分别进行预处理,获得多个人脸特征;针对每个第一实体标签,对第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集;将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,若两个第一人脸特征集的相似度比对结果满足预设条件,则将任意两个第一人脸特征集进行合并处理;基于新获得的多个第一人脸特征集以及第一实体标签集,构建人脸特征库。本公开可以高效构建更完备的人脸特征库,提高人物实体标签的识别准确率。



1. 一种人脸特征库构建方法,其特征在於,包括:

针对第一实体标签集中的每个第一实体标签,获取所述第一实体标签对应的多个第一人物图像,对所述多个第一人物图像分别进行预处理,获得多个人脸特征;其中,所述预处理至少包括人脸检测以及人脸特征提取;

针对每个第一实体标签,对所述第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集;

将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将所述任意两个第一人脸特征集进行合并处理;

基于新获得的多个第一人脸特征集以及所述第一实体标签集,构建人脸特征库。

2. 根据权利要求1所述的方法,其特征在於,所述对所述第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集,包括:

对所述第一实体标签对应的多个人脸特征进行聚类处理,获得多个第一人脸特征簇,每个第一人脸特征簇包括一个聚类中心;

针对每个所述聚类中心,若所述聚类中心分别与各个其它聚类中心的相似度均不满足第一相似度条件,则删除所述聚类中心所在的第一人脸特征簇;

基于剩余的多个第一人脸特征簇,获得所述第一人脸特征集。

3. 根据权利要求2所述的方法,其特征在於,所述基于剩余的多个第一人脸特征簇,获得所述第一人脸特征集,包括:

针对剩余的每个第一人脸特征簇,确定所述第一人脸特征簇中的每个人脸特征与对应的聚类中心的相似度,将相似度不满足第二相似度条件的人脸特征从所述第一人脸特征簇中剔除;

基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集。

4. 根据权利要求3所述的方法,其特征在於,所述基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集,包括:

针对新获得的每个第一人脸特征簇,若所述第一人脸特征簇中的人脸特征的数量不满足数量条件,则将所述第一人脸特征簇删除。

5. 根据权利要求3所述的方法,其特征在於,所述基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集,包括:

针对新获得的每个第一人脸特征簇,确定所述第一人脸特征簇中的每个人脸特征分别与多个其它人脸特征的相似度,将相似度满足第三相似度条件的其它人脸特征从所述第一人脸特征簇中剔除;

基于再次新获得的多个第一人脸特征簇,获得所述第一人脸特征集。

6. 根据权利要求2所述的方法,其特征在於,所述将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将所述两个第一人脸特征集进行合并处理,包括:

针对每两个第一人脸特征集,执行以下操作:

获取一个第一人脸特征集的多个聚类中心,以及获取另一第一人脸特征集的多个聚类中心;

确定所述一个第一人脸特征集的每个聚类中心,分别与所述另一第一人脸特征集的多个聚类中心的相似度,获得相似度向量;

基于获得的多个相似度向量构建类间相似度矩阵,确定所述类间相似度矩阵中达到第一相似度阈值的目标相似度,将所述目标相似度的数量作为所述相似度比对结果;

若所述相似度比对结果满足预设条件,则将所述一个第一人脸特征集与所述另一第一人脸特征集进行合并处理。

7. 根据权利要求6所述的方法,其特征在于,所述将所述一个第一人脸特征集与所述另一第一人脸特征集进行合并处理之前,还包括:

基于所述一个第一人脸特征集的每个聚类中心,分别与所述一个第一人脸特征集的其他各个聚类中心的相似度,构建第一类内相似度矩阵;

基于所述另一第一人脸特征集的每个聚类中心,分别与所述另一第一人脸特征集的其他各个聚类中心的相似度,构建第二类内相似度矩阵;

若基于所述第一类内相似度矩阵、所述第二类内相似度矩阵以及所述类间相似度矩阵,确定所述一个第一人脸特征集中的目标聚类中心与所述另一第一人脸特征集的各个聚类中心满足第一预设关系,则将所述目标聚类中心所在的第一人脸特征簇剔除;

若基于所述第一类内相似度矩阵、所述第二类内相似度矩阵以及所述类间相似度矩阵,确定所述另一第一人脸特征集中的目标聚类中心与所述一个第一人脸特征集的各个聚类中心满足第二预设关系,则将所述目标聚类中心所在的第一人脸特征簇剔除。

8. 一种人脸特征库构建装置,其特征在于,包括:

第一获取模块,用于针对第一实体标签集中的每个第一实体标签,获取所述第一实体标签对应的多个第一人物图像,对所述多个第一人物图像分别进行预处理,获得多个人脸特征;其中,所述预处理至少包括人脸检测以及人脸特征提取;

筛选模块,用于针对每个第一实体标签,对所述第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集;

比对模块,用于将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将所述任意两个第一人脸特征集进行合并处理;

构建模块,用于基于新获得的多个第一人脸特征集以及所述第一实体标签集,构建人脸特征库。

9. 一种电子设备,其特征在于,包括处理器和存储器,其中,所述存储器存储有可在所述处理器上运行的计算机程序,当所述计算机程序被所述处理器执行时,使得所述处理器执行权利要求1~7中任一所述方法的步骤。

10. 一种计算机可读存储介质,其特征在于,其存储有计算机程序,当所述计算机程序在电子设备上运行时,使得所述电子设备执行权利要求1~7中任一所述方法的步骤。

人脸特征库构建方法、装置、电子设备和存储介质

技术领域

[0001] 本申请涉及互联网技术领域,尤其涉及一种人脸特征库构建方法、装置、电子设备和存储介质。

背景技术

[0002] 本部分旨在为权利要求书中陈述的本公开的实施方式提供背景或上下文。此处的部分描述不因为包括在本部分中就承认是现有技术。

[0003] 随着移动视频产业的发展,每天都会有海量的视频被发布到网络上,这些视频中存在大量的影视混剪、人物合集等内容,但是,很多此类视频中没有明确人物的实体标签(例如人名),可以通过对视频图像中的人物进行人脸识别,得到具体的实体标签;这些实体标签一方面有利于视频平台对视频的检索和分发,向用户推荐其可能喜欢的视频,另一方面也可以剔除一些具有不良影响的人物的视频,避免这些视频进入视频平台。

[0004] 针对人物的人脸识别,通常存在妆容多变、环境多变、角度多变等问题,因此,为了更好地识别视频图像中的人物实体标签,需要构建更完备的人脸特征库,为每个需要识别的人物的实体标签提供更多的人物图像。

[0005] 但是,人脸特征库的构建耗时费力,如何提高人脸特征库的构建效率以及丰富性是需要解决的问题。

发明内容

[0006] 本公开实施例提供一种人脸特征库构建方法、装置、电子设备和存储介质,用于高效构建更完备的人脸特征库,在后续识别视频中人物的实体标签时,可以为每个人物的实体标签提供更多的人物图像,提高识别准确率。

[0007] 第一方面,本公开实施例提供了一种人脸特征库构建方法,包括:

[0008] 针对第一实体标签集中的每个第一实体标签,获取所述第一实体标签对应的多个第一人物图像,对所述多个第一人物图像分别进行预处理,获得多个人脸特征;其中,所述预处理至少包括人脸检测以及人脸特征提取;

[0009] 针对每个第一实体标签,对所述第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集;

[0010] 将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将所述任意两个第一人脸特征集进行合并处理;

[0011] 基于新获得的多个第一人脸特征集以及所述第一实体标签集,构建人脸特征库。

[0012] 可选的,所述对所述第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集,包括:

[0013] 对所述第一实体标签对应的多个人脸特征进行聚类处理,获得多个第一人脸特征簇,每个第一人脸特征簇包括一个聚类中心;

- [0014] 针对每个所述聚类中心,若所述聚类中心分别与各个其它聚类中心的相似度均不满足第一相似度条件,则删除所述聚类中心所在的第一人脸特征簇;
- [0015] 基于剩余的多个第一人脸特征簇,获得所述第一人脸特征集。
- [0016] 可选的,所述基于剩余的多个第一人脸特征簇,获得所述第一人脸特征集,包括:
- [0017] 针对剩余的每个第一人脸特征簇,确定所述第一人脸特征簇中的每个人脸特征与对应的聚类中心的相似度,将相似度不满足第二相似度条件的人脸特征从所述第一人脸特征簇中剔除;
- [0018] 基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集。
- [0019] 可选的,所述基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集,包括:
- [0020] 针对新获得的每个第一人脸特征簇,若所述第一人脸特征簇中的人脸特征的数量不满足数量条件,则将所述第一人脸特征簇删除。
- [0021] 可选的,所述基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集,包括:
- [0022] 针对新获得的每个第一人脸特征簇,确定所述第一人脸特征簇中的每个人脸特征分别与多个其它人脸特征的相似度,将相似度满足第三相似度条件的其它人脸特征从所述第一人脸特征簇中剔除;
- [0023] 基于再次新获得的多个第一人脸特征簇,获得所述第一人脸特征集。
- [0024] 可选的,所述将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将所述两个第一人脸特征集进行合并处理,包括:
- [0025] 针对每两个第一人脸特征集,执行以下操作:
- [0026] 获取一个第一人脸特征集的多个聚类中心,以及获取另一第一人脸特征集的多个聚类中心;
- [0027] 确定所述一个第一人脸特征集的每个聚类中心,分别与所述另一第一人脸特征集的多个聚类中心的相似度,获得相似度向量;
- [0028] 基于获得的多个相似度向量构建类间相似度矩阵,确定所述类间相似度矩阵中达到第一相似度阈值的的目标相似度,将所述目标相似度的数量作为所述相似度比对结果;
- [0029] 若所述相似度比对结果满足预设条件,则将所述一个第一人脸特征集与所述另一第一人脸特征集进行合并处理。
- [0030] 可选的,所述将所述一个第一人脸特征集与所述另一第一人脸特征集进行合并处理之前,还包括:
- [0031] 基于所述一个第一人脸特征集的每个聚类中心,分别与所述一个第一人脸特征集的其它各个聚类中心的相似度,构建第一类内相似度矩阵;
- [0032] 基于所述另一第一人脸特征集的每个聚类中心,分别与所述另一第一人脸特征集的其它各个聚类中心的相似度,构建第二类内相似度矩阵;
- [0033] 若基于所述第一类内相似度矩阵、所述第二类内相似度矩阵以及所述类间相似度矩阵,确定所述一个第一人脸特征集中的目标聚类中心与所述另一第一人脸特征集的各个聚类中心满足第一预设关系,则将所述目标聚类中心所在的第一人脸特征簇剔除;

[0034] 若基于所述第一类内相似度矩阵、所述第二类内相似度矩阵以及所述类间相似度矩阵,确定所述另一第一人脸特征集中的目标聚类中心与所述一个第一人脸特征集的各个聚类中心满足第二预设关系,则将所述目标聚类中心所在的第一人脸特征簇剔除。

[0035] 可选的,所述方法还包括:

[0036] 获取第二实体标签集以及第二人物图像集,所述第二实体标签集中的每个第二实体标签与所述第二人物图像集中的至少一个第二人物图像相对应;

[0037] 针对每个第二人物图像,若从所述人脸特征库中检索到与所述第二人物图像中的人脸特征相匹配的目标人脸特征,则将所述第二人物图像中的人脸特征保存至所述人脸特征库中所述目标人脸特征对应的目标实体标签下。

[0038] 可选的,所述若从所述人脸特征库中检索到与所述第二人物图像中的人脸特征相匹配的目标人脸特征,则将所述第二人物图像中的人脸特征保存至所述人脸特征库中所述目标人脸特征对应的目标实体标签下,包括:

[0039] 若从所述人脸特征库中检索到与所述第二人物图像中的人脸特征相匹配的目标人脸特征,则确定所述目标人脸特征是否满足以下条件中的至少一个:所述目标人脸特征对应的目标实体标签与所述第二人物图像对应的第二实体标签一致,所述目标人脸特征与所述第二人物图像中的人脸特征的相似度小于第二相似度阈值;

[0040] 若满足,则将所述第二人物图像中的人脸特征保存至所述人脸特征库中的所述目标实体标签下。

[0041] 可选的,所述方法还包括:

[0042] 针对每个第二人物图像,若从所述人脸特征库中未检索到与所述第二人物图像中的人脸特征相匹配的目标人脸特征,则将所述第二人物图像作为待入库的候选人物图像;

[0043] 基于获得的多个候选人物图像各自对应的第二标签,获得新的第二实体标签集,以及对所述多个候选人物图像分别进行预处理,获得第二人脸特征集;其中,所述预处理至少包括人脸检测以及人脸特征提取;

[0044] 将所述第二人脸特征集以及所述新的第二实体标签集保存至人脸特征临时库,所述新的第二实体标签集中的每个第二实体标签与所述第二人脸特征集中的至少一个人脸特征相对应。

[0045] 可选的,所述将所述第二人脸特征集以及所述新的第二实体标签集保存至人脸特征临时库,包括:

[0046] 将所述第二人脸特征集中的多个人脸特征进行聚类处理,获得多个第二人脸特征簇,每个第二人脸特征簇包括一个聚类中心;

[0047] 针对每个第二人脸特征簇,若所述第二人脸特征簇中的候选人脸特征与对应的聚类中心的相似度不满足第四相似度条件,则将所述候选人脸特征从所述第二人脸特征簇中剔除;

[0048] 基于新获得的多个第二人脸特征簇,获得新的第二人脸特征集,将所述新的第二人脸特征集以及所述新的第二实体标签集对应保存至所述人脸特征临时库。

[0049] 可选的,所述方法还包括:

[0050] 针对所述新的第二实体标签集中的每个第二实体标签,对所述第二实体标签对应的多个人脸特征进行所述预设筛选处理,获得第三人脸特征集;

[0051] 将获得多个第三人脸特征集以及所述多个第三人脸特征集各自对应的第二实体标签,保存至所述人脸特征库。

[0052] 可选的,所述获取所述第一实体标签对应的多个第一人物图像,包括:

[0053] 针对任一第一网络页面,对所述第一网络页面中的页面内容进行识别,若识别到人物图像以及与任一所述第一实体标签匹配的人物实体信息,则将识别到的人物图像作为任一所述第一实体标签对应的第一人物图像;

[0054] 从多个所述第一网络页面中,识别到每个所述第一实体标签对应的多个第一人物图像。

[0055] 可选的,所述获取第二实体标签集和第二人物图像集,包括:

[0056] 针对任一第二网络页面,对所述第二网络页面中的页面内容进行识别,若识别到人物图像以及人物实体信息,则基于所述人物实体信息获得第二实体标签,以及将识别到的人物图像作为所述第二实体标签对应的第二人物图像;

[0057] 基于由多个所述第二网络页面获得的各个第二实体标签,获得所述第二实体标签集,以及基于每个所述第二实体标签对应的至少一个第二人物图像,获得所述第二人物图像集。

[0058] 第二方面,本公开实施例还提供了一种人脸特征库构建装置,包括:

[0059] 第一获取模块,用于针对第一实体标签集中的每个第一实体标签,获取所述第一实体标签对应的多个第一人物图像,对所述多个第一人物图像分别进行预处理,获得多个人脸特征;其中,所述预处理至少包括人脸检测以及人脸特征提取;

[0060] 筛选模块,用于针对每个第一实体标签,对所述第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集;

[0061] 比对模块,用于将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将所述任意两个第一人脸特征集进行合并处理;

[0062] 构建模块,用于基于新获得的多个第一人脸特征集以及所述第一实体标签集,构建人脸特征库。

[0063] 可选的,所述筛选模块还用于:

[0064] 对所述第一实体标签对应的多个人脸特征进行聚类处理,获得多个第一人脸特征簇,每个第一人脸特征簇包括一个聚类中心;

[0065] 针对每个所述聚类中心,若所述聚类中心分别与各个其它聚类中心的相似度均不满足第一相似度条件,则删除所述聚类中心所在的第一人脸特征簇;

[0066] 基于剩余的多个第一人脸特征簇,获得所述第一人脸特征集。

[0067] 可选的,所述基于剩余的多个第一人脸特征簇,获得所述第一人脸特征集时,所述筛选模块还用于:

[0068] 针对剩余的每个第一人脸特征簇,确定所述第一人脸特征簇中的每个人脸特征与对应的聚类中心的相似度,将相似度不满足第二相似度条件的人脸特征从所述第一人脸特征簇中剔除;

[0069] 基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集。

[0070] 可选的,所述基于新获得的多个第一人脸特征簇,获得所述第一人脸特征集时,所

述筛选模块还用于：

[0071] 针对新获得的每个第一人脸特征簇，若所述第一人脸特征簇中的人脸特征的数量不满足数量条件，则将所述第一人脸特征簇删除。

[0072] 可选的，所述基于新获得的多个第一人脸特征簇，获得所述第一人脸特征集时，所述筛选模块还用于：

[0073] 针对新获得的每个第一人脸特征簇，确定所述第一人脸特征簇中的每个人脸特征分别与多个其它人脸特征的相似度，将相似度满足第三相似度条件的其它人脸特征从所述第一人脸特征簇中剔除；

[0074] 基于再次新获得的多个第一人脸特征簇，获得所述第一人脸特征集。

[0075] 可选的，所述比对模块还用于：

[0076] 针对每两个第一人脸特征集，执行以下操作：

[0077] 获取一个第一人脸特征集的多个聚类中心，以及获取另一第一人脸特征集的多个聚类中心；

[0078] 确定所述一个第一人脸特征集的每个聚类中心，分别与所述另一第一人脸特征集的多个聚类中心的相似度，获得相似度向量；

[0079] 基于获得的多个相似度向量构建类间相似度矩阵，确定所述类间相似度矩阵中达到第一相似度阈值的目标相似度，将所述目标相似度的数量作为所述相似度比对结果；

[0080] 若所述相似度比对结果满足预设条件，则将所述一个第一人脸特征集与所述另一第一人脸特征集进行合并处理。

[0081] 可选的，还包括剔除模块，用于：

[0082] 基于所述一个第一人脸特征集的每个聚类中心，分别与所述一个第一人脸特征集的其它各个聚类中心的相似度，构建第一类内相似度矩阵；

[0083] 基于所述另一第一人脸特征集的每个聚类中心，分别与所述另一第一人脸特征集的其它各个聚类中心的相似度，构建第二类内相似度矩阵；

[0084] 若基于所述第一类内相似度矩阵、所述第二类内相似度矩阵以及所述类间相似度矩阵，确定所述一个第一人脸特征集中的目标聚类中心与所述另一第一人脸特征集的各个聚类中心满足第一预设关系，则将所述目标聚类中心所在的第一人脸特征簇剔除；

[0085] 若基于所述第一类内相似度矩阵、所述第二类内相似度矩阵以及所述类间相似度矩阵，确定所述另一第一人脸特征集中的目标聚类中心与所述一个第一人脸特征集的各个聚类中心满足第二预设关系，则将所述目标聚类中心所在的第一人脸特征簇剔除。

[0086] 可选的，所述装置还包括：

[0087] 第二获取模块，用于获取第二实体标签集以及第二人物图像集，所述第二实体标签集中的每个第二实体标签与所述第二人物图像集中的至少一个第二人物图像相对应；

[0088] 第一更新模块，用于针对每个第二人物图像，若从所述人脸特征库中检索到与所述第二人物图像中的人脸特征相匹配的目标人脸特征，则将所述第二人物图像中的人脸特征保存至所述人脸特征库中所述目标人脸特征对应的目标实体标签下。

[0089] 可选的，所述第一更新模块还用于：

[0090] 若从所述人脸特征库中检索到与所述第二人物图像中的人脸特征相匹配的目标人脸特征，则确定所述目标人脸特征是否满足以下条件中的至少一个：所述目标人脸特征

对应的目标实体标签与所述第二人物图像对应的第二实体标签一致,所述目标人脸特征与所述第二人物图像中的人脸特征的相似度小于第二相似度阈值;

[0091] 若满足,则将所述第二人物图像中的人脸特征保存至所述人脸特征库中的所述目标实体标签下。

[0092] 可选的,所述装置还包括保存模块,用于:

[0093] 针对每个第二人物图像,若从所述人脸特征库中未检索到与所述第二人物图像中的人脸特征相匹配的目标人脸特征,则将所述第二人物图像作为待入库的候选人物图像;

[0094] 基于获得的多个候选人物图像各自对应的第二标签,获得新的第二实体标签集,以及对所述多个候选人物图像分别进行预处理,获得第二人脸特征集;其中,所述预处理至少包括人脸检测以及人脸特征提取;

[0095] 将所述第二人脸特征集以及所述新的第二实体标签集保存至人脸特征临时库,所述新的第二实体标签集中的每个第二实体标签与所述第二人脸特征集中的至少一个人脸特征相对应。

[0096] 可选的,所述将所述第二人脸特征集以及所述新的第二实体标签集保存至人脸特征临时库时,所述保存模块还用于:

[0097] 将所述第二人脸特征集中的多个人脸特征进行聚类处理,获得多个第二人脸特征簇,每个第二人脸特征簇包括一个聚类中心;

[0098] 针对每个第二人脸特征簇,若所述第二人脸特征簇中的候选人脸特征与对应的聚类中心的相似度不满足第四相似度条件,则将所述候选人脸特征从所述第二人脸特征簇中剔除;

[0099] 基于新获得的多个第二人脸特征簇,获得新的第二人脸特征集,将所述新的第二人脸特征集以及所述新的第二实体标签集对应保存至所述人脸特征临时库。

[0100] 可选的,所述装置还包括第二更新模块,用于:

[0101] 针对所述新的第二实体标签集中的每个第二实体标签,对所述第二实体标签对应的多个人脸特征进行所述预设筛选处理,获得第三人脸特征集;

[0102] 将获得多个第三人脸特征集以及所述多个第三人脸特征集各自对应的第二实体标签,保存至所述人脸特征库。

[0103] 可选的,所述第一获取模块还用于:

[0104] 针对任一第一网络页面,对所述第一网络页面中的页面内容进行识别,若识别到人物图像以及与任一所述第一实体标签匹配的人物实体信息,则将识别到的人物图像作为任一所述第一实体标签对应的第一人物图像;

[0105] 从多个所述第一网络页面中,识别到每个所述第一实体标签对应的多个第一人物图像。

[0106] 可选的,所述第二获取模块还用于:

[0107] 针对任一第二网络页面,对所述第二网络页面中的页面内容进行识别,若识别到人物图像以及人物实体信息,则基于所述人物实体信息获得第二实体标签,以及将识别到的人物图像作为所述第二实体标签对应的第二人物图像;

[0108] 基于由多个所述第二网络页面获得的各个第二实体标签,获得所述第二实体标签集,以及基于每个所述第二实体标签对应的至少一个第二人物图像,获得所述第二人物图

像集。

[0109] 第三方面,本公开实施例还提供了一种电子设备,包括处理器和存储器,所述存储器上存储有可在所述处理器上运行的计算机程序,当所述计算机程序被所述处理器执行时,使得所述处理器实现第一方面任一种人脸特征库构建方法的步骤。

[0110] 第四方面,本公开实施例还提供了一种计算机可读存储介质,其存储有计算机程序,当所述计算机程序在电子设备上运行时,使得所述电子设备执行第一方面任一种人脸特征库构建方法的步骤。

[0111] 第五方面,本公开实施例提供一种计算机程序产品,其包括计算机程序,所述计算机程序存储在计算机可读存储介质中;当电子设备的处理器从所述计算机可读存储介质读取所述计算机程序时,所述处理器执行该计算机程序,使得所述电子设备执行上述任一种人脸特征库构建方法的步骤。

[0112] 本公开实施例提供的人脸特征库构建方法至少具有如下有益效果:

[0113] 根据本公开实施例提供的方案,在构建人脸特征库时,首先获得包含多个人物的第一实体标签的第一实体标签集,针对每个第一实体标签,获取该第一实体标签对应的多个第一人人物图像,对这些第一人人物图像分别进行人脸检测以及人脸特征提取,获得多个人脸特征;然后,对每个第一实体标签对应的多个人脸特征进行预设筛选处理,获得每个第一实体标签对应的第一人人脸特征集,以完成每个第一实体标签内部的人脸特征筛选;并且,考虑到一个人物可能具有多个第一实体标签,因此需要确定是否将多个第一实体标签的第一人脸特征集进行合并,具体的,将多个第一人人脸特征集中的每两个第一人人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人人脸特征集的相似度比对结果满足预设条件,则将任意两个第一人人脸特征集进行合并处理;最后,基于新获得的多个第一人人脸特征集以及第一实体标签集,构建人脸特征库。

[0114] 基于本公开实施例的上述方案,可以高效构建更完备的人脸特征库,在后续识别视频中人物的实体标签时,可以为每个人物的实体标签提供更多的人物图像,提高识别准确率。

[0115] 本公开的其它特征和优点将在随后的说明书中阐述,并且,部分地从说明书中变得显而易见,或者通过实施本公开而了解。本公开的目的和其他优点可通过在所写的说明书、权利要求书、以及附图中所特别指出的结构来实现和获得。

附图说明

[0116] 为了更清楚地说明本公开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公开的一些实施例,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,还可以根据这些附图获得其他的附图。

[0117] 图1为本公开实施例提供的一种人脸特征库构建方法的流程图;

[0118] 图2为本公开实施例提供的另一种人脸特征库构建方法的流程图;

[0119] 图3为本公开实施例提供的另一种人脸特征库构建方法的流程图;

[0120] 图4为本公开实施例提供的另一种人脸特征库构建方法的流程图;

[0121] 图5为本公开实施例提供的另一种人脸特征库构建方法的流程图;

- [0122] 图6为本公开实施例提供的一种人脸特征库构建的逻辑示意图；
- [0123] 图7为本公开实施例提供的一种人脸特征库构建的逻辑示意图；
- [0124] 图8为本公开实施例提供的一种人脸特征库的构建过程示意图；
- [0125] 图9为本公开实施例提供的一种人脸特征集的筛选过程示意图；
- [0126] 图10为本公开实施例提供的另一种人脸特征库构建装置的示意图；
- [0127] 图11为本公开实施例提供的一种电子设备的结构示意图。

具体实施方式

[0128] 为了使本公开的目的、技术方案和优点更加清楚，下面将结合附图对本公开作进一步地详细描述，显然，所描述的实施例仅仅是本公开的一部分实施例，而不是全部的实施例。基于本公开中的实施例，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，都属于本公开保护的范围。

[0129] 为了便于本领域技术人员更好地理解本申请的技术方案，下面对本申请涉及的一些名词进行介绍。

[0130] 人脸检测：通过算法将视频图像中的人脸区域进行提取，得到只有人脸的图像，通常是一个正方形矩形。

[0131] 人脸识别：对一个人脸图像，通过将其与库中的人像进行比对，判断其身份标签。

[0132] 人脸特征提取：对人脸图像进行分析，提取其特征，特征通常是一个固定维度的向量。

[0133] 人脸特征库：将一批有身份标签的人脸图像通过人脸特征提取得到特征后进行存储得到的数据库。

[0134] 需要说明的是，本公开的说明书和权利要求书中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。

[0135] 此外，术语“包括”和“具有”以及它们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0136] 下面对本公开实施例的设计思想进行介绍。

[0137] 如前文，视频中的人物的实体标签有着重要的作用，针对没有明确人物的实体标签的视频，可以通过对视频图像中的人物进行人脸识别，得到具体的实体标签。针对人物的人脸识别，通常存在妆容多变、环境多变、角度多变等问题，为了更好地识别视频图像中的人物实体标签，需要构建更完备的人脸特征库，但是，人脸特征库的构建耗时费力，如何提高人脸特征库的构建效率以及丰富性是需要解决的问题。

[0138] 有鉴于此，本公开实施例提供一种人脸特征库构建方法，利用当前成熟的人脸识别技术，自动化完成大规模的人脸特征库的构建，提高了构建效率以及人脸特征库的丰富性，从而提高视频中的人物的实体标签的识别准确率。

[0139] 本公开技术方案中，对数据的采集、传播、使用等，均符合国家相关法律法规要求，

例如:数据包括人物图像等。

[0140] 下面结合附图以及具体实施方式对本申请实施例的人脸特征库构建方法进行介绍。

[0141] 本公开实施例的人脸特征库构建方法可以由终端设备或者服务器单独执行,也可以由服务器和终端设备配合执行。终端设备包括但不限于桌面计算机、手机、电脑、智能家电、智能语音交互设备、车载终端等电子设备。服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0142] 下面对本公开实施例的人脸特征库构建方法进行介绍。

[0143] 参阅图1所示,本公开实施例提供的一种人脸特征库构建方法,包括以下步骤S101-S104:

[0144] 步骤S101,针对第一实体标签集中的每个第一实体标签,获取第一实体标签对应的多个第一人物图像,对多个第一人物图像分别进行预处理,获得多个人脸特征;其中,预处理至少包括人脸检测以及人脸特征提取。

[0145] 其中,第一实体标签集可以包含预先设定的多个人物的第一实体标签,该第一实体标签可以是人物的姓名或称呼,每个人物可以对应一个或多个第一实体标。每个第一实体标签对应的各第一人物图像,可以通过设定方式获取,例如从网络中获取或者通过其它合法方式获取。

[0146] 可选的,当从网络中获取每个第一实体标签对应的多个第一人物图像时,可以执行以下步骤A1-A2:

[0147] A1、针对任一第一网络页面,对第一网络页面中的页面内容进行识别,若识别到人物图像以及与任一第一实体标签匹配的人物实体信息,则将识别到的人物图像作为任一第一实体标签对应的第一人物图像。

[0148] 本公开实施例中,第一网络页面可以是任一具有访问权限的网络平台中的网络页面,如果在某个第一网络页面中识别到人物图像,同时,基于自然语言处理分析提取到与任一第一实体标签匹配的人物实体信息,则可以将识别到的人物图像作为该第一实体标签对应的第一人物图像。

[0149] 需要说明的是,当在某个第一网络页面中识别到人物图像,同时识别到两个或多个第一实体标签时,可以将识别到的人物图像同时作为这些第一实体标签对应的第一人物图像,因为后续会对每个第一实体标签对应的各个第一人物图像进行筛选。

[0150] A2、从多个第一网络页面中,识别到每个第一实体标签对应的多个第一人物图像。

[0151] 基于上述步骤A1,可以分别对大量第一网络页面中的页面内容进行识别,并从其中的多个第一网络页面中分别获得相应的第一实体标签以及对应的第一人物图像,最终,获得每个第一实体标签对应的多个第一人物图像。

[0152] 在获得每个第一实体标签对应的多个第一人物图像后,针对每个第一实体标签,对多个第一人物图像分别进行人脸检测以及人脸特征提取等,获得多个人脸特征。通过对多个第一人物图像分别进行人脸检测,可以剔除无人脸图像,获取具有人脸的每个第一人物图像的人脸图像以及人脸关键点。基于每个人脸图像的人脸关键点对该人脸图像进行校

正;例如:基于人脸关键点模板,将人脸图像中的人脸关键点映射到这个模板上。进一步地,针对校正后的每个人脸图像,采用人脸特征提取算法进行特征提取,得到人脸特征。

[0153] 步骤S102,针对每个第一实体标签,对第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集。

[0154] 考虑到通过上述步骤S102获取的每个第一实体标签对应的各个人脸特征,可能存在与第一实体标签不匹配的人脸特征或者重复的人脸特征,需要删除这些人脸特征。

[0155] 本公开实施例中,针对每个第一实体标签,可以基于预设筛选方式对第一实体标签对应的多个人脸特征进行筛选处理,删除与第一实体标签不相关的人脸特征,以保证人脸特征的一致性和多样性;将筛选出的各个人脸特征组成第一人脸特征集。下面实施例中将进一步详细介绍筛选过程。

[0156] 步骤S103,将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将任意两个第一人脸特征集进行合并处理。

[0157] 该步骤中,考虑到一个人物对应的第一实体标签可能不止一个,如果上述第一实体标签集中包含同一人物的多个第一实体标签,则需要将这些第一实体标签的第一人脸特征集进行合并。

[0158] 具体的,基于预设方式将每两个第一人脸特征集进行相似度比对,如果某两个第一人脸特征集的相似度比对结果满足预设条件,则将这两个第一人脸特征集进行合并处理,以保证最终获得的每个人物对应一个第一人脸特征集。其中,每两个第一人脸特征集的相似度比对方式,以及上述预设条件均可以根据需要设置,本公开下面实施例中将进行示例性介绍。

[0159] 步骤S104,基于新获得的多个第一人脸特征集以及第一实体标签集,构建人脸特征库。

[0160] 通过本公开实施例的上述方案,可以高效构建更完备的人脸特征库,在后续识别视频中人物的实体标签时,可以为每个人物的实体标签提供更多的人物图像,提高识别准确率。

[0161] 下面对上述步骤S102中对第一实体标签对应的多个人脸特征进行预设筛选处理的过程进行介绍。

[0162] 在一些实施例中,如图2所示,上述步骤S102对第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集,可以包括以下步骤S1021-S1023:

[0163] 步骤S1021,对第一实体标签对应的多个人脸特征进行聚类处理,获得多个第一人人脸特征簇,每个第一人人脸特征簇包括一个聚类中心。

[0164] 其中,可以采用聚类算法对多个人脸特征进行聚类处理,聚类算法包括但不限于meanshift、K-Means等。具体的,以meanshift算法为例,使用较小的半径 r ,例如为0.8,对 N 个人脸特征 $\{F_1, F_2, F_3, F_4 \dots F_N\}$ 进行聚类,得到 K 个聚类中心,其中, $K < N$, K 的值由meanshift算法计算得到,将其它非聚类中心的人脸特征归类到一个聚类中心 F_k ,形成一组人脸特征簇 C_k (即第一人人脸特征簇),将不属于任何一个人脸特征簇的人脸特征删除。

[0165] meanshift是一种迭代聚类的算法,通过在一个样本点(即一个人脸特征)的半径 r 范围内,计算该样本点的均值漂移向量,不断移动聚类中心,使其向密度最大的地方移动,

最终找到概率密度的局部最大点,这些局部最大点就是聚类中心 F_k ,同时每个聚类中心 F_k 移动过程中,所经过的人脸特征组成了该聚类中心 F_k 所在的人脸特征簇 C_k 。

[0166] 步骤S1022,针对每个聚类中心,若聚类中心分别与各个其它聚类中心的相似度均不满足第一相似度条件,则删除聚类中心所在的第一人脸特征簇。

[0167] 例如,第一相似度条件可以是达到第一阈值,例如为0.5,对此不作限定。

[0168] 该步骤中,计算每个聚类中心分别与各个其它聚类中心的相似度,如果某个聚类中心与其它聚类中心的相似度均未超过(小于或等于)第一阈值,可以认为这个聚类中心所在的第一人脸特征簇,与其它聚类中心的第一人脸特征簇差距过大,需要删除。上述每两个聚类中心之间的相似度可以通过相似度算法进行计算,下面以相似度算法为计算余弦相似度为例。

[0169] 假设为K个聚类中心,将这些聚类中心两两计算余弦相似度,形成维度是K*K的类内相似度矩阵 M_{intra} ,其中,聚类中心a的特征向量 F_a 和聚类中心b的特征向量 F_b 的余弦相似度Similarity可以通过以下式(1)计算:

$$[0170] \quad \text{Similarity} = \frac{F_a \cdot F_b}{\|F_a\| \|F_b\|} = \frac{\sum_{i=1}^M f_{ai} f_{bi}}{\sqrt{\sum_{i=1}^M f_{ai}^2} \sqrt{\sum_{i=1}^M f_{bi}^2}} \quad (1)$$

[0171] 逐行遍历类内相似度矩阵 M_{intra} ,若某个聚类中心和其它聚类中心的余弦相似度均未超过第一阈值 T_1 ,则将这个聚类中心所在的第一人脸特征簇删除。

[0172] 通过上述实施方式,可以将每个第一实体标签的多个人脸特征进行内部数据清洗,删除不满足条件的第一人脸特征簇,使得筛选出的第一人脸特征簇与对应的第一实体标签相匹配。

[0173] 步骤S1023,基于剩余的多个第一人脸特征簇,获得第一人脸特征集。

[0174] 通过上述步骤S1022,将与其它第一人脸特征簇差距较大的第一人脸特征簇删除后,可以将剩余的多个第一人脸特征簇组成第一人脸特征集;还可以针对剩余的多个第一人脸特征簇继续进行筛选,以进一步保证筛选出的第一人脸特征簇与对应的第一实体标签相匹配。

[0175] 可选的,针对剩余的多个第一人脸特征簇继续进行筛选时,上述步骤S1023基于剩余的多个第一人脸特征簇,获得第一人脸特征集,可以包括以下步骤A1-A2:

[0176] A1、针对剩余的每个第一人脸特征簇,确定第一人脸特征簇中的每个人脸特征与对应的聚类中心的相似度,将相似度不满足第二相似度条件的人脸特征从第一人脸特征簇中剔除。

[0177] 例如,第二相似度条件可以是达到第二阈值,例如为0.6,对此不作限定。

[0178] A2、基于新获得的多个第一人脸特征簇,获得第一人脸特征集。

[0179] 在上述步骤S1022中,将多个聚类中心中不满足条件的聚类中心所在的第一人脸特征簇删除后,针对剩余的每个第一人脸特征簇,还可以将第一人脸特征簇内与聚类中心的相似度未达到第二阈值的人脸特征删除,以剔除每个第一人脸特征簇内与其聚类中心的差距较大的人脸特征。

[0180] 例如:对于一个第一人脸特征簇 C_k ,计算其中的每个人脸特征与聚类中心 F_k 的相似度,删除相似度未达到第二阈值 T_2 的人脸特征,以剔除噪声数据。

[0181] 为了使得每个第一人脸特征簇 C_k 中的人脸特征达到一定数量,以保证第一人脸特征簇 C_k 中的人脸特征彼此之间的距离在合理范围内,可以设置第一人脸特征簇的最小人脸特征数量,例如为2,如果第一人脸特征簇 C_k 中的人脸特征的数量小于最小人脸特征数量,则可以删除这个第一人脸特征簇 C_k 。

[0182] 可选的,上述步骤A2中基于新获得的多个第一人脸特征簇,获得第一人脸特征集时,针对新获得的每个第一人脸特征簇,若该第一人脸特征簇中的人脸特征的数量不满足数量条件,则将该第一人脸特征簇删除。

[0183] 在一些实施例中,上述步骤A2基于新获得的多个第一人脸特征簇,获得第一人脸特征集,还可以包括以下步骤:

[0184] A21、针对新获得的每个第一人脸特征簇,确定第一人脸特征簇中的每个人脸特征分别与多个其它人脸特征的相似度,将相似度满足第三相似度条件的其它人脸特征从第一人脸特征簇中剔除。

[0185] 例如,第三相似度条件可以是达到第三阈值,例如0.95,对此不作限定。

[0186] A22、基于再次新获得的多个第一人脸特征簇,获得第一人脸特征集。

[0187] 在上述步骤A1中将每个第一人脸特征簇内与聚类中心的差距较大的人脸特征删除后,考虑到每个第一人脸特征簇内,多个人脸特征之间可能存在十分相似的人脸特征,为了避免重复,对于十分相似的多个人脸特征可以只保留一个。

[0188] 例如:对于一个第一人脸特征簇 C_k 中的多个人脸特征,两两计算相似度,得到簇内相似度矩阵Mcluster,逐行遍历Mcluster,若存在与某个人脸特征的相似度达到第三阈值 T_3 的人脸特征,则将该人脸特征删除,以去除非常相似的人脸特征,避免重复。

[0189] 基于本申请上述实施例的筛选过程,可以获得每个第一实体标签对应的最终的第一人脸特征集。为了保证每个第一人脸特征集具有一定的数据量,可以设置第一人脸特征集的最小人脸特征数量,例如为5,如果某个第一人脸特征集中的人脸特征的数量小于最小人脸特征数量,认为可靠性不够,可以删除该第一人脸特征集,从而保证每个第一人脸特征集中的人脸特征既具有一致性,又具有多样性。

[0190] 下面实施例对上述步骤S103中将两个第一人脸特征集进行合并处理的具体过程进行介绍。

[0191] 通过上述实施例将每个第一实体标签对应的多个人脸特征进行内部数据清洗之后,还可以对第一实体标签之间进行数据清洗,具体的,将一个第一实体标签的第一人脸特征集与其它第一实体标签的第一人脸特征集进行分析比对,将属于同一人物的第一人脸特征集进行合并,进一步提高构建的人脸特征库中数据的准确性。

[0192] 在一些实施例中,如图3所示,上述步骤S103中将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将两个第一人脸特征集进行合并处理,可以包括以下步骤S1031-S1035:

[0193] 步骤S1031,针对每两个第一人脸特征集,执行以下步骤S1032-S1035。

[0194] 其中,获取通过上述步骤S1021-S1023进行筛选之后获得的多个第一人脸特征集,针对每两个第一人脸特征集执行下述步骤。

[0195] 步骤S1032,获取一个第一人脸特征集的多个聚类中心,以及获取另一第一人脸特

征集的多个聚类中心。

[0196] 该步骤中,每个第一人脸特征集的多个聚类中心,是通过上述步骤S1022对第一人脸特征集中的多个聚类中心进行筛选之后获得的。

[0197] 步骤S1033,确定一个第一人脸特征集的每个聚类中心,分别与另一第一人脸特征集的多个聚类中心的相似度,获得相似度向量。

[0198] 步骤S1034,基于获得的多个相似度向量构建类间相似度矩阵,确定类间相似度矩阵中达到第一相似度阈值的目标相似度,将目标相似度的数量作为相似度比对结果。

[0199] 例如,假设一个第一实体标签的第一人脸特征集A包含 K_1 个聚类中心,另一个第一实体标签的第一人脸特征集B包含 K_2 个聚类中心,针对第一人脸特征集A中的每个聚类中心,逐个计算与第一人脸特征集B中的 K_2 个聚类中心的相似度,得到一个维度为 K_2 的相似度向量,将获得的 K_1 个维度为 K_2 的相似度向量组成类间相似度矩阵 M_{class}^{AB} ,则 M_{class}^{AB} 的维度是 $K_1 * K_2$;遍历 M_{class}^{AB} ,统计 M_{class}^{AB} 中相似度达到相似度阈值的数量,作为类间相似度矩阵 M_{class}^{AB} 中达到相似度阈值的数量,例如相似度阈值为0.6,对此不作限定。

[0200] 步骤S1035,若相似度比对结果满足预设条件,则将一个第一人脸特征集与另一第一人脸特征集进行合并处理。

[0201] 该步骤中,预设条件可以是:类间相似度矩阵中达到相似度阈值的数量占比达到预设比值等。例如:目标相似度的数量大于80%,则认为上述两个第一实体标签对应同一人物,将对应的两个第一人脸特征集进行合并处理。

[0202] 如果上述目标相似度的数量不满足预设条件,可以认为上述两个第一实体标签之间的人脸特征无交叉,继续比对其中一个第一实体标签的第一人脸特征集与下一个第一实体标签的第一人脸特征集。

[0203] 在一些实施例中,考虑到一个第一人物图像中可能包含多个人物,导致不同的第一实体标签可能对应相同的人脸特征,需要将相同的人脸特征删除。针对通过上述步骤S1031-S1035确定需要合并处理的两个第一人脸特征集,可以继续内部数据清洗,以保证合并后的第一人脸特征集中的人脸特征不重复。

[0204] 可选的,上述步骤S1035中将一个第一人脸特征集与另一第一人脸特征集进行合并处理之前,还可以执行以下步骤B1-B4:

[0205] B1、基于一个第一人脸特征集的每个聚类中心,分别与一个第一人脸特征集的其他各个聚类中心的相似度,构建第一类内相似度矩阵;

[0206] 其中,第一类内相似度矩阵与上述实施例中的类内相似度矩阵 M_{intra} 的构建方式类似,假设上述一个第一人脸特征集包括 K 个聚类中心,则将这些聚类中心两两计算余弦相似度,形成维度是 $K * K$ 的第一类内相似度矩阵。

[0207] B2、基于另一第一人脸特征集的每个聚类中心,分别与另一第一人脸特征集的其他各个聚类中心的相似度,构建第二类内相似度矩阵。

[0208] 第二类内相似度矩阵与上述第一类内相似度矩阵的计算方式类似,在此不再赘述。

[0209] B3、若基于第一类内相似度矩阵、第二类内相似度矩阵以及类间相似度矩阵,确定一个第一人脸特征集中的目标聚类中心与另一第一人脸特征集的各个聚类中心满足第一

预设关系,则将目标聚类中心所在的第一人脸特征簇剔除;

[0210] 其中,类间相似度矩阵通过上述步骤S1034中获得,即针对第一人脸特征集A中的每个聚类中心,逐个计算与第一人脸特征集B中的K₂个聚类中心的相似度,得到一个维度为K₂的相似度向量,将获得的K₁个维度为K₂的相似度向量组成类间相似度矩阵 M_{class}^{AB} 。

[0211] 例如:第一人脸特征集A的第一类内相似度矩阵为 M_{intra}^A ,第一人脸特征集B的第二类内相似度矩阵为 M_{intra}^B ,类间相似度矩阵为 M_{class}^{AB} 。分别计算 M_{class}^{AB} 中每一行的均值和每一列的均值,得到 $\overline{(M_{class}^{AB})_{row}}$ 和 $\overline{(M_{class}^{AB})_{col}}$,再计算 M_{intra}^A 中每一行的均值(除去与当前行相同序号的列的元素),得到 $\overline{M_{intra}^A}$,计算 M_{intra}^B 中每一行的均值(除去与当前行相同序号的列的元素),得到 $\overline{M_{intra}^B}$ 。

[0212] 依次将 $\overline{(M_{class}^{AB})_{row}}$ 中的第i个元素与 $\overline{M_{intra}^A}$ 中的第i个元素比较,如果大于,则说明第一人脸特征集A的第i个聚类中心与第一人脸特征集B的相似度更高,删除第一人脸特征集A中的第i个聚类中心所在的第一人脸特征簇。

[0213] B4、若基于第一类内相似度矩阵、第二类内相似度矩阵以及类间相似度矩阵,确定另一第一人脸特征集中的目标聚类中心与一个第一人脸特征集的各个聚类中心满足第二预设关系,则将目标聚类中心所在的第一人脸特征簇剔除。

[0214] 与上述步骤B3类似,依次将 $\overline{(M_{class}^{AB})_{col}}$ 中的第j个元素与 $\overline{M_{intra}^B}$ 中的第j个元素比较,若大于,则说明第一人脸特征集B中的第j个聚类中心与第一人脸特征集A的相似度更高,删除第一人脸特征集B中的第j个聚类中心所在的第一人脸特征簇。

[0215] 基于本公开上述实施例,构建出一个多样的可靠的人脸特征库。下面实施例中可以实时获取新的人物图像以及对应的新的实体标签,定时更新构建的人脸特征库,以使人脸特征库更丰富。

[0216] 在一些实施例中,如图4所示,在上述实施例中的步骤S101-S104的基础上,还可以执行以下步骤S105-S106:

[0217] 步骤S105,获取第二实体标签集以及第二人物图像集,第二实体标签集中的每个第二实体标签与第二人物图像集中的至少一个第二人物图像相对应;

[0218] 可选的,步骤S105中获取第二实体标签集和第二人物图像集,可以包括以下步骤C1-C2:

[0219] C1、针对任一第二网络页面,对第二网络页面中的页面内容进行识别,若识别到人物图像以及人物实体信息,则基于人物实体信息获得第二实体标签,以及将识别到的人物图像作为第二实体标签对应的第二人物图像。

[0220] 本公开实施例中,第二网络页面可以是任一具有访问权限的网络平台中的网络页面,上述实施例中的第一网络页面可以是历史网络页面,这里的第二网络页面可以是新的网络页面。如果在某个第二网络页面中识别到人物图像,同时,基于自然语言处理分析提取到人物实体信息,则基于人物实体信息和人物图像,获得第二实体标签及其对应的第二人物图像。

[0221] C2、基于由多个第二网络页面获得的各个第二实体标签,获得第二实体标签集,以及基于每个第二实体标签对应的至少一个第二人物图像,获得第二人物图像集。

[0222] 步骤S106,针对每个第二人物图像,若从人脸特征库中检索到与第二人物图像中的人脸特征相匹配的目标人脸特征,则将第二人物图像中的人脸特征保存至人脸特征库中的目标人脸特征对应的目标实体标签下。

[0223] 其中,针对每个第二人物图像进行人脸检测,若检测到人脸,则对人脸图像进行人脸特征提取,获得人脸特征,从人脸特征库中检索是否存在与该人脸特征匹配的目标人脸特征,例如:两个人脸特征的相似度达到第三相似度阈值,可以认为这两个人脸特征相匹配。第二相似度阈值可以根据需要设置,例如为0.6,对此不作限定。

[0224] 可选的,为了保证上述第二人物图像中的人脸特征与目标实体标签相匹配,当从人脸特征库中检索到与第二人物图像中的人脸特征相匹配的目标人脸特征时,还可以确定目标人脸特征是否满足以下两种条件中的至少一个:

[0225] 第一种条件,目标人脸特征对应的目标实体标签与第二人物图像对应的第二实体标签一致;

[0226] 第二种条件,目标人脸特征与第二人物图像中的人脸特征的相似度小于第二相似度阈值;其中,第二相似度阈值可以设置的较高,例如为0.9或者0.95等,对此不作限定,通过该条件的限制,可以避免将十分相似的人脸特征重复入库。

[0227] 如果目标人脸特征满足上述两种条件中的至少一个,则将第二人物图像中的人脸特征保存至人脸特征库中的目标实体标签下。

[0228] 步骤S107,针对每个第二人物图像,若从人脸特征库中未检索到与第二人物图像中的人脸特征相匹配的目标人脸特征,则将第二人物图像作为待入库的候选人物图像。

[0229] 步骤S108,基于获得的多个候选人物图像各自对应的第二标签,获得新的第二实体标签集,以及对多个候选人物图像分别进行预处理,获得第二人脸特征集;其中,预处理至少包括人脸检测以及人脸特征提取。

[0230] 该步骤中,对多个候选人物图像分别进行预处理的过程,与本公开上述实施例中步骤S101的实施过程类似,在此不再赘述。

[0231] 步骤S109,将第二人脸特征集以及新的第二实体标签集保存至人脸特征临时库,新的第二实体标签集中的每个第二实体标签与第二人脸特征集中的至少一个人脸特征相对应。

[0232] 可选的,如图5所示,步骤S109中将第二人脸特征集以及新的第二实体标签集保存至人脸特征临时库,可以包括以下步骤S1091-S1093:

[0233] 步骤S1091,将第二人脸特征集中的多个人脸特征进行聚类处理,获得多个第二人脸特征簇,每个第二人脸特征簇包括一个聚类中心。

[0234] 其中,可以基于聚类算法对多个人脸特征进行聚类处理,例如聚类算法为kmeans,设置聚类中心的数量与第二实体标签的数量相同,即每个第二实体标签对应一个第二人脸特征簇。

[0235] 步骤S1092,针对每个第二人脸特征簇,若第二人脸特征簇中的候选人脸特征与对应的聚类中心的相似度不满足第四相似度条件,则将候选人脸特征从所述第二人脸特征簇中剔除。

[0236] 其中,第四相似度条件可以是达到第四阈值,例如0.6,对此不作限定。

[0237] 步骤S1093,基于新获得的多个第二人脸特征簇,获得新的第二人脸特征集,将新的第二人脸特征集以及新的第二实体标签集对应保存至人脸特征临时库。

[0238] 在一些实施例中,针对上述人脸特征临时库中的新的第二人脸特征集以及新的第二实体标签集,可以定期保存至构建的人脸特征库中,具体可以执行以下步骤D1-D2:

[0239] D1、针对新的第二实体标签集中的每个第二实体标签,对第二实体标签对应的多个人脸特征进行预设筛选处理,获得第三人脸特征集。

[0240] 该步骤与本公开上述实施例中的步骤S102的实施过程类似,即将多个人脸特征进行聚类处理,获得多个第三人脸特征簇,再继续执行筛选处理,在此不再赘述。

[0241] D2、将获得多个第三人脸特征集以及多个第三人脸特征集各自对应的第二实体标签,保存至人脸特征库。

[0242] 本公开实施例中,可以每隔固定时间,例如几天、一周等,对人脸特征临时库中的新的第二人脸特征集以及新的第二实体标签集,执行上述步骤D1和D2。为了保证准确性,在执行上述步骤D1时,可以设定上述第三人脸特征簇中的最小人脸特征数量,例如为5,对此不作做限定;在执行上述步骤D2时,设定每个第二实体标签对应的第三人脸特征集中的最小人脸特征数量,例如为10,对此不作做限定。

[0243] 下面结合具体实例对本公开实施例的人脸特征库的构建方法以及更新方法进行示例性介绍。

[0244] 如图6所示,在构建阶段,使用大量的互联网历史数据(历史网络页面中的页面内容),基于人脸特征库构建系统,构建初始的人脸特征库,后续每隔固定时间,对于互联网新增数据(新的网络页面中的页面内容),基于人脸特征库更新系统,若判定新增的人物图像中的人脸特征,属于人脸特征库中已有的人物(具有上述第一实体标签),则将该人物图像中的人脸特征加入人脸特征库,或者,当该人物图像中的人脸特征与人脸特征库中已有的人脸特征重复时,丢弃该人物图像中的人脸特征;若判定新增的人物图像中的人脸特征不属于人脸特征库中已有的人物,则先将该人物图像存入人脸特征临时库,定期对人脸特征临时库中的人物图像进行汇总,当人物图像的数量达到设定数量时,可以将人脸特征临时库合并入人脸特征库,完成人脸特征库的更新。

[0245] 下面对人脸特征库构建系统的构建过程进行示例性介绍。

[0246] 人脸特征库构建系统基于一份预先获取的人物名单,将每个人物的姓名或称呼生成一个人物实体标签(对应上述实施例中第一实体标签),具体的,从互联网历史数据中获取大量人物图像,构成每个人物实体标签对应的人物图像集,获得的大量人物图像通常存在以下问题:

[0247] 1、人物图像中无人脸,需要删除;

[0248] 2、多个重复相同的人物图像,需要去重;

[0249] 3、人物图像中的人物与人物实体标签不符,需要删除;

[0250] 4、一个人物可能不止有一个常用名,导致一个人物对应多个人物图像集,需要合并处理;

[0251] 5、一个人物图像中同时存在多个人物,导致不同人物对应的人物图像集之间有混合,需要删除合并。

[0252] 针对以上问题,本公开实施例通过以下流程对多个人物实体标签对应的人物图像集进行清洗筛选,完成可靠的人脸特征库的构建,如图7所示,具体包括以下流程:

[0253] 1) 人脸检测:对每个人物实体标签对应的人物图像集进行人脸检测,剔除无人脸图像,获取人脸图像集,并基于每个人脸图像中的人脸关键点对人脸图像进行校正;

[0254] 2) 人脸特征提取:针对每个人物实体标签对应的多个矫正后的人脸图像,使用人脸特征提取算法对校正后的每个人脸图像进行特征提取,得到人脸特征;

[0255] 3) 人物实体标签内部清洗:对每个人物实体标签的多个人脸特征进行聚类处理,得到多个聚类中心,同时删除与该人物实体标签不相关的人脸特征,以保证每个人物实体标签的人脸特征的一致性和多样性;

[0256] 4) 人物实体标签间清洗:计算每个人物实体标签的人脸特征集与其它人物实体标签之间的人脸特征集相似度,应该合并的不同人脸特征集进行合并,同时将存在一定相似性的混杂人脸特征进行剔除,得到最终的人脸特征库。

[0257] 下面对上述人脸检测及人脸特征提取过程进行示例性介绍。

[0258] 针对每一个人物实体标签对应的人物图像集,对人物图像集使用人脸检测器和特征提取器,提取到一组数量为 N 的人脸特征集合,记为 $\{F_1, F_2, F_3, F_4 \dots F_N\}$ 。其中每一个 F_n 是一个一维的特征向量 $(f_1, f_2, f_3, \dots, f_M)$,向量长度 M 由特征提取器确定,例如为256或512。

[0259] 下面对人物实体标签内部数据清洗过程进行示例性介绍。

[0260] 基于上述过程,针对每一个人物实体标签,初步获得 N 个人脸特征,这些人脸特征大部分属于同一个人物,但由于网络数据的杂乱,可能存在一部分噪声数据,首先对这 N 个人脸特征进行分析筛选以剔除噪声数据。如图8所示,主要筛选流程如下:

[0261] a1、采用聚类算法,例如meanshift算法,使用较小的半径 r ,如0.8,对一个人物实体标签的 N 个人脸特征 $\{F_1, F_2, F_3, F_4 \dots F_N\}$ 进行聚类,到 K 个聚类中心,其中 $K < N$,是由meanshift算法计算得到,将其它的人脸特征归类到一个聚类中心 F_k ,形成一组人脸特征簇 C_k ,将不属于任何一个人脸特征簇的人脸特征删除。

[0262] a2、将这 K 个聚类中心,两两计算相似度,例如余弦相似度,形成维度是 $K * K$ 的类内相似度矩阵 M_{intra} ,其中特征向量 F_a 和特征向量 F_b 的相似度计算公式如上述实施例中的式(1)。

[0263] 逐行遍历矩阵 M_{intra} ,若某个聚类中心和其它聚类中心的相似度均小于第一阈值 $T_1 = 0.5$,认为这个聚类中心所在的人脸特征簇中的人脸特征,与该人物实体标签下的其它人脸特征差距过大,删除这个聚类中心所在的人脸特征簇;

[0264] a3、对于一个人脸特征簇 C_k ,计算其中的每一个人脸特征到聚类中心 F_k 的相似度,删除相似度小于第二阈值 $T_2 = 0.6$ 的人脸特征,以剔除噪声数据,若此时该人脸特征簇中的人脸特征数量小于簇内的最小人脸特征数量,例如为2,则删除这个人脸特征簇,以保证人脸特征簇中的人脸特征之间的距离在合理范围内;

[0265] a4、对人脸特征簇 C_k 中剩下的人脸特征,两两计算相似度,得到簇内相似度矩阵 $M_{cluster}$,逐行遍历 $M_{cluster}$,若存在与当前人脸特征的相似度大于第三阈值 $T_3 = 0.95$ 的人脸特征,则删除该人脸特征,以去除非常相似的人脸特征,避免重复入库;

[0266] 通过以上筛选,确定最后获得的每个人物实体标签对应的人脸特征的数量,若小于类内的最小人脸特征数量,例如为5,则认为可靠性不够,删除该人物实体标签对应的人

脸特征集。

[0267] 通过以上步骤,保留下来的每一个人物实体标签对应的人脸特征集,都有一定的数据量,且具有一定的一致性和多样性。

[0268] 下面对人物实体标签间数据清洗过程进行示例性介绍。

[0269] 人物实体标签之间的数据清洗,指将一个人物实体标签的人脸特征集与其它人物实体标签的人脸特征集进行分析比对,将属于同一个人物的人脸特征集合并,不同人物之间的人脸特征无交叉,进一步提高人脸特征库中数据的准确性。具体流程包括:

[0270] b1、获取一个人物实体标签的人脸特征集A的多个聚类中心,设为聚类中心的数量为 K_1 ,将人脸特征集A的每个聚类中心,逐个与其它人物实体标签的人脸特征集B的多个聚类中心进行余弦相似度计算,得到类间相似度矩阵 M_{class}^{AB} ,设人脸特征集B的聚类中心的数量为 K_2 ,则 M_{class}^{AB} 的维度是 $K_1 * K_2$;

[0271] b2、遍历类间相似度矩阵 M_{class}^{AB} ,统计 M_{class}^{AB} 中元素值大于第二阈值 $T_2=0.6$ 的数量,若所有元素的值都小于0.6,则认为这两个人脸特征集的人脸特征无交叉,继续比对下一个人物实体标签对应的人脸特征集,若 M_{class}^{AB} 中的元素值大于0.6的数量大于或等于80%,则认为这两个人脸特征集属于同一个人物,对这两个人脸特征集进行合并,对合并后的人脸特征集再次进行人物实体标签内部数据清洗操作;

[0272] b3、对类间相似度矩阵 M_{class}^{AB} 分别求每一行的均值和每一列的均值,得到 $(\overline{M_{class}^{AB}})_{row}$ 和 $(\overline{M_{class}^{AB}})_{col}$,再对人脸特征集A的类内相似度矩阵 M_{intra}^A 和人脸特征集B的类内相似度矩阵 M_{intra}^B ,分别求每一行的均值(除去与当前行相同序号的列的元素),得到 $\overline{M_{intra}^A}$ 和 $\overline{M_{intra}^B}$ 。依次将 $(\overline{M_{class}^{AB}})_{row}$ 中的第i个元素与 $\overline{M_{intra}^A}$ 中的第i个元素比较,若前者大于后者,则说明人物A的第i个聚类中心与B相似度更高,直接删除A中第i个聚类中心及其对应的特征簇。同样的,依次将 $(\overline{M_{class}^{AB}})_{col}$ 中的第j个元素与 $\overline{M_{intra}^B}$ 中的第j个元素比较,若前者大于后者,则说明人脸特征集B中的第j个聚类中心与人脸特征集A相似度更高,直接删除人脸特征集B中的第j个聚类中心及其对应的人脸特征簇。

[0273] 基于以上步骤,一个多样的可靠的人脸特征库构建完成。

[0274] 下面对人脸特征库更新系统的更新过程进行示例性介绍。

[0275] 网络平台中每日都有海量的新增内容,产生一些新的人物图像,基于新的人物图像更新人脸特征库,可以进一步提高人物人脸识别的时效性和准确性。人脸特征库更新系统,通过网络平台中获取新的人物图像和对应的新的人物实体标签,定时对人脸特征库进行更新。具体流程包括:

[0276] c1、定期从具有访问权限的网络平台中获取网络页面,针对一个网络页面的页面内容,基于自然语言处理分析提取人物实体标签信息,若提取到人物实体标签信息,同时识别到人物图像,得到新的人物实体标签集 $labelA = \{label_1, label_2 \dots label_n\}$,以及新的人物图像集 $imgA = \{img_1, img_2 \dots img_m\}$ 。

[0277] c2、将imgA中的人物图像依次送入人脸特征库进行检索,设定识别阈值为0.6,若

检索到标签label_i,且label_i属于labelA,则在labelA中删除label_i,若识别置信度小于第三阈值 $T_3=0.95$,则将当前人物图像中的人脸特征保存至人脸特征库中;若label_i不属于labelA,则直接删除当前人物图像。

[0278] c3、将T_2=0.6,则删除当前人脸特征。将剩下的人脸特征保存至人脸特征临时库中,同时,每一个人脸特征对应labelB中的一个实体标签。

[0279] c4、每隔固定时间(如几天、一周),对人脸特征临时库中的每个人物实体标签对应的人脸特征集,使用人物实体标签内部数据清洗过程进行清洗,为了保证准确性,设置人脸特征簇中的最小人脸特征数量,例如为5,以及设置每个人物实体标签的人脸特征集中的最小人脸特征数量,例如为10。将每个人物实体标签对应的人脸特征集,进行人物实体标签内部数据清洗后,若剩余的人脸特征的数量达到10,则将该人物实体标签及其对应的人脸特征集保存至人脸特征库中,同时,删除人脸特征临时库中的该人物实体标签及其对应的人脸特征集。

[0280] 通过上述更新过程,实现人物实体内部多样性的更新以及新晋人物入库的时效性,进一步提高了人物人脸特征库的准确性。

[0281] 本公开实施例提出了一种自动化快速构建和更新大规模人脸特征库的方案,可以实现在无人工或极少人工参与下的人脸特征库的构建和更新。通过人物实体标签内部清洗和人物实体标签间清洗的双维度过滤筛选方案,通过聚类算法和多维度相似度矩阵分析,有效实现数据筛选,实现高准确率、多样性的人脸特征库的构建。

[0282] 基于相同的发明构思,本公开实施例还提供了一种人脸特征库构建装置,该装置解决问题的原理与上述实施例的方法相似,因此该装置的实施可以参见方法的实施,重复之处不再赘述。

[0283] 如图9所示,本公开实施例还提供了一种人脸特征库构建装置,包括:

[0284] 第一获取模块91,用于针对第一实体标签集中的每个第一实体标签,获取第一实体标签对应的多个第一人物图像,对多个第一人物图像分别进行预处理,获得多个人脸特征;其中,预处理至少包括人脸检测以及人脸特征提取;

[0285] 筛选模块92,用于针对每个第一实体标签,对第一实体标签对应的多个人脸特征进行预设筛选处理,获得第一人脸特征集;

[0286] 比对模块93,用于将多个第一人脸特征集中的每两个第一人脸特征集进行相似度比对,在每次相似度比对中,若两个第一人脸特征集的相似度比对结果满足预设条件,则将任意两个第一人脸特征集进行合并处理;

[0287] 构建模块94,用于基于新获得的多个第一人脸特征集以及第一实体标签集,构建人脸特征库。

[0288] 可选的,筛选模块92还用于:

[0289] 对第一实体标签对应的多个人脸特征进行聚类处理,获得多个第一人脸特征簇,每个第一人脸特征簇包括一个聚类中心;

[0290] 针对每个聚类中心,若聚类中心分别与各个其它聚类中心的相似度均不满足第一相似度条件,则删除聚类中心所在的第一人脸特征簇;

[0291] 基于剩余的多个第一人脸特征簇,获得第一人脸特征集。

[0292] 可选的,基于剩余的多个第一人脸特征簇,获得第一人脸特征集时,筛选模块92还用于:

[0293] 针对剩余的每个第一人脸特征簇,确定第一人脸特征簇中的每个人脸特征与对应的聚类中心的相似度,将相似度不满足第二相似度条件的人脸特征从第一人脸特征簇中剔除;

[0294] 基于新获得的多个第一人脸特征簇,获得第一人脸特征集。

[0295] 可选的,基于新获得的多个第一人脸特征簇,获得第一人脸特征集时,筛选模块92还用于:

[0296] 针对新获得的每个第一人脸特征簇,确定第一人脸特征簇中的每个人脸特征分别与多个其它人脸特征的相似度,将相似度满足第三相似度条件的其它人脸特征从第一人脸特征簇中剔除;

[0297] 基于再次新获得的多个第一人脸特征簇,获得第一人脸特征集。

[0298] 可选的,基于再次新获得的多个第一人脸特征簇,获得第一人脸特征集时,筛选模块92还用于:

[0299] 针对再次新获得的每个第一人脸特征簇,若第一人脸特征簇中的人脸特征的数量不满足数量条件,则将第一人脸特征簇删除。

[0300] 可选的,比对模块93还用于:

[0301] 针对每两个第一人脸特征集,执行以下操作:

[0302] 获取一个第一人脸特征集的多个聚类中心,以及获取另一第一人脸特征集的多个聚类中心;

[0303] 确定一个第一人脸特征集的每个聚类中心,分别与另一第一人脸特征集的多个聚类中心的相似度,获得相似度向量;

[0304] 基于获得的多个相似度向量构建类间相似度矩阵,确定类间相似度矩阵中达到第一相似度阈值的目标相似度,将目标相似度的数量作为相似度比对结果;

[0305] 若相似度比对结果满足预设条件,则将一个第一人脸特征集与另一第一人脸特征集进行合并处理。

[0306] 可选的,如图10所示,还包括剔除模块95,用于:

[0307] 基于一个第一人脸特征集的每个聚类中心,分别与一个第一人脸特征集的其它各个聚类中心的相似度,构建第一类内相似度矩阵;

[0308] 基于另一第一人脸特征集的每个聚类中心,分别与另一第一人脸特征集的其它各个聚类中心的相似度,构建第二类内相似度矩阵;

[0309] 若基于第一类内相似度矩阵、第二类内相似度矩阵以及类间相似度矩阵,确定一个第一人脸特征集中的目标聚类中心与另一第一人脸特征集的各个聚类中心满足第一预设关系,则将目标聚类中心所在的第一人脸特征簇剔除;

[0310] 若基于第一类内相似度矩阵、第二类内相似度矩阵以及类间相似度矩阵,确定另一第一人脸特征集中的目标聚类中心与一个第一人脸特征集的各个聚类中心满足第二预

设关系,则将目标聚类中心所在的第一人脸特征簇剔除。

[0311] 可选的,装置还包括:

[0312] 第二获取模块96,用于获取第二实体标签集以及第二人物图像集,第二实体标签集中的每个第二实体标签与第二人物图像集中的至少一个第二人物图像相对应;

[0313] 第一更新模块97,用于针对每个第二人物图像,若从人脸特征库中检索到与第二人物图像中的人脸特征相匹配的目标人脸特征,则将第二人物图像中的人脸特征保存至人脸特征库中的目标人脸特征对应的目标实体标签下。

[0314] 可选的,第一更新模块97还用于:

[0315] 若从人脸特征库中检索到与第二人物图像中的人脸特征相匹配的目标人脸特征,则确定目标人脸特征是否满足以下条件中的至少一个:目标人脸特征对应的目标实体标签与第二人物图像对应的第二实体标签一致,目标人脸特征与第二人物图像中的人脸特征的相似度小于第二相似度阈值;

[0316] 若满足,则将第二人物图像中的人脸特征保存至人脸特征库中的目标实体标签下。

[0317] 可选的,装置还包括保存模块98,用于:

[0318] 针对每个第二人物图像,若从人脸特征库中未检索到与第二人物图像中的人脸特征相匹配的目标人脸特征,则将第二人物图像作为待入库的候选人物图像;

[0319] 基于获得的多个候选人物图像各自对应的第二标签,获得新的第二实体标签集,以及对多个候选人物图像分别进行预处理,获得第二人脸特征集;其中,预处理至少包括人脸检测以及人脸特征提取;

[0320] 将第二人脸特征集以及新的第二实体标签集保存至人脸特征临时库,新的第二实体标签集中的每个第二实体标签与第二人脸特征集中的至少一个人脸特征相对应。

[0321] 可选的,将第二人脸特征集以及新的第二实体标签集保存至人脸特征临时库时,保存模块98还用于:

[0322] 将第二人脸特征集中的多个人脸特征进行聚类处理,获得多个第二人脸特征簇,每个第二人脸特征簇包括一个聚类中心;

[0323] 针对每个第二人脸特征簇,若第二人脸特征簇中的候选人脸特征与对应的聚类中心的相似度不满足第四相似度条件,则将候选人脸特征从所述第二人脸特征簇中剔除;

[0324] 基于新获得的多个第二人脸特征簇,获得新的第二人脸特征集,将新的第二人脸特征集以及新的第二实体标签集对应保存至人脸特征临时库。

[0325] 可选的,装置还包括第二更新模块99,用于:

[0326] 针对新的第二实体标签集中的每个第二实体标签,对第二实体标签对应的多个人脸特征进行预设筛选处理,获得第三人脸特征集;

[0327] 将获得多个第三人脸特征集以及多个第三人脸特征集各自对应的第二实体标签,保存至人脸特征库。

[0328] 可选的,第一获取模块91还用于:

[0329] 针对任一第一网络页面,对第一网络页面中的页面内容进行识别,若识别到人物图像以及与任一第一实体标签匹配的人物实体信息,则将识别到的人物图像作为任一第一实体标签对应的第一人物图像;

[0330] 从多个第一网络页面中,识别到每个第一实体标签对应的多个第一人物图像。

[0331] 可选的,第二获取模块96还用于:

[0332] 针对任一第二网络页面,对第二网络页面中的页面内容进行识别,若识别到人物图像以及人物实体信息,则基于人物实体信息获得第二实体标签,以及将识别到的人物图像作为第二实体标签对应的第二人物图像;

[0333] 基于由多个第二网络页面获得的各个第二实体标签,获得第二实体标签集,以及基于每个第二实体标签对应的至少一个第二人物图像,获得第二人物图像集。

[0334] 为了描述的方便,以上各部分按照功能划分为各模块分别描述。当然,在实施本申请时可以把各模块的功能在同一个或多个软件或硬件中实现。

[0335] 关于上述实施例中的装置,其中各个模块的具体执行方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0336] 基于相同的发明构思,本公开实施例还提供了一种电子设备,该电子设备解决问题的原理与上述实施例的方法相似,因此该电子设备的实施可以参见方法的实施,重复之处不再赘述。

[0337] 参阅图11所示,电子设备可以包括处理器1102和存储器1101。存储器1101向处理器1102提供存储器1101中存储的程序指令和数据。在本公开实施例中,存储器1101可以用于存储本公开实施例中人脸特征库构建的程序。

[0338] 处理器1102通过调用存储器1101存储的程序指令,处理器1102用于执行上述任意方法实施例中的方法,例如图2所示的实施例提供的一种人脸特征库构建方法。

[0339] 本公开实施例中不限定上述存储器1101和处理器1102之间的具体连接介质。本公开实施例在图11中以存储器1101和处理器1102之间通过总线1103连接,总线1103在图11中以粗线表示,其它部件之间的连接方式,仅是进行示意性说明,并不引以为限。总线1103可以分为地址总线、数据总线、控制总线等。为便于表示,图11中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0340] 存储器可以包括只读存储器(Read-Only Memory,ROM)和随机存取存储器(Random Access Memory,RAM),还可以包括非易失性存储器(Non-Volatile Memory,NVM),例如至少一个磁盘存储器。可选地,存储器还可以是至少一个位于远离前述处理器的存储装置。

[0341] 上述处理器可以是通用处理器,包括中央处理器、网络处理器(Network Processor,NP)等;还可以是数字指令处理器(Digital Signal Processing,DSP)、专用集成电路、现场可编程门阵列或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。

[0342] 本公开实施例还提供了一种计算机存储介质,该计算机可读存储介质内存储有计算机程序,电子设备的处理器从计算机可读存储介质读取该计算机程序,处理器执行该计算机程序,使得该电子设备执行上述任意方法实施例中的脸特征库构建方法。

[0343] 在具体的实施过程中,计算机存储介质可以包括:通用串行总线闪存盘(USB,Universal Serial Bus Flash Drive)、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的存储介质。

[0344] 与上述方法实施例基于同一发明构思,本申请实施例提供了一种计算机程序产品

或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述任意一种人脸特征库构建方法的步骤。

[0345] 程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以是但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPR0M或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0346] 本领域内的技术人员应明白,本公开的实施例可提供为方法、系统、或计算机程序产品。因此,本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本公开可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0347] 本公开是参照根据本公开的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0348] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0349] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0350] 显然,本领域的技术人员可以对本公开进行各种改动和变型而不脱离本公开的精神和范围。这样,倘若本公开的这些修改和变型属于本公开权利要求及其等同技术的范围之内,则本公开也意图包含这些改动和变型在内。

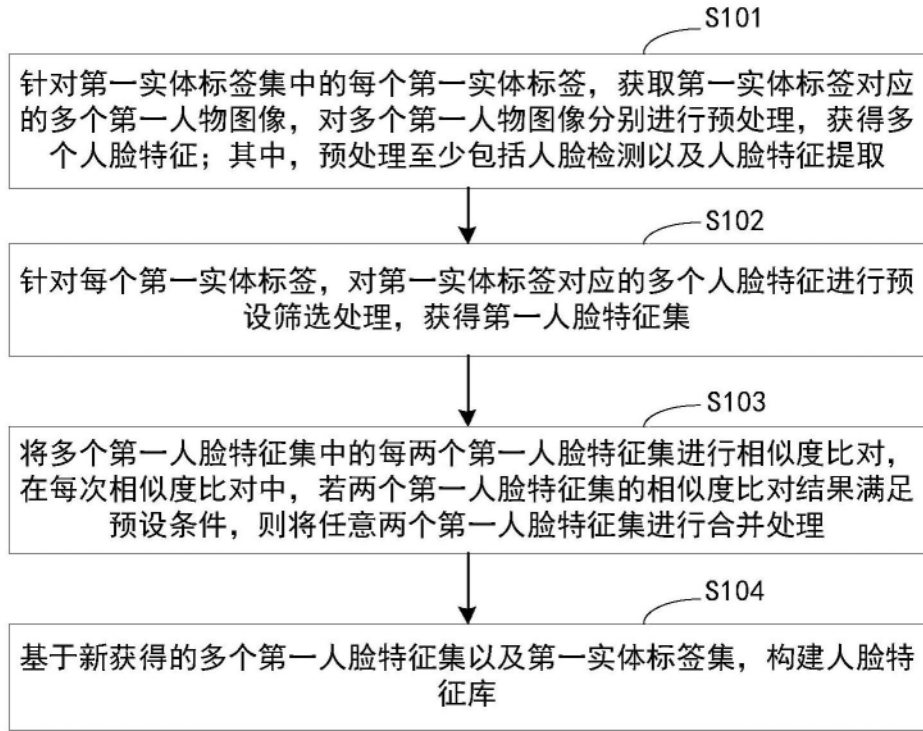


图1

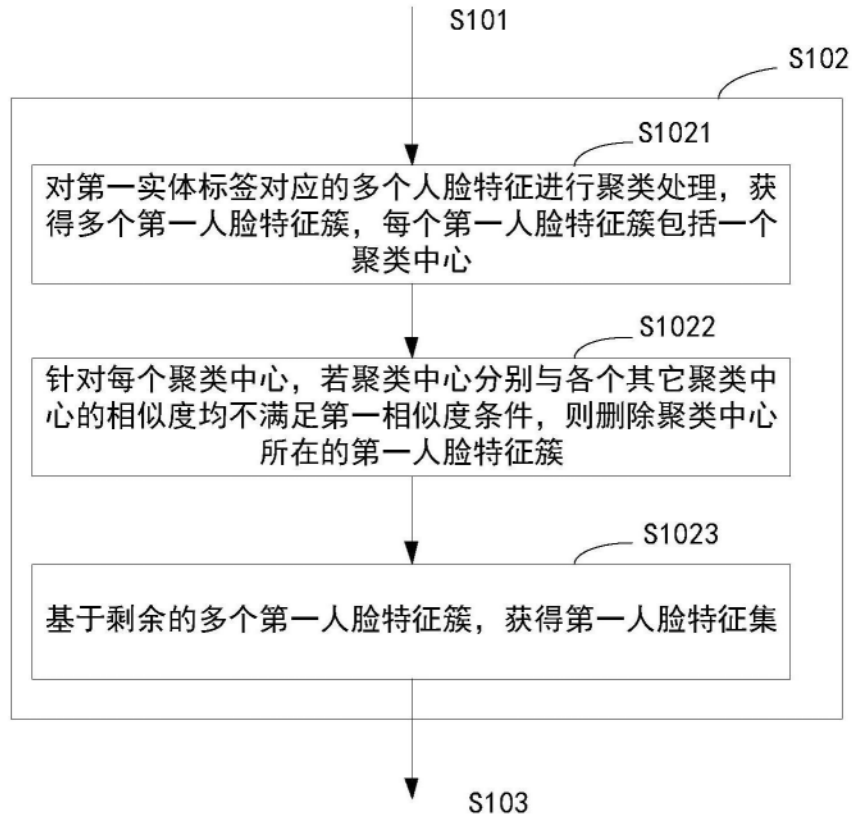


图2

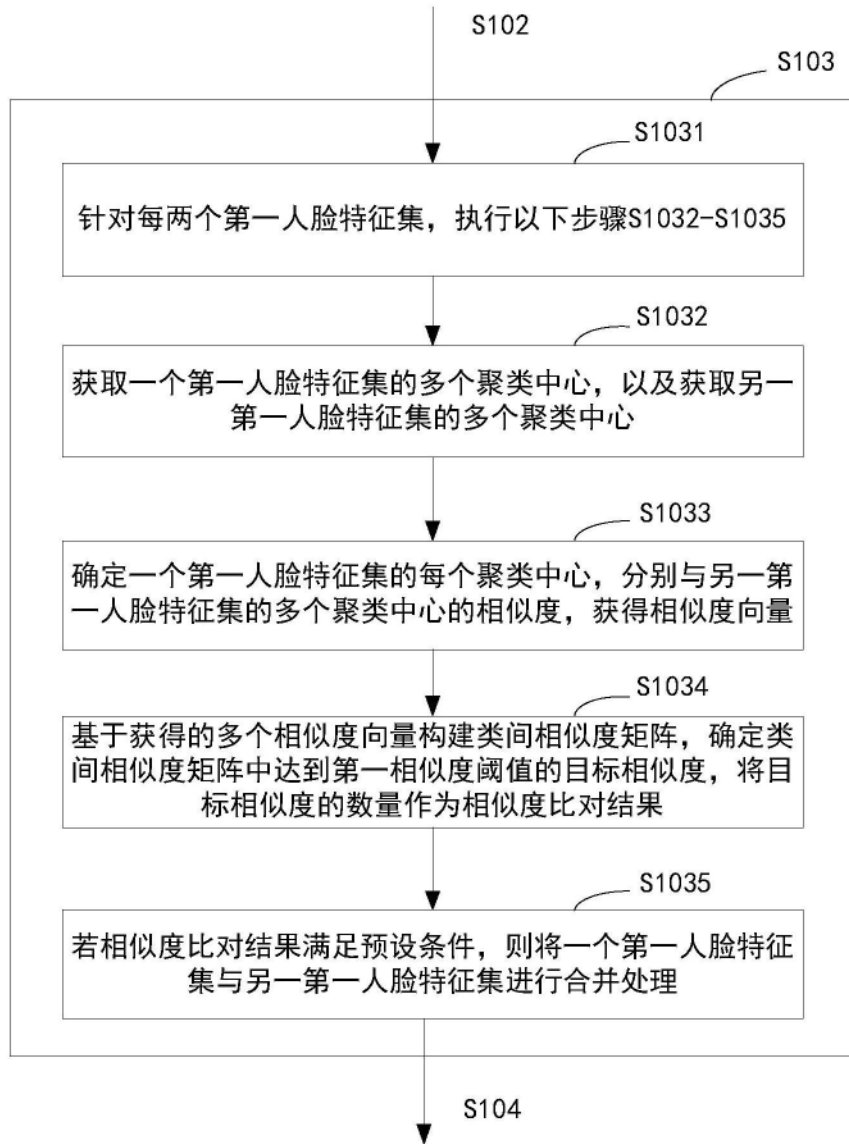


图3

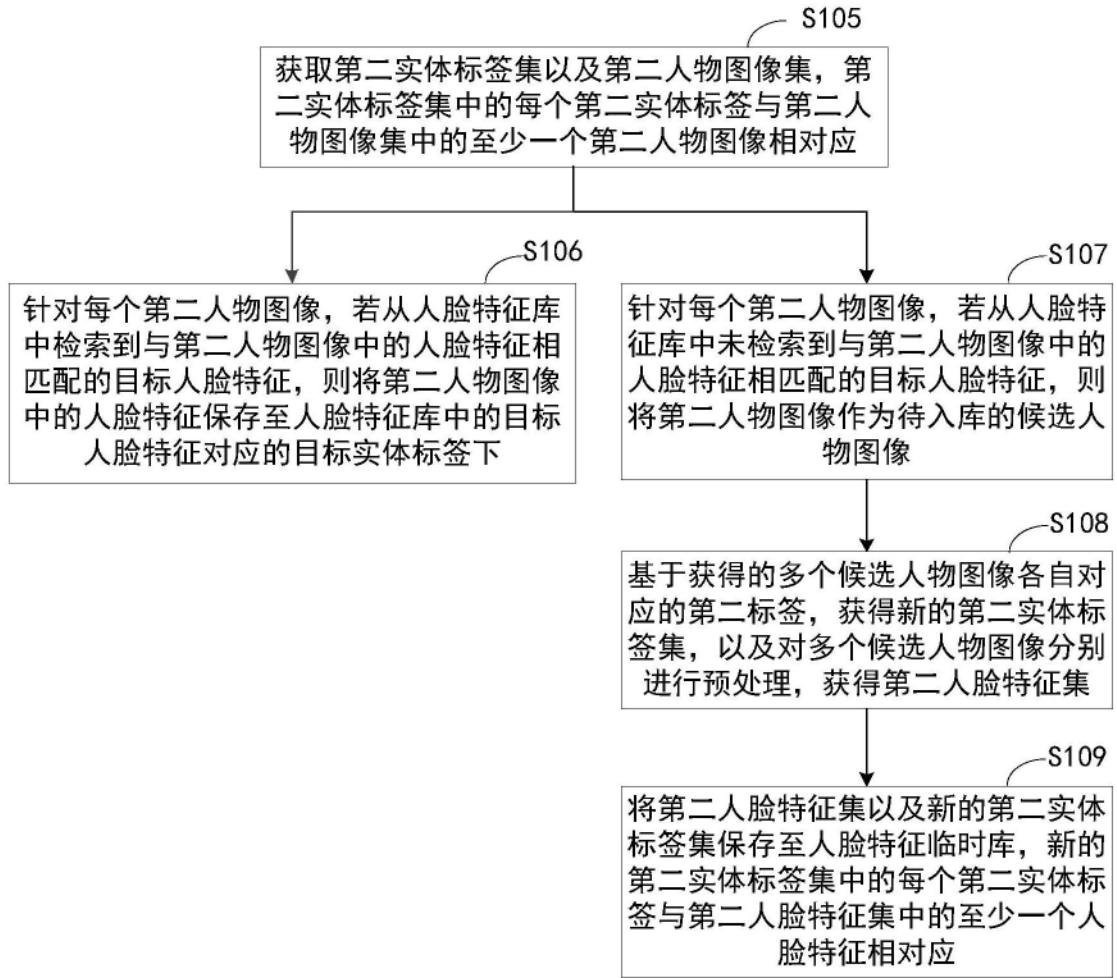


图4

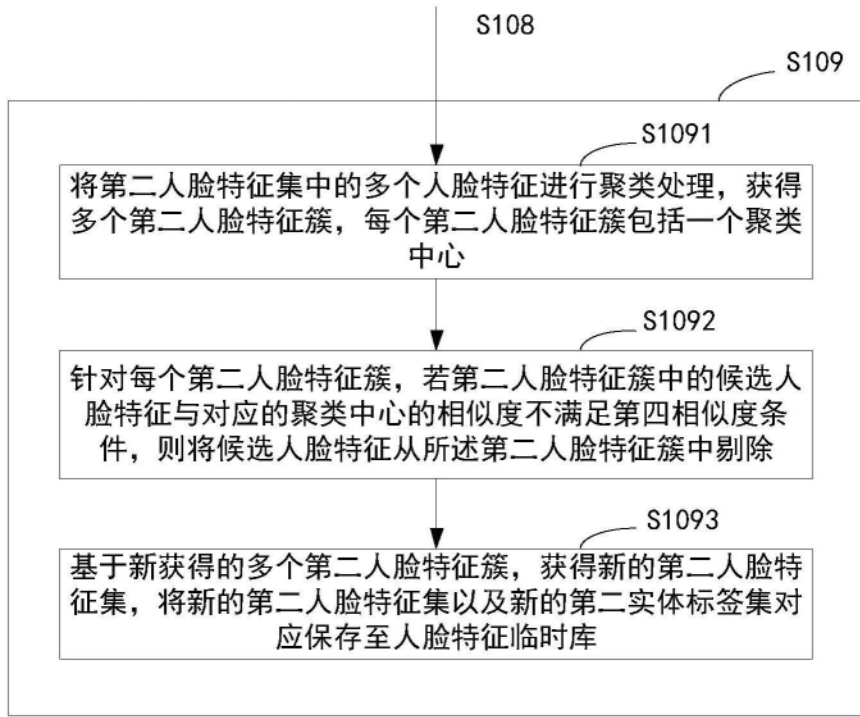


图5

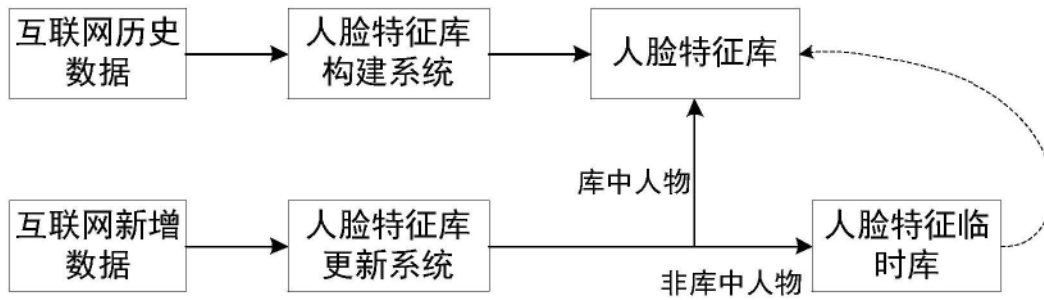


图6

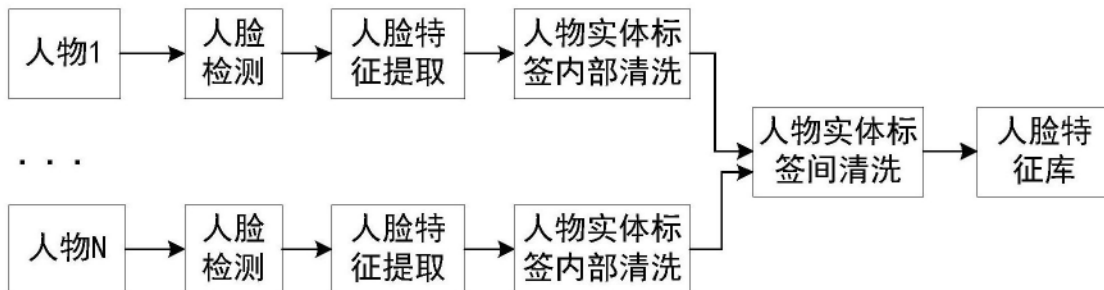


图7

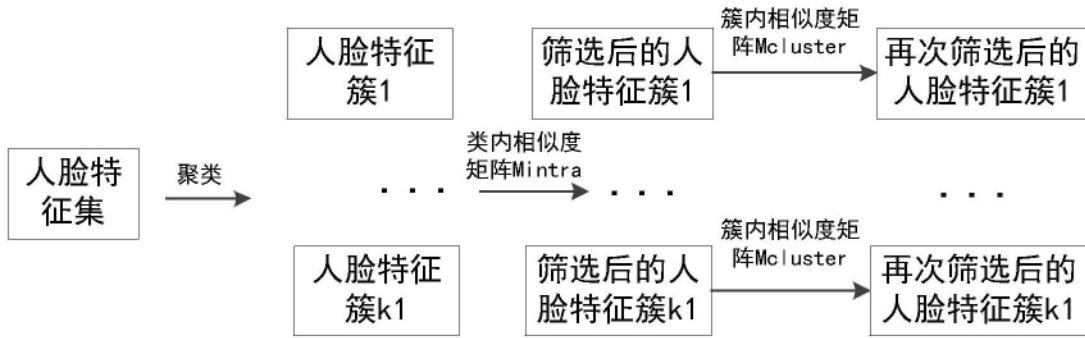


图8

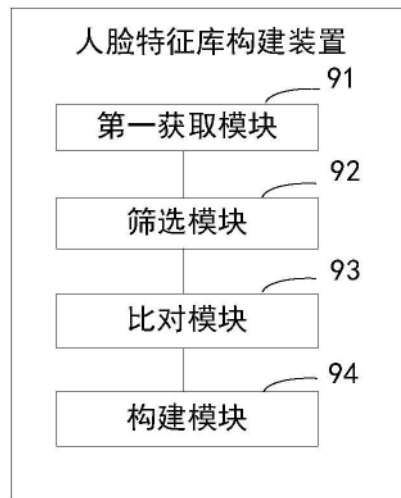


图9

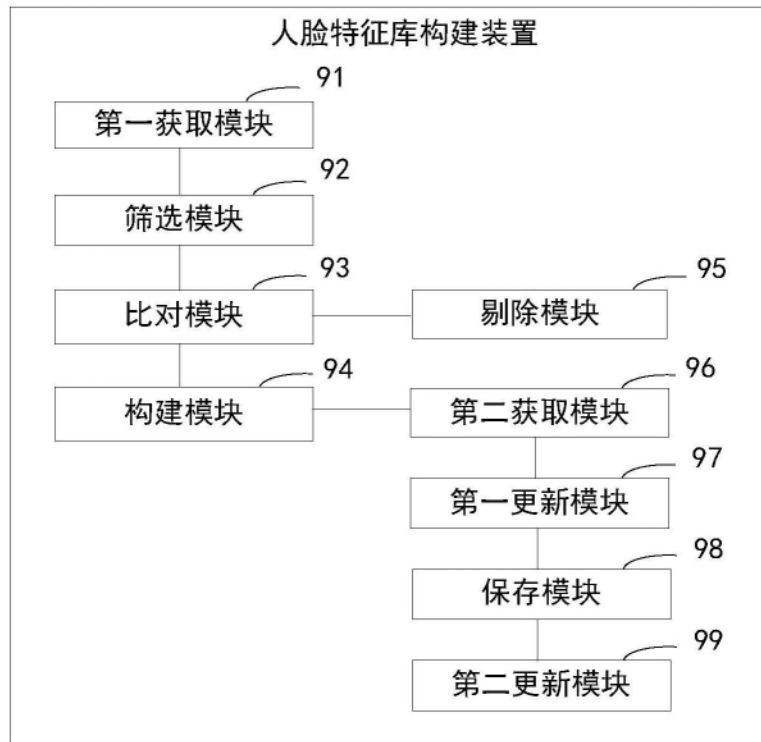


图10

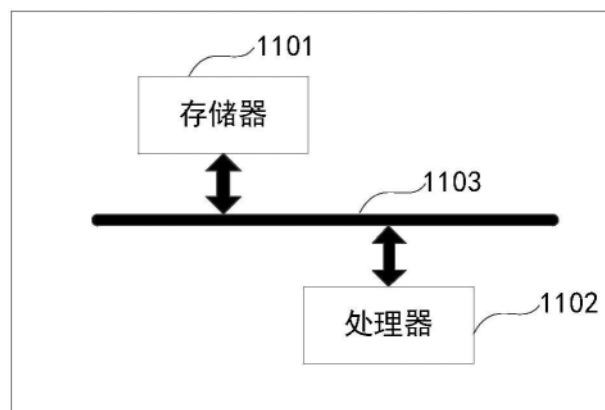


图11