

融合人脸跟踪和聚类的人脸图像去重方法

林增敏 洪朝群 庄蔚蔚

厦门理工学院计算机与信息工程学院 福建 厦门 361024

(954688123@qq.com)

摘要 人脸图像去重处理对智能监控系统的人脸识别有着重要意义。针对视频中人脸检测环节会产生大量重复的人脸图像的问题,提出了一种融合人脸跟踪和聚类的人脸图像去重方法。在视频中,利用 Multi-task Convolutional Neural Network 中的人脸检测算法提取人脸框及其对应的坐标。根据人脸跟踪算法构建人脸轨迹和约束矩阵,同时引入人脸质量评估算法,从人脸轨迹中挑选人脸姿态和图像清晰度较好的人脸图像作为人脸轨迹的代表图片;再由约束矩阵和无监督聚类算法对人脸代表图像进行聚类以获取同一个人的人脸图像;最后对每一个人的人脸图像再次进行评估,得到去重结果。实验结果表明,通过人脸跟踪与无监督聚类融合的方法获取同一个人的人脸图像,再结合人脸质量评估算法,能够快速有效地从一段视频中获取每个人不重复的高质量人脸图片。

关键词:人脸检测;人脸跟踪;人脸聚类

中图法分类号 TP391

Face Image Deduplication Based on Fusion of Face Tracking and Clustering

LIN Zeng-min, HONG Chao-qun and ZHUANG Wei-wei

College of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian 361024, China

Abstract Face image deduplication is of great significance to face recognition in intelligent surveillance systems, since face detection in videos will produce a large number of repeated face images. In this paper, a method of face image deduplication in videos by integration of face tracking and clustering is proposed. In a video, use the face detection algorithm in the Multi-task Convolutional Neural Network to extract the face frame and its corresponding coordinates. Face tracking is used to construct the face trajectory and the constraint matrix, and the face quality evaluation algorithm is introduced to select the face pose and image clarity from the face trajectory. An optimal face image is used as a representative of the face trajectory. Combined with the constraint matrix and unsupervised clustering algorithm, the representative images of the faces are clustered to obtain the face image of the same person. Finally, the face image of each person is evaluated again to obtain the deduplication. Experimental results show that, through face tracking and unsupervised clustering, the face image deduplication method in videos can quickly and efficiently obtain high-quality face images that are not repeated for each person from a video.

Keywords Face detection, Face track, Face clustering

1 引言

随着人工智能技术的不断发展,其中的人脸识别技术逐渐被应用到安防、金融、交通等监控领域中^[1]。人脸识别的流程包括人脸检测和人脸比对等一系列流程^[2],而人脸检测环节会产生大量重复的人脸图片,从而导致计算量过大和内存占用过高的问题^[3]。因此,在人脸检测过程中需要对检测到的人脸图像进行去重处理,挑选出同一个人的人脸图像中质量较好的一张作为代表人脸,这样不仅可以减少计算量,还可

以提高人脸识别的准确率^[4]。对于人脸图像去重算法的实现,目前主流的思路是人脸质量评估算法,选取人脸图像质量较高的图像作为去重结果。Chen 等^[5]在获取一个人的人脸图像的基础上,运用了人脸姿态、图像清晰度、对比度等指标的图像评估算法对单个人的一系列人脸图像进行去重处理,实验结果表明,基于人脸图像评估算法的人脸去重具有良好的性能,挑选出的人脸图像具有代表性和说服力。但是,在实际场景中获取到的人物个数通常是未知的,因此这种算法不适用于大规模人流场景。

基金项目:国家自然科学基金(61871464,61836002);福建省自然科学基金(2018J01573);福建省高校杰出青年科研人才培养计划;福建省高校新世纪优秀人才项目(2018J01573)

This work was supported by the National Natural Science Foundation of China (61871464,61836002), Fujian Provincial Natural Science Foundation of China (2018J01573), Distinguished Young Scientific Research Talents Plan in Universities of Fujian Province and Program for New Century Excellent Talents in University of Fujian Province(2018J01573).

通信作者:洪朝群(cqhong@xmut.edu.cn)

在视频中无监督获取同一个人的人脸图像的主要方法是人脸跟踪,而人脸跟踪方法在目标被遮挡以及存在大面积相似颜色背景的干扰下,会导致跟踪错误或者跟踪丢失,同时,算法还需要初始化跟踪窗口以及设定处理边界等预处理^[6]。此外,Wang等采用视频中人脸聚类的方法来代替人脸跟踪算法^[7],将检测到的所有人脸聚类以划分人脸归属,但是存在聚类时间长且人脸特征提取算法复杂度高的问题。因此,本文提出一种人脸跟踪与无监督聚类融合的方法来获取视频中同一个人的人脸图像。通过人脸跟踪算法构建人脸轨迹^[8-9],根据人脸轨迹内的人脸属于同一个人以及人脸质量评估算法,挑选出人脸轨迹内姿态、图像清晰度较好的人脸图像作为人脸轨迹的代表图像。再根据人脸轨迹信息构建约束矩阵,用无监督聚类方法对人脸轨迹代表图像进行聚类,聚类结果即为每一个人的脸图像。

2 人脸跟踪与聚类融合的人脸图像去重算法

2.1 算法流程

本文算法主要由3个模块组成。首先是人脸检测和人脸轨迹构建模块,通过MTCNN算法^[10]提取输入图片中的人脸框并回归检测出人的左眼、右眼、鼻子、左嘴角、右嘴角的坐标,再结合人脸位置信息和哈希感知算法对检测到的人脸进行人脸跟踪,对比相邻的两帧画面中的人脸图片,将存在较高重叠面积和哈希感知相似度的图片归于同一轨迹。其次是人脸图像去重模块和人脸轨迹聚类模块,人脸图像的去重是为了更好地进行人脸轨迹聚类,人脸轨迹聚类的结果为人脸去重提供同一个人的图像。通过人脸图片质量评估挑选出人脸轨迹中每个人较为清晰和姿态较正的人脸图片,根据人脸轨迹的先验信息构建约束矩阵,以无监督聚类方法得到聚类结果。最后再对聚类结果进行人脸质量评估,得到的结果即为去重结果。系统流程图如图1所示。

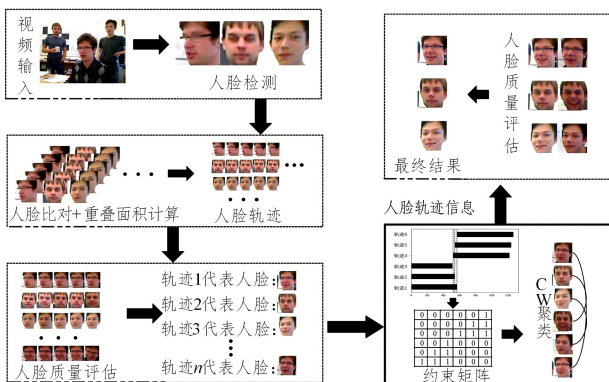


图1 系统流程图

Fig. 1 System flow chart

2.2 人脸轨迹与先验信息

本文参考了文献^[11]所采用的人脸跟踪方法创建人脸轨迹。其主要使用了两种度量方法判断相邻帧的人脸属于同一个人:1)图像相似性:使用人脸的灰度直方图判断人脸的相似性;2)重叠面积:使用人脸在相邻帧画面中的人脸重叠面积。

本文同样采用这两种度量方式,但是对于第一种方法,采用哈希感知方法代替直方图交叉,主要考虑到两点,一是哈希

感知算法的运算量比较小,二是监控视频中相邻两帧画面中人脸的变化并不大,使用哈希感知算法提取的特征向量足够区分不同人脸所属的轨迹。本文对传统的基于均值的哈希感知算法做了改进,将只计算一张图片的像素均值改为计算该帧画面中所有人脸图片的像素均值,由实验结果可知,全局均值更能体现不同人脸图片的差异性。对于第二种方法,本文采用相邻帧画面中人脸区域的交并比来计算图片的重叠程度,在传统的帧重叠算法上,将分母的值固定能够增加算法的鲁棒性。

公式如下:

$$\begin{cases} d_{spatial} = \frac{w * h}{w_1 * h_1 + w_2 * h_2} \\ d_{appearance} = \frac{(256 - f(a, b))}{256} \\ d_{detect} = d_{spatial} * 0.8 + d_{appearance} * 0.2 \end{cases} \quad (1)$$

其中, (x_1, y_1, w_1, h_1) 和 (x_2, y_2, w_2, h_2) 分别代表先前帧和当前帧的位置、宽度和高度。 $w = \max(x_1 + w_1, x_2 + w_2) - \min(x_1, x_2)$, $h = \max(y_1 + h_1, y_2 + h_2) - \min(y_1, y_2)$ 。 a 和 b 分别表示当前帧和先前帧的人脸哈希感知值, $f(a, b)$ 表示取 a , b 之间的汉明距离,汉明距离的最大值为 256。

将相似度阈值设置为 0.75,当不同的人脸之间的相似度超过阈值时,则表示属于同一人的人脸,将其划分到同一个人的人脸轨迹中,反之新建一个轨迹。

根据以上人脸跟踪方法构建监控视频片段中的人脸轨迹以及人脸图片中帧序号信息,如图2所示,横坐标表示按时间顺序的帧画面,纵坐标表示人脸轨迹。其中轨迹1,2,3和轨迹4,5,6在大量的帧画面中都有重叠,可以推断其人脸是同时被检测到的,即一定不属于同一个人。同理,轨迹1、轨迹4、轨迹5有部分重叠,可以断定它们不属于同一个人。可以获取人脸轨迹的先验信息:1)同一轨迹内的人脸图像一定属于同一个人;2)在同一帧上有重叠记录的人脸轨迹一定不属于同一人的人脸。

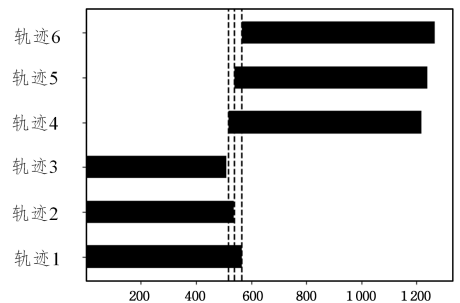


图2 人脸轨迹示例

Fig. 2 Face track example

2.3 人脸图像质量评估

在单一的人脸轨迹内,本文使用了两种度量方法来衡量人脸图片的质量:1)人脸模糊程度,采用拉普拉斯方差算法计算人脸图像的模糊度^[12];2)人脸特征点的偏移差,根据MTCNN算法回归得到的5个人脸特征点,分别计算关键点坐标值在 x 轴和 y 轴上的偏移差。

方法1)流程如算法1所示,最终输出结果为拉普拉斯滤

波计算后的人脸图像均值,该值越大,表示人脸图像越清晰,反之越模糊。

算法1 拉普拉斯方差算法

输入:人脸矩阵 Mat

输出:人脸均值 Mean

1. 高斯模糊去噪
2. 转换灰度图
3. 在此图像上利用拉普拉斯算子滤波

$$Offset = \sqrt{(y_1 - y_2)^2 + (y_4 - y_5)^2 + \left(x_3 - \frac{x_1 + x_2}{2}\right)^2 + \left(x_3 - \frac{x_4 + x_5}{2}\right)^2} \quad (2)$$

由于两种计算方法在数值上无关联且大小不一致,简单的以相加值作为比较结果并不准确,所以需要对其进行归一化处理。先求取所有人脸图像的人脸均值和人脸偏移差的最大值,其次求取人脸偏移差与其最大值的差值再分别将差值与人脸均值除以各自的最大值,最后的评估分数由二者加权相加得到,如式(3)所示。

$$\begin{cases} Q_{laplace} = Mean / max (Mean) \\ Q_{posture} = (\max(Offset) - Offset) / \max (Offset) \\ Q = Q_{laplace} * 0.5 + Q_{posture} * 0.5 \end{cases} \quad (3)$$

根据式(3)计算,得到该人脸轨迹内所有人脸图像的评估分数,选取分数最高的人脸图片作为单一人脸轨迹的代表图片。每条人脸轨迹都按此标准选取人脸图片,以达到人脸去重的目的。

2.4 人脸轨迹聚类

Chinese Whispers 动态聚类算法^[13]是一种无监督的聚类算法,可以自动生成聚类类别。本文采用的聚类方法是基于人脸轨迹先验信息的 Chinese Whispers 动态聚类。

根据先验信息(1),本文设定以人脸质量评估的方法挑选一张图片代表一条人脸轨迹,以减少人脸轨迹之间进行比较的计算量,同时完成一部分的人脸图像去重任务。

根据先验信息(2),将一定不属于同一个人的人脸轨迹之间的相似信号设置为 0,其余设置为 1,可以得到人脸轨迹的约束矩阵如图 3 所示。

Track	1	2	3	4	5	6
1	0	1	1	1	1	1
2	1	0	1	1	1	1
3	1	1	0	1	1	1
4	1	1	1	0	1	1
5	1	1	1	1	0	1
6	1	1	1	1	1	0

(a)相似矩阵

Track	1	2	3	4	5	6
1	0	0	0	0	0	1
2	0	0	0	0	1	1
3	0	0	0	1	1	1
4	0	0	1	0	0	0
5	0	1	1	0	0	0
6	1	1	1	0	0	0

(b)约束后的矩阵

图3 约束矩阵

Fig. 3 Constraint matrix

4. 直方图归一化映射到 0-255

5. 求均值 Mean

方法 2) 中人脸偏移差为特征点坐标值在 x 轴与 y 轴方向上差值的累加,其中 $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)$, 分别表示人脸的左眼、右眼、鼻子、左嘴角、右嘴角的坐标, $Offset$ 越小表示人脸姿态的偏移程度越小,越符合算法选择图片的标准。其计算公式如式(2)所示。

轨迹运算公式如式(4)所示。

$$S(x, y) = \begin{cases} 0, & \text{if } x \cap y \neq \emptyset \\ 1, & \text{else} \end{cases} \quad (4)$$

其中, x, y 表示任意两条人脸轨迹, $S(x, y)$ 表示轨迹之间的相似信号。

Chinese Whispers 动态聚类算法结合人脸轨迹约束矩阵,算法流程如算法 2 所示。

算法2 结合约束矩阵的动态聚类算法

输入:人脸轨迹代表图片 $V = \{v_1, v_2, \dots, v_n\}$ 和约束矩阵 M_c

输出:人脸轨迹代表图片所属类别

1. 设定相似度阈值 e , 迭代次数 n 和分块大小 P ;
2. 计算图片集 V 的相似度矩阵 M_s , 结合约束矩阵 M_c 生成最近邻矩阵 M_n ;
3. 将图片集 V 分割成大小为 P 的若干个块 $P_j \in V$;
4. 在每个块内初始化类别 $\text{class}(v_i) = i, v_i \in P_j$;
5. 随机选取一个未遍历过的节点 v_i , 从节点开始, 从最近邻矩阵 M_n 中选取权重最大的节点 v_j , 将节点 v_j 归于节点 v_i 所在的类;
6. 重复步骤 5, 遍历所有节点, 得到聚类结果 $\text{class}(v_i), v_i \in P_j$;
7. 将所有 P_j 块内产生的类中心视作 v_i , 再重复步骤 3-1 步骤 6;
8. 合并类中心被聚为一类的类, 输出聚类结果。

3 实验

3.1 实验环境及数据

本文算法以 Linux 系统和 OpenCV3.4 作为软件开发环境, CPU 主频为 2.2 GHz, 摄像头的分辨率为 640×480 , 开发语言为 C++。

本文采用的数据集有两个: 1) 人脸跟踪测试视频, 每 30 帧截取一幅画面, 该数据集包含 3 个人的 68 张人脸; 2) 从一段现场拍摄视频中获得的数据集, 每 30 帧截取一张人脸, 该数据集包含 13 个人的 100 张图片。

3.2 人脸检测与人脸跟踪

在实际场景中, 人脸检测和人脸跟踪是同步进行的。本文在视频中运用 MTCNN 人脸检测算法, 提取出人脸框后, 依据式(1), 如果相似度小于阈值, 则将人脸图像归入同一轨迹中, 大于阈值, 则新建一个人脸轨迹。部分结果如图 4 所示。

可以发现, 每一条人脸轨迹内的人脸图像属于同一个人的, 但是由于人在移动过程产生重叠和漏检, 导致新建了多条同一个人的人脸轨迹, 这是人脸跟踪中容易出现的问题。



图4 部分人脸轨迹结果

Fig. 4 Part of face track results

3.3 人脸图像质量评估

在构建完人脸轨迹之后,通过式(3)选取分数最高的图片作为每个人脸轨迹的代表。对单一人脸轨迹中的人脸图像,以人脸均值和人脸特征点的偏移差作为横轴和纵轴构建决策图,结果如图5所示,图中的每个圆圈代表一张人脸图像,其中的数字表示通过式(3)得到的分数。从图中可以发现,分数越高越靠近图的左上方,即证明通过式(3)得到的分数能够有效选取人脸清晰度较高以及人脸偏移程度较小的人脸图像。

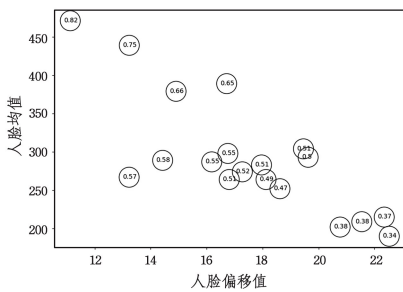


图5 人脸质量评估结果

Fig. 5 Face quality assessment

3.4 人脸聚类与结果分析

本文采用文献[14]中的聚类效果评价方法。该方法的评价标准主要有3个:均一性(记作 Homogeneity)、完整性(记作 Completeness)以及均一性和完整性的加权平均 F-measure。对于任意图片 i , $C(i)$ 表示图片 i 经过聚类后所在的类别, $K(i)$ 表示图片应该被分到的类别, n 表示图片的个数。均一性、完整性和 F-measure 的定义如式(5)所示:

$$\begin{cases} H = \frac{1}{n} \sum_{i=1}^n \frac{N(C_i = K_i)}{N(K_i)} \\ C = \frac{1}{n} \sum_{i=1}^n \frac{N(C_i = K_i)}{N(C_i)} \\ F = \frac{2 * H * C}{H + C} \end{cases} \quad (5)$$

经过人脸图像质量评估方法挑选之后的人脸图片可以代表一条人脸轨迹,所以对人脸轨迹的聚类也可以视作对所有人脸图像的聚类。本文采用 Dlib 库提供的 ResNet 模型提取

人脸图像的特征向量,用欧氏距离比较人脸特征向量之间的相似性,并以此构建人脸图像的相似矩阵,结合人脸轨迹中包含人脸图片的帧序号信息,根据式(4)可以构建相应的约束矩阵。将约束矩阵和人脸轨迹代表图片输入算法2中,得到聚类结果。

图6和图7采用参数的阈值作为坐标系的横轴,通过式(5)计算得到的 F-measure 值作为坐标系纵轴。从图6、图7中可以看出,本文算法相比 Chinese Whispers 动态聚类算法在聚类效果上有显著提高。同时,可以发现在样本相对简单的数据集1中,两种算法都能够达到比较好的效果, F 值接近1。在样本较为复杂的数据集2中,效果下降得比较明显,但是在特定阈值上,本文算法的 F-measure 值还是可以接近0.9。

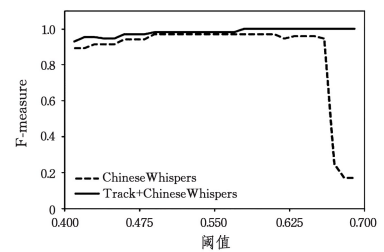


图6 聚类算法的 F 值(数据集 1)

Fig. 6 F-measure of clustering algorithm(dataset 1)

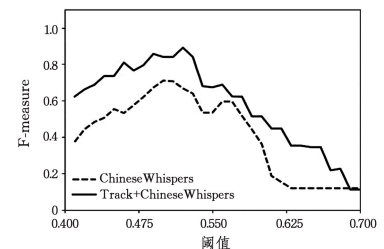


图7 聚类算法的 F 值(数据集 2)

Fig. 7 F-measure of clustering algorithm(dataset 2)

由于本文设计以人脸轨迹中人脸质量较好图像代表人脸轨迹,因此总体聚类的数目大为减少,在进行聚类比较时,可

以更快得到聚类结果。如表 1 所列,本文的算法在同一个数据集上完成聚类所耗费的时间上比传统的聚类算法大大减少。

表 1 算法时间
Table 1 Algorithm time

算法	数据集 1/s	数据集 2/s
Chinese Whispers	28.12	37.14
Trail + Chinese Whispers	7.23	12.88

结束语 人脸检测是人脸识别中必不可少的环节,在大规模人流场景中,重复检测人脸图像会增加整个系统的计算量。本文提出的结合人脸轨迹信息和聚类的人脸图像去重算法能够快速准确地获取同一个人的人脸图像,并且通过人脸质量评估得到清晰度较高和姿态较好的人脸图像,以此作为去重的结果。通过本算法可以减少整个人脸识别系统中的计算量 and 提高人脸检测的成功率。本文提出的算法仍然存在一些问题,如人脸轨迹构建过程中的人脸比对算法鲁棒性较差,一旦人脸图片变化较大,人脸比对的正确率就会比较低。后期会设计更好的人脸比对算法,提高系统的鲁棒性。

参考文献

- [1] LI X F. Development status and changes of the face recognition market [J]. *China Security*, 2020(21): 39-42.
- [2] ZHOU J, LU C Y, ZHANG C S, et al. A review of automatic face recognition methods [J]. *Acta Electronica Sinica*, 2000(4): 102-106.
- [3] LIANG L H, AI H Z, XU G Y, et al. Review of research on face detection [J]. *Journal of Computer Science*, 2002(5): 449-458.
- [4] FANG G N, HU Q H, FANG S Y, et al. Face image quality evaluation in video stream and its application in classroom attendance system [J]. *Computer Applications and Software*, 2018, 35(10): 140-146, 251.
- [5] CHEN Q. Face detection and face deduplication in video surveillance [D]. *China Jiliang University*, 2018.
- [6] CHEN Z H, WU Y D, CAI G R, et al. Face quality assessment algorithm based on texture feature fusion [J]. *Journal of Jimei University (Natural Science)*, 2018, 23(4): 312-320.
- [7] WANG C, CUI R Y, JIN J X, et al. Research on the optimization method of face images in videos based on face clustering [J].

- Progress in Laser and Optoelectronics*, 2019, 56(20): 243-251.
- [8] CAO X, ZHANG C, ZHOU C, et al. Constrained multi-view video face clustering [J]. *IEEE Transactions on Image Processing*, 2015, 24(11): 4381-4393.
- [9] WU B, ZHANG Y, HU B, et al. Constrained clustering and its application to face clustering in videos [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland Oregon: IEEE, 2013: 3507-3514.
- [10] ZHANG K, ZHANG Z, LI Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks [J]. *IEEE Signal Processing Letters*, 2016, 23(10): 1499-1503.
- [11] ORTIZ E G, WRIGHT A, SHAH M. Facerecognition in movie trailers via mean sequence sparse representation-based classification [J]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 3531-3538.
- [12] PECH-PACHECO J L, CRISTOBAL G, CHAMORRO-MARTINEZ J, et al. Diatom autofocusing in brightfield microscopy: a comparative study [C]// *15th International Conference on Pattern Recognition*. Barcelona, Spain. New York: IEEE, 2002: 6894564.
- [13] ZHANG C H, SHI Y Z. Dynamic face clustering based on Chinese Whispers [J]. *Computer Technology and Development*, 2019, 29(11): 92-96.
- [14] AMIG, ENRIQUE, GONZALO, et al. A comparison of extrinsic clustering evaluation metrics based on formal constraints [J]. *Information Retrieval*, 2009, 12(5): 461-486.



LIN Zeng-min, born in 1996, postgraduate. His main research interests include computer vision and image processing.



HONG Chao-qun, born in 1984, Ph.D. professor, is a member of China Computer Federation. His main research interests include computer vision, high-performance computing and Internet of Things.

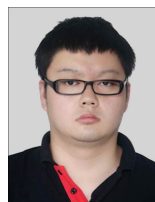
(上接第 592 页)

- [4] DENG Y, SHI W K, ZHU Z F. Efficient combination approach of conflict evidence [J]. *Journal of Infrared Millimeter Wave*, 2004, 23(1): 27-32.
- [5] JIANG L P, ZHOU Y. An evidence synthesis method based on importance weight and its application [J]. *Journal of Naval Engineering University*, 2012, 24(4): 24-28.
- [6] WEI Y C. Improved D-S evidence synthesis method based on KL distance* [J]. *Telecommunication Technology*, 2011, 51(1): 27-30.
- [7] WANG X X, YANG F F. A synthetic method for dealing with conflict evidence [J]. *Journal of Missile and Guidance*, 2007, 27(5): 255-257.
- [8] SUN Q, YE X Q, GU W K. A new synthesis formula based on

evidence theory [J]. *Acta Electronica Sinica*, 2000, 28(8): 117.



XU Jiang-jun, born in 1978, master, associate professor. His main research interests include fault diagnosis and information fusion.



PENG Xu, born in 1987, Ph.D. lecturer. His main research interests include multi-level converters and converter topologies.