

基于自适应编辑距离的颜料光谱匹配识别方法

王可^{1,2*}, 王慧琴^{1,2**}, 殷颖², 毛力², 张毅²

¹西安建筑科技大学管理学院, 陕西 西安 710055;

²西安建筑科技大学信息与控制工程学院, 陕西 西安 710055

摘要 为解决传统光谱匹配算法对同色系不同颜料物质光谱数据匹配识别精度不高的问题,提出了一种自适应阈值的编辑距离光谱匹配算法;研究了如何利用编辑距离对光谱反射率曲线差别敏感的特性来提高匹配精度的方法,同时通过自适应设定编辑距离的判定条件来减小算法匹配同种物质时在不同条件下光谱数据的误差。结果表明:与传统光谱匹配算法相比,自适应编辑距离算法的匹配精度更高,对颜料的识别结果更好。

关键词 光谱学; 光谱匹配; 编辑距离; 自适应阈值; 颜料识别

中图分类号 O433.4

文献标识码 A

doi: 10.3788/LOP55.113004

Pigment Spectral Matching Recognition Method Based on Adaptive Edit Distance

Wang Ke^{1,2*}, Wang Huiqin^{1,2**}, Yin Ying², Mao Li², Zhang Yi²

¹ School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China;

² School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China

Abstract To solve the problem that the traditional spectral matching algorithms have low accuracy in matching spectral data of different pigment materials in the same color system, we propose an adaptive threshold edit distance spectral matching algorithm. The edit distance is researched to improve the matching accuracy by using its characteristics of being sensitive to the spectral reflectance difference. At the same time, by adaptively setting the judging conditions of the edit distance, we reduce the error of this algorithm in matching the spectral data of the same pigment materials under different conditions. The results show that the matching accuracy of the adaptive edit distance algorithm is higher than that of the traditional spectral matching algorithms, and the recognition results of the adaptive edit distance algorithm for the pigment is better than that of the traditional algorithms.

Key words spectroscopy; spectral matching; edit distance; adaptive threshold; pigment recognition

OCIS codes 300.6550; 240.6645; 110.2350

1 引言

光谱匹配技术是通过将所测得的光谱反射率与光谱反射率数据库中已有的数据进行比较,求得 2 条曲线之间的相似性或者差异性的技术^[1],该技术能精确地在光谱图像上进行像元识别。对壁画颜料物质进行无损检测的识别方法是利用颜料的光谱

“指纹”特性^[2-4],将壁画中待检测颜料的光谱反射率曲线与标准颜料光谱数据库中样本的光谱反射率曲线进行匹配,根据匹配结果的相似度来判断识别颜料的种类^[5]。这种检测识别方法的关键是光谱匹配结果的精确度。

目前已有多种光谱匹配算法被用于解决实际问题,如:光谱数据编码匹配算法^[6]、光谱角度匹配

收稿日期: 2018-04-26; 修回日期: 2018-05-22; 录用日期: 2018-05-28

基金项目: 国家自然科学基金青年科学基金(61701388)、住房和城乡建设部科学技术计划(2017-K2-014)、陕西省国际科技合作计划(2017KW-036)、陕西省教育厅科研计划(17JK0431)、西安市科学技术局软科学研究计划(2016043SF/RK06(3))、西安市碑林区科技计划(GX1606)、西安建筑科技大学科技基金(QN1628,JC1514)

* E-mail: wangke@xauat.edu.cn; ** E-mail: hqwang@xauat.edu.cn

(SAM)算法^[7]、光谱信息散度(SID)算法^[8]、光谱相似度匹配(SCF)算法^[9]等。由于壁画中相同色系不同颜料的物质的光谱反射率特性比较接近,因此上述算法在同色系不同物质光谱匹配时容易出现误匹配的情况。针对这一情况,本文将编辑距离(ED)算法引入到光谱匹配中,利用编辑距离对曲线数据差别更加敏感的优势,实现光谱反射率曲线数据的高精度匹配,从而来识别同色系不同颜料的物质;并针对颜料光谱反射率曲线在不同采集条件和不同浓度情况下仅幅值改变而形状不变的特点,改进编辑距离的匹配判定条件,通过自适应设定的差值阈值来减小匹配误差,进一步提高ED算法在壁画颜料物质光谱匹配识别时的精度。

2 光谱匹配方法

2.1 SAM 算法

SAM法又称为光谱角度填图法^[6]。设有2个 n 波段的光谱矢量 $\mathbf{T}=(t_1, t_2, \dots, t_n)$ 和 $\mathbf{R}=(r_1, r_2, \dots, r_n)$, \mathbf{T} 与 \mathbf{R} 之间的广义夹角 θ 定义为

$$\theta = \arccos \frac{\sum_{i=1}^n t_i r_i}{\sqrt{\sum_{i=1}^n t_i^2} \sqrt{\sum_{i=1}^n r_i^2}}, \theta \in \left[0, \frac{\pi}{2}\right]. \quad (1)$$

θ 的值越接近于0,说明2条光谱曲线的相似性越大。

2.2 SID 算法

SID法是一种基于信息论来衡量2条光谱之间差异的光谱识别方法。 x 和 y 这2条光谱的概率向量分别为 $\mathbf{a}=(a_1, a_2, \dots, a_N)$ 和 $\mathbf{b}=(b_1, b_2, \dots, b_N)$,其中 $a_i = x_i / \sum_{i=1}^N x_i$, $b_i = y_i / \sum_{i=1}^N y_i$, N 为光谱波长范围。由信息理论可知, y 关于 x 的相对熵为 $D(x \parallel y)$, x 关于 y 的相对熵为 $D(y \parallel x)$, x 和 y 的光谱信息散度为

$$\alpha(x, y) = D(x \parallel y) + D(y \parallel x). \quad (2)$$

$\alpha(x, y)$ 的值越小,则说明2条光谱曲线越匹配。

2.3 SCF 算法

SCF法是利用相关系数进行测度的方法,相关系数 k 定义为

$$k = \frac{\delta_{oi}^2}{\delta_{oo}\delta_{ii}} = \frac{\sum_{k=1}^n [x_o(k) - \bar{x}_o] [x_i(k) - \bar{x}_i]}{\sqrt{\sum_{k=1}^n [x_o(k) - \bar{x}_o]^2} \sqrt{\sum_{k=1}^n [x_i(k) - \bar{x}_i]^2}}, \quad (3)$$

式中: \bar{x}_o 、 \bar{x}_i 为平均光谱; δ_{oi} 为协方差; δ_{oo} 、 δ_{ii} 为标准差。相关系数的值在 $(-1, 1)$ 内,越接近于1说明2种光谱越相似。

通过随后的实验验证可知,上述3种算法在对同色系不同颜料物质的光谱进行匹配时,准确率较低,容易出现误匹配的情况,因此需要利用一种精度更高的算法进行光谱匹配。

3 ED 算法

ED算法最初应用在字符串相似性衡量方面,通过计算2个字符串之间转化所需的最少的编辑操作数量,来衡量2个字符串之间的相似性^[10]。

编辑距离的核心是利用动态规划来求解最优问题,用合适的递归方程把距离定义为代价函数。代价函数与相应的基本操作成本相关,每种基本操作的代价都是1。一般情况下,各基本操作的代价可以不同^[11]。假设 $\delta_{del}(a)$ 、 $\delta_{ins}(a)$ 和 $\delta_{subs}(a, b)$ 分别为删除字符 a 、插入字符 a 和用字符 b 代替 a 相应的代价,则在代价都为1的简单情况下,代价方程为

$$\begin{cases} \delta_{ins}(a) = 1 \\ \delta_{del}(a) = 1 \\ \delta_{subs}(a, b) = 1, a \neq b \\ \delta_{subs}(a, b) = 0, a = b \end{cases} \quad (4)$$

$C(i, j)$ 代表将 $A[1, \dots, i]$ 转变为 $B[1, \dots, j]$ 所需的代价。则 $C(i, j)$ 可以递归表示为

$$\begin{cases} C(0, 0) = 0 \\ C(i, 0) = C(i-1, 0) + \delta_{del}(A[i]), i > 0 \\ C(0, j) = C(0, j-1) + \delta_{ins}(B[j]), j > 0 \\ C(i, j) = \begin{cases} C(i-1, j-1) + \delta_{subs}(A[i], B[j]) \\ C(i-1, j) + \delta_{del}(A[i]) \\ C(i, j-1) + \delta_{ins}(B[j]) \end{cases}, i > 0, j > 0 \end{cases} \quad (5)$$

基于代价方程,利用矩阵等辅助数据结构,使用动态规划算法建立自底向上的求解过程。计算长度为 n 和 n' 的 2 个字符串之间编辑距离的时间复杂度为 $O(nn')$ 。对于 2 个字符串 φ_m 和 φ_n ,构造 $(m+2) \times (n+2)$ 型匹配关系矩阵 \mathbf{D} 来计算字符串之间的编辑距离。假设 $m=n=3$,则矩阵 \mathbf{D} 如图 1 所示。

		φ_1	φ_2	φ_3
	0	1	2	3
φ_1	1			
φ_2	2			
φ_3	3			

图 1 匹配关系矩阵 \mathbf{D} Fig. 1 Matching relationship matrix \mathbf{D}

利用表达式 d_{ij} 填充矩阵 \mathbf{D} 。 d_{ij} 的表达式为

$$d_{ij} = \begin{cases} i, & j = 1 \\ j, & i = 1 \\ \min\{d_{i-1,j-1}, d_{i-1,j}, d_{i,j-1}\} + a_{ij}, & i, j > 1 \end{cases}, \quad (6)$$

其中

$$a_{ij} =$$

$$\begin{cases} 0, & \phi_i = \varphi_j \\ 1, & \phi_i \neq \varphi_j \end{cases} \quad i = 2, \dots, m+1; j = 2, \dots, n+1. \quad (7)$$

匹配关系矩阵右下角的元素 $d_{m+2,n+2}$ 为字符串 ϕ_m 与 φ_n 之间的编辑距离。从(7)式中可以看出,编辑距离的判定条件是当 2 个字符差值为 0 时返回 0,差值不为 0 时返回 1,实际上相当于对差值设置了门限为 0 的阈值,即零阈值。这个判定条件使 ED 算法对于不同形状的光谱反射率曲线之间的差别更敏感,匹配区分精度更高。但是如果进行匹配的 2 组数据存在微小的幅值差距,ED 算法的判定条件就会将这个差距放大,给出低相似度的匹配率。在实际情况下,不同采集条件下得到的同种壁画颜料的光谱数据会有误差,不同浓度的同种颜料的光谱反射率曲线的幅值也有差别。图 2 所示为相同颜料在 3 组不同采集条件下的光谱反射率 R_1 、 R_2 和 R_3 ,光谱数据分辨率为 400~900 nm 时的采样个数为 662。

利用 ED 算法对图 2 中 3 组相同颜料的光谱反射率曲线进行匹配实验,选取 R_1 为样本数据, R_2 、 R_3 为待测数据, e_d 为 2 条光谱反射率曲线之间的

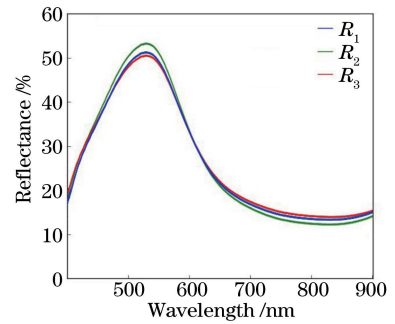


图 2 不同采集条件下同种颜料的光谱反射率曲线
Fig. 2 Spectral reflectance curves of same pigment under different collection conditions

编辑距离。匹配结果如表 1 所示。由于待测数据和样本数据都是同一物质的光谱反射率曲线,因此理想的编辑距离应接近 0。由表 1 可知, R_1 与 R_2 的编辑距离为 186, R_1 与 R_3 的编辑距离为 119,与实际不符。ED 算法的判定条件为:2 组数据相同时,判定操作数为 0;2 组数据不同时,操作数判定为 1。该判定条件对幅值的变化非常敏感,因此,ED 算法对幅值差别判定敏感的特性会导致其对一种颜料物质在不同状态下的光谱曲线匹配产生误判定的结果。针对这一问题,需要对 ED 算法进行改进,以提高其匹配准确率。

表 1 ED 算法的匹配结果

Table 1 Matching results of ED algorithm

Sample data	Test data	Sampling point number	e_d
R_1	R_2	662	186
R_1	R_3	662	119

4 自适应编辑距离(ATED)算法

ED 算法对不同状态下的同种颜料物质的光谱曲线进行匹配时会产生误判定,针对这一情况,本课题组对 ED 算法进行改进:1) 对光谱反射率数据进行归一化,将反射率数据的幅值全部限定在 $[0,1]$ 区间,在不改变反射率曲线形状的情况下,将幅值归一到相同的区间内,以方便后续设置编辑距离的阈值。2) 利用自适应差值阈值替代零阈值作为新的判定条件,控制 ED 算法对曲线幅值的敏感反应程度。

4.1 光谱数据归一化

在获得光谱反射率数据后,对其进行归一化处理,使数据值都集中在 $[0,1]$ 区间。对光谱向量 \mathbf{M} 进行归一化,得到向量 \mathbf{S} ,其表达式为

$$\mathbf{S}(i) = \frac{\mathbf{M}(i) - \min(\mathbf{M})}{\max(\mathbf{M}) - \min(\mathbf{M})}, \quad (8)$$

式中: $\min(\mathbf{M})$ 与 $\max(\mathbf{M})$ 分别为向量 \mathbf{M} 的最小值

与最大值。

4.2 自适应阈值设定

通过设定自适应阈值的方式改变编辑距离的

原始判定条件,控制 ED 算法在不同匹配情况下的敏感度。增加自适应阈值后的编辑距离判定条件为

$$a_{ij} = \begin{cases} 0, & |\phi_i - \varphi_j| < T \\ 1, & |\phi_i - \varphi_j| > T \end{cases}, i = 2, \dots, m + 1; j = 2, \dots, n + 1, \quad (9)$$

式中: T 为阈值,自适应选取。自适应阈值的方法为类别方差法^[12],将 2 组光谱反射率数据的差值取绝对值,即: $|\phi_i - \varphi_j| = \{X_1, X_2, \dots, X_K\}$, K 为 $m \times n$, X_K 为 2 组光谱反射率的差值。通过阈值 T 可以将曲线分为 2 组不同的数据:其中的一组表示光谱反射率数据一致的情况,即

$$C_0 = \{X_1, X_2, \dots, X_t\}, X_i \leq T, \quad (10)$$

集合 C_0 中的 X 为小于阈值 T 的差值;另外一组表示光谱反射数据不一致的情况,即

$$C_1 = \{X_{t+1}, X_{t+2}, \dots, X_K\}, X_i > T, \quad (11)$$

集合 C_1 中的 X 为大于阈值的差值。 X_K 出现的概率为 $p_i = 1/N$ 。

C_0 、 C_1 类出现的概率 w_0 和 w_1 分别为

$$w_0 = \sum_{i=1}^t p_i, \quad (12)$$

$$w_1 = \sum_{i=t+1}^K p_i. \quad (13)$$

C_0 、 C_1 类出现的均值 μ_0 和 μ_1 分别为

$$\mu_0 = \sum_{i=1}^t x_i p_i, \quad (14)$$

$$\mu_1 = \sum_{i=t+1}^K x_i p_i. \quad (15)$$

C_0 、 C_1 类出现的方差 σ_0 和 σ_1 分别为

$$\sigma_0^2 = \sum_{i=1}^t (X_i - \mu_0)^2, \quad (16)$$

$$\sigma_1^2 = \sum_{i=t+1}^K (X_i - \mu_1)^2. \quad (17)$$

由此可以计算出 C_0 、 C_1 类的类内方差 σ_w^2 、类间方差 σ_b^2 、总方差 σ_t^2 分别为

$$\sigma_w^2 = w_0 \sigma_0^2 + w_1 \sigma_1^2, \quad (18)$$

$$\sigma_b^2 = w_0 w_1 (\mu_1 - \mu_2)^2, \quad (19)$$

$$\sigma_t^2 = \sigma_b^2 + \sigma_w^2. \quad (20)$$

当 σ_b^2/σ_t^2 最大时,对应的 T 值就是最佳阈值。

ATED 算法输出的结果与光谱反射率曲线采样点的个数相关,最终的值越接近 0,说明 2 条曲线越相似。当 2 条曲线完全不相似时,返回的值为这 2 条曲线元素的个数。因此,ATED 算法的匹配精度转换为百分比后的公式为

$$\frac{L - L_d}{L} \times 100\%, \quad (21)$$

式中: L 为 2 条光谱反射率曲线的长度,即元素个数; L_d 为 2 条光谱反射率曲线间的编辑距离。

为了验证 ATED 算法匹配的准确率,分别利用 ED 算法和 ATED 算法对图 2 中的光谱数据进行匹配率计算,光谱数据的分辨率为 400~900 nm 时采样个数为 662,匹配率结果如表 2 所示。由表 2 可知,利用 ATED 算法对图 2 中的光谱反射率数据进行匹配,匹配率比 ED 算法的匹配率明显提高。这说明,ATED 算法可以较精确地对同种颜料物质在不同状态下的光谱数据进行匹配和识别。

表 2 ED 算法和 ATED 算法的匹配结果

Table 2 Matching results using ED algorithm and ATED algorithm

Sample data	Algorithm	Test data	Sampling point number	e_d	Matching rate / %
R_1	ED	R_2	662	186	71.98
R_1		R_3		119	82.15
R_1	ATED	R_2	662	43	93.51
R_1		R_3		22	96.67

考虑到不同的光谱分辨率可能会对匹配率精度造成影响,同时选取反射率 R_1 、 R_2 、 R_3 在另外 4 种采样精度时的数据来验证 ATED 算法。4 种采样精度如下:光谱数据的分辨率为 400~900 nm 时,采样个数分别为 331、166、83 和 42。验证结果如表

3 所示。由表 3 可知,在不同的光谱采样分辨率下,ATED 算法的匹配结果差别比较小,随着采样分辨率降低,匹配率精度会稍微下降,能够在不同采样分辨率尺度下保证匹配的精度。但是,由于计算过程中存在进位精度误差,根据(21)式可知,过小的

表 3 不同光谱采样分辨率下 ATED 算法的匹配精度

Table 3 Matching accuracy of ATED algorithm at different spectral resolutions

Sample data	Test data	Sampling point number	e_d	Matching rate /%
R_1	R_2	331	22	93.35
R_1	R_3		11	96.67
R_1	R_2	166	11	93.37
R_1	R_3		6	96.39
R_1	R_2	83	6	92.77
R_1	R_3		3	96.39
R_1	R_2	42	3	92.86
R_1	R_3		2	95.24

采样分辨率会导致计算中编辑距离的微小误差,进而将产生较大的匹配精度偏移。因此,在实验条件和采集设备允许的情况下,应尽量使用高精度的光谱采样分辨率来计算 ATED 算法的匹配率。

5 结果分析及应用验证

5.1 实验数据

从颜料标准样本库中选取 15 种颜料进行测试,分别为白蛤 (clammeal)、大红 (dahong)、花青

(indigo)、石绿 (malachite)、石青 (azurite)、曙红 (blush)、藤黄 (gamboge)、天蓝 (azure)、胭脂 (madder)、赭石 (hematite)、朱膘 (cinnabar)、铅丹 (minium)、群青 (ultramarine)、土红 (laterite)、银朱 (vermilion)。具体颜料如图 3 所示。

采用分光光度计 SpectroSuite 测量每种颜料的光谱反射率数据,测量范围为 400~900 nm,包括可见光至近红外范围。15 种颜料的光谱反射率曲线如图 4 所示。



图 3 颜料样本示例

Fig. 3 Pigment samples

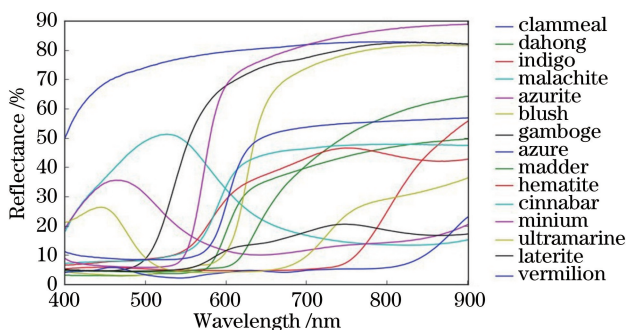


图 4 15 种颜料的光谱反射率曲线

Fig. 4 Spectral reflectance curves of fifteen kinds of pigments

5.2 算法准确率对比

光谱匹配算法的输出值并不统一,值域范围也不相同。为了更直观地比较各类算法的匹配结果,将 SAM、SCF、SID、ATED 算法的输出结果统一转化为百分数,以反映待测颜料与库中标准颜料的匹配率,值越接近 100%,表示 2 种颜色是同一种物质的可能性越大。

SAM 算法输出的区间为 $[0, \pi/2]$,将其输出转换为百分比的公式为

$$\left(1 - \frac{\theta}{\pi/2}\right) \times 100\%。 \quad (22)$$

用 1 减去 SAM 算法计算出的 2 条光谱反射率曲线的广义夹角 θ 与输出值中的最大值 $\pi/2$ 的商,再乘以 100%,就可以得到最终的相似度。

SCF 算法输出的区间为 $(-1, 1)$,将其输出转换为百分比的公式为

$$k + 1/2 \times 100\%。 \quad (23)$$

先将整个输出的区间向右平移 1 个单位,同时把相关系数加 1,再对这 2 个值求商,得到的值乘以

100%,就可以得到最终的相似度。

SID 算法最终输出的结果没有明确的范围,通过对已有的全部颜料的光谱反射率数据进行两两匹配可以看出,SID 算法的取值都不会超过 1。为了统一最终的输出结果,将 SID 算法输出的最大值设置为 1,最终结果是越接近于 0 证明越相似,则转换为百分比的公式为

$$(1 - \alpha) \times 100\%。 \quad (24)$$

使用 4 种算法进行对比实验,选择的标准样本为标准颜料库中已有的石青、石绿、铅丹、白蛤、大红、花青、曙红、藤黄、天蓝、胭脂、赭石、朱膘 12 种颜料,进行同种颜料间的匹配,结果如表 4 所示。由表 4 可知;在对同种颜料进行光谱匹配时,SAM 算法的匹配率为 94%~97%,SCF 算法的匹配率为 95%~98%,SID 算法的匹配率为 96%~98%,ATED 算法的匹配率为 98%~100%;在每组颜料匹配的对比中,ATED 算法的匹配率均高于其他 3 种算法的匹配率。由此可知,ATED 算法匹配结果的精度比 SAM、SCF、SID 算法匹配结果的精度更高。

表 4 4 种算法光谱匹配结果

Table 4 Spectral matching results of four kinds of algorithms

Sample pigment	Test pigment	Matching ratio / %			
		SAM	SCF	SID	ATED
Malachite	Malachite	94.98	97.63	97.78	99.13
Azurite	Azurite	96.87	97.74	97.95	99.36
Minium	Minium	95.38	95.35	96.85	100.00
Clammeal	Clammeal	97.33	96.43	96.38	99.15
Dahong	Dahong	95.36	95.34	98.24	98.95
Indigo	Indigo	96.35	98.46	96.13	99.33
Blush	Blush	95.11	97.58	97.34	98.86
Gamboge	Gamboge	96.35	96.37	96.37	99.26
Azure	Azure	97.51	97.56	97.28	98.67
Madder	Madder	96.23	96.23	96.74	100.00
Hematite	Hematite	95.24	96.34	98.16	99.84
Cinnabar	Cinnabar	97.43	97.56	97.68	98.17

5.3 同色系颜料匹配识别验证

中国古代颜料中的红色颜料有大红、铅丹、曙红、胭脂、赭石、朱膘、土红、银朱、深红等。古代壁画中常用的多为矿物颜料铅丹、赭石、银朱,这些颜料在颜色上相近,光谱反射率曲线也较相似,多为“S”型曲线,如图 5 所示。

对同色系颜料的近似光谱曲线进行匹配时,为有效区分识别不同的颜料物质,需要使用高精度的光谱匹配算法。为了验证本文提出的 ATED 算法识别同色系颜料光谱反射率曲线的精度,选取 6 种红色颜料——大红、铅丹、曙红、胭脂、赭石、朱膘、土

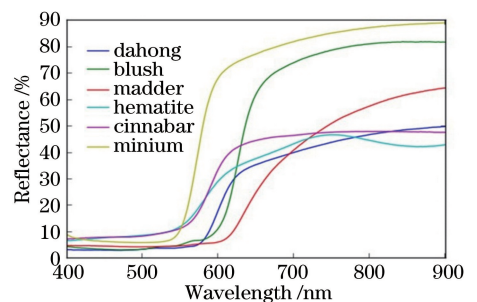


图 5 不同红色颜料的光谱反射率

Fig. 5 Spectral reflectance of different red pigments
红进行实验,分别以大红、赭石、曙红作为待测样本,匹配结果分别如表 5~7 所示。

表 5 待测颜料大红的匹配结果

Table 5 Matching results of test pigment dahong

Sample pigment	SAM algorithm		SCF algorithm		SID algorithm		ATED algorithm	
	Matching rate / %	Result	Matching rate / %	Result	Matching rate / %	Result	Matching rate / %	Result
Minium	89.41		94.75		97.27		91.54	
Dahong	92.06		97.62		98.52		100.00	
Blush	92.56	Blush	98.43	Blush	98.39	Dahong	96.98	Dahong
Madder	86.59		94.66		96.04		86.40	
Hematite	88.70		97.37		95.76		96.07	
Hematite	89.41		93.48		96.55		95.02	

由表 5 可知:当待测样本颜料为大红时,SAM 算法与 SCF 算法的匹配结果均为颜料曙红,即认为标准颜料库中的大红颜料与待测的曙红颜料的反射率更相

似,出现了误匹配的现象;SID 算法与 ATED 算法的匹配结果与真实情况一致,ATED 的匹配率为 100%,认定为完全匹配,SID 算法的匹配率为 98.52%。

表 6 待测颜料赭石的匹配结果

Table 6 Matching results of test pigment hematite

Sample pigment	SAM algorithm		SCF algorithm		SID algorithm		ATED algorithm	
	Matching rate / %	Result	Matching rate / %	Result	Matching rate / %	Result	Matching rate / %	Result
Minium	92.79		96.34		93.64		94.34	
Dahong	88.16		97.24		95.12		95.15	
Blush	83.80	Cinnabar	95.17	Hematite	94.31	Madder	96.21	Hematite
Madder	86.75		94.58		97.95		94.36	
Hematite	94.05		98.55		92.35		99.18	
Cinnabar	96.13		96.12		95.45		96.38	

由表 6 可知:当待测样本颜料为赭石时,SAM 算法的匹配结果为朱膘,SID 算法的匹配结果为胭脂,出现了误匹配;SCF 算法与 ATED 算法的匹配

结果与真实情况一致,并且 ATED 算法的匹配精度高于 SCF 算法的匹配精度,ATED 的匹配率为 99.18%,SCF 的匹配率为 98.55%。

表 7 待测颜料曙红的匹配结果

Table 7 Matching results of test pigment blush

Sample pigment	SAM algorithm		SCF algorithm		SID algorithm		ATED algorithm	
	Matching rate / %	Result	Matching rate / %	Result	Matching rate / %	Result	Matching rate / %	Result
Minium	93.26		93.82		93.78		96.42	
Dahong	95.42		95.43		98.14		98.31	
Blush	92.18	Dahong	96.13	Cinnabar	96.45	Dahong	99.75	Blush
Madder	93.33		96.07		95.23		97.23	
Hematite	90.24		95.39		93.28		95.13	
Cinnabar	91.12		97.48		97.35		96.14	

由表 7 可知:当待测样本颜料为曙红时,SAM 算法及 SID 算法的匹配结果为大红,SCF 算法的匹配结果为朱膘,均出现了误匹配;ATED 算法给出了正确的识别结果,且匹配率达到了 99.75%。

通过上述 3 组实验可以看出,在处理同色系相似颜料光谱匹配问题时,SAM、SCF、SID 算法都存在误匹配现象,而且在匹配正确时,ATED 算法的匹配率比其他 3 种算法的匹配率更高,识别结果更

精确。同时,ATED 算法针对同种颜料在不同浓度和采集条件下光谱反射率曲线形状一致、幅值有差别的情况下也可以得到较好的匹配识别结果。

6 结 论

本课题组研究了基于自适应阈值编辑距离的光谱匹配算法,基于 ED 算法的基本理论,结合壁画颜料光谱数据的特点,利用自适应设定数据差值的阈

值方式对 ED 算法的判定条件进行改进,改进的判定条件能够克服基于编辑距离的光谱匹配算法对光谱数据幅值过于敏感的缺点。实验验证的结果表明,基于自适应阈值 ED 算法的光谱匹配率比传统 SAM、SCF、SID 算法的光谱匹配率有所提高,在壁画颜料的检测识别应用中取得了更好的识别效果。

参 考 文 献

- [1] Du X P, Liu H, Chen H, *et al.* Research of satellite shape inversion matching algorithm based on photometric characteristic [J]. *Acta Optica Sinica*, 2016, 36(8): 0820001.
杜小平, 刘浩, 陈杭, 等. 基于光度特性的卫星形状反演匹配算法研究 [J]. *光学学报*, 2016, 36(8): 0820001.
- [2] Li W J, Wang T Y, Zhou Y, *et al.* Terahertz non-destructive inspection of air defect within adhesive layers of multi-layer bonded structure [J]. *Acta Optica Sinica*, 2017, 37(1): 0111002.
李文军, 王天一, 周宇, 等. 多层胶接结构胶层空气缺陷的太赫兹无损检测 [J]. *光学学报*, 2017, 37(1): 0111002.
- [3] Yang S S, Mi L, Zhu R, *et al.* Non-invasive investigation of microstructure of Ding kiln porcelain with optical coherence tomography [J]. *Chinese Journal of Lasers*, 2016, 43(2): 0208002.
杨珊珊, 米磊, 朱锐, 等. 光学相干断层成像技术在定窑瓷器检测中的应用 [J]. *中国激光*, 2016, 43(2): 0208002.
- [4] Zheng J W, Yang T W. Classification method of biological tissues based on Raman spectrum features [J]. *Laser & Optoelectronics Progress*, 2017, 54(5): 053001.
郑家文, 杨唐文. 基于拉曼光谱特征的生物组织识别方法 [J]. *激光与光电子学进展*, 2017, 54(5): 053001.
- [5] Gong M T, Feng P L. Preliminary study on the application of hyperspectral imaging in the classification of and identification Chinese traditional pigments classification: a case study of spectral angle mapper [J]. *Sciences of Conservation and Archaeology*, 2014, 26(4): 76-83.
巩梦婷, 冯萍莉. 高光谱成像技术在中国画颜料分类和识别上的应用初探: 以光谱角填图 (SAM) 为例 [J]. *文物保护与考古科学*, 2014, 26(4): 76-83.
- [6] Xu Z, He M X, Li P F, *et al.* Spectral matching algorithm applied to identify the terahertz spectrum of *caulis spatholobi* and *caulis sargentodoxae* [J]. *Spectroscopy and Spectral Analysis*, 2017, 37(1): 42-47.
徐哲, 何明霞, 李鹏飞, 等. 光谱匹配算法应用于鸡血藤与大血藤的太赫兹光谱区分 [J]. *光谱学与光谱分析*, 2017, 37(1): 42-47.
- [7] Hu B, He Z W, Zhang J J, *et al.* Research on alteration information extraction with multispectral data [J]. *Geospatial Information*, 2013, 11(4): 104-106.
胡滨, 何政伟, 张佳佳, 等. 多光谱数据矿化蚀变信息提取研究 [J]. *地理空间信息*, 2013, 11(4): 104-106.
- [8] Wang W C, Wang H Q, Wang K, *et al.* An identification method of spectral information divergence pigment based on statistical manifold [J]. *Laser & Optoelectronics Progress*, 2018, 55(1): 013002.
王伟超, 王慧琴, 王可, 等. 一种基于统计流形的光谱信息散度颜料识别方法 [J]. *激光与光电子学进展*, 2018, 55(1): 013002.
- [9] Guo T, Hua W S, Liu X, *et al.* Comprehensive evaluation of optical camouflage effect based on hyperspectra [J]. *Laser & Optoelectronics Progress*, 2016, 53(10): 101002.
郭彤, 华文深, 刘恂, 等. 一种基于高光谱的光学伪装效果综合评价方法 [J]. *激光与光电子学进展*, 2016, 53(10): 101002.
- [10] Xue Y W, Shen J Y, Zhang Y. Modified edit distance algorithm and its application in web search [J]. *Journal of Xi'an Jiaotong University*, 2008, 42(12): 1450-1454.
薛晔伟, 沈钧毅, 张云. 一种编辑距离算法及其在网页搜索中的应用 [J]. *西安交通大学学报*, 2008, 42(12): 1450-1454.
- [11] Xu J Y, Wang H Q, Wu M, *et al.* A crack detection algorithm based on editing distance in pictorial carved stone [J]. *Laser & Optoelectronics Progress*, 2017, 54(8): 081006.
许君扬, 王慧琴, 吴萌, 等. 一种基于编辑距离的汉代画像石裂缝的检测算法 [J]. *激光与光电子学进展*, 2017, 54(8): 081006.
- [12] Huang X D, Sun L, Liu S L. A retrieval method of relevance feedback images based on discriminative extreme learning [J]. *Journal of Xi'an Jiaotong University*, 2016, 50(8): 96-102.
黄晓冬, 孙亮, 刘胜蓝. 一种判别极端学习的相关反馈图像检索方法 [J]. *西安交通大学学报*, 2016, 50(8): 96-102.