

文章编号:1004-9037(2013)03-0307-06

# 一种视频流特定人物检测方法

高毫林<sup>1</sup> 陈 刚<sup>1</sup> 罗 宁<sup>2</sup> 李弼程<sup>1</sup>

(1. 解放军信息工程大学信息工程学院, 郑州, 450002; 2. 南海舰队指挥所, 湛江, 524000)

**摘要:**为实现视频流特定人物检测,提出了一种两阶段检测方法,即基于稳健哈希签名的流级检测和基于子图局部特征的文件级检测。第一阶段利用签名快速发现与查询图像相似度较高的关键帧,第二阶段采用人脸子图的局部特征搜索与查询图像相似度次高的关键帧,它主要包括人脸检测、提取人脸子图的局部特征和相似计算三部分。实验结果证明该方法既可以快速检测到特定人物,也可以提升人物检测性能。

**关键词:**视频流;特定人物检测;稳健哈希签名;人脸子图

**中图分类号:** TN391.4

**文献标志码:** A

## Specific Person Detection Method in Video Stream

Gao Haolin<sup>1</sup>, Chen Gang<sup>1</sup>, Luo Ning<sup>2</sup>, Li Bicheng<sup>1</sup>

(1. Institute of Information System Engineering, PLA Information Engineering University, Zhengzhou, 450002, China;

2. Command Center of the South Sea Fleet, Zhanjiang, 524000, China)

**Abstract:** To detect a specific person in video stream, a two-step detection method is proposed, ie stream level detection based on robust Hash signature and file level detection based on local features of sub-image. The first step aims to find quite similar key frames to query image with image signature, and the second step is to find less similar key frame matching local features of face sub-image, including face detection, local feature extraction of face sub-image and similarity measure. The experimental results show that the method can detect specific person and improve the performance of human detection.

**Key words:** video stream; specific person detection; robust Hash signature; face sub-image

## 引 言

视频流人物检测主要是对视频流进行不间断的检测,实时发现与查询人物相似的关键帧,并定位关键帧所在的镜头。人物检测在安检、视频监控、视频检索等方面都有着重要的作用。人物检测算法主要可以分为基于子窗口的方法和基于部分人体的方法<sup>[1]</sup>。基于子窗口的方法使用的特征有方向梯度直方图<sup>[2]</sup>(Histogram of oriented gradients, HOG),协方差矩阵<sup>[3]</sup>等。基于部分人体的方法把人体分为几部分分别进行检测,最后将结果进行整合,Mikolajczyk等<sup>[4]</sup>把人体分为7个部分,对每部分分别进行检测,Shet等<sup>[5]</sup>把逻辑推理应

用于研究低级检测器输出信息的内容扩充。Chen等<sup>[6]</sup>为了在新闻视频中找到特定人,综合利用文本搜索、节目主持人信息和人脸识别等方法。Schwartz等<sup>[1]</sup>将人脸检测和个人检测器综合起来用于人物检测。

人物检测中的重要组成部分是对人脸的检测。人脸检测、跟踪与识别是3个相互独立而又统一的问题<sup>[7]</sup>。人脸检测方法主要有基于特征的方法、模板匹配法、基于表象的方法。人脸跟踪等价于在连续帧间建立基于位置、颜色、形状、纹理、色彩等特征的对应匹配问题<sup>[8]</sup>。人脸识别的方法包括特征脸识别方法、弹性图匹配法、小波分析法等。

人脸特征提取所使用的经典方法之一是PCA,但它是基于一维的,在处理图像时,需要将二

维图像矩阵转换成一维向量,当图像较多时运算量大。Yang 等人<sup>[9]</sup>提出了基于图像矩阵的二维 PCA 方法。文献[10]将 L1 范式结合 2DPCA 用于人脸识别。非负矩阵分解<sup>[11]</sup>(Non-Negative matrix factorization, NMF)在矩阵中所有元素非负的条件下对矩阵进行分解,由于它具有可解释性和明确的物理意义在人脸识别上有较多应用。文献[12]利用 2DPCA 和 NMF 的基本思想,提出非负二维主成分分析的方法进行人脸识别。在变换域方法中,Hafed 等<sup>[13]</sup>首先提出基于离散余弦变换器(Discrete cosine transform, DCT)的人脸识别方法,Chen 等<sup>[14]</sup>在 DCT 域和空域上利用主成分分析和线性判别分析进行人脸识别,文献[15]将 DCT 与线性判别分析结合的方法进行改进用于人脸识别。

为了从视频流检测特定人物,提出了两阶段的检测方法,即基于稳健哈希签名的视频流上的检测和基于人脸子图局部特征的文件级的检测。该方法首先在视频流上进行检测,主要是采用稳健哈希签名的方法检索与查询人物所在图像相似度较高的关键帧,然后在图像数据库上进行基于子图的检测,也就是提取查询图像包含人脸的子图,并与图像集中各图像经人脸检测提取出的子图进行局部特征匹配,检测与查询图像相似度次高的关键帧,从而实现视频流特定人物检测。其中基于稳健哈希签名的视频流特定人物实时检测方法。该方法先计算目标图像的签名特征,主要包括 DCT 变换和哈希运算两部分,然后,在实时接收的视频流上计算镜头分割所得关键帧图像的签名特征;最后,计算这两个签名特征之间的汉明距离,若距离小于给定的阈值,则认为检测到目标人物出现,并提取关键帧图像所在的镜头。实验结果表明,该方法用于实时视频流人物检测可以达到较高的查全率。基于人脸子图局部特征的文件级的检测方法,首先提取查询图像人脸子图并计算其局部特征,然后对各关键帧图像进行人脸检测,得出包含人脸的子图。再计算这些子图的局部特征,最后计算它们与查询子图的相似性得出检测结果。

## 1 基于稳健哈希签名的流级检测

图像二维 DCT 变换已经将图像的主要信息集中于矩阵左上角,只需要取出左上角固定大小的子阵就可以代表图像主要信息<sup>[16]</sup>。利用合适的哈希函数对图像 DCT 结果进行哈希就可以生成签名。

将一幅图像的二维 DCT 变换矩阵记为  $C(u, v)$ ,由于 DCT 已经把图像的能量集中于矩阵的左上角,因此,可以提取  $C(u, v)$  矩阵的子矩阵  $S(m, n)$  用来生成图像签名。

$$S(m, n) = \text{Truc}(C(u, v)) = C(m, n) \quad m, n = 0, 1, \dots, M \quad (1)$$

式中:  $M$  为子矩阵  $S$  的行数。

### 1.1 稳健哈希

稳健哈希将一个信号映射到一个较短的二进制字符串,它的稳健性是指即使对信号进行小的处理,该信号的哈希输出保持不变,只有当信号内容发生变化时输出才有变化。

设输入图像用  $I$  代表,哈希函数用  $H(*)$  表示,输出向量为  $h$ ,那么  $h = H(I)$ ,  $h$  来自于集合  $\{0, 1\}^N$ ,该集合的势为  $2^N$ 。 $H(*)$  可以用来描述感知相等的图像,感知相等的两幅图像输出应该相同,感知不等的图像应产生不相关的值。而感知相等的图像不一定有同样的数字特征,所以  $H(*)$  是一个多对一映射<sup>[17]</sup>。它有随机化、两两独立和稳健性三个性质。

在图像处理中,稳健图像哈希将一幅图像用一个短二进制向量表示,对于内容相同的图像哈希结果保持不变,对于内容明显不同的图像哈希结果变化较大。它可以用于图像鉴定、图像检索和模式识别等。稳健哈希多被用来检测对图像的篡改和非法操作。在人物检测中,稳健哈希可以用来在不同的位置、条件和背景等拍摄情况下检测到同一个人。

### 1.2 图像签名的生成

哈希运算对 DCT 变换后的子矩阵  $S(m, n)$  进行。 $H(*)$  表示哈希函数,  $\text{Sig}(I)$  表示生成的签名。采用的稳健哈希函数定义如下

$$\text{Sig}(I) = H(I) = h(0)h(1)\dots h(M^2) \quad (2)$$

$$h(i) = \begin{cases} 1 & \mathbf{S}\left(\lfloor \frac{1}{M} \rfloor, i - \lfloor \frac{i}{M} \rfloor\right) \geq m_e \\ 0 & \text{其他} \end{cases} \quad i = 0, 1, \dots, M^2 \quad (3)$$

式中:  $m_e = \frac{1}{M^2} \sum_{m, n < M} S(m, n)$  表示子矩阵的均值。

该哈希函数先将子矩阵  $S(m, n)$  进行量化,然后将量化结果串成二进制向量,最后得到图像签名。该哈希实际上就是取子矩阵量化后的值,这样每个哈希值  $h(i)$  的产生是随机的,而且两幅图像的哈希结果相互独立,同时也满足稳健性的要求,

因为 DCT 变换本身的量化过程就对干扰有一定的容忍能力。

图像签名生成的过程如下:

(1)对图像进行 2 维 DCT 变换得矩阵  $C(u, v)$ ;

(2)取  $C(u, v)$  子矩阵得矩阵  $S(m, n)$ ;

(3)对子矩阵  $S(m, n)$  进行哈希,得到图像全局签名  $\text{Sig}(I)$ 。

这样,每幅图像就可以得到一串 64 Byte 的二进制序列。图 1 显示了一幅图像及其对应的全局签名。由于基于块匹配的方法常用于图像匹配,也计算该图像的分块签名。分块签名由各图像块



```
10110100110000110000110000001100
100000001000000000000000000000
```

图 1 一幅图像及其对应的全局签名

### 1.3 图像相似度度量

得出签名后,各图像之间的相似度计算就成了它们对应的签名之间的相似度计算,这里采用汉明距离来表示签名的相似度。这两个等长字符串之间的汉明距离是他们对应位置的不同字符的个数。也就是将一个字符串变换成另外一个字符串所需要替换的字符个数。即

$$d(\text{Sig}(I_1), \text{Sig}(I_2)) = \text{Sig}(I_1) \oplus \text{Sig}(I_2) \quad (4)$$

它表示两个签名的汉明距离。为验证签名特征进行相似计算的效果,以图 1 作为基准图像,并选取了与它相似的 8 幅图像作为测试图像,如图 2 所示。它们以基准图像为标准像按相似度依次排列。这些测试图像与基准图像全局签名和分块签名的汉明距离在图 3 中显示。可见分块签名距离区分性更强,全局签名距离相同的图像对应的分块签名距离就可能不相同。



图 2 测试图像

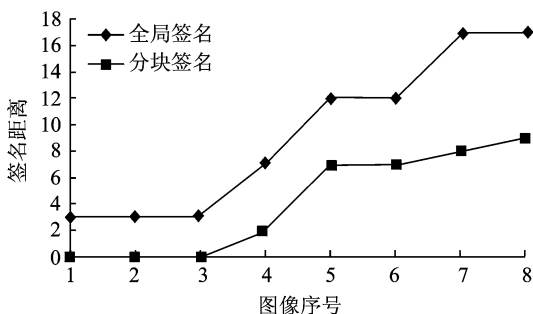


图 3 不同签名的图像距离

## 2 基于人脸子图局部特征的文件级检测

在近似重复检测中,使用子图的方法<sup>[18]</sup>可以充分利用查询目标主要信息,减小背景噪声的影响。人脸是人物最有区别性的一部分,人脸检测的技术也已经得到了充分的发展。所以,完全可以通过图像的人脸子图进行特定人物检测。另外,从语义概念的角度讲,人物是图像中的语义概念之一,因此常用的用于语义搜索的局部特征 PCA-SIFT<sup>[19]</sup>,尺度不变特征转换(Scale-invariant feature transform, SIFT)<sup>[20]</sup>等对于特定人物检测也适用。采用这种方法可以初步解决出现背景、姿势等不同时人物检测准确率较低的问题。近年来,已有学者将 SIFT 特征应用于人脸识别<sup>[21]</sup>和人物检测<sup>[22]</sup>。

### 2.1 人脸子图提取

人脸是人体最明显的生物特征,利用人脸的识别进行特定人物检测是有效的方法之一。而如果不

不对图像区域加以限制,局部特征在提取时针对整幅图像进行,这样不仅特征点多,增加了图像相似度衡量的计算代价,同时由于背景的影响也会降低检测的准确度。显然,只对查询图像和待检测图像的人脸部分进行特征提取和相似度衡量,可以同时提高检索的效率和准确率。

人脸检测方法主要可分为两大类,基于统计的方法和基于知识的方法。目前比较流行的是基于统计的方法,它将人脸检测问题转换为模式识别问题,将人脸区域看作一类模式,对人脸和非人脸进行训练,构造分类器。主要包括基于神经网络的方法、基于支持向量机的方法、基于 Haar 特征 Adaboost 的方法等。基于知识的方法利用人的先验知识建立规则进行人脸检测,如人脸的形状、五官分布和人眼特征,这类方法检测效果依赖于特征

提取和预先定义的规则。当图像质量较差特征不容易检测或者采用的规则不全面不准确时都会影响检测效果。

基于 Haar 特征 Adaboost 的方法可以实时、准确地检测到人脸,它采用的 Adaboost 算法从多个弱分类器构造强分类器。Haar 可以通过积分图像快速计算得出,从而保证了检测的实时性。图 4 是该方法对 Caltech256 部分图像检测结果。



图 4 人物图像及其对应的人脸子图

## 2.2 SIFT 局部特征

SIFT 特征对旋转、平移、亮度变化、尺度缩放和噪声具有较好的不变性,对视角变化、仿射变换保持一定程度的稳定性。它很好地符合评价局部特征性能的两个指标:稳定性和独特性,其匹配性能优于同类型其他局部特征,在目标识别、图像检索、图像拼接和场景分类等领域已经得到了成功的应用。人脸虽然表情多样,但不同的人脸具有较强的独特性,同一个人的人脸也保持了一定的稳定性。所以,可以用 SIFT 特征描述人脸。图 5 给出了人物图像及其对用的人脸图像的 SIFT 特征检测结果。



图 5 人物图像及其对应的人脸子图的 SIFT 特征点

## 2.3 检测流程

基于子图局部特征的文件级检测主要流程如下:

(1)用基于 Haar 特征的 Adaboost 方法对查询图像和待检测图像进行人脸区域检测,得出人脸子图;

(2)对查询图像人脸子图进行 SIFT 特征提

取;

(3)对待检测图像提取出的各人脸子图进行 SIFT 特征提取;

(4)计算查询子图与待检测子图的相似程度,如果有某一子图相似度大于给定阈值,就返回该子图所在的图像,并认为检测到了查询人物出现。

(5)重复(3),(4),直到完成所有图像的检测。

图 6 给出了对查询子图进行检测的结果。可见,当背景变化时该方法仍能检测到包含查询人物的图像。



图 6 人物图像及检测结果

## 3 视频流特定人物检测方法实验结果

### 3.1 视频流特定人物检测方法

本文将视频流特定人物检测方法分为流级粗过滤和文件级细筛选两阶段。第 1 阶段采用基于稳健哈希签名的方法对实时视频流进行检测,特征提取和相似度计算的速度较快,适合在线检测,主要发现与查询图像相似度较高的关键帧。第 2 阶段采用基于人脸子图局部特征的方法对图像库进行检测,先提取人脸子图,再提取 SIFT 特征进行相似计算,主要发现与查询图像不同场景的关键帧。

### 3.2 流级检测实验结果

使用两种签名在不同相似门限下进行检测的准确率是不同的,图 7 给出了检测结果。分块签名的准确率要高于全局签名,而且在选取合适的门限时,准确率可以达到 80% 以上。当然,这里用来计算准确率的图像需要进行人工标定。

在查询人物设定时,选取视频节目中经常出现的国家领导人或重要人物进行检测。该实验在视频流数据采集的平台上进行,该平台对常用卫星频段进行扫描,发现并获取视频数据,进一步对视频数据进行分析,得出视频关键帧图像。本实验选取已获取人物图像,对该图像进行在线检测,实验中

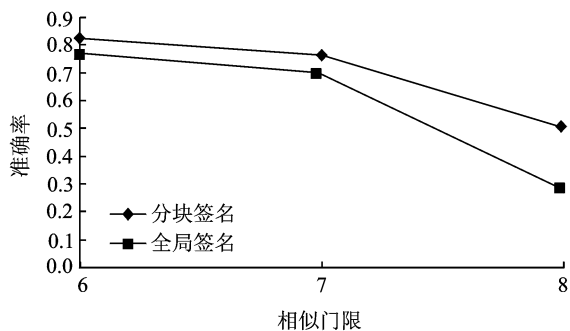


图 7 3 种签名在不同门限下的查准率

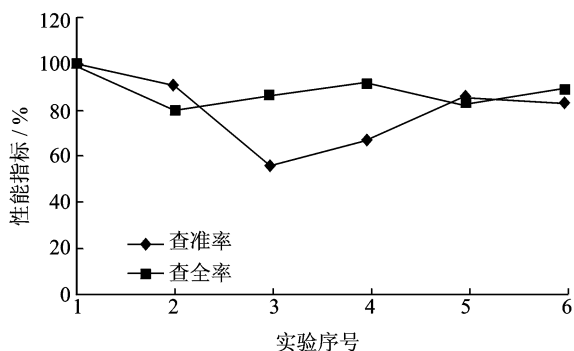


图 8 视频流人物检测性能指标

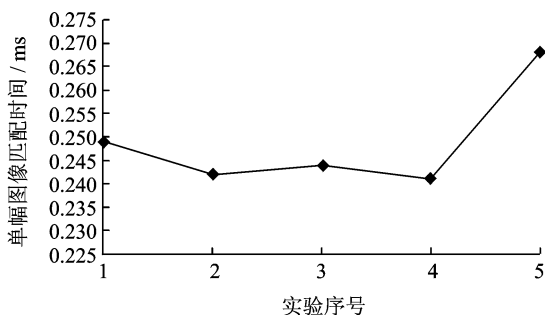


图 9 单幅图像匹配时间

选取 5 副特定人物图像,对 PressTV 进行了 12 h 的在线检测,检测结果如图 8 所示。检测中,单幅图像匹配时间如图 9 所示,匹配时间包括特征提取和相似计算两部分,大约在 0.24~0.27 ms 之间,而这样的关键帧所在的镜头长度远远大于这个范围,所以完全可以达到实时的效果。

### 3.3 文件级检测实验结果

文件级检测采用 SIFT 和人脸子图相结合的方法,检测数据采用的是 Caltech256 库中的人脸图像,结果如图 10 所示。该方法要高于直接采用 SIFT 特征对全图进行检测的方法。对不同人物进行检测期检出率是不同的,这与图像数据本身

有关。如果人脸表情变化较大,检出率就会降低。

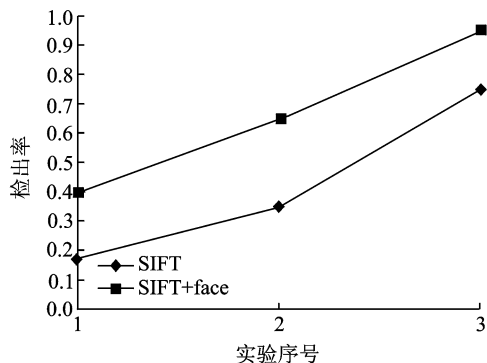


图 10 文件级检测结果

## 4 结束语

人脸识别和人物检测近年来已经有很多研究,但在视频流特定人物检测方面仍有待提高。本文将人物检测方法分为基于稳健哈希签名的实时的流级粗过滤和基于人脸子图局部特征的非实时的文件级细筛选两阶段。流级检测采用快速特征提取和快速相似计算的方法,适用于在线检测,主要发现与查询图像相似度较高的人物图像。文件级检测先提取人脸子图再计算子图 SIFT 特征,可以达到较高的准确率,主要发现相似度次高的不同场景人物图像。

实际上,即使采用了人脸提取和性能较好的局部特征 SIFT 进行人物检测,仍然不能做到对所有的人保持较高的检出率。这是因为目前使用 SIFT 特征并没有利用各特征点之间的关系,这样,人脸各器官的相对位置等信息被抛弃了,而这对人脸识别是很重要的。其实,SIFT 特征更适合于刚性物体的检索,因为刚性物体在不同图像中虽然有角度、背景、旋转等变化,但其特征点相对位置的变化并不明显。而人脸等人体组成部分或动物组成部分等非刚性物体的局部特征点的相对位置变化较大,所以仅仅用 SIFT 特征是难以达到令人满意的效果的。可行的改进方法之一是将 SIFT 特征与人脸器官相对位置、人脸轮廓等信息结合进行检测,另外一个可行的方法是采用非刚性物体匹配的方法<sup>[23]</sup>进行检索。

### 参考文献:

[1] Schwartz W R, Gopalan R. Robust human detection under occlusion by integrating face and person detectors [C]//International Conference on Biometrics. Berlin, Germany: Springer, 2009:970-979.

- [2] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2005:886-893.
- [3] Tuzel O, Porikli F, Meer P. Human detection via classification on riemannian manifolds [C]//IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis: IEEE, 2007:1-8.
- [4] Mikolajczyk K, Schmid C, Zisserma A. Human detection based on a probabilistic assembly of robust part detectors [C]//The European Conference on Computer Vision. Czech: Springer, 2004:69-82.
- [5] Shet V, Neumann J, Ramesh V, et al. Bilattice-based logical reasoning for human detection [C]//IEEE Conference on Computer Vision and Pattern Recognition. Brazil: IEEE, 2007:243-250.
- [6] Chen Mingyu, Hauptmann A. Searching for a specific person in broadcast news video[C]//The IEEE International Conference on Acoustics, Speech and Signal Processing. Montreal, Canada: IEEE, 2004: 1036-1039.
- [7] 夏思宇. 彩色图像序列的人脸检测、跟踪与识别研究 [D]. 南京:东南大学自动化学院, 2006.  
Xia Siyu. Study on face detection, tracking and recognition in color image sequence [D]. Nanjing: School of Automation, Southeast University, 2006.
- [8] 夏思宇, 潘泓, 金立左, 等. 基于特征组合的人脸跟踪方法[J]. 数据采集与处理, 2011, 26(1): 15-19.  
Xia Siyu, Pan Hong, Jin Lizuo, et al. Face tracking based on feature combination[J]. Journal of Data Acquisition and Processing, 2011, 26(1): 15-19.
- [9] Yang Jian, Zhang D, Frangi A F. Two dimensional PCA: a new approach to appearance based representation and recognition[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2004, 26(1): 131-137.
- [10] 郑豪. 基于 L1 范式的分块 2DPCA 人脸识别方法 [J]. 数据采集与处理, 2011, 26(6): 666-670.  
Zheng Hao. Method of face recognition based on L1-norm and block two-dimensional principal component analysis[J]. Data Acquisition and Processing, 2011, 26(6): 666-670.
- [11] Lee D D, Seung H S. Learning the parts of objects with non-negative matrix factorization [J]. Nature, 1999, 401: 788-791.
- [12] 严慧, 金忠, 杨静宇. 非负二维主成分分析及在人脸识别中的应用[J]. 模式识别与人工智能, 2009, 22(6): 809-814.  
Yan Hui, Jin Zhong, Yang Jingyu. Non-negative two-dimensional principal component analysis and its application to face recognition[J]. Pattern Recognition and Artificial Intelligence, 2009, 22(6): 809-814.
- [13] Hafed Z M, Levine M D. Face recognition using the discrete cosine transform[J]. International Journal of Computer Vision, 2001, 43(3): 167-188.
- [14] Chen W, Meng J E, Wu S. PCA and LDA in DCT domain[J]. Pattern Recognition Letters, 2005, 26(15): 2474-2482.
- [15] 伊洪涛, 付平, 沙学军. 基于 DCT 和线性判别分析的人脸识别[J]. 电子学报, 2009, 37(10): 2211-2214.  
Yin Hongtao, Fu Ping, Sha Xuejun. Face recognition based on DCT and PCA[J]. Acta Electronica Sinica, 2009, 37(10): 2211-2214.
- [16] Xavier N, Patrick G. A fast shot matching strategy for detecting duplicate sequences in a television stream [C]//International Workshop on Computer Vision Meets Databases. Baltimore, USA: ACM, 2005: 121-128.
- [17] Kamil Senel. A learning framework for robust hashing of face images[D]. Turkey: Electrical and Electronics Engineering Department, Bogazici University, 2010.
- [18] Ke Y, Sukthankar R, Huston L. An efficient parts-based near-duplicate and sub-image retrieval system [C]//ACM Multimedia. New York, USA: ACM, 2004: 869-876.
- [19] Ke Y, Sukthankar R. PCA-sift: a more distinctive representation for local image descriptors[C]//IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE, 2004: 506-513.
- [20] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal on Computer Vision, 2004, 60(2): 91-110.
- [21] Bicego M, Lagorio A, Grico E, et al. On the use of SIFT features for face authentication[C]//The Conference on Computer Vision and Pattern Recognition Workshop. New York, USA: IEEE, 2006: 35-39.
- [22] Luo Jun, Ma Yong, Takikawa E, et al. Person-specific features for face recognition [C]//The International Conference on Acoustics, Speech, and Signal Processing. New York, USA: IEEE, 2007: 593-596.
- [23] Shekhovtsov A, Kovtun I, Hlavac V. Efficient MRF deformation model for non-rigid image matching[J]. Computer Vision and Image Understanding, 2008, 112(1): 91-99.

作者简介: 高毫林(1979-), 男, 博士生, 研究方向: 图像检索, E-mail: holygao@126.com; 陈刚(1979-), 男, 讲师, 研究方向: 模式分析与网络信息处理; 罗宁(1975-), 女, 高级工程师, 研究方向: 图像处理; 李弼程(1970-), 男, 教授, 研究方向: 智能信息处理。

