

服务于区域光伏预测的天空图像 K-means 云空辨识模型

苏 适¹, 李红萍², 严玉廷¹, 陆 海¹, 王 飞^{2,3},
任 惠², N. A. Engerer⁴, 米增强²

(1. 云南电网有限责任公司电力科学研究院, 云南 昆明 650217; 2. 新能源电力系统国家重点实验室, 河北 保定 071003, 3. 美国伊利诺伊大学厄巴纳-香槟分校, 伊利诺伊州 厄巴纳 61801, 美国;
4. 澳大利亚国立大学环境与社会学院, 堪培拉 2601, 澳大利亚)

摘要: 地基天空图像的云空辨识及云团预测是研究区域电网光伏发电功率分布与变化的前提, 对支撑调度提高光伏发电消纳比例具有重要意义。首先分别采用较高的红蓝分量比值和较低的红蓝分量比值作为固定阈值分割地基天空图像, 依次提取辨识结果中的天空像素点和云像素点的位置信息并获取原图像中对应位置的天空像素点和云像素点的 RGB 值; 其次对获得的天空像素和云像素求均值并将各自均值中的红蓝分量相除获取初始聚类中心; 然后使用 K-means 算法, 利用加权欧氏距离计算每一个聚类样本与聚类中心之间的距离, 通过数次迭代得到聚类结果, 进而将聚类结果还原成矩阵得到地基天空图像的云空辨识结果图; 最后利用云南某光伏电站全天空成像仪 TSI-VIS-J1006 采集的天空图像进行仿真, 结果表明该方法较固定阈值法的收敛速度更快、聚类精度更高, 能够有效实现地基天空图像的云空辨识。

关键词: 初始聚类中心; 天空图像; 云空识别; 区域光伏预测

中图分类号: TM615 **文献标识码:** A **文章编号:** 1007-2691 (2017) 06-0061-08

Cloud Identification Model Based on K-means for Regional Photovoltaic Forecasting

SU Shi¹, LI Hongping², YAN Yuting¹, LU Hai¹, WANG Fei^{2,3},
REN Hui², N. A. Engerer⁴, MI Zengqiang²

(1. Electric Power Research Institute of Yunnan Power Grid Co., Ltd Kunming 650217, China;

2. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources,
North China Electric Power University, Baoding 071003, China;

3. University of Illinois at Urbana-Champaign, Urbana 61801, USA;

4. Fenner School of Environment and Society, The Australian National University, Canberra ACT 2601, Australia)

Abstract: Cloud space identification of ground sky image and cloud prediction are the premises of studying the distribution and change of photovoltaic power in regional power grid. They are important for supporting scheduling and increasing the proportion of consumptive PV power generation. Firstly, the high and the low ratio of red and blue compo-

收稿日期: 2016-09-29.

基金项目: 国家自然科学基金资助项目 (51577067, 51277075); 北京市自然科学基金资助项目 (3162033); 河北省自然科学基金资助项目 (E2015502060); 河北省科技支撑计划重点项目 (12213913D); 云南省新能源重大科技专项 (2013ZB005); 新能源电力系统国家重点实验室开放课题 (LAPS15009, LAPS16007); 中央高校基本科研业务费重点项目 (2014ZD29); 云南电网有限责任公司科技项目 (YNKJQQ00000280).

nents were adopted respectively as fixed thresholds to segment ground sky images. We sequentially extracted the location of the sky and the cloud pixels in the identification results and got the RGB value of the corresponding cloud and the sky pixels in the original image. And then we calculate the average of the sky and cloud pixels that were obtained and make mean value of the red and blue components divided to get the initial cluster center. Then basing on the K-means algorithm and Weighted Euclidean distance, we calculate the distance between each cluster sample and cluster center. We get the clustering results by several iterations. Then the clustering results are reverted to a matrix so as to obtain the cloud space identification results of ground sky images. Finally, we use the sky images collected by the total sky imager TSI-VIS-J1006 of a PV power station in Yunnan to carry out simulation. And the results show that comparing with the fixed threshold method, this method is faster and its clustering accuracy is higher, and it can effectively realize the cloud space identification of the ground sky image.

Key words: initial cluster center; sky image; cloud identification; regional PV power forecasting

0 引言

气候问题、资源短缺使得充分利用可再生能源已经成为全球共识。以太阳能、风能为代表的大规模可再生能源并网发电已经成为新型电力系统不可阻挡的发展趋势,对电力系统深层次的影响正在凸显^[1]。然而光伏发电由于受太阳辐照度、环境温度等不确定因素的影响导致其出力具有间歇性和随机波动性。为降低光伏发电并网对电网带来的冲击,提高太阳能的利用率,准确的功率预测是首要任务。目前,光伏发电功率预测方法主要根据历史气象要素数据和光伏电站输出功率数据进行统计分析或神经网络方法进行预测,如基于人工神经网络的预测模型^[2]和基于支持向量机的预测模型^[3]。上述方法在晴空条件下功率预测误差较小、精度较高。但在多云天气下由于未涉及云团移动造成的辐照度变化导致在多云天气下光伏发电预测精度有限。为解决上述问题,文献[4]通过云团分类识别与云团运动预测环节预测未来云团运动图像预测水平面总辐射值。文献[5]利用数值天气预报和地基云图相结合,根据预测时间的范围采用不同的云提取模型和辐射衰减模型。由于云团形变在短时间内变化缓慢,未来时刻的运动可通过云团的跟踪学习对云团移动速度和移动方向进行预测从而得到未来时刻此云团的位置,计算出云团未来时刻对辐照度的影响。上述两种方法由于涉及云团提取、运动预测和云团遮挡提高了光伏发电功率预测的准确性和可靠性。由于地基观测设备拍摄的天空图像具有观测点灵活、易操作、方便快捷和成本较低等优势,因此云提取多采用地基天空图像。现阶段国内关于地基天空图像云团提取的研究尚处于起步阶段。

天空图像的云空辨识涉及一种图像分割技术,阈值分割技术是最经典和流行的图像分割方法之一^[6]。文献[7]对 RGB 图像进行处理选取红蓝波段比值 0.77 为分割阈值,即大于 0.77 为云,小于 0.77 为天空,获得较好的云检测效果。但是,固定阈值法由于所选阈值的单一性不能适用于所有天空图像。为了克服这一缺点,文献[8]提出了一种自适应算法计算分割阈值。该算法选取使类间方差最大的最优阈值分割天空图像取得了较好的云空辨识效果。但是此算法处理少云、多云类型的天空图像时因云空像素点数量的不平衡导致辨识结果误差很大。同时,对于云空辨识结果分析通常是通过人眼对比,始终缺乏一个通用的评价标准,这将给天空图像的云空辨识带来一个巨大的挑战。

本文在阈值法的基础使用 K-均值聚类算法进行云空辨识。K-均值聚类算法,又称为 K-means 算法,是一种无监督模式识别学习方法,初始聚类中心依据相似性和相邻性构造分类器,从而将给定的数据对象分割为若干不同的类^[9]。该算法对初始聚类中心具有很强的依赖性。文献[10,11]提出了初始聚类中心的选取方法。通过对图像中云和天空像素点的颜色特征的提取得到初始聚类中心,同时引入云图像素归一化红蓝比值^[12],利用 K-means 算法实现对天空图像的云空辨识。将此方法应用于云南地区光伏电站中的地基设备 TSI-VIS-J1006 采集的天空图像,结果表明该辨识方法能够较准确的辨识各种类型的天空图像,具有很高的应用价值。

1 天空图像采集

近年来,电荷耦合器件(Charge Coupled De-

vice, CCD)等硬件技术的不断发展,很多地基遥感测云仪器研制成功,其中以全天空成像仪(Total Sky Imager, TSI)为代表。TSI 具有成本低、易集成、易维护性的优点,因此在大多数地基天空图像采集系统中选均用 TSI 作为对云进行观测的硬件设备。本文使用型号为 TSI-VIS-J1006 的天空成像仪采集实时天空彩色图像,其分辨率为 1083×1083 , 采样频率为 1 min/张。TSI-VIS-J1006 由 EKO Instruments 公司研制的自动、全彩色天空成像系统,能够满足各种天气下对局部地区天空的观测需要^[13]。该设备采用具有数字图像处理功能的新型摄云机 Total Sky Camera J1006。摄云机上的半球形遮光屏可以避免阳光直射起到保护 TSI 的作用。图 1 为 TSI-VIS-J1006 的外观。

由于 TSI 采用球面成像,天空图像的边缘部分会产生较大的畸变并且会受到 TSI 安装位置周边建筑物的影响导致天空图像中建筑倒影的存在。TSI VIS-J1006 自带畸变校正和噪声去除的软件,将天空图像复原。图 2(a)是 TSI 拍摄的天空图像,图 2(b)是复原之后的天空图像。文中采用的云空辨识算法均是在复原图像的基础上进行操作。



图 1 Total Sky Imager VIS-J1006
Fig.1 Total Sky Imager VIS-J1006

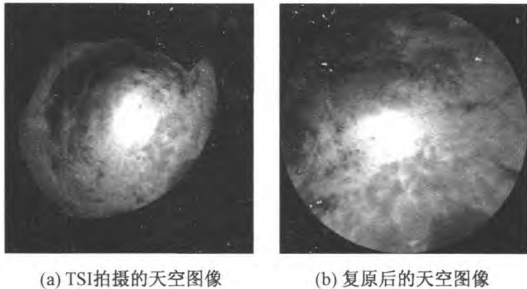


图 2 TSI 拍摄的天空图像以及复原图像
Fig.2 Sky image and restored image

2 基于 K-means 算法的图像分割算法

K-means 算法用于图像分割是通过提取天空图

像的颜色特征构成聚类样本,根据选定的聚类数和聚类中心按照选定的距离测度将每一个样本划分到某一类别中,使得分类结果中同类别中样本点的相似度较高而不同类别中样本点的相似度较低,从而实现图像分割,具体流程图见图 3。由于传统的 K-means 算法随机选择聚类中心使得每次聚类都可能产生不同的分类结果。文中针对每一幅图像的颜色特征选择聚类中心,使用加权欧式距离对样本进行分类,使 K-means 算法更具针对性。

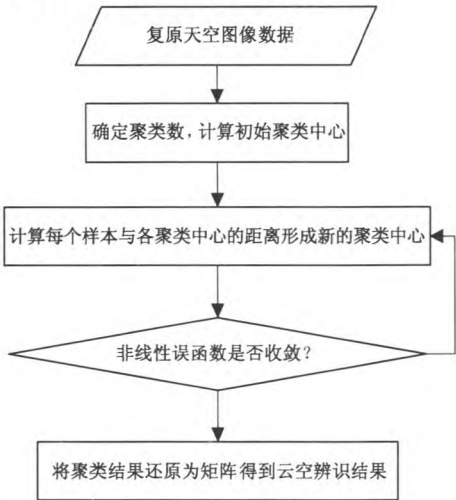


图 3 云空图辨识流程
Fig.3 Cloud identification flow chart

2.1 提取复原天空图像的特征矩阵

TSI 拍摄的天空复原图像是真彩色图像,图像由红、绿、蓝三个通道构成,即图像中每个像素点都采用相互独立的红(R)、绿(G)、蓝(B)三色光相互叠加来实现混色。真彩色图像采用 3 个维度相同的二维数组表示并且各数组元素均用 0 至 255 的整数表示各基色分量的强度。提取复原天空图像的 R 矩阵(红色分量矩阵)、 G 矩阵(绿色分量矩阵)、 B 矩阵(蓝色分量矩阵)。 R 矩阵、 G 矩阵、 B 矩阵中采用行和列体现天空复原图像中像素点的位置信息,元素值为对应基色分量的在此像素点的像素值。然后根据公式(1)提取的红蓝分量归一化差值矩阵。四矩阵为同型矩阵,分别将四矩阵中的每一列按列编号顺序排列成一个列向量,根据公式(2)获得复原天空图像的特征矩阵 f 。图 2(b)复原图像中存在黑色背景区域,为去除背景区域的影响需将复原天空图像通过 `rgb2gray` 函数转为灰度图像并提取灰度矩阵,灰度矩阵中灰度值为 0 的像素点构成的区域为黑色背景区域,提取 0 灰度值像素点的位置信息并将特征矩阵中对应位

置行向量设为空值即可去除背景区域的影响。应用 K-means 算法对四个列向量构成的矩阵聚类,得到聚类结果。若复原天空图像分辨率较高,可适当进行缩放以提高云空辨识效率。

$$d = \frac{b-r}{r+b} \quad (1)$$

式中: r 为红色分量像素值; b 为蓝色分量像素值; d 为红蓝分量归一化差值。

$$f = [rc, gc, bc, dc] \quad (2)$$

式中: rc 为红色分量列向量; gc 为绿色分量列向量; bc 为蓝色分量列向量; dc 为红蓝分量归一化差值列向量。

2.2 K-means 算法原理

K-means 算法是基于硬划分准则的常用聚类算法,其核心思想是通过不断迭代更新聚类中心,使非线性目标函数聚类误差平方和最小,从而得到最优的聚类效果。非线性目标函数聚类误差平方和 J 的计算公式为

$$J = \sum_{i=1}^k \sum_{j=1}^{n_k} (x_j - y_k)^2 \quad (3)$$

式中: k 为分类的类数; n_k 为第 k 个聚类中的数据总数; x_j 为第 k 类的数据; y_k 为此类的聚类中心。K-means 算法的具体步骤如下:

步骤一:确定需要分类的类数 k 。将复原天空图像进行云空辨识即将天空图像分成云和天空两部分,因此聚类数为 2。

步骤二:选取 k 个类的初始聚类中心。 k 为 2 即分别选取云像素点的聚类中心和天空像素点的聚类中心。太阳光由波长不同的红、橙、黄、绿、青、蓝、紫七种颜色构成。根据瑞利散射散射光强度与波长的四次方成反比,当太阳光透过大气时,波长较短的蓝光和紫光碰到大气分子、冰晶、水滴很容易发生散射现象,由于人眼对紫光不敏感使得晴朗天空呈蓝色。云是由小水滴和空气中的粉尘组成,它的直径大于任何一种颜色的光的波长,因此云可以反射所有光线,使得云在天空中呈现出白色。采用红蓝通道的比值运算可以突出不同波段间光谱的差异,提高对比度。文献[7]中用红蓝波段比值 0.77 为固定阈值分割地基天空图像,大于 0.77 为云小于 0.77 为天空。以此方法为依据获得 K-means 算法的初始聚类中心,具体步骤如下:

(1)提取复原地基天空图像的 R 矩阵、 G 矩阵、 B 矩阵;将 R 矩阵和 B 矩阵对应位置元素的相

除得到红蓝波段比值矩阵 I_t ;

(2)以 0.77 为阈值分割 I_t 时由于所选阈值的单一性会造成云的误识或漏识,为确保初始聚类中心云像素点和天空像素点具有较高的可信度,选取较大阈值进行云像素点的选取,选取较小的阈值进行天空像素点的选取。当阈值较大时,辨识结果图像中检测出的云的范围将缩小,对云的误识概率大大降低;反之,当选取的阈值较小时,天空像素的误识的概率降低,因此用较小阈值分割 I_t 所得的辨识结果中的天空像素的可信度较高。综上所述,选取较大的红蓝波段比矩阵的分割阈值 T_h 满足 $0.77 < T_h < 1$,选取较小的分割阈值 T_l 满足 $0 < T_l < 0.77$;通过对大量天空图像的处理,将 T_h 依次设置为 0.8、0.85、0.9、0.95 进行云像素点的提取,提取结果中云量随 T_h 的增大而降低。 T_h 为 0.8、0.85、0.9 时得到的提取结果中云量虽大,但是存在将天空误辨识为云的像素点,为提取可靠性较高的云像素点将 T_h 定为 0.95。同理,天空像素点提取选取的 T_l 小于或等于 0.5 时,由于阈值过低导致天空图像云空辨识结果中无天空像素点或存在很少的天空像素点,应选择较大的 T_l 值。依次选取 T_l 为 0.55、0.6、0.65、0.7,辨识结果中天空的范围随着阈值的增大而增大,对比辨识结果和原图中的天空范围,采用 0.7 作为提取天空像素点的分割阈值可以保证提取到较大数量的可信度较高的天空像素,从而保证得到的天空部分的聚类中心具有较高的可信度。

(3)根据公式(4)选取较高可信度的云像素点和天空像素点:

$$\begin{cases} i \in \text{云像素点} & I_t^i > T_h \\ i \in \text{天空像素点} & I_t^i < T_l \end{cases} \quad (4)$$

其中 I_t^i 为 R/B 特征空间中 i 像素点的红蓝分量比值。 T_h 设定为 0.95, T_l 设定为 0.7。

(4)提取步骤 3 中选取的云像素点位置信息,通过此位置信息获得复原图像中对应位置像素点的颜色特征即红、绿、蓝分量值,然后分别求取云像素点红、绿、蓝三分量的均值,求取红蓝均值的归一化差值,并将其组成一个一行四列的行向量作为云像素点的初始聚类中心。同理提取与步骤 3 中获得天空像素点同位置处的复原天空图像中像素点的红、绿、蓝分量值,然后分别求取均值并求取红蓝均值的归一化差值,组成一个一行四列

的行向量作为天空像素点的初始聚类中心。综上所述,复原天空图像特征矩阵的聚类中心是 2 行 4 列的矩阵,矩阵中第一行为天空像素点的聚类中心,第二行为云像素点的聚类中心。对聚类中心矩阵的规定能使复原天空图像中采用 K-means 算法辨识出的天空像素点的类型标识符为“1”,云像素点的类型标识符为“2”。

步骤三:基于加权欧式距离的 K-means 算法,根据初始聚类中心对复原天空图像的特征矩阵进行聚类计算,计算特征矩阵中各行向量与当前聚类中心的距离,找到离它最近的聚类中心,并将该行向量分配到该类;加权欧式距离的计算公式为

$$d = \left(\sum_{i=1}^4 (a_i(x_i - y_i))^2 \right)^{1/2} \quad (5)$$

式中: x_i 和 y_i 为像素点特征矩阵中行向量第 i 维数值, y_i 为当前聚类中心第 i 维数值; a_i 为特征行向量第 i 维数值的权重系数。权重系数的设定对于云空辨识结果影响较大。由于红蓝波段归一化差值在云空辨识中精确度较高,因此需设定较高的权重系数来凸显这一优势。然而,若权值过大则会使加权欧式距离的计算只依赖与红蓝波段归一化差值,红绿蓝通道的影响大大降低,造成云空辨识采用的天空图像特征量较少使辨识结果存在较大偏差。为了把红绿蓝三个彩色分量作为一个整体进行聚类,并体现红蓝波段归一化差值的优势,通过大量的仿真验证,将红蓝波段归一化差值的权重设定为 2.5,红绿蓝分量的权重均为 1。

步骤四:分别计算新生成的各类中所有对象的均值,并作为新的聚类中心。聚类中心的计算公式为

$$c_i = \frac{1}{w} \sum_{j=1}^w x_{ij} \quad (6)$$

式中: w 为第 k 类聚类的特征行向量的个数; x_{ij} 为

第 j 个特征行向量的第 i 维数值; c_i 为新的聚类中心的第 i 维数值。

步骤五:计算非线性目标函数聚类误差平方和,若其收敛则停止聚类否则重复步骤三~步骤五。

2.3 聚类结果还原

K-means 算法聚类的结果是只包含每个样本类别信息的列向量。采用 K-means 算法将不包含黑色背景区域的复原天空图像像素点分成两类,因此聚类结果中只包含类型标签 1 和类型标签 2。为使云空辨识结果图中包含黑色背景区域,建立一个与 R 矩阵同型的零矩阵,根据聚类结果确定矩阵中复原天空图像各像素点所属区域的类型,将该矩阵中每个元素(矩阵中只包含 0、1、2 三种元素值)乘 127.5,使矩阵元素分布在 0 至 255 之间得到云空辨识图像。聚类结果中天空像素点的类型标识符为“1”,云像素点的类型标识符为“2”,正比例变换后天空像素点的像素值为 127.5,云像素点的像素值为 255,因此辨识结果图像中天空呈现灰色,云呈现白色。

在用此方法处理图 4(f1)时,由于图 4(f1)中存在由于强光照射白光掩盖了蓝色,使天空呈现白色。在处理此类型图像中此方法会将由于强光照射造成的白色区域误辨识为云。为解决这一问题,首先提取由于强光造成天空呈现白色区域(下述中统称为误辨识区域)的红绿蓝三通道的值;然后提取天空图像中云的 RGB 值。采集两种类型的像素点各 254 个,绘制折线图如图 3 所示。通过对比分析数据发现,误辨识区域的 B 分量值较大,并且各通道之间存在较大的差值,然而云像素点的红绿蓝各通道之间存在较小的差值。因此误辨识特征为

蓝色分量像素值 $\in [180, 255]$

蓝色分量像素值 - 红色分量像素值 $\in [20, 90]$

绿色分量像素值 - 红色分量像素值 ≥ 15

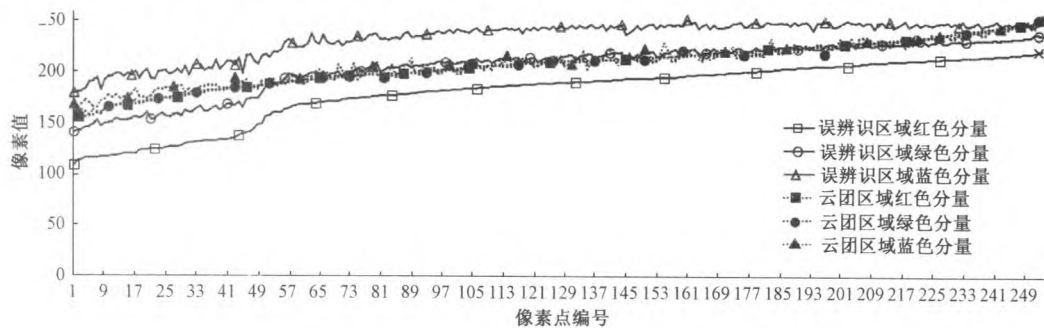


图 3 误辨识天空区域与云区域对比

Fig.3 Mistaken identification area sky and clouds contrast

在进行云空辨识之前首先找出满足以上三个条件的像素点,并将此像素点标记为天空像素点。即找出满足条件的像素点在聚类结果中的对应位置,将其类别编号设置为 1;然后提取剩余像素点构成的复原图像的特征矩阵,采用上述方法选取聚类中心,使用 K-means 算法得到云空辨识结果。

3 仿真结果与分析

基于上述云空辨识方法针对云南某光伏电站 TSI 设备采集并进行形变矫正、去噪处理后的复原图像进行云空辨识。根据天空图像中云量的不同从云南地区 2015 年 7 月 22 日至 8 月 16 日地基天空图像中选取 200 幅进行云空辨识。在本文中只选取较有代表性的 6 幅图像进行仿真示例。图 4(a1)的拍摄时间为 7 月 22 日 19 时 55 分 2 秒,天空图像中无云。图 4(b1)拍摄于 8 月 12 日 8 时 29 分 2 秒,天空图像中云量较少,图 4(c1)拍摄于同一天的 15 时 11 分 2 秒属于中等云量类型的天空图像。图 4(d1)拍摄于 7 月 30 日 10 时 3 分 2 秒是多云类型的地基天空图像。图 4(e1)拍摄时间为 8 月 16 日 16 时 47 分 2 秒,图中全部为云,是多云类型地基天空图像的特例。图 4(f1)拍摄时间为 8 月 7 日 18 时 27 分 2 秒,图中左侧天空区域由于强光照射呈现白色,此区域易误辨识为云。图 4(a2)~(f2)分别为图 4(a1)~(f1)的云空辨识结果图。图 4(a2)~(f2)中白色区域为使用上述云空辨识方法提取出的对应图 4(a1)~(f1)的云,灰色区域为对应图 4(a1)~(f1)的天空区域。

对全部为云和全部为天空的特殊天空图像进行云空辨识,通过上述算法产生一组聚类中心,聚类结果中类别编号仅有一类,因此,图 4(a2)准确的检测出图 4(a1)中的全部天空,图 4(e2)准确的检测出图 4(e1)中的全部的云。对比辨识结果图 4(b2)~(d2)与天空图像图 4(b1)~(d1),图 4(b2)~(d2)中均出现较薄云层漏检的情况。这是由于较薄的云层各通道像素值低于普通云层各通道像素值并且高于天空各通道像素值,聚类过程中由于较薄云层像素值距离天空聚类中心的距离较近,聚类结果中便将较薄云层的像素点归类为天空像素点,从而出现了对较薄云层的误检。图 4(f1)中左侧天空区域在较强太阳光的影响下呈现白色,此处天空区域极易被误辨识为云,本文

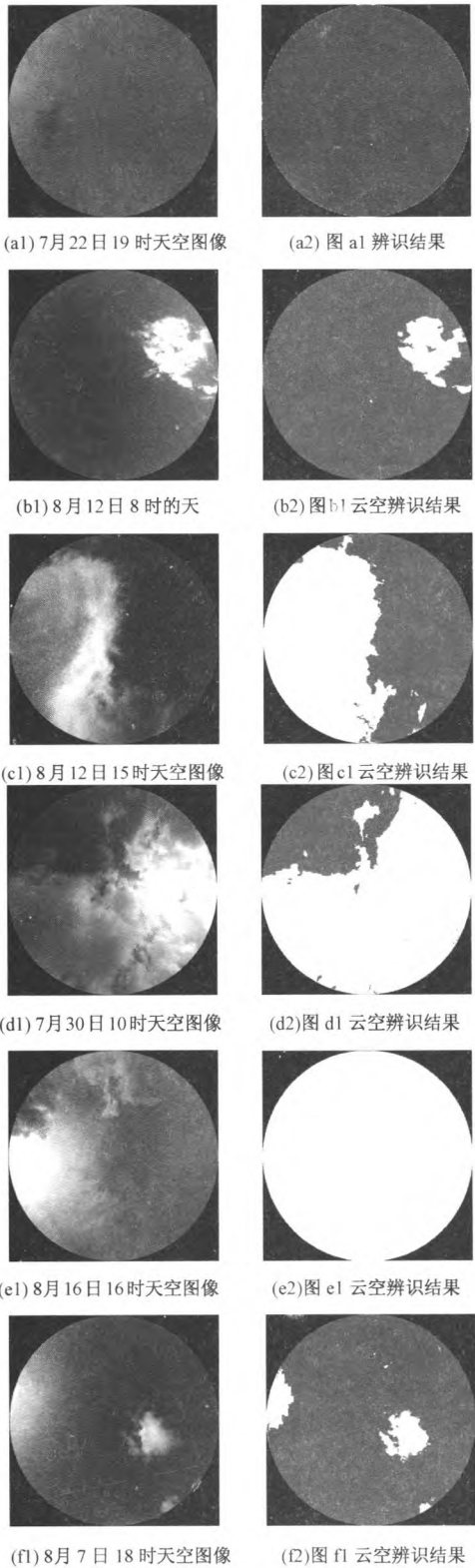


图 4 复原天空图像与云空辨识结果图
Fig. 4 Restoration sky images and cloud identification results

采用 K-means 算法通过三个限制条件的设定识别出此区域并将其标记为天空大大降低了此类天空像素点的误辨识率。

4 算法对比

将 K-means 算法与文献[7]中用红蓝波段比值 0.77 为固定阈值分割地基天空图像的方法进行对比。图 5 为采用固定阈值 0.77 分割图 4(a1)~(f1)的红蓝波段比值矩阵所得的云空辨识结果。图 5(a)和图 5(e)云空辨识结果均不存在云像素点的误检和漏检现象。图 5(b)~(d)和图 5(f)存在不同程度的云像素点的误检和漏检现象。

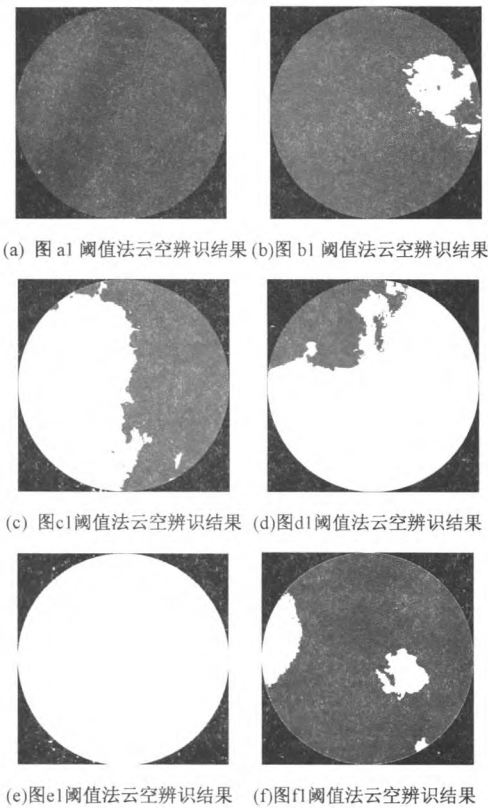


图 5 阈值法云空辨识结果图

Fig. 5 Threshold cloud identification results

由于对云空辨识效果缺乏统一的评价标准给结果分析带来巨大的挑战,同时由于使用云覆盖百分比作为云空辨识效果的评价标准并不能反应针对云或天空像素点同时存在误判和漏判时导致云覆盖百分比无变化这一情况下的云空辨识效果,为解决这一问题需要结合辨识结果图边缘与人工提取结果边缘的叠加图像反应该情况下的云空辨识效果。图 6 中白色较粗线条为人工提取图 4(b)~(d)和图 4(f)的云边界线,分别提取 K-means 算法和固定阈值法所得云空辨识结果图中云边界线并与人工分割结果叠加形成对比图。图 6(a1)~(d1)中白色细线为使用 K-means 算法所得云空辨识结果图的云团边界线,图 6(a2)~(d2)中白色细线为采用固定阈值进行云空辨识所

得辨识结果中云团的边界线。通过对图 6(a1)和图 6(a2)的观察可得两种算法均存在云像素点的漏检,图 6(a2)中存在将天空像素误检为云的情况。对于图 4(c1)采用 K-means 算法和固定阈值法进行云空辨识所得结果与人工分割结果较为相似,图 6(b1)和图 6(b2)更直观的体现出这一较好的处理结果。K-means 算法和固定阈值法对图 4(d1)和图 4(f1)的处理结果存在较大偏差,采用 K-means 算法进行云空辨识误差较小,和人工分割结果相近。图 6(c1)~(d1)和图 6(c2)~(d2)体现了这一处理效果。表 1 对比了 K-means 算法、固定阈值法处理结果与人工分割结果中云覆盖百分比(云覆盖百分比=辨识为云像素个数/天空图像总像素数),将表 1 中两种算法得到的云覆盖百

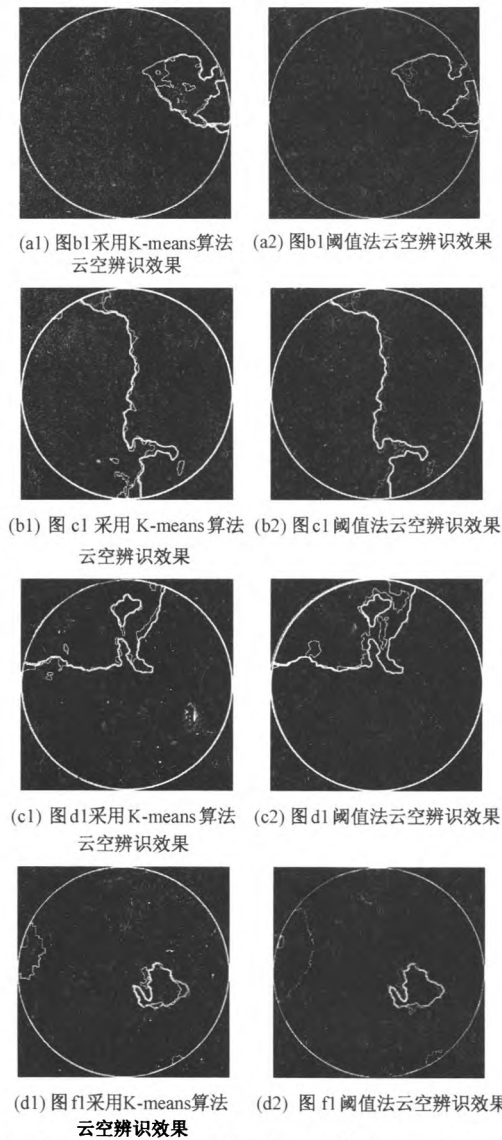


图 6 K-means 算法与固定阈值法云空辨识效果对比图

Fig.6 K-means algorithm and fixed threshold method cloud identification results comparison images

分比分别与人工分割结果的云空覆盖百分比相减得到表 2。表 1 和表 2 体现了上述对比分析,K-means 算法处理图 4(b1)所得结果与固定阈值的处理结果相比较差,但两种方法误差相差 0.15% 表示其处理结果相似。K-means 算法处理云量较大的图 4(c1)时云覆盖百分比的误差为 1.09% 明显优于固定阈值法 6.54% 的误差值。对于图 4(f1)的处理,两种方法差异明显,K-means 算法的云空辨识结果虽仍存在将呈现白色的天空区域误辨识为云的现象,但相较固定阈值法处理结果误辨识比例明显降低。

表 1 云覆盖百分比对比

Tab.1 Comparison of percentage of cloud cover (%)			
图像编号	K-means 算法	固定阈值	人工分割结果
图 4(b1)	7.99	9.04	8.59
图 4(c1)	51.68	51.72	50.46
图 4(d1)	76.10	81.55	75.01
图 4(f1)	7.00	11.60	3.95

表 2 误差对比

Tab.2 Error comparison (%)		
图像编号	K-means 算法	固定阈值
图 4(b1)	-0.60	0.45
图 4(c1)	1.22	1.26
图 4(d1)	1.09	6.54
图 4(f1)	3.05	7.65

对比 K-means 算法和固定阈值法进行云空辨识性很难实现各类型天空图像较准确的云空辨识效果。本文基于 K-means 算法针对每一幅地基天空图像求取聚类中心,克服了传统 K-means 算法由于聚类中心的随机选择造成聚类结果不稳定,聚类准确率低的缺点。在提取效率方面两种方法的云空辨识时间相差无几。综上所述,K-means 算法与以红蓝波段比 0.77 为固定阈值的云空辨识方法相比更具有普遍适用性。

5 结 论

本文提出了基于 K-means 算法的云空辨识方法,通过设定固定阈值更加合理的选择出云聚类中心和天空聚类中心。并且通过对强光影响的天空区域颜色特征的提取总结此类型区域的特征,实现此类型区域较好的云空辨识。通过对比 K-means 算法与固定阈值法的云空辨识结果、分析各自的优缺点可以得出:K-means 算法处理地基天空图像取得较好的效果。但是,由于 K-means 算法中加权欧式距离的权重根据经验设定,造成云空

辨识效果稳定性较差。如何更加合理的选择权重系数将是下一步的研究重点。

参考文献:

[1] 丁明,王伟胜,王秀丽,等. 大规模光伏发电对电力系统影响综述 [J]. 中国电机工程学报, 2014, 34 (1): 1-14.

[2] ADEL M, ALESSANDRO M P. A 24 h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy [J]. Solar Energy, 2010, 85 (50): 807-821.

[3] 王飞,米增强,甄钊,等. 基于天气状态模式识别的光伏电站发电功率分类预测方法 [J]. 中国电机工程学报, 2013, 33 (34): 75-82.

[4] 丁宇宇,丁杰,周海,等. 基于全天空成像仪的光伏电站水平面总辐射预报 [J]. 中国电机工程学报, 2014, 33 (1): 50-56.

[5] 朱想,居蓉蓉,程序,等. 组合数值天气预报与地基云图的光伏超短期功率预测模型 [J]. 电力系统自动化, 2015, 39 (6): 4-10.

[6] 杨丹,赵海斌,龙哲,等. MATLAB 图像处理实例详解 [M]. 北京:清华大学出版社, 2015.

[7] KREUTER A, ZANGERL M, SCHWARZMANN M, et al. All-sky imaging: A simple, versatile system for atmospheric research [J]. Appl Opt, 2009, 48 (6): 1091-1097.

[8] YANG J, LV W T, MA Y, et al. An automatic ground based cloud detection method based on adaptive threshold [J]. Journal of Applied Meteorological Science, 2009, 20 (6): 713-721.

[9] 李翠,冯冬青. 基于改进 K-均值聚类的图像分割算法研究 [J]. 郑州大学学报(理学版), 2011, 43 (1): 109-113.

[10] 吴晓蓉. K-均值聚类算法初始中心选取相关问题的研究 [D]. 长沙:湖南大学, 2008.

[11] 张靖,段富. 优化初始聚类中心的改进 k-means 算法 [J]. 计算机工程与设计, 2013, 34 (5): 1691-1694.

[12] 杨俊,吕伟涛,马颖,等. 基于自适应阈值的地基云自动检测方法 [J]. 应用气象学报, 2009, 20 (6): 713-721.

[13] 朱想,周海,丁杰,等. 全天空云图图像复原算法 [J]. 计算机辅助设计与图形学学报, 2014, 26 (6): 896-902.

作者简介: 苏适 (1972-), 男, 高级工程师, 研究方向为智能电网与新能源电力系统以及光伏风电功率预测。