

文章编号: 1001-0920(2009)12-1917-04

一种特征加权 FCM 算法的图像重建技术研究

肖满生, 吕 勇, 曾 嵘
(湖南工业大学 理学院, 湖南 株洲 412008)

摘 要: 针对传统 FCM(Fuzzy C-means)算法中初始聚类中心选取的随机性以及初始值敏感的问题, 提出一种基于进化策略的色彩空间加权的 FCM 聚类算法. 通过在 RGB(Red Green Blue)色彩空间矢量中设置加权矩阵来补偿各色彩的非均匀性, 并采用一种类内最小距离最大的统计聚类算法来初始化聚类中心. 实验结果表明, 该算法能有效减少颜色量化后的均方差值, 保持重建图像的整体层次和局部特征细节, 对研究图像处理技术有较强的实际意义.

关键词: 特征加权; 色彩量化; 模糊 C 均值; 进化策略; 初始聚类中心

中图分类号: TP391 **文献标识码:** A

Research on image reconstruction technology by FCM of weighted characteristic

XIAO Man-sheng, LV Yong, ZENG Rong

(College of Science, Hu'nan University of Technology, Zhuzhou 412008, China. Correspondent: XIAO Man-sheng, E-mail: xiaomansheng@tom.com)

Abstract: Aiming at a defect on randomness of the initial clustering center choosing and sensitivity of initial value in tradition FCM (fuzzy C-means) algorithm, a clustering algorithm about FCM of weighted color space based on evolutionary strategy is proposed. By interposing weighted matrix in RGB(Red Green Blue) color space, the color's inhomogeneous is compensated. And by using a statistics clustering algorithm of minimal maximal distance, clustering center is initiated. The experimental results show that the algorithm can decrease effectively the mean square deviation of color quantization, keep overall arrangement of ideas and part characteristic detail in image reconstruction, and has practical value to the study of the image process technology.

Key words: Weighted characteristic; Color quantization; FCM; Evolutionary strategy; Initial clustering center

1 引 言

图像重建的关键是色彩量化技术, 即根据人的视觉特性将原图像中的多种颜色归类到较少数目的颜色. 该技术在图像压缩、模糊识别等领域有很强的实际意义. 传统的 FCM(Fuzzy C-means)算法存在初始聚类中心选取的随意性, 算法对噪音数据敏感, 容易收敛到局部极值点, 因而不能达到全局最优^[1,2]. 表现在色彩量化上, 对颜色样本的初始聚类中心选取敏感, 局部颜色着色不正, 特别是色彩分布极不均匀的图像, 其重建效果不够理想, 如“万绿丛中一点红”, 其重要的红色会着色不正或丢失. Liew 等^[3]提出的改进 ASFCM 算法在一定程度上克服了上述缺点, 但该算法的运算量过大, 处理效率不高. 文献^[4]给出了像素色彩模型, 两种色彩在空间的距

离越短, 则它们相互的隶属度值越大, 但由于人的视觉特性对不同颜色的敏感度不一样, 特别是在颜色分布均匀的平滑区域或过渡部分, 其对不同颜色的敏感度差别更大. 因此, 如何根据人的视觉特性对图像各个区域颜色进行适当量化, 使量化后的颜色尽量保留原图像的主体色彩且均匀分布, 以致人眼不易观察到量化前后的图像偏差等问题具有实际应用价值, 但目前还没有一套完整成熟的解决方案.

本文在传统 FCM 算法的基础上, 提出一种基于特征加权 FCM 算法的图像色彩量化方法. 初始聚类中心的选取借鉴了文献^[5]中的类内最小距离最大的聚类算法以代替传统的随机选取方法. 在聚类过程中, 通过对色彩空间加权, 并采用一种基于进化策略的 FCM 聚类权值优化算法, 来解决图像重

收稿日期: 2009-01-21; 修回日期: 2009-03-30.

基金项目: 湖南省教育厅科研项目(08C292).

作者简介: 肖满生(1968—), 男, 湖南邵东人, 副教授, 硕士, 从事数据库和数据挖掘等研究; 吕勇(1966—), 男, 重庆人, 副教授, 从事模糊数学理论等研究.

建后局部重要颜色的失真、整体色彩分布不均匀以及不符合人的视觉特性等问题。

2 像素色彩特征加权

基于RGB(Red Green Blue)模式的像素色彩空间中,每一个像素颜色由3个色彩空间坐标(R, G, B)决定. 设一组颜色样本集 $X = \{x_1, x_2, \dots, x_n\} \subset R^s$, R^s 为色彩空间,其中包含 n 种颜色样本. 每个样本为一个由 R, G, B 三种色彩组成的3维矢量,即 $x_k = (x_{kR}, x_{kG}, x_{kB})^T \in R^s$. 现要求把这 n 种颜色样本量化成 l 种颜色;同时由于人眼对不同色彩的敏感度不同,使得样本中各维色彩对量化颜色的贡献不均匀,因此,采用FCM量化聚类时必须考虑各维色彩权重的不同影响.

假设已知色彩空间的每维权重分别为 w_R, w_G, w_B ,且满足 $w_R + w_G + w_B = 1$,其权重矩阵表示为

$$W = \begin{bmatrix} w_R & 0 & 0 \\ 0 & w_G & 0 \\ 0 & 0 & w_B \end{bmatrix},$$

则加权后的样本为

$$x'_k = Wx_k, k = 1, 2, \dots, n. \quad (1)$$

加权后任意两个颜色样本的欧氏距离可表示为

$$D_E^2(x'_k, x'_j) = \|x'_k - x'_j\|^2 = (x'_k - x'_j)^T(x'_k - x'_j), \forall k, j. \quad (2)$$

将式(1)代入(2),并令 $W = w^T w$,得

$$D_E^2(x'_k, x'_j) = (x_k - x_j)^T W (x_k - x_j) = D_W^2(x_k, x_j), \forall k, j. \quad (3)$$

因为 W 是对角矩阵,由上式可知,加权处理后任意两个颜色样本矢量的欧氏距离等于原始样本的加权欧氏距离,所以,加权处理后的基于欧氏距离的FCM算法就变成了原始颜色样本的基于加权欧氏距离的FCM算法.然而在实际应用中, W 矩阵的权值并不是预先确定的,而应在聚类过程中不断调整优化,以使重建后的图像尽量保持原图像的视觉效果,因此,基于特征加权FCM聚类量化的目标除了寻求最优的模糊聚类颜色划分 U^* 外,还包括如何确定最佳加权矩阵 W^* . 本文提出的基于进化策略的色彩量化方法为上述问题的解决提供了一个较好的方案.

3 基于特征加权FCM聚类的色彩量化方法

3.1 初始聚类中心的选取

为了解决传统FCM算法中随机选取初始聚类中心而造成图像重建后局部重要颜色丢失或偏色等问题,本文采用一种基于类内最小距离最大的方法来选取初始聚类中心.即在3维色彩空间中,把 n 种样本颜色 $X = \{x_1, x_2, \dots, x_n\}$ 量化为 l ($l < n$) 种颜

色.首先在其中指定 p 个样本为初始类的代表点,记为

$$V^0 = \{v_k^0 \mid k = 1, 2, \dots, p, 1 \leq p \leq l\}.$$

这 p 个样本应是 X 集合中频度最高的前 p 种颜色样本.然后在剩余的样本集 $V^1 = X - V^0 = \{v_k^1 \mid k = 1, 2, \dots, n - p\}$ 中依次找出与初始类代表点最小距离最大的样本点作为补充的新类代表点,直到找够 l 个代表点为止,这样FCM算法的初始聚类中心的选取工作就完成了.具体实现如下:

1) 利用下式统计每种样本的颜色直方图频度^[6]:

$$h(i) = \frac{k(i)}{M \times N}, i = 0, 1, \dots, n. \quad (4)$$

其中: $h(i)$ 为样本 x_i 对应的频度, $k(i)$ 表示颜色为 i 的像素点在图像中出现的次数, n 为样本颜色数目, $M \times N$ 为图像尺寸.式(4)满足

$$\sum_{i=0}^{n-1} h(i) = 1.$$

2) 按颜色样本频度选取 p 个初始类代表点.即选取 $h(i)$ 值最大的前 p 个样本点作为初始类代表点构成集合 V^0 ,剩余的集合为 $V^1 = X - V^0$. p 值的大小可根据量化颜色数或图像重建的实际需要确定,一般情况下取 $p = l/2$.

3) 按最小距离最大的准则选取剩下的 $l - p$ 个初始类代表点,即:

① 设 V^0, V^1 中的元素个数分别为 l_0, l_1 ;

② 计算 V^1 中每个样本点 v_i^1 与 V^0 中各样本点 v_j^0 的欧几里德距离 d_{ij} ;

③ 求 V^1 中每个样本点 v_i^1 与 V^0 中所有样本点的最小距离 d_i ,即

$$d_i = \min(d_{ij}), i = 1, 2, \dots, l_1, j = 1, 2, \dots, l_0;$$

④ 求 V^1 中各样本点 v_i^1 与 V^0 中所有样本点的最小距离的最大值 d_k ,即

$$d_k = \max(d_i), i = 1, 2, \dots, l_1, k = 1, 2, \dots, l_1;$$

⑤ 将 V^1 中最大值 d_k 所对应的点加入 V^0 中,即 $V^0 = V^0 \cup \{v_k^1\}$, $l_0 = l_0 + 1, l_1 = l_1 - 1$;

⑥ 重复以上步骤,直到 $l_0 = l$ 为止.

3.2 色彩量化实现算法

基于类内最小距离最大的初始聚类中心选取算法较好地解决了色彩量化中局部颜色的丢失或失真问题,但没有考虑整体颜色层次.为保证图像量化后全局色彩的均匀分布以符合人的视觉特性,本文在上述色彩加权处理的基础上,结合类内最小距离最大算法选取的初始聚类中心,采用一种基于进化策略的FCM优化算法^[7,8]对图像进行量化处理.其具体实现过程如图1所示.

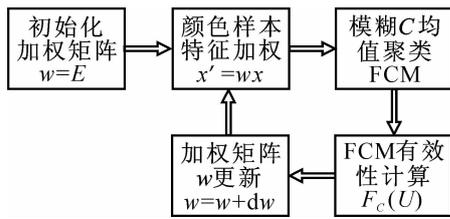


图 1 基于进化策略的特征加权 FCM 算法

算法步骤描述如下：

1) 初始化加权矩阵. 首先假设色彩空间中 R, G, B 各维具有同等的重要性, 其加权矩阵为 $w = E$, 此时的聚类算法即为传统的 FCM 算法.

2) 颜色样本特征加权. 考虑到 R, G, B 各种颜色对人的视觉敏感性的不同, 用给定的加权矩阵对样本集中各样本进行加权, 即 $x'_k = wx_k, k = 1, 2, \dots, n$.

3) 调用模糊 C 均值算法. 即用传统的 FCM 算法对加权后的样本进行聚类, 样本初始聚类中心的选取采用前面所述的类内最小距离最大算法, 目的是使局部重要色彩点(极值点)不失真.

4) 计算步骤 3) 中聚类的有效性. 在用 FCM 算法得到给定条件的最优 l 类色彩加权划分 U_w^* 后, 为了衡量其有效性, 用下式所定义的划分系数^[9]来计算 FCM 聚类的有效性:

$$F(U, l, w) = \frac{1}{n} \sum_{i=1}^l \sum_{k=1}^n (u_{ik})^2 = \frac{\text{Tr}(U, U)}{n} = \frac{\langle U, U \rangle}{n} = \frac{\|U\|^2}{n} \quad (5)$$

显然 $F < 1$. 一般情况下, 模糊划分越分明, 则有效性值 F 越大, 但 F 值不可能等于 1, 因为 $F = 1$ 代表硬划分.

5) 更新加权矩阵. 加权矩阵的更新可用下式中的 Gauss 噪音来扰动:

$$w_i = w_i + N(0, r\delta), i \in \{1, 2, 3\}. \quad (6)$$

其中: r 为用户定义的比例常数; $\delta = 1 - F(U, l, w)$ 为学习率, 其大小由步骤 4) 的有效性确定; $N(0, r\delta)$ 为一个 Gauss 随机变量值.

6) 由步骤 5) 可知, 聚类有效性越大, 则 δ 越小, 更新权值的变化也越小. 当聚类的有效性趋近于 1 时, δ 趋近于 0, 随机变量值 $N(0, r\delta)$ 也趋近于 0, 权值趋于稳定, 算法停止; 否则, 转步骤 2) 继续执行.

上述算法中, 代表颜色敏感性的特征权值是逐步进化迭代确定下来的, 而不是人为规定的, 因此, 当算法收敛时, 权值也达到最佳匹配. 故经过该进化策略的聚类量化后, 图像的颜色分布均匀, 整体感好, 符合人的视觉特性, 同时也保留了局部极值点的颜色不偏色.

4 图像重建实验

为检验上述算法的实际应用效果, 本文采用两幅 256 色图像进行实验, 其中一幅为颜色分布较为均匀的图像, 另一幅为颜色分布极不均匀的图像. 分别对两幅图像进行 128, 64, 32, 16 色量化重建. 图 2 为色彩均匀分布的图像进行 128 色量化重建后的效果图, 图 3 为颜色分布不均匀的图像进行 128 色量化重建后的效果图(限于篇幅, 原图像及量化后的其他图像没在文中给出). 通过实验可知, 量化后的图像能较好地保持原图像色彩, 特别是颜色分布不均匀的图像, 其局部颜色偏色少, 细节突出.

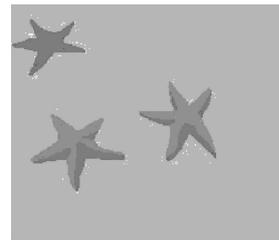


图 2 均匀图像重建效果



图 3 不均匀图像重建效果

另外, 为了定量检验该算法的有效性, 实验中分别采用了由 Yang 和 Filev^[10] 提出的基于山函数的 FCM 算法(简称算法 1)和本文提出的算法(简称算法 2)对上述图像进行了多级量化对比实验. 山函数法是一个有名的模糊聚类算法, 主要用来获得一个良好的初始聚类中心, 从而得到模糊聚类的全局最优解; 另外, 山函数法的计算量会随特征空间的维数呈指数增长, 而本文的图像处理是固定的 3 维特征空间, 故不考虑其计算量.

实验条件为: FCM 算法中模糊加权指数 m 取经验值 2, 其终止条件为两次迭代之后隶属度 $|U^{(t+1)} - U^{(t)}| < 0.01$; R, G, B 色彩初始权值均为 $1/3$; 两图像的初始颜色样本数均为 256; 式(6)中定义的比例常数 r 取为 1. 实验中对量化的均方差值进行了比较分析, 其结果如图 4 和图 5 所示. 其中: 图 4 为实验中第 1 个颜色均匀分布的图像量化后的均方差值对比曲线, 图 5 为第 2 个颜色分布不均匀图像量化后的均方差值对比曲线.

比较图 4 中的曲线可知, 由于图像在色彩空间中分布较为均匀, 本文算法的效果比基于山函数算

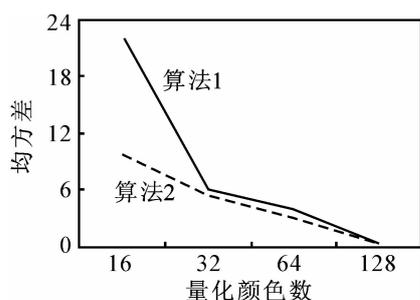


图4 颜色均匀分布图像量化均方差值对比

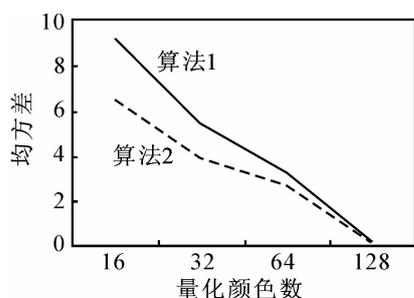


图5 颜色非均匀分布图像量化均方差值对比

法的效果有明显改善,如其在16级量化时,均方差值小1倍多,而且重建的图像差别较小,细节部分也较理想.从图5可以看出,即使在图像颜色分布极不均匀的情况下,该算法也得到了较理想的量化效果;在同样条件下,其均方差值也比山函数法的均方差值小.

综上所述,该算法不仅保留了原图像整体颜色的层次,而且避免了局部细节处颜色的丢失或着色不正的情况,从而较好地解决了图像重建中整体层次与局部细节之间的矛盾.

5 结 论

图像重建可通过对图像中的颜色量化来提取关键的颜色种类,进而组合出新的图像,并保持新旧图像不失真.本文通过对色彩空间加权,采用一种进化策略的FCM聚类算法来完成图像色彩的量化工作,并通过一种基于类内最小距离最大的方法来完成初始聚类中心的选取.该算法较好地解决了图像中某些关键局部细节颜色以及全局图像的整体色彩层次的均匀化等问题.但是,在量化过程中还有很多问题没有得到很好解决.如初始聚类中心选取中,初始中心的个数 p 值为多少最合适;FCM中模糊指数

m 值的取值为多大等.这些都是人为凭经验确定的,并没有相关的理论依据.这些问题是作者下一步的研究目标.

参考文献(References)

- [1] 张敏,于剑.基于划分的模糊聚类算法[J].软件学报,2004,15(6):858-868.
(Zhang M, Yu J, Fuzzy partitioned clustering algorithms [J]. J of Software, 2004, 15(6): 858-868.)
- [2] Kim D W, Lee K H, Lee D. On cluster validity index for estimation of the optimal number of fuzzy clusters [J]. Pattern Recognition, 2004, 37(10): 2009-2025.
- [3] Liew A W C, Yan H, Law N F. Image segmentation based on adaptive cluster prototype estimation [J]. IEEE Trans on Fuzzy Systems, 2005, 13(4): 444-453.
- [4] 肖满生,于训全,周丽娟.基于多维向量模型模糊聚类的图像识别研究[J].计算机工程与设计,2008,29(15):4001-4002.
(Xiao M S, Yu X Q, Zhou L J. The research on image recognition based on fuzzy clustering for multi-dimensional vector model [J]. Computer Engineering and Design, 2008, 29(15): 4001-4002.)
- [5] 凌卫新.聚类分析在彩色图像量化中的应用[J].计算机工程与应用,2002,12(1):88-89.
(Ling W X. Application of clustering analysis in color image quantization [J]. Computer Engineering and Application, 2002, 12(1): 88-89.)
- [6] 高新波.模糊聚类分析及其应用[M].西安:西安电子科技大学出版社,2004,92-108,158-160.
(Gao X B. Fuzzy clustering analysis and application [M]. Xi'an: Xidian University Press, 2004: 92-108, 158-160.)
- [7] Angeline P J, Saunders G M, Pollack J B. A evolutionary algorithm that constructs recurrent neural network [J]. IEEE Trans on NN, 1994, 5(1): 54-65.
- [8] Babu G P, Murty M N. Clustering with evolution strategies [J]. Pattern Recognition, 1994, 2(27): 321-329.
- [9] Bezdek J C. Clustering validity with fuzzy sets [J]. J of Mathematical Biology, 1974, 1(1): 57-71.
- [10] Yang R R, Filev D P. Approximate clustering via the mountain method [J]. IEEE SMC, 1994, 24(8): 1279-1284.