

基于 SURF 特征的羊绒羊毛鉴别算法

柴新玉¹, 路 凯¹, 钟跃崎^{1,2}

(1.东华大学 纺织学院, 上海 201620; 2.东华大学 纺织面料技术教育部重点实验室, 上海 201620)

摘要: 为了实现羊绒与羊毛纤维的快速自动鉴别,提出了基于 SURF 特征的羊毛羊绒鉴别算法。首先采用扫描电子显微镜获取两种纤维的表面形貌图像,再对图像进行预处理,去除背景并对纤维区域进行增强处理,然后提取 SURF 特征并聚类,将原始图像转换成相对应的高维向量,最后对数据进行分类。试验结果表明,该方法有效,在两种纤维不同混合比例的情况下,识别率都超过 90%。

关键词: 羊毛; 羊绒; 鉴别; SURF 特征; 扫描电子显微镜

中图分类号: TS101.92

文献标识码: B

文章编号: 1001-2044(2018)06-0025-04

Identification algorithm of wool and cashmere based on SURF feature

CHAI Xinyu¹, LU Kai¹, ZHONG Yueqi^{1,2}

(1.College of Textiles, Donghua University, Shanghai 201620, China)

(2.Key Laboratory of Textile Science & Technology, Ministry of Education, Donghua University, Shanghai 201620, China)

Abstract: To achieve the quick and automatic identification of wool and cashmere, a practical algorithm based on speed-up robust features (SURF) is proposed. Firstly, the surface images of the two fibers are taken by scanning electron microscope (SEM), and the image is preprocessed, the background is removed and the fiber region is enhanced. Then the SURF features are extracted and clustered, and the original images are converted into corresponding high-dimensions vector data. Finally, the data is categorized by support vector machine (SVM). The results show that the proposed algorithm is an effective approach for identification of wool and cashmere with average recognition rates over 90% with different mixing ratios of wool and cashmere.

Key words: wool; cashmere; identification; SURF feature; SEM

羊绒、羊毛纤维外观相似,理化性能差别较小,但价格相差巨大,因此对于两种纤维的准确鉴别就显得格外重要。目前比较常见的鉴别方法主要有扫描电子显微镜法(Scanning Electron Microscope, SEM)^[1-2]、光学显微镜法^[3]、计算机图像分析法^[4-6]、DNA 分析法^[7]、近红外光谱法^[8]等。

检测机构在实际测试中仍然采用光学显微镜对纤维外观形貌进行人工鉴别,这种方法依赖于检测人员的主观判断,并且通常需要一年或更长的时间才能培训出一名合格的检测人员。通过提取纤维的线粒体 DNA^[9-10],使用聚合酶链式反应技术(Polymerase Chain Reaction, PCR)扩增然后测序,根据山羊和绵羊两个种类在碱基序列上的差别来确定纤维的种属,这种方法可以达到较高的准确率,但是由于时间和成本的原因没有大量普及。也有研究人员采用图像处理法处理纤维图像,之后再提取鳞片的周长、面积等几何特征参数进行纤维鉴别,这种方法对于图像质量的要求

较高,很难达到理想的效果,并且可获得的特征参数不能代表所有的纤维种类。

一些研究者探索了利用光镜图像鉴别纤维的方法,光镜图像获取方便,但是相较于 SEM,图像的清晰度比较差,之后的图像处理和提取特征也存在较大的难度。本文比较了纤维 SEM 图像在不同混合比例下的识别效果,并比较了直接使用原始图像和使用经处理的图像进行分类的效果差别。结果发现,使用羊绒、羊毛纤维的 SEM 图像进行研究,结果稳定可靠,识别准确度较高,并且去除背景后识别精度也有所提升。

1 研究方法

图 1 给出了算法的主要流程。

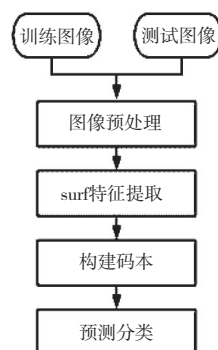


图 1 算法整体流程

收稿日期: 2017-09-27

基金项目: 国家自然科学基金资助项目(61572124);上海市自然科学基金资助项目(14ZR1401100)

作者简介: 柴新玉(1993—),男,硕士生,主要从事基于图像的纤维鉴别技术研究。

通信作者: 钟跃崎。E-mail: zhyq@dhu.edu.cn。

本文基于两种纤维图像在外观形态特征上的不同来进行分类。首先使用显微成像系统得到羊绒和羊毛纤维的图像,然后通过特征提取算子得到两者局部特征的集合,构建区别两者的识别模型,最后对测试图像加以验证。

1.1 图像背景去除

直接获取的原始图像中会存在一些干扰因素,比如背景亮度不匀、杂质和噪音等。为了测试图像背景对分类结果的影响,需要对直接拍摄的图像进行预处理,最后只保留纤维主干部分。

对 SEM 图像进行处理的过程见图 2。

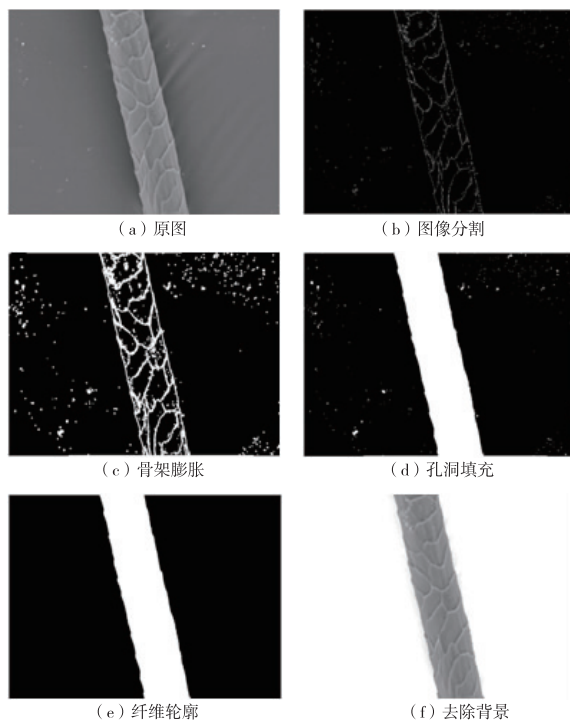


图2 SEM 图像背景剔除

如图 2 所示,首先调整原始图像灰度值,增大纤维区域和背景的对比度,再使用边缘检测算子获取纤维轮廓^[11],得到如图 2(b)所示的骨架图像;然后检测骨架区域和图像边缘的交点并分别连接,再对骨架进行膨胀,得到图 2(c)中封闭的纤维区域;接下来填充孔洞并去掉面积较小的封闭区域,得到图 2(e)中的纤维轮廓模版;最后通过模版去除原始图像中的背景得到图 2(f)。

1.2 SURF 特征提取

在计算机处理图像时,希望可以模仿人类通过一种算法自动找到某类图像中的稳定特征,提出描述方法,并与其他种类的图像进行区别。SURF 算法提出了一种解决办法,通过该算法找到某类图像内部的关

键点,然后对两种图像进行区分^[12-13]。

SURF 算法可以检测图像中的一些特征点,比如角点、边缘点、亮点等。此方法使用了积分图像,加快了特征点获取的速度。在特征提取时,使用海瑟检测算子,在图像的 X 点处,定义尺度为 σ 的海瑟矩阵,见式(1):

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \quad (1)$$

式中, $L_{xx}(X, \sigma)$ 指高斯二阶微分 $\frac{\partial^2}{\partial x^2} g(\sigma)$ 在 X 处和图像 $I(x, y)$ 卷积的结果, $L_{xy}(X, \sigma)$ 和 $L_{yy}(X, \sigma)$ 的计算方法与 $L_{xx}(X, \sigma)$ 类似。其中 $g(\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ 。整个过程主要包括 4 个步骤:(1) 计算积分图像;(2) 建立图像的尺度空间;(3) 在建立的尺度空间上进行特征点定位;(4) 特征点表征。根据上述方法试验的结果见图 3,圆圈代表提取出的 SURF 特征。可以发现它们主要集中在纤维的两侧边缘。

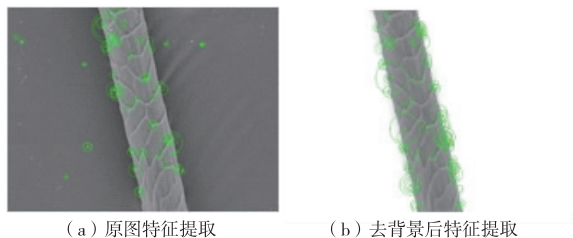


图3 SURF 特征检测

1.3 构建码本

SURF 特征虽然可以描述图像,但由于提取出的数据量巨大,需要找到一种方法降低计算量。本文采用聚类的方法对数据进行处理^[14],再将聚类后的每个中心映射为视觉单词。通过这种方法,将图像中的二维像素信息转化为更易于分类的视觉单词。一般使用 k -means 作为聚类的方法,首先在之前提取的 SURF 特征 $X\{x^{(1)}, \dots, x^{(M)}\}$ 里随机确立 k 个聚类的中心 $U\{u^{(1)}, \dots, u^{(K)}\}$,通过不断迭代,目标是 minimized 各个样本点到每个中心的欧式距离,见式(2):

$$D(X, U) = \sum_{m=1}^M \sum_{k=1}^K \|x_m - u_k\|^2 \quad (2)$$

式中: M ——SURF 特征;

K ——聚类中心的总数;

x ——SURF 提取的特征;

u ——聚类中心点

聚类的中心点即表述纤维特征的视觉单词 W

$\{w^{(1)}, \dots, w^{(n)}\}$, 然后把图像中的每个 SURF 特征都映射到离它最近的中心。最终, 单幅图像通过充分提取 SURF 特征, 并聚类得到视觉词, 再计算每个视觉词的出现次数, 得到能够描述单根纤维的语义信息, 综合整理出一类纤维的语义特征模型。将单幅图像都转化为一个包含纤维关键信息的视觉词直方图, 见图 4, 图中横坐标表示聚类中心, 即视觉词将此直方图表示成数字矢量, 归一化后, 码本矢量就能用于训练分类器了。

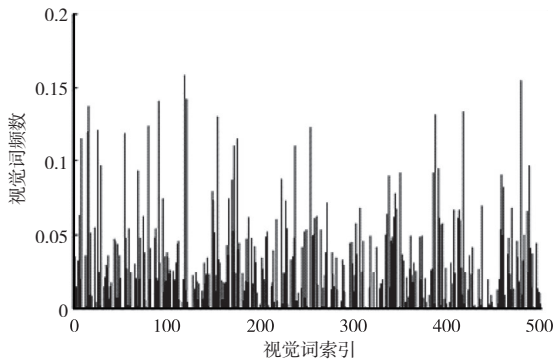


图4 视觉词频统计

1.4 分类器介绍

本文采用支持向量机进行图像的分类^[15], 支持向量机是一种监督学习理论, 对于小样本和高维的数据分类问题有较好的适用性, 普遍应用于各类模型分类算法中, 其基本思想见图 5。

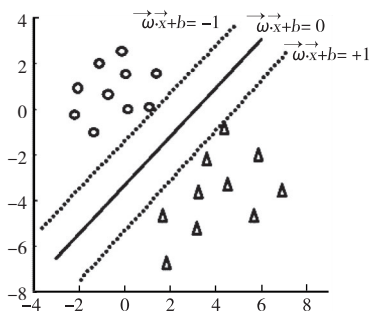


图5 SVM 分类示意图

图 5 中的圆圈和三角分别表示两类物体, 中间的实线为分类超平面。通过对事先标定好种类的数据集进行分析, 形成一个最佳的分类超平面 $\omega^T x + b = 0$, 把两种训练样本分开, 其中, x 表示数据点, ω 是 n 维特征向量。在纤维分类时, x 就是每张图像所提取出的单词频率矢量, 首先使用一个映射函数把原始样本映射到高维空间中, 再用线性回归的方法进行分析, 得到回归函数, 见式(3):

$$f(x) = \sum_{i=1}^N a_i y_i K(x_i, x) + b \quad (3)$$

式中: N ——样本个数;

x ——特征;

y ——结果标签;

$K(x_i, x)$ ——核函数

通过训练集图像计算出分隔两个种类的超平面, 根据距离的远近对测试集图像进行预测。

2 试验部分

2.1 试验条件与流程

试验设备: 所用计算机的 CPU 为 Intel Core i7-6500U, 主频为 2.50 GHz, 内存为 4 GB, 操作系统为 Windows 10, 软件开发环境为 MATLAB R2016b。

本文所用纤维试样为国产白绒和土种毛, 由鄂尔多斯集团提供。分别拍摄这两种纤维的 SEM 照片。制样方法参考 GB/T 14593—2008《山羊绒、绵羊毛及其混合纤维定量分析方法》。采用的测试设备为 HITACHI TM3000 型台式扫描电镜, 放大倍数为 1 000 倍。具体试验步骤如下: (1) 采用图像背景去除算法对原始 SEM 图像进行预处理; (2) 用 SURF 算法对训练图像进行特征提取和描述, 得到图像的描述符; (3) 对得到的所有特征点和描述符用 k-means 方法得到聚类中心; (4) 用得到的聚类中心来统计每幅图像中特征点的频数, 获得视觉词矢量; (5) 对数据进行分类。

本文选择的纤维图像总数为 1 000 张, 其中羊绒和羊毛各为 500 张。试验时选择总数的 70% 作为训练集, 20% 作为验证集, 剩下 10% 为测试集。

2.2 试验结果与讨论

本文识别率的定义见式(4):

$$A_c = R_c / T_c$$

$$A_w = R_w / T_w$$

$$A_t = (A_c + A_w) / 2 \quad (4)$$

式中: A_c ——羊绒识别率;

A_w ——羊毛识别率;

A_t ——平均识别率;

R_c ——羊绒正确识别的根数;

R_w ——羊毛正确识别的根数;

T_c ——羊绒实际根数;

T_w ——羊毛实际根数

图 6 为对两种纤维的 SEM 图像进行分类的结果, 羊绒和羊毛按照不同的根数比例混合后进行测试。

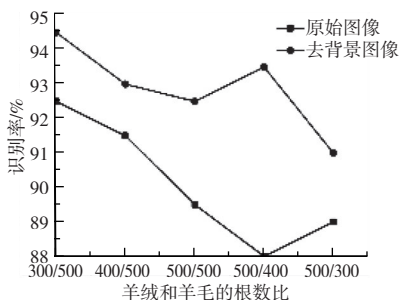


图6 原始图像与去背景后的识别结果比较

从图6可以发现,经过背景去除后的图像在识别精度上均好于原始图像。这是因为在去掉背景的图像中,SURF特征均集中在纤维区域,而对于原始图像,由于存在背景噪声,SURF算子会在非纤维区域提取到一些特征点,最终降低了分类的精度,这说明去除图像背景对识别效果会产生积极的影响。

为了验证样本根数对识别精度的影响,首先固定羊绒和羊毛的混合比例,然后逐渐改变总样本量进行试验,识别结果见图7。

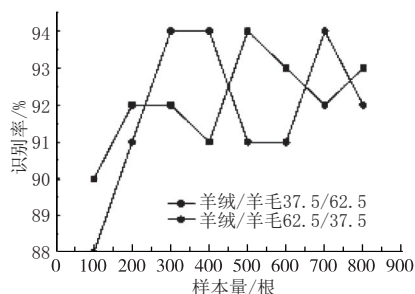


图7 不同样本总量的识别率

由图7可以发现,在混合比例为37.5/62.5时,平均识别率均不低于90%;当混合比例为62.5/37.5时,除样本量为100时平均识别率为88%,其他情况均大于90%,表明了本方法在小样本量时的有效性。

图8为两种纤维在不同混合比时的识别结果,试验样本选用的是去除背景后的图像。

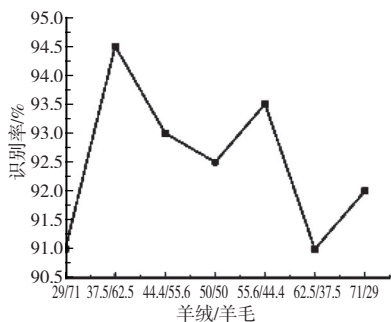


图8 不同混合比的识别率

从图8可以发现,平均识别率最低为91%,最高为

94.5%,CV值为0.014,说明本文方法不受混合比的影响。经过统计,当使用1000根纤维进行试验时,平均训练时间为120s,测试时间为16s。

3 结语

在SEM图像中,羊绒、羊毛两种纤维的表观结构存在着细微差别,利用SURF特征算子可以快速有效地提取出同种类纤维的特征用于分类辨别。试验结果证明,本文方法在小样本、不同混合比情况下有较高的识别率,并且通过对图像进行背景去除,可以进一步提高识别精度。

本文的研究是对SURF特征点提取算法在羊绒羊毛纤维鉴别的探索。以后的工作将采用该方法对比测试更多种类的纤维样品,逐步发现影响分类效果的其他因素,并对识别算法进一步优化,进而获得更好的识别效果。

GSTR

参考文献:

- [1] WORTMANN F J. Quantitative fiber mixture analysis by scanning electron microscopy. III. Round trial results on mohair/wool blends [J]. Textile Research Journal, 1991, 61(7): 371-374.
- [2] WORTMANN F J, AUGUSTIN P. Quantitative fiber mixture analysis by scanning electron microscopy. VII. Modeling the microscopic analysis of binary animal fiber blends [J]. Textile Research Journal, 2004, 56(3): 248-252.
- [3] 何兰芝,陈莉萍,王雪梅.山羊绒与羊毛纤维鉴别检测综述[J].上海纺织科技,2008,36(10):44-46.
- [4] 陶晨.利用图像处理和分析技术测定羊毛/羊绒混纺比[J].毛纺科技,2009,37(2):47-50.
- [5] 石先军,胡新荣,等.基于鳞片纹图基因码的羊绒理论识别精度及正误判率[J].纺织学报,2014,35(4):5-10.
- [6] 石先军,于伟东,袁子厚.基于贝叶斯方法的山羊绒与细羊毛的鉴别[J].纺织学报,2008,29(1):26-28.
- [7] YOICHIRO Y. DNA analysis for animal fibers; The 4th International Cashmere Determination Technique Symposium Paper Collection [C]. 2008.
- [8] 吴桂芳.基于红外光谱和场发射扫描电镜技术的羊绒原料品质分析的研究[D].杭州:浙江大学,2009.
- [9] 金美菊,阮勇,李翔,等.基于基因技术的羊绒与羊毛纤维定性鉴别方法[J].纺织学报,2012,33(8):19-23.
- [10] TANG M, ZHANG W, ZHOU H, et al. real-time PCR method for quantifying mixed cashmere and wool based on hair mitochondrial DNA [J]. Textile Research Journal, 2014, 84(15): 1612-1621.
- [11] LIU H, LIU Y, LI Q, et al. Medical image segmentation based on contourlet transform and watershed algorithm; International Symposium on It in Medicine and Education [C]. IEEE, 2012.

☞(下转第45页)

整经绞数为7绞,第1绞为762根,第2~6绞为738根,第7绞为147根。

由于经纱是涤纶长丝,要求集束性好,故所选浆料的黏附力要强,同时为防止静电产生,应添加抗静电剂。浆料配方为:水100%、聚丙烯酸2.5%、PVA205MB 1.5%、浸透剂0.3%、抗静电剂0.2%。

具体的上浆工艺为:上浆率4%,回潮率1.0%,伸长率0.5%,车速80 m/min, pH 6,含固率5%,浆液粘度8 Pa·s,上浆温度80℃,靠近筒子架压浆辊气压为0.4 MPa,靠近烘房压浆辊的气压为0.6 MPa。合成纤维不应使用碱性较强的浆液,因此将pH调整为6。

由于纱线是化纤长丝,不容易断头,车速可稍快,以提高生产效率;其次要在浆料中添加抗静电剂,以避免产生静电,导致经纱相互黏连,影响生产的顺利进行;加大边纱张力,避免在织造过程中形成松边;加强纱线通道的检查,防止刮伤涤纶长丝;加强生产管理,做好回潮率、上浆率、伸长率、好轴率等浆纱质量的考核;织轴盘片要垂直,避免出现凹凸边等疵轴。

2.3 穿经工序

穿经工序工艺如下:箱号为138.5齿/10 cm,地组织和边组织采用2入/箱,银丝1入/箱,每花箱齿数为63齿,全幅箱齿数为2 355齿,上机箱幅为170 cm。综页数采用5页,第1~4页综穿地组织和边组织经纱,其中地组织经纱采用顺穿法。由于涤纶长丝强力较高,所以不需要使用停经片。由于银丝不耐高温,且不需上浆,故将111根银丝筒子分别放在两排自制的简易筒子架上,并预设一定张力,不经整浆联合机的上浆与干燥等装置,按色经排列要求在整浆联合机车头处插入到伸缩箱中,从而满足生产需要^[1]。银丝穿入第5页综。

2.4 织造工序

采用ZAX9100型喷气织机进行织造。为确保织

造的顺利进行,主要从上机张力、开口工艺、引纬工艺等方面进行设置。

上机张力的对织物风格和织机效率有着直接的影响。上机张力大,有利于开清梭口。由于经纱是涤纶长丝,强力比较高,上机张力应偏大掌握,经过调试,将上机张力调整为900 N^[2]。开口工艺主要涉及开口时间及开口量、后梁高低、停经架高低等,由于织物是平纹组织,为有利于打紧纬纱和消除箱痕,应采用早开口和高后梁,开口时间为295°,后梁高度为85 mm,深度为50 mm,停经架高度为65 mm,深度为15 mm,开口量为85、75、65、60、55 mm。

为确保纬纱能顺利引入梭口,以减少纬缩疵点,应合理调节引纬工艺参数,主喷启闭时间为65°~195°,第1~6组辅助喷嘴的启闭时间分别为65°~120°、90°~140°、120°~170°、150°~210°、170°~250°、190°~290°。由于纬纱是两种不同捻向的强捻纱,为防止其退捻产生断头或形成纬缩疵点,主、辅助喷嘴的压力要求适当调高,主喷嘴的压力设置为0.3 MPa,辅助喷嘴的压力设置为0.35 MPa,同时为防止纬纱回缩滑出主喷嘴,主喷嘴的微风应调大,设置为0.1 MPa。

此外,挡车工应经常检查引纬通道,防止飞花等杂物堵塞引纬通道,造成停台。

3 结语

合理的工艺是生产嵌银丝涤长丝/棉交织裙用织布的基础。通过工艺、设备、操作等方面的优化,使得该织物织机效率达到93%,下机一等品率达到85%,入库一等品率达到99%,达到了预期目的。

参考文献:

- [1] 蔡永东. 纯棉特细紧密纺双织物的生产体会[J]. 棉纺织技术, 2012, 40(6): 60-62.
- [2] 马顺彬, 蔡永东, 葛龙德. 涤纶包覆纱纬弹性织面料的生产技术要点[J]. 棉纺织技术, 2011, 39(3): 60-62.

(上接第28页)

- [12] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded Up Robust Features. Computer Vision-ECCV 2006[M]. Springer Berlin Heidelberg, 2006.
- [13] 赵春晖, 王莹, MASAhide, 等. 一种基于词袋模型的图像优化分类方法[J]. 电子与信息学报, 2012, 34(9): 2064-2070.
- [14] LIN W C, TSAI C F, CHEN Z Y, et al. Keypoint selection for efficient

bag-of-words feature generation and effective image classification [J]. Information Sciences, 2016(329): 33-51.

- [15] KINGSKY G, EDWARD Y C, BEITAO L. Using one class and two class SVMs for multiclass image annotation[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(10): 1333-1346.