可见-近红外多光谱和多种算法模型融合的血迹年龄预测

戎念慈，黄梅珍*

上海交通大学电子信息与电气工程学院仪器科学与工程系，上海 200240

摘 要 精确的血迹的年龄估计在刑侦法医鉴定中有着重要的意义。构建了以8个LED为照明光源、以黑白CCD相机为成像单元的可见-近红外多光谱成像系统，利用以k最近邻算法、支持向量机算法和随机森林算法为基模型的融合模型分析预测血迹年龄，研究了利用可见-近红外反射多光谱精确估计人体血液年龄的可行性，并与前人利用高光谱进行血迹年龄预测的研究结果进行了对照，还检验了血液特异性对模型的影响。实验记录了11个人体血液样本在1～20 d的在400～940 nm之间的8个波长通道的图像，使用标准正态变换校正(SNV)对光谱进行预处理，以消除基线平移和散射作用带来的样本光谱差异。随机选择经过预处理后的7个样本用作训练集以建立模型，剩余的4个样本用作测试集以验证模型，建立了基于 k 最近邻算法、支持向量机算法和随机森林算法的融合模型，并与 k 最近邻算法模型，支持向量机算法模型，随机森林算法模型进行比较，结果显示融合模型的实验结果更好。基于该融合模型得到的实验结果为：0～2 d之间预测样本的正确分类率(CCR)为80%，平均误差为0.053 d，2～20 d之间预测样本的CCR为69%，平均误差为0.442 d，与利用高光谱获得的结果相当甚至更好。为测试该方法的适用性，检验了血液特异性对本模型的影响，实验样本取自8个不同捐献者的20个血迹，将其中4个捐献者的10个样本加入原有模型，另外4个捐献者的10个样本作为测试集以检验血液特异性对模型影响。实验结果为：在0～2 d之间的CCR为75.6%，平均误差为0.063 d，2～20 d之间预测样本的CCR为65.6%，平均误差为0.467 d。CCR无显著降低，表明该模型对不同来源的血迹样本，仍然有着较好的适应性。与采用高光谱的研究结果相比，多光谱结合模型融合算法同样可以获得较好的血迹年龄估计结果，并具有结构简单，成本低廉，稳定性好的优点，有望成为一种高精度的血迹年龄预测手段，在法医学领域中有重要应用前景。

关键词 多光谱；血迹；年龄估计；模型融合

中图分类号：O433.4 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2020)01-0168-06

引 言

暴力犯罪现场的血迹是刑事侦查中的重要物证。血液从离开身体的那一刻开始老化，可以通过研究血液老化的规律来估算血液年龄，估计发生创伤事件的时间。对于犯罪现场办案人员来说，精确的血迹年龄预测可以用来推测出犯罪发生的时间，有助于确定犯罪嫌疑人[1]。文献中记载的血迹年龄估计方法可以追溯至80多年前。近年来出现了更多利用光谱技术，例如荧光寿命[2]，近红外(NIR)光谱[3]，高光谱成像[4]，拉曼光谱[5]等来预测血迹年龄的报道，但这些技术大多数测试条件比较苛刻，设备复杂而昂贵，且大多不能现场分析，需要在犯罪现场收集样本后送去实验室分析。监测血迹年龄的最简单方法是观察血迹颜色随时间的变化。当血液离开人体时，氧合血红蛋白(HbO_2)会快速氧化成高铁血红蛋白(MetHb)，而高铁血红蛋白又会缓慢变成血红蛋白(HC)，该反应会引起血液颜色变化，从而使得可见光谱估计血迹年龄成为可能[7]。可见近红外反射光谱技术相比其他血迹年龄估计技术，具有无损检测，装置简单，价格低廉等优势，受到了越来越多的关注。

1960年，Patterson使用色度计对血迹的反射率进行测量并将CIE色度指数的变化与血迹的年龄联系起来，取得了一些成果。Bremmer等[8]使用漫反射光谱法跟踪了0到60天之间血迹的老化过程，提出了基于 HbO_2 转化为MetHb然后转化为HC的血液老化的动力学模型。董永芳等[9]使用了基于遗传区间偏最小二乘法对血迹年龄进行估计。Li

收稿日期：2018-11-30，**修订日期**：2019-03-21

基金项目：国家重大仪器设备开发专项(2012YQ180132)和国家自然科学基金项目(61775133)资助

作者简介：戎念慈，1994年生，上海交通大学电子信息与电气工程学院硕士研究生 e-mail：15358570321@163.com

*通讯联系人 e-mail：mzhuang@sjtu.edu.cn

等[10-11]使用了线性判别分析与可见近红外反射光谱相互结合的方法预测血迹年龄。上述研究基本都基于价格比较昂贵的高光谱成像系统或高光谱相机进行，Thanakiatkrai 等[12]则使用智能手机对血迹斑点进行拍照，通过图像分析预测血迹年龄。

目前可见光谱法估计血迹年龄的精度普遍不够理想，还有待提高。Li 等[10]使用显微分光光度计 TIDAS MSP 400 进行光谱采集，利用线性判别分析(linear discriminant analysis，LDA)模型对血迹年龄进行预测，预测结果在 2～20 d 内的平均误差为 0.923 d，正确分类率(correct classification rate，CCR)为 47.7%，当容许误差为 1 d 时，CCR 到达 80.7%，容许误差为 2 d 时，CCR 可以达到 92.3%。Thanakiatkrai 等[12]使用智能手机对血迹斑点进行拍照，通过 RGB 三个波段进行血迹年龄预测，平均误差为 0.61 d。2013 年，Li 等[11]利用双高光谱系统采集光谱，使用改进的 LDA 预测血迹年龄，在前 7 d，平均误差为 0.27 d，在 30 d 内时平均误差为 1.17 d，容许误差为 1 d 时，CCR 达到 89.3%，其测试数据集在 2～20 d 的平均误差为 0.85 d，CCR 为 61.6%。董永芳等[9]使用的基于遗传区间偏最小二乘法预测血迹年龄，0～2 d 的平均误差为 0.063 d，2～20 d 的平均误差为 1.185 d。Edelman 等[4]使用最小二乘进行血迹年龄估计，平均误差在 0～2 和 2～20 d 分别为 1.65 和 3.5 d。

Bremermer 等[8]的工作未考虑血迹特异性对模型的影响，而 Li[10]等使用 LDA 模型预测血迹年龄时，发现当使用一个新的血迹样本验证模型，CCR 就从 91.5% 下降至 37.3%，表明，血迹的特异性对血迹时间模型可能有着很大的影响。本文使用的机器学习模型，如 k 最近邻算法(k-Nearest Neighbor，k-NN)、支持向量机算法(support vector machine，SVM)和随机森林算法(random forest，RF)有着很强的抗干扰能力，能够更好地估计来源不同的血迹的年龄，在应对血液特异性对血迹年龄估计上有着很强的适应性。

构建了以 8 个 LED 为照明光源、以黑白 CCD 相机为成像单元的可见-近红外多光谱成像系统，研究了利用可见-近红外反射多光谱精确估计人体血液年龄的可行性，使用了融合 k-NN，SVM 和 RF 的融合模型方法进行血迹年龄估计，建立了血迹预测模型并验证了血液特异性对模型的影响。相比于其他方法，平均误差更小，稳定性更好，所建模型的准确率得到了提升。

1　实验部分

1.1　样本采集

实验用的 11 个血液样本采集自健康志愿捐献者，采集时间为上午 10:00—10:20。分别取 20 μL 滴在白色纯棉布上，制得 11 个血迹样本。储存于常温的黑暗环境下。11 个血迹样本随机分成两部分，其中 7 个血迹作为训练集样本，4 个血迹作为测试集样本。

验证不同个体的血迹特异性对模型影响的实验采集了 8 名健康志愿捐献者的 20 个血迹样本．采集时间为 15:20—15:30。分别取 20 μL 滴在白色纯棉布上，制得 20 个血迹样本，并储存于常温的黑暗环境下。20 个血迹样本随机分成两部分，其中 10 个血迹作为训练集样本加入模型建立，10 个血迹作为测试集样本。

1.2　仪器

自主研制的以 LED 为光源的可见-近红外多光谱系统框图如图 1 所示，系统由两部分组成，包括照明模块和图像采集模块。照明模块由 LED、驱动电源及其控制软件和光纤组成，通过照明控制软件控制不同波长的 LED 发光，通过光纤传导，从而实现令不同波长的光照明被测物的功能。LED 的额定电压为 3 V，电流为 1.5 A，其中心波长及带宽如表 1 所示，发射光谱如图 2 所示。

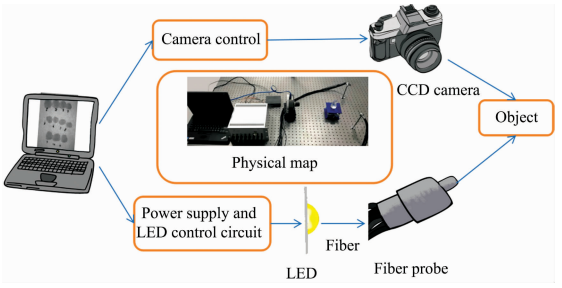

图 1　可见-近红外多光谱系统

Fig. 1　Visible-near infrared multi-spectrum system

表 1　LED 的中心波长及带宽

Table 1　Wavelengths and bandwidths of LEDs

中心波长/nm	带宽/nm
395	15
430	20
490	20
525	30
590	15
625	15
835	25
925	50

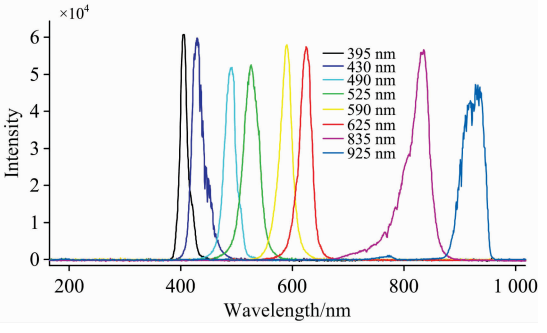

图 2　LED 发射光谱

Fig. 2　Emission spectra of LEDs

图像采集模块由黑白 CCD 相机及相机控制软件组成，通过软件设置曝光时间、采样频率、焦距等参数，控制 CCD 相机实现采集图像和存储功能。其中 CCD 相机为福州鑫图光电有限公司的 TCC-1.4LICE-N 相机。多光谱系统通过图

像采集模块采集被测物在不同波长的 LED 照射下的图像，实现多光谱的采集。

1.3 反射光谱采集及模型评价标准

采用漫反射方式测量样本多光谱。使用白布参考区域比对多次测量时 LED 亮度。所有采集过程均在暗室中进行。采集时间为 1，2，3，4，5，7，9，11，13，17，21，25，30，35，46，49，60，73.5，77，82，100，107，117，126，131，142.5，147，153，165，170.5，174，197.5，220，251.5，271，296.5，346.5，366.5，418.5 和 439.5 h，共获取 320 幅图像。

模型评价标准：使用 CCR 与平均误差指标对模型进行评价。CCR 越接近 1，平均误差越小，模型的预测能力越好。

2 结果与讨论

2.1 光谱预处理

2.1.1 反射率计算

首先，采集相机的暗噪声 I_{dark}，并通过记录未染有血迹的空白布的参考图像光强(I_0)进行多光谱反射率的计算。被测样品图光强(I_s)是在同等光照条件下通过相机采集，依次采集血迹在 8 个不同 LED 照明下的各时段反射率。多光谱反射率(R)通过式(1)计算得出[6]

$$R = \frac{I_s - I_{dark}}{I_0 - I_{dark}} \tag{1}$$

2.1.2 标准正态变换校正

图 3(a)为血迹年龄在 1.00 h 时的 11 个血迹斑点的反射率折线图。由图可知，由于存在基线平移和散射影响，同样年龄的血迹反射率有着较大差异[13]。因此，需要对光谱进行预处理，本文采用标准正态变换校正(standard normal variate transformation，SNV)对光谱进行预处理，对每组反射率进行标准化预处理[13]。预处理结果如图 3(b)所示。SNV 校正后，反射率差异显著降低，有效消除了基线平移和散射作用带来的光谱差异。

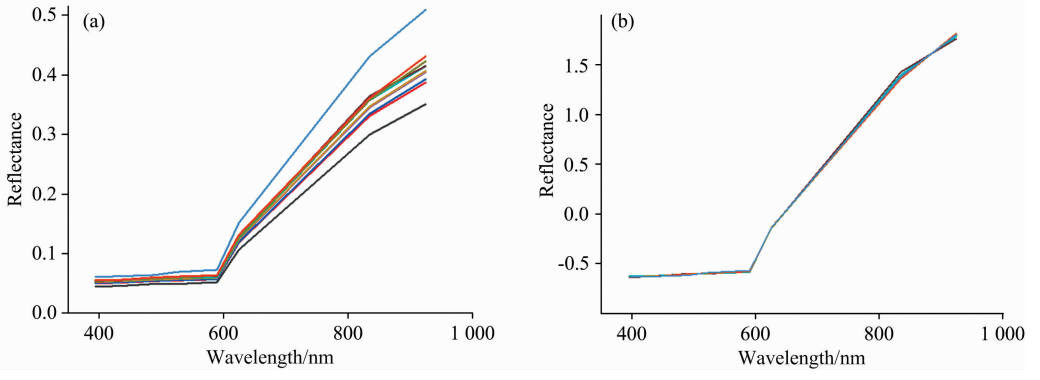

图 3 年龄相同的 11 个血迹的光谱

(a)：原始光谱；(b)：SNV 预处理后光谱

Fig. 3 Spectra of eight bloodstains with the same age

(a)：Raw reflectance spectra；(b)：Spectra after SNV pretreatment

2.2 基于 k-NN，SVM，RF 的模型融合法估计血迹年龄

2.2.1 模型结果与分析

k-NN 是在给定的训练数据集上，寻找与测试集的欧几里德距离最小的 k(k=4)个实例，并以其中的多数决定测试样本的分类[14]。SVM 是由 Vapnik 首先提出的一种基于结构风险最小化的分类器，鲁棒性较好，计算效率较高，并且具有过拟合控制策略以及良好的抗干扰和噪声能力[15]。RF 是采用构造多颗“决策树”的方式进行分类[16]，图 4(a)—(c)分别为使用 k-NN，SVM 和 RF 得到的血迹估计年龄。

由于血液年龄估计的准确性在 0～2 和 2～20 d 两个时间段之间存在明显的差异，因此将数据集的评价分成两个时间段。表 2 为各模型的预测结果，三种模型的预测误差都较小，有较准确的预测能力。但三种模型对短时间与长时间有着不同的表现，SVM 在短时间内有着更好的预测能力，而 RF 则对长时间有更为出色的预测能力。

为了找出一种兼具 SVM 的短期预测优势和 RF 长期预测优势的模型，采用模型融合方法；模型融合是一种对模型的集成策略。不同的模型，从不同的角度观测数据集，k-NN 更加关注样本点之间的距离关系；RF 更加关注分裂节点时候的不纯度变化；SVM 则注重于寻找不同类别之间的分界面。模型融合结合了不同模型的观测角度，得到一个更加全面的结果。

模型融合步骤如下：把训练集分为不交叉的三份 train1，train2，train3。分别以 train1，train2，train3 作为测试集，剩下的两份作为训练集建模，将预测结果作为新模型的训练集。将多模型对测试集进行预测，将预测结果取平均，作为测试集的新表达。分别使用 k-NN，SVM，RF 作为模型融合的基模型，将 RF 作为模型融合的第二层模型进行建模预测。图 4(d)为融合模型得到的血迹估计年龄。在 0～2 d 内的平均误差为 0.053 d，CCR 达到 80%，在 2～20 d 的平均误差为 0.442 d，CCR 达到 69%。在 0～2 d 内若容许误差为 2 h 时，CCR 可达到 88%，在 2～20 d 内若容许误差为 1 d 时，CCR 可达到 92%。同时拥有短时间与长时间的较好的预测能力。根据 Li 等[12, 13]论文中的血迹预测结果，在 1～20 d

内，CCR 为 65%，平均误差 0.85 d，相比之下，本模型具有更好的预测能力和稳健性。

表 2 列出了部分不同文献研究采用高光谱相机和建模方法得到的预测结果与本工作的结果对比表明，采用多光谱系统结合模型融合方法，得到了较满意的血迹年龄预测结果。

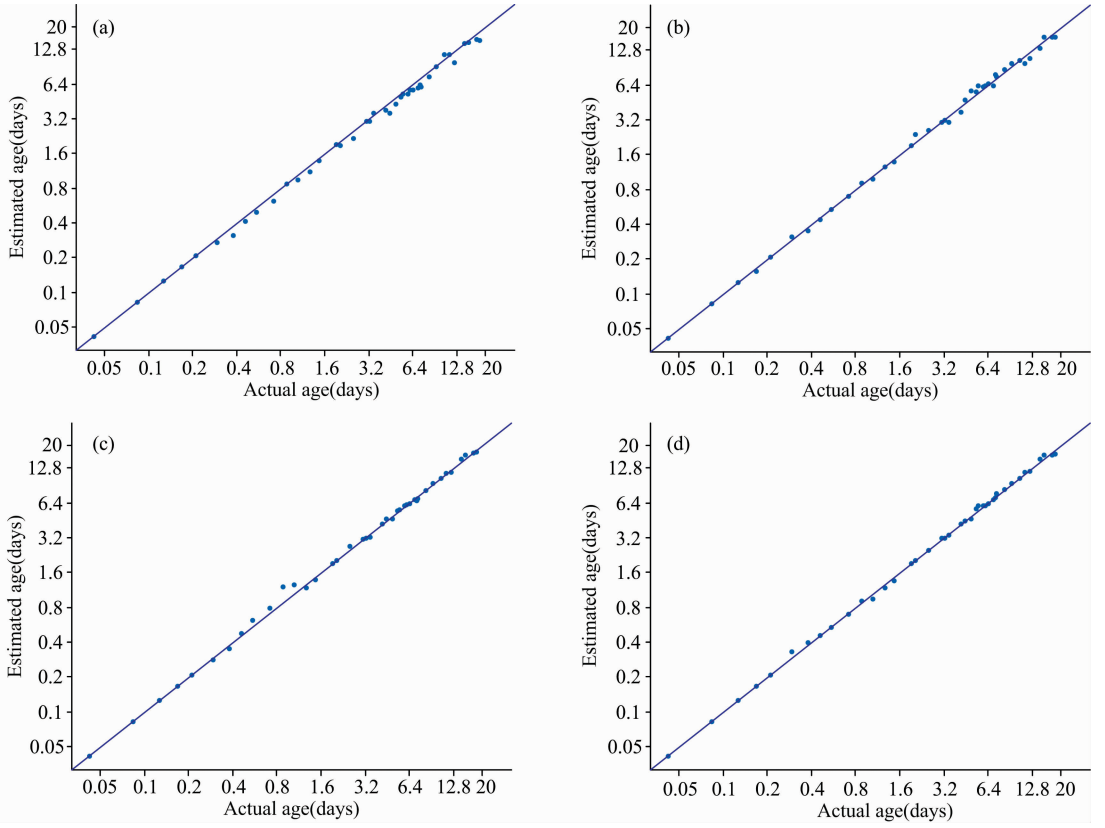

图 4　采用不同模型预测白布上血迹年龄的结果

(a)：*k*-NN 模型；(b)：SVM 模型；(c)：RF 模型；(d)：整体模型

Fig. 4　Results of the age estimation of blood stains on white cotton by different models

(a)：*k*-NN；(b)：SVM；(c)：RF；(d)：Ensembling

表 2　本模型与其他模型血迹年龄预测结果对照

Table 2　Results of the age estimation of blood stains by different instruments and models

模型	0～2 d		2～20 d		仪器
	平均误差(d)	CCR/%	平均误差(d)	CCR/%	
k-NN	0.065	73	0.686	50	实验室自主研发的多光谱系统，拥有 8 个照明 LED，中心波长分别为 395，430，490，525，590，625，835 和 925 nm，带宽分别为 15，20，20，30，15，15，25 和 50 nm
SVM	0.054	75	0.586	60	
RF	0.059	75	0.514	62	
模型融合	0.053	80	0.442	65	
线性判别分析[10]	—	—	0.71	54	加入 Olympus BX51 显微镜的 J&M Tidas MSP400 光谱仪系统，光谱范围为 380～780 nm，光谱分辨率约为 1 nm。
改进线性判别分析[11]	—	—	0.85	61.6	光谱范围为 400～720 nm，全谱最大光谱带宽为 20 nm 的简单低成本高光谱系统。
遗传区间偏最小二乘[9]	0.066 3	—	1.815	—	美国海洋光学的 USB-4000 微型光纤光谱仪，光谱范围 200～1 100 nm。
结合血红蛋白估计[6]	约 1.65	—	约 3.5	—	线扫描光谱成像系统，光谱范围为 400～1 000 nm。

2.2.3　血液特异性对模型影响

为检验血液特异性对模型影响，采集了来自 8 名不同志愿捐献者的 20 个血迹样本。将其中 10 个来自 4 名捐献者的血迹样本加入原模型增强对不同来源血液的稳定性，剩下 10

个来自其余 4 名捐献者的测试集样本，对短期血迹估计模型进行验证。验证结果为 0～2 d 内，*k*-NN 的 CCR 为 70.2%，平均误差为 0.069 4 d，SVM 的 CCR 为 72.8%，平均误差为 0.063 9 d，RF 的 CCR 为 67.9%，平均误差为 0.069 8 d，使用模型融合方法，得到的 CCR 为 75.6%，平均误差为 0.063 1 d；2～20 d 之间预测样本的 CCR 为 65.6%，平均误差为 0.467 d。对比表 2 中的结果，模型应对血液特异性影响的能力较强。董永芳等[9]建立的基于遗传区间和最小二乘模型应对血液特异性有较好的表现，在 0～2 d 内的平均误差为 0.062 5 d，2～20 d 内的平均误差为 0.467 d。相比前人建立的血迹年龄估计模型，使用基模型为 *k*-NN，SVM 和 RF 的融合模型有着更好的表现。

3 结 论

相比于昂贵的高光谱系统，本工作建立的 LED 光源和单色 CCD 相机组成的多光谱系统价格低廉，结构简单，同样可以达到快速无损估计血迹年龄的目的。原始光谱图像经过 SNV 预处理，使用了将 *k*-NN，SVM 和 RF 作为基模型的模型融合方法，得到了更加准确的预测结果。实验中将 11 个人体血液样本中的 7 个样本作为训练集建立模型，对其余 4 个血迹样本进行预测，在 0～2 d 内的平均误差为 0.053 d，CCR 达到 80%，在 2～20 d 的平均误差为 0.442 d，CCR 达到 65%。还验证了血液特异性对本模型的影响，在加入来自不同捐献者的血迹样本时，CCR 无显著降低，表明使用的多种算法融合模型对血液特异性有着较好的抗干扰能力。与参考文献的研究结果相比，所建预测模型的平均误差显著减小，预测能力显著提升。因此，可见-近红外多光谱和多种算法融合的模型可以成为一种快速无损且高精度的血迹年龄预测手段，将会在法医学领域中有重要应用价值。

References

[1] Edelman G J, Gaston E, van Leeuwen T G, et al. Forensic Science International, 2012, 223(1-3): 28.
[2] Shine S M, Suhling K, Beavil A, et al. Analytical Methods, 2017, 9(13): 2007.
[3] Edelman G, Manti V, van Ruth S M, et al. Forensic Science International, 2012, 220(1-3): 239.
[4] Edelman G, van Leeuwen T G, Aalders M C. Forensic Science International, 2012, 223(1-3): 72.
[5] Premasiri W R, Lee J C, Ziegler L D. Journal of Physical Chemistry B, 2012, 116(31): 9376.
[6] Edelman G, van Leeuwen T G, Aalders M C. Forensic Science International, 2012, 223(1-3): 72.
[7] Bremmer R H, de Bruin D M, De J M, et al. Plos One, 2011, 6(7): e21845.
[8] Bremmer R H, Nadort A, van Leeuwen T G, et al. Forensic Science International, 2011, 206(1): 166.
[9] DONG Yong-fang, MENG Yao-yong, ZHANG Ping-li, et al(董永芳, 孟耀勇, 张平丽, 等). Acta Optica Sinica(光学学报), 2015, 35(8): 369.
[10] Li B, Beveridge P, O'Hare W T, et al. Forensic Science International, 2011, 212(1-3): 198.
[11] Li B, Beveridge P, O'Hare W T, et al. Science & Justice, 2013, 53(3): 270.
[12] Thanakiatkrai P, Yaodam A, Kitpipit T. Forensic Science International, 2013, 233(1-3): 288.
[13] Barnes R J, Dhanoa M S, Lister S J. Applied Spectroscopy, 2016, 43(5): 772.
[14] Chang C C, Lin C J. LIBSVB: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology, 2011, 2: 27.
[15] MA Jun-cheng, DU Ke-ming, ZHENG Fei-xiang, et al(马浚诚, 杜克明, 郑飞翔, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(6): 1863.
[16] Jia J, Liu Z, Xiao X, et al. Journal of Theoretical Biology, 2016, 394: 223.

Age Estimation of Bloodstains Based on Visible-Near Infrared Multi-Spectrum Combined Ensembling Model

RONG Nian-ci, HUANG Mei-zhen*
Department of Instrumentation Science and Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China

Abstract The accurate estimation of blood age is of great significance in the forensic identification of criminal investigation. In this paper, a visible-near-infrared multispectral imaging system with 8 LEDs as illumination source and monochrome CCD camera as image input unit is constructed. The ensembling model based on k nearest neighbor method, support vector machine and random forest method is used to analyze and estimate the age of bloodstains. The feasibility of using the visible-near-infrared reflectance multispectral to accurately estimate the age of human blood was investigated, and the results were compared with the previous studies using hyperspectral techniques for blood age estimation. The influence of blood specificity was also tested. The experiment recorded images of 8 channels from 400 to 940 nm on days 1 to days 20 of 11 human blood samples, and the spectra were preprocessed using standard normal variate transformation (SNV) to eliminate spectral differences due to the baseline shift and scattering. Seven preprocessed samples were randomly selected as training set to build the model, and the remaining four samples were used as test sets to test the model, a model ensembling model based on k nearest neighbor method, support vector machine and random forest method was built. Compared with the results by *k*-NN model, SVM model and RF model the result is better. The correct classification rate (CCR) of the samples between 0 and 2 d is 80%, the average error is 0.053 d, and the CCR between 2 and 20 d is 69%. The average error is 0.442 d, which is comparable or better than that obtained by using hyperspectral techniques. In order to test the practical applicability of the method, this paper tested the effect of blood specificity on the model. The test sample was 20 blood samples taken from 8 different donors, 10 of which from 4 donors were used to refine the original model, and 10 samples from another 4 donors were used as test sets to test the effect of blood specificity. The estimated age of blood from different sources is: CCR is 75.6% between 0 and 2 d, and the average error is 0.063 1 d. After adding blood samples from different donors, there was no significant decrease in CCR, indicating that the model still has good adaptability to blood samples from different sources. The results show that compared with the previous research results, multispectral technology combined with model ensembling algorithm could obtain more accurate age estimation results, and has the advantages of simple set-ups, low-cost and good stability, which might be a high-precision blood age estimation method and have important application value in the field of forensic science.

Keywords Multispectral; Bloodstains; Age estimation; Ensembling model

(Received Nov. 30, 2018; accepted Mar. 21, 2019)

* Corresponding author