

K-harmonic means clustering algorithm using feature weighting for color image segmentation

Zhiping Zhou¹ · Xiaoxiao Zhao¹ · Shuwei Zhu²

Received: 24 September 2016 / Revised: 13 July 2017 / Accepted: 9 August 2017 /
Published online: 18 August 2017
© Springer Science+Business Media, LLC 2017

Abstract This paper mainly proposes K-harmonic means (KHM) clustering algorithms using feature weighting for color image segmentation. In view of the contribution of features to clustering, feature weights which can be updated automatically during the clustering procedure are introduced to calculate the distance between each pair of data points, hence the improved versions of KHM and fuzzy KHM are proposed. Furthermore, the Lab color space, local homogeneity and texture are utilized to establish the feature vector to be more applicable for color image segmentation. The feature group weighting strategy is introduced to identify the importance of different types of features. Experimental results demonstrate the proposed feature group weighted KHM-type algorithms can achieve better segmentation performances, and they can effectively distinguish the importance of different features to clustering.

Keywords K-harmonic means · Color image segmentation · Feature weighting · Homogeneity · Feature group

✉ Zhiping Zhou
zpz@jiangnan.edu.cn

Xiaoxiao Zhao
6151905019@vip.jiangnan.edu.cn

Shuwei Zhu
zswjiang@163.com

¹ Engineering Research Center of Internet of Things Technology Applications Ministry of Education, Jiangnan University, Wuxi 214122, China

² College of Electronics and Information Engineering, Tongji University, Shanghai, 201804, China

1 Introduction

Image segmentation is a process of dividing an image into several non-overlapping, consistent regions which are homogeneous with respect to some characteristics. It plays an important role in image processing and pattern recognition. Results of image segmentation are very useful in applications such as object recognition, medical image processing, face recognition, content based image retrieval as well as other computer vision applications. In the past decades, many methods have been proposed for image segmentation which can be categorized into four groups: threshold-based, clustering-based, edge-based and region-based methods [18]. Among them, clustering is the method of creating groups of objects based on the similarity of relevant features, which has been widely used for image segmentation. Fuzzy C-means (FCM) algorithm has been a popular image segmentation method due to its simplicity of implementation and clustering validity. In order to resolve the outlier or noise sensitivity problem, many improved methods of FCM were proposed in the literature, the ones integrated with local spatial information were widely researched [1, 4, 9]. However, the FCM-type algorithms are usually sensitive to the selection of the initial cluster centers. Hence some study has been carried out and the K-harmonic means (KHM) [23] algorithms is a very effective method to cope with the problem mentioned above, which is also researched for image segmentation [15]. Many improved algorithms based on KHM have been proposed. The most popular methods are the combination of KHM with other swarm intelligence algorithms, which can take full advantage of global search ability of heuristic optimization and local search ability of KHM. Such as combining KHM with Particle Swarm Optimization (PSO) [21], Ant Colony Optimization (ACO) [13], Variable neighborhood search (VNS) [2, 5], Candidate groups search (CGS) [10], firefly algorithm (FA) [24], simple swarm optimization (SSO) [22]. However, these swarm intelligence based clustering algorithms are usually time-consuming and the improvements of performance are insignificant for many data sets.

For the above-mentioned KHM and its improved version, it is assumed that all features make equal contributions during the cluster processing, which can be a limitation in practical clustering situations. Because some features can have higher relevance in the clustering information than others. Recently the feature weighting strategy has gained much attention and been used to improve different kinds of clustering algorithms. The W-K-means algorithm [11] utilizes the feature weight to measure the importance of features and introduces the parameter β to further control its compact. But the appropriate value of β is hardly determined for different datasets due to the need of some prior knowledge. Then the E-W-K-means [12] was proposed to overcome the drawback of W-K-means, it uses the between-cluster information as the denominator of the dissimilarity measure. Xing [20] proposed an improved FCM method called IFWFCM, which was based on feature weighted distance and the feature weights were adaptively updated during the clustering process. But there is no relevant work with respect to KHM algorithm.

In this paper, we propose novel KHM-type clustering algorithms using adaptive feature weighting to identify the importance of different features. Some elements including Lab color space, local homogeneity and texture of the color image are extracted to constitute a more comprehensive feature vector. Moreover, the feature group weighting strategy is introduced to new clustering algorithm for color image segmentation. Experiments on some benchmark datasets from the UCI Repository and Berkeley segmentation datasets [3] demonstrate that the proposed feature weighting methods outperform others.

The rest of the paper is organized as follows: Section 2 will present KHM algorithm and feature weighting method. Section 3 will describe the proposed methods. Section 4 will analyze the experimental results on benchmark datasets and color images. Section 5 will conclude the work in this paper.

2 Preliminaries

2.1 K-harmonic means clustering

The KHM algorithm, similar to K-means and FCM, is a center based partitional clustering algorithm, but the most obvious difference is that the harmonic means (HM) of the distances from each data point to the centers as components to its object function, allowing KHM to be less sensitive to the initial centers and achieve better clustering performance. Assuming that a data set $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{id})$, which has n data points and each contains d features, is partitioned into k clusters with the centers denoted as \mathbf{c}_j ($j = 1, 2, \dots, k$). The objective function of KHM is expressed as follows.

$$J_{KHM}(X, C) = \sum_{i=1}^n \frac{k}{\sum_{j=1}^k \frac{1}{d_{ij}^p}}, \forall i = 1, 2, \dots, n. \tag{1}$$

where d_{ij} is the Euclidian distance between data point \mathbf{x}_i and cluster center \mathbf{c}_j , p is an input parameter and can affect the performance of clustering, typically set as $p \geq 2$ [22].

During the clustering process, the objective function is minimized and keeps steady until the end. Then \mathbf{x}_i is assigned to cluster j with the largest membership value. The new cluster center can be calculated according to (2).

$$\mathbf{c}_j^{new} = \frac{\sum_{i=1}^n m_{KHM}(\mathbf{c}_j / \mathbf{x}_i) \times w_{KHM}(\mathbf{x}_i) \times \mathbf{x}_i}{\sum_{i=1}^n m_{KHM}(\mathbf{c}_j / \mathbf{x}_i) \times w_{KHM}(\mathbf{x}_i)} \tag{2}$$

where $m_{KHM}(\mathbf{c}_j / \mathbf{x}_i)$ is the membership function.

$$m_{KHM}(\mathbf{c}_j / \mathbf{x}_i) = \frac{d_{ij}^{-p-2}}{\sum_{j=1}^k d_{ij}^{-p-2}} \tag{3}$$

$w_{KHM}(\mathbf{x}_i)$ is the weighting function.

$$w_{KHM}(\mathbf{x}_i) = \frac{\sum_{j=1}^k d_{ij}^{-p-2}}{\left(\sum_{j=1}^k d_{ij}^{-p}\right)^2} \tag{4}$$

In [16, 19], some study was carried out on the fuzzy version of KHM that was similar to FCM, fuzzy K-harmonic means (FKHM), where the fuzzy membership u_{ij} ($i = 1, 2, \dots, n, j = 1, 2, \dots, k$) was added in the dissimilarity computation that can replace $m_{KHM}(\mathbf{c}_j / \mathbf{x}_i)$ to calculate the assignment of the dataset. The objective function of FKHM is expressed as follows.

$$J_{FKHM}(X, C) = \sum_{i=1}^n \frac{k}{\sum_{j=1}^k \frac{1}{u_{ij}^m d_{ij}^p}}, \forall i = 1, 2, \dots, n. \tag{5}$$

where m is the fuzzy parameter that is set as 2 here.

2.2 Feature weighting clustering algorithm

Certain features of datasets may exhibit higher relevance than others. Feature weighting clustering methods have been proposed in recent years. To overcome the problem that the elements in a feature weight vector cannot be adaptively updated during the clustering process. In [20], an improved FCM method called IFWFCM was proposed, the updated equation can be derived from the revised object function based on the weighted distance. The objective function of IFWFCM is shown as (6).

$$F(U, V, w; D) = \sum_{j=1}^K \sum_{i=1}^n u_{ij}^m \sum_{q=1}^d [w_q(x_{iq} - c_{jq})]^2 \tag{6}$$

where $u_{ij} \in [0, 1]$, $\sum_{j=1}^k u_{ij} = 1$, $w_q \in [0, 1]$, $\sum_{q=1}^d w_q = 1$, and m is the fuzzy parameter that is usually set as 2.

3 The proposed algorithms

3.1 Feature weighted K-harmonic means

As described above, all features make equal contributions in the traditional KHM algorithm that maybe a limitation to the clustering performance, the importance of certain features to the dynamic cluster information cannot be appropriately manifested. But the updated feature weights can more accurately reflect the relevance of each feature in the clustering process. Hence, the feature weighting strategy like that of W-K-means and IFWFCM can be introduced into KHM to alleviate this problem. An input parameter β of W-K-means should be set in advance and the appropriate value β is hardly determined for different datasets. In view of the parameter p that is included in KHM, the feature weighted Euclidian distance $d_{ij}^{(w)}$ with no parameter can be utilized to calculate the dissimilarity between each pair of the data point and the center. Hence the feature weighted K-harmonic means (WKHM) is proposed and thereby its objective function is defined as (7).

$$J_{WKHM}(X, C) = \sum_{i=1}^n \frac{k}{\sum_{j=1}^k \frac{1}{[d_{ij}^{(w)}]^p}}, \forall i = 1, 2, \dots, n. \tag{7}$$

where $d_{ij}^{(w)} = \|\text{diag}(\mathbf{w})(\mathbf{x}_i - \mathbf{c}_j)\| = \sqrt{\sum_{q=1}^d w_q^2(x_{iq} - c_{jq})^2}$ with the feature weight vector $\mathbf{w} = (w_1, w_2, \dots, w_d)^T$ and $\text{diag}(\mathbf{w}) = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & w_d \end{bmatrix}$, $w_q \in [0, 1]$, $\sum_{q=1}^d w_q = 1$.

The clustering procedure of WKHM is carried out by minimizing the objective function, hence it can be regarded as an optimization problem. In terms of the constraint of feature weights, we can utilize the Lagrange multiplier technology to achieve an unconstrained

optimization problem and formulate the function L expressed as (8), where λ is the Lagrange multiplier.

$$L = \sum_{i=1}^n \frac{k}{\sum_{j=1}^k \frac{1}{[d_{ij}^{(w)}]^p}} - \lambda \left(\sum_{q=1}^d w_q - 1 \right) \tag{8}$$

For each iteration, the update equations of cluster centers $c_j (j = 1, 2, \dots, k)$ and feature weights $w_q (q = 1, 2, \dots, d)$ can be obtained by taking the derivatives of the function L with respect to each of them. The detailed theoretical derivations are provided in Appendix A and B.

The update equation of $w_q (q = 1, 2, \dots, d)$ is shown as (9). Appendix A provides the detailed theoretical derivations of (9).

$$w_q^{new} = \left(\sum_{l=1}^d \frac{\sum_{i=1}^n D_{ilq}}{\sum_{i=1}^n D_{il}} \right)^{-1} \tag{9}$$

where $D_{iq} = \frac{\sum_{j=1}^k \left([d_{ij}^{(w)}]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k [d_{ij}^{(w)}]^{-p} \right)^2}$.

The overall procedure of WKHM is summarized as Algorithm 1. It should be noted that during the initialization step, the values of the elements in feature weight vector w are initialized with the same value.

Algorithm 1 WKHM

Input: The dataset $X = (x_1, x_2, \dots, x_n)$;

Output: The cluster labels.

Step1 Initialize the parameters of the algorithm: feature weight vector w , number of clusters k , number of iterations, randomly choose the initial cluster centers.

Step2 Compute the objective function according to (7).

Step3 Compute the new cluster centers according to (2) by replacing the d_{ij} with $d_{ij}^{(w)}$ in the membership function $m_{KHM}(c_j/x_i)$ and weight function $w_{KHM}(x_i)$.

Step4 Update the feature weights according to (9).

Step5 Go to **Step6** if the predefined number of iterations is reached or the value of the objective function does not change obviously, otherwise go back to **Step2**.

Step6 Assign the data point x_i to cluster j by the maximum of membership function.

3.2 Feature weighted fuzzy K-harmonic means

In this section, $d_{ij}^{(w)}$ is also introduced into FKHM to form the feature weighted fuzzy K-harmonic means (WFKHM) algorithm, whose objective function is defined as (10).

$$J_{WFKHM}(X, C) = \sum_{i=1}^n \frac{k}{\sum_{j=1}^k \frac{1}{u_{ij}^m [d_{ij}^{(w)}]^p}}, \forall i = 1, 2, \dots, n. \tag{10}$$

where fuzzy membership $u_{ij} \in [0, 1]$, $\sum_{j=1}^k u_{ij} = 1$ and feature weight $w_q \in [0, 1]$, $\sum_{q=1}^d w_q = 1$, m is the fuzzy parameter that is set as 2 here.

Similar to WKHM, the update formulas of fuzzy membership, cluster center and feature weight are respectively expressed as (11)–(13) by utilizing the Lagrange multiplier. In view of the same principle to derive these variables of WFKHM with respect to WKHM, we do not provide the detailed analysis here.

$$u_{ij} = \left[\sum_{s=1}^k \left(\frac{[d_{ij}^{(w)}]^p}{[d_{is}^{(w)}]^p} \right)^{\frac{1}{m+1}} \right]^{-1} \tag{11}$$

$$c_j^{new} = \frac{\sum_{i=1}^n \frac{u_{ij}^{-m} [d_{ij}^{(w)}]^{-p-2}}{\left(\sum_{j=1}^k \left(u_{ij}^{-m} [d_{ij}^{(w)}]^{-p} \right) \right)^2} \mathbf{x}_i}{\sum_{i=1}^n \frac{u_{ij}^{-m} [d_{ij}^{(w)}]^{-p-2}}{\left(\sum_{j=1}^k \left(u_{ij}^{-m} [d_{ij}^{(w)}]^{-p} \right) \right)^2}} \tag{12}$$

$$w_q^{new} = \left(\sum_{l=1}^d \frac{\sum_{i=1}^n D_{iq}}{\sum_{i=1}^n D_{il}} \right)^{-1} \tag{13}$$

where $D_{iq} = \frac{\sum_{j=1}^k \left(u_{ij}^{-m} [d_{ij}^{(w)}]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k \left(u_{ij}^{-m} [d_{ij}^{(w)}]^{-p} \right) \right)^2}$.

The overall procedure of WFKHM is similar to that of WKHM, which is shown as Algorithm 2.

Algorithm 2 WFKHM

Input: The dataset $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$;

Output: The cluster labels.

Step1 Initialize the parameters of the algorithm: feature weight vector \mathbf{w} , number of clusters k , number of iterations, fuzzy membership u , randomly choose the initial cluster centers.

Step2 Compute the objective function according to (10).

Step3 Calculate the new fuzzy memberships according to the (11).

Step4 Calculate the new cluster centers according to the (12)

Step5 Update the feature weights according to (13).

Step6 Go to **Step7** if the predefined number of iterations is met or the value of the objective function does not change obviously, else go back to **Step2**.

Step7 Assign the data point \mathbf{x}_i to cluster j by the maximum of u_{ij} .

3.3 Feature extraction of color image

To deal with color image segmentation, the IFWFCM algorithm [20] has confirmed the effectiveness of adaptive feature weighting strategy. However, the overall segmentation performance of IFWFCM on color images is not superior to that of FCM according to the

experimental results, and even become worse in some cases. Moreover, IFWFCM only chose Lab color space, so the feature weighting strategy cannot be fully taken advantage of. In this study, the feature vector comprised of color space, local homogeneity and texture is used, where eight elements are included. Generally, there are different types of color spaces in the literature such as RGB, Lab, HSV color spaces, and the second one is chosen to describe the pixels in our method due to its popularity and effectiveness. The local homogeneity that quantifies the uniformity of a region in the image is calculated in the HSV color space [7, 14, 17]. Two efficient methods: Singular Value Decomposition (SVD) and QR factorization in the RGB color space are utilized to calculate the texture components of the image, where the spacing parameter is $N_s = 8$, the size of the window is 32, and the theta angles to rotate the window are $\theta = [0, 22.5, 45, 67.5]$. Then, the homogeneity feature (HF) with 3 elements, 3 feature components in Lab color space and the texture feature (TF) with 2 elements are integrated to construct the 8 dimensional feature vector $\{HF_{hue}, HF_{sat}, HF_{val}, L, a, b, TF_{SVD}, TF_{QR}\}$, hence the segmentation process can be carried out.

Here we mainly introduce the local homogeneity, it substantially makes use of the local information from images. Homogeneity is defined as a composition of standard deviation and discontinuity. For an image with the size $M \times N$, P_{ij} denotes the pixel at the location (i, j) , hence its normalized standard deviation is calculated as follows.

$$V_{ij} = \frac{v_{ij}}{v_{\max}} \text{ and } v_{ij} = \sqrt{\frac{1}{d^2} \sum_{p=i-((d-1)/2)}^{i+((d-1)/2)} \sum_{q=j-((d-1)/2)}^{j+((d-1)/2)} (I_{pq} - \mu_{ij})^2} \tag{14}$$

where $v_{\max} = \max\{v_{ij}\}$, $0 \leq i, p \leq M - 1, 0 \leq j, q \leq N - 1$. I_{pq} is the color component of a pixel at the location (p, q) .

w_{ij} is a size $d \times d$ (for example 5×5) window centered at (i, j) for the computation of deviation. μ_{ij} is the mean color component of a pixel P_{ij} within the window w_{ij} , whose equation is shown as follows.

$$\mu_{ij} = \frac{1}{d^2} \sum_{p=i-((d-1)/2)}^{i+((d-1)/2)} \sum_{q=j-((d-1)/2)}^{j+((d-1)/2)} I_{pq} \tag{15}$$

The discontinuity of color component at each position (i, j) can be calculated by its magnitude e_{ij} of the gradient that is obtained by Sobel operator within a $t \times t$ (for example 3×3) window, hence the normalized magnitude of the gradient is calculated as follows.

$$E_{ij} = \frac{e_{ij}}{e_{\max}} \text{ and } e_{ij} = \sqrt{G_x^2 + G_y^2} \tag{16}$$

where $e_{\max} = \max\{e_{ij}\}$, G_x and G_y are the gradient of color component in the x and y directions respectively.

Therefore, the local homogeneity of a pixel P_{ij} in an image is calculated as (17).

$$H_{ij} = 1 - E_{ij} \times V_{ij} \tag{17}$$

The value of local homogeneity at each location is in the interval $[0, 1]$. The more uniform the local region around a pixel is, the larger the homogeneity value of the pixel.

3.4 Feature group weighting based clustering approaches

During the process of color image segmentation, in order to distinguish different types of features more effectively, the feature group weighting strategy is introduced into WKHM and WFKHM, the feature group weighted K-harmonic means (GWKHM) and the feature group weighted fuzzy K-harmonic means (GWFKHM) are further proposed. Note that, in [6, 8] the feature groups are assigned the weights in each cluster that may not be applicable for our study, hence we adopt the global feature group weighting approach here. The features of local homogeneity, Lab color space and texture are respectively represented as three groups: $G(1) = \{HF\}$, $G(2) = \{FLab\}$, $G(3) = \{TF\}$, a weight value $v_t (t = 1, 2, 3)$ is assigned to the corresponding feature group. The feature group weighted Euclidian distance $d_{ij}^{(gw)}$ is calculated as (18).

$$d_{ij}^{(gw)} = \sqrt{v_1 \sum_{q \in HF} w_q^2 (x_{iq} - c_{jq})^2 + v_2 \sum_{q \in FLab} w_q^2 (x_{iq} - c_{jq})^2 + v_3 \sum_{q \in TF} w_q^2 (x_{iq} - c_{jq})^2}$$

$$= \sqrt{\sum_{t=1}^T v_t \sum_{q \in G(t)} w_q^2 (x_{iq} - c_{jq})^2}$$
(18)

where $\sum_{t=1}^T v_t = 1, T = 3$ and $\sum_{q \in G(t)} w_q = 1, w_q \in [0, 1], 1 \leq t \leq T$.

For the algorithm GWKHM, the Lagrange function L_2 formulated as (19) is similar to that of WKHM, where $\lambda_t (t = 1, 2, 3)$ is the Lagrange multiplier.

$$L_2 = \sum_{i=1}^n \frac{k}{\sum_{j=1}^k \frac{1}{[d_{ij}^{(gw)}]^p}} - \sum_{t=1}^T \lambda_t \left(\sum_{q \in G(t)} w_q - 1 \right)$$
(19)

The update equation of feature weights of $G(1)$ is expressed as (20). The detailed theoretical derivations of (20) are shown in Appendix B.

$$w_q^{new} = \left(\sum_{l \in G(1)} \frac{\sum_{i=1}^n D_{iq}}{\sum_{i=1}^n D_{il}} \right)^{-1}$$
(20)

where $D_{iq} = \frac{\sum_{j=1}^k \left([d_{ij}^{(gw)}]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k [d_{ij}^{(gw)}]^{-p} \right)^2}$

Similar to the case of $q \in G(1)$, the update formulas of feature weights in the case of $q \in G(2)$ and $q \in G(3)$ are respectively $w_q^{new} = \left(\sum_{l \in G(2)} \frac{\sum_{i=1}^n D_{iq}}{\sum_{i=1}^n D_{il}} \right)^{-1}$ and $w_q^{new} = \left(\sum_{l \in G(3)} \frac{\sum_{i=1}^n D_{iq}}{\sum_{i=1}^n D_{il}} \right)^{-1}$, where D_{iq} is the same as that in (20).

For the algorithm GWFKHM, we can derive the update equations with the same principle described above, hence the computations of fuzzy membership and cluster center are the same forms as (11) and (12), but it needs to be noted that $d_{ij}^{(w)}$ should be replaced by $d_{ij}^{(gw)}$.

Table 1 The characters of experimental datasets

Dataset	K	d	n
Iris	3	4	150
Wine	3	13	178
Sonar	2	60	208
Ionosphere	2	33	351
WDBC	2	30	569
Australian	2	14	690
Vehicle	4	18	846
Satellite	6	33	635

The update equation of feature weight is shown as (21) in the cases of $q \in G(1)$, $q \in G(2)$ and $q \in G(3)$.

$$w_q^{new} = \left(\sum_{t \in G(t)} \frac{\sum_{i=1}^n D_{iq}}{\sum_{i=1}^n D_{it}} \right)^{-1}, \quad 1 \leq t \leq T \tag{21}$$

where $D_{iq} = \frac{\sum_{j=1}^k \left(u_{ij}^{-m} [d_{ij}^{(w)}]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k \left(u_{ij}^{-m} [d_{ij}^{(w)}]^{-p} \right) \right)^2}$

The overall procedures of GWKHM and GWFKHM are respectively identical to those of WKHM and WFKHM, where the major difference lies in the update equations of feature weights in different groups. That means in the **Step4** of Algorithm 1, the (9) is replaced by (20) with $q \in G(1)$, $q \in G(2)$, $q \in G(3)$, while in the **Step5** of Algorithm 2, the (13) is replaced by (21).

Moreover, the update equation of feature weight of IFWFCM is formulated as (22) if the $d_{ij}^{(g^w)}$ is utilized, and the new method is named as GIFWFCM.

$$w_q^{new} = \left(\sum_{t \in G(t)} \frac{\sum_{i=1}^n \sum_{j=1}^k u_{ij}^m (x_{iq} - v_{jq})^2}{\sum_{i=1}^n \sum_{j=1}^k u_{ij}^m (x_{it} - v_{jt})^2} \right)^{-1}, \quad 1 \leq t \leq T \tag{22}$$

3.5 Computational complexity

The time complexity of WKHM and WFKHM are both $O(2TnKD_1)$, where T is the number of iterations, n is the number of data points, K is the number of clusters, and $D_1 = 3$ is the number of features in the Lab color space. Meanwhile, the time complexity of GWKHM and GWFKHM are both $O(2TnKD_2)$, where $D_2 = 8$ is the number of extracted features. However, it should be noted that during the preprocessing phase, the time complexity of calculating the local homogeneity is $O(2MN D_3)$, where $M \times N$ is the size of the image, and $D_3 = 3$ is the number of features in local homogeneity; and the time complexity of calculating the texture elements is $O(4(M/N_s)(N/N_s))$, where $N_s = 8$ that has been described in the Section 3.3.

Table 2 The results of *Acc* and *RI* of the five algorithms on eight datasets (Mean \pm Standard deviation)

Datasets	Algorithms	<i>Acc</i>	<i>RI</i>	Running time/s
Iris	K-means	0.856 \pm 0.112	0.862 \pm 0.049	0.0095
	KHM	0.893 \pm 0.000	0.880 \pm 0.000	0.0081
	WKHM	0.953 \pm 0.000	0.942 \pm 0.000	0.0118
	FKHM	0.893 \pm 0.000	0.880 \pm 0.000	0.0474
	WFKHM	0.907 \pm 0.000	0.892 \pm 0.000	0.0548
Wine	K-means	0.584 \pm 0.047	0.615 \pm 0.010	0.0195
	KHM	0.556 \pm 0.000	0.683 \pm 0.000	0.0171
	WKHM	0.534 \pm 0.000	0.593 \pm 0.000	0.0174
	FKHM	0.697 \pm 0.000	0.705 \pm 0.000	0.0590
	WFKHM	0.534 \pm 0.000	0.589 \pm 0.000	0.0647
Sonar	K-means	0.543 \pm 0.006	0.502 \pm 0.001	0.0136
	KHM	0.553 \pm 0.000	0.503 \pm 0.000	0.0185
	WKHM	0.530 \pm 0.002	0.499 \pm 0.001	0.0585
	FKHM	0.558 \pm 0.000	0.504 \pm 0.000	0.0868
	WFKHM	0.606 0.000	0.520 \pm 0.000	0.0666
Sphere	K-means	0.701 \pm 0.021	0.582 \pm 0.015	0.0098
	KHM	0.687 \pm 0.000	0.568 \pm 0.000	0.0066
	WKHM	0.687 \pm 0.000	0.568 \pm 0.000	0.0275
	FKHM	0.709 \pm 0.000	0.587 \pm 0.000	0.0787
	WFKHM	0.695 \pm 0.000	0.575 \pm 0.000	0.1031
WDBC	K-means	0.854 \pm 0.000	0.750 \pm 0.000	0.0163
	KHM	0.803 \pm 0.000	0.683 \pm 0.000	0.0239
	WKHM	0.852 \pm 0.000	0.748 \pm 0.000	0.0436
	FKHM	0.837 \pm 0.000	0.726 \pm 0.000	0.1152
	WFKHM	0.856 \pm 0.000	0.753 \pm 0.000	0.1301
Australian	K-means	0.562 \pm 0.000	0.507 \pm 0.000	0.0158
	KHM	0.559 \pm 0.000	0.506 \pm 0.000	0.0516
	WKHM	0.733 \pm 0.000	0.608 \pm 0.000	0.0718
	FKHM	0.559 \pm 0.000	0.506 \pm 0.000	0.0871
	WFKHM	0.793 \pm 0.000	0.671 \pm 0.000	0.0896
Vehicle	K-means	0.437 \pm 0.015	0.623 \pm 0.025	0.1074
	KHM	0.396 \pm 0.000	0.615 \pm 0.000	0.0519
	WKHM	0.439 \pm 0.029	0.646 \pm 0.014	0.0605
	FKHM	0.448 \pm 0.005	0.630 \pm 0.002	0.1519
	WFKHM	0.449 \pm 0.004	0.648 \pm 0.009	0.2432
Satellite	K-means	0.658 \pm 0.039	0.842 \pm 0.026	1.3746
	KHM	0.673 \pm 0.007	0.856 \pm 0.001	0.0504
	WKHM	0.726 \pm 0.000	0.858 \pm 0.000	2.1792
	FKHM	0.634 \pm 0.001	0.830 \pm 0.001	1.2959
	WFKHM	0.705 \pm 0.001	0.851 \pm 0.001	2.4567

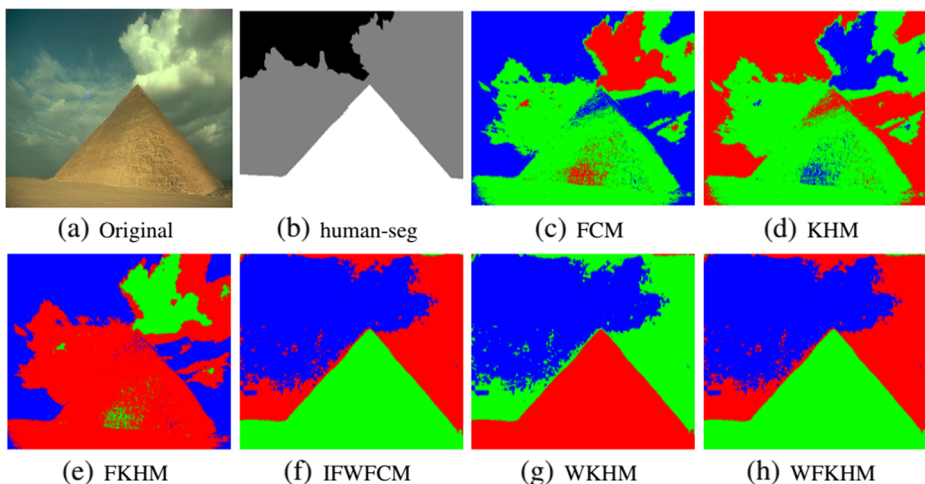
Table 3 Error rates of different algorithms for color images with the Lab features

Image	FCM	KHM	FKHM	IFWFCM	WKHM	WFKHM
3096	1.7021%	1.7409%	1.5460%	1.9365%	1.9365%	1.9365%
135069	0.7902%	0.7804%	0.7811%	0.6179%	0.6043%	1.1755%
238011	18.3561%	12.0991%	37.4136%	11.2493%	11.0861%	31.4195%
299091	47.4498%	48.1512%	45.2115%	28.3793%	28.3793%	28.2459%
113016	24.7583%	24.5659%	25.4176%	40.2167%	15.3186%	40.6843%
163062	42.2743%	38.0153%	46.6687%	44.3469%	41.3760%	52.1847%
124084	28.4441%	26.9862%	25.3697%	42.5451%	36.6047%	48.0275%
80099	46.0891%	3.7372%	4.3290%	0.3459%	0.3407%	0.3491%
113044	24.5160%	40.5580%	25.1514%	14.8386%	17.1294%	16.8632%
24063	23.4441%	24.9694%	28.8482%	24.4830%	18.4902%	24.6688%
108073	54.4883%	57.2593%	58.5339%	55.8021%	53.0800%	57.0217%
134052	51.5178%	50.3883%	48.9751%	47.8514%	45.2257%	48.1286%

4 Experiment and analysis

4.1 The experiment on benchmark datasets

In this section, the performances of the two proposed methods WKHM and WFKHM are evaluated by conducting some experiments on several benchmark datasets, which are also compared with K-means, KHM and FKHM. The five algorithms are applied to eight common real-life datasets from the UCI Machine Learning Repository, namely Iris, Wine, Sonar, Ionosphere, WDBC, Australian, Vehicle and Satellite. The characteristics of all datasets are shown in Table 1, where K is the number of cluster, d is the number of dimensions and n is the number of data points.

**Fig. 1** The segmentation results according to different algorithms on image 299091

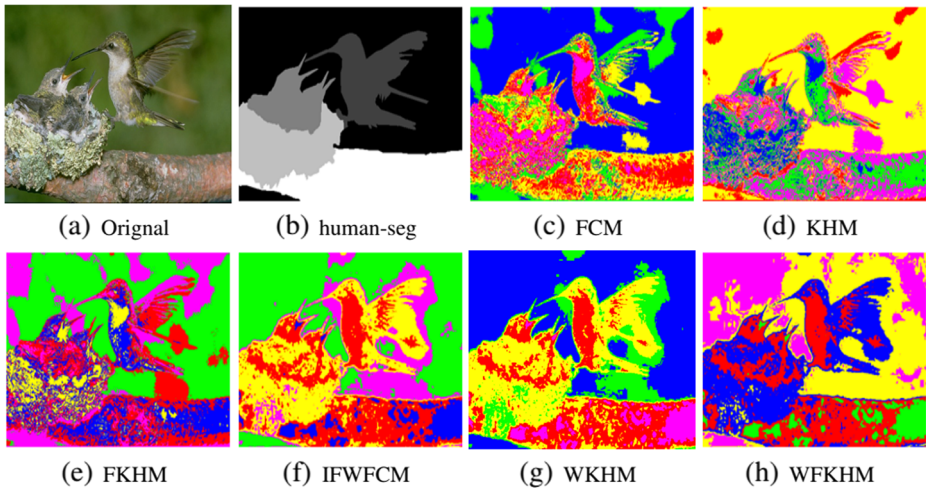


Fig. 2 The segmentation results according to different algorithms on image 163062

The experiments are conducted on a computer with Intel i7-4770, CPU 3.40GHz and 16GB RAM by using MATLAB2014a. To evaluate the performance of all algorithms, two well-known external cluster validity indexes (CVIs): accuracy (*Acc*) and rand index (*RI*) are adopted here, their values are in the interval $[0,1]$, the larger they are, the better clustering result. The parameters of the above five algorithms are set as: the total iteration number is $Maxiter = 100$, the fuzzy membership index in FKHM and WFKHM is $m = 2$. Note that, if the value of the objective function does not change obviously, the algorithm will stop in advance with the iteration number smaller than $Maxiter$. The parameter p is set as 3. All algorithms are repeatedly conducted 20 times independently, then their performances are compared in terms of the means and the standard deviations of *Acc* and *RI* shown in Table 2. For each dataset, the best value of each evaluation metric is highlighted in bold type.

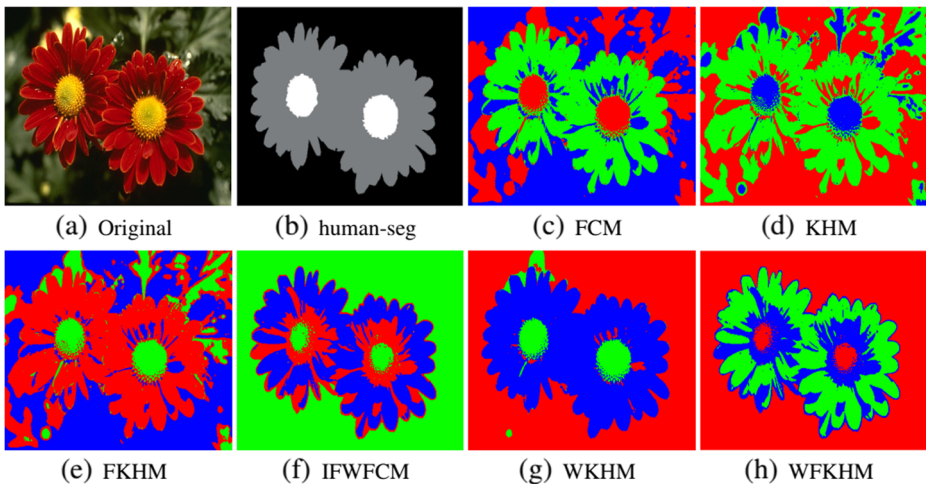


Fig. 3 The segmentation results according to different algorithms on image 124084

Table 4 Error rates of different algorithms for color images with the extracted features

Image	FCM	KHM	FKHM	GIFWFCM	GWKHM	GWFKHM
3096	1.1062%	1.1114%	1.1807%	20.5063%	0.7532%	0.6664%
135069	0.4462%	0.4611%	0.4566%	0.4333%	0.4275%	0.4152%
238011	15.7635%	14.2382%	16.8865%	11.5731%	11.1580%	14.5167%
299091	30.3651%	36.6209%	11.8911%	28.3793%	25.3703%	6.3154%
113016	6.4870%	7.1075%	6.8724%	13.6204%	3.1807%	3.2267%
163062	24.3988%	26.6663%	44.9570%	25.6048%	26.0419%	32.5762%
124084	34.1610%	4.7791%	34.8294%	24.0536%	3.5699%	4.4838%
80099	0.4430%	0.4404%	0.4495%	0.6496%	0.3258%	0.4035%
113044	8.4559%	9.4216%	8.7266%	8.1683%	4.2079%	4.1062%
24063	22.8224%	24.0199%	24.8125%	24.3843%	13.6612%	18.0925%
108073	37.3178%	36.1623%	47.7374%	37.9324%	34.2122%	39.5224%
134052	34.0587%	35.1714%	34.1643%	34.6623%	33.3509%	31.6177%

It can be seen from Table 2 that the performances of two feature weighting approaches have shown obvious superiority over the K-means, standard KHM and FKHM for most datasets, which verifies the effectiveness of adaptive feature weighting strategy. Except for datasets Wine and Sphere, the improvement of *Acc* on the datasets Iris, Sonar, WDBC, Australian, Vehicle and Satellite are respectively 6.72%, 7.92%, 22.70%, 41.86%, 0.22% and 7.88%. For the proposed algorithms, the variances of *Acc* and *RI* are 0 in most cases, which has demonstrated the insensitiveness of the algorithms, it can solve the initialization problem of K-means.

4.2 The experiment on color images

To verify the performance of proposed methods with feature weight for color image segmentation, we carried out our experiments on the well-known Berkeley Segmentation Dataset.

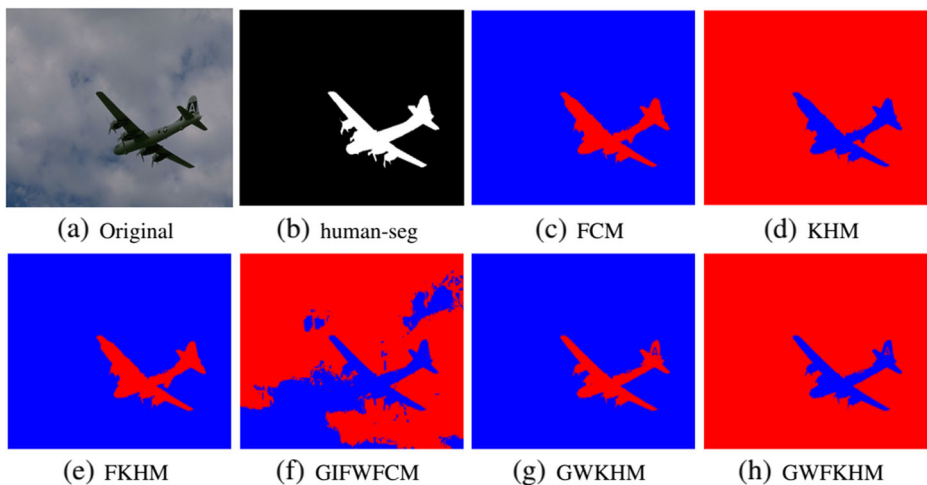


Fig. 4 The segmentation results according to different algorithms on image 3096

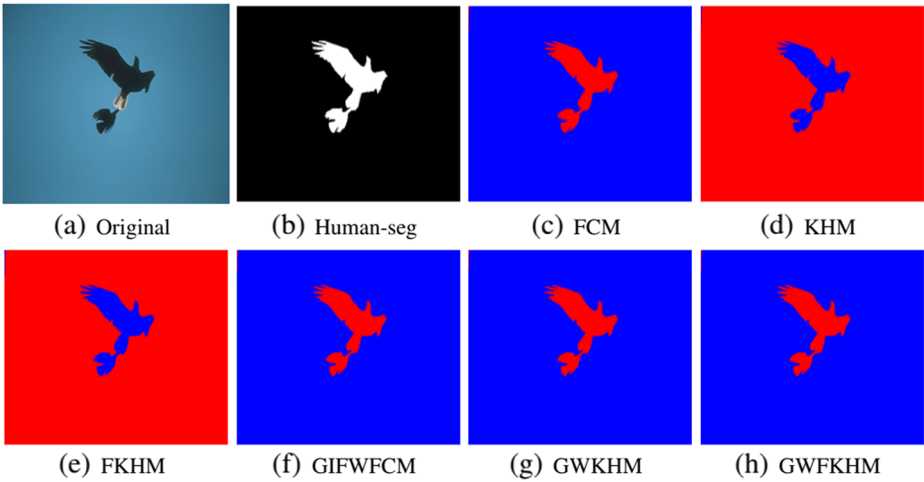


Fig. 5 The segmentation results according to different algorithms on image 135069

The human segmentation results are provided for the sake of evaluating the segmentation performance, but it should be noted that they are merely provided for rough analysis because the completely true segmentation results are not achieved. As the quantitative measurement for segmentation results, the ratio of misclassification *Error* is adopted here, whose equation is shown as (23).

$$Error = \frac{\sum_{i=1}^K e_i}{M \times N} \times 100\% \tag{23}$$

where, *K* is the number of color regions, and *e_i* is the color error of the *i*th region.

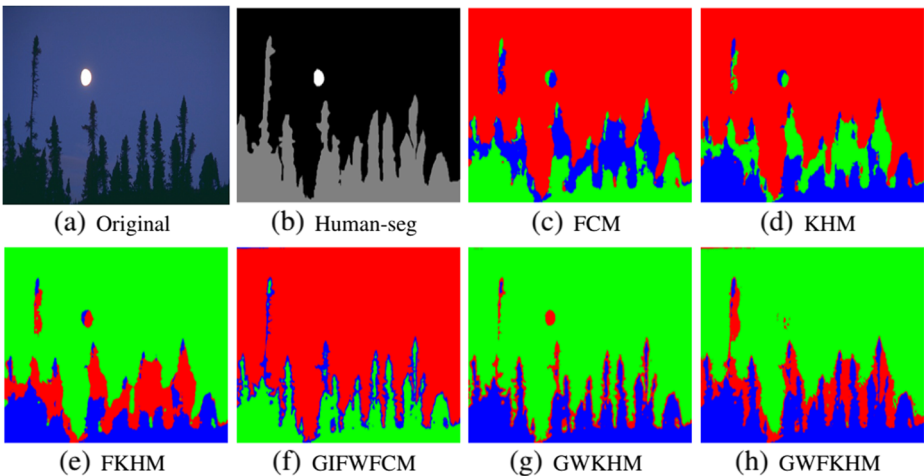


Fig. 6 The segmentation results according to different algorithms on image 238011

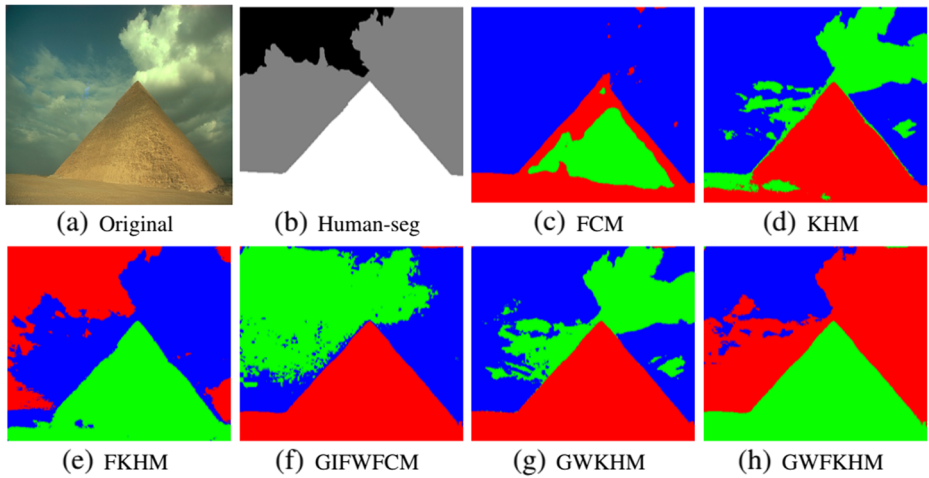


Fig. 7 The segmentation results according to different algorithms on image 299091

4.2.1 Experiments in the Lab color space

In this section, the segmentation performances of FCM, KHM, FKHM, IFWFCM [20], WKHM and WFKHM in the Lab color space are presented. The parameters of the clustering algorithms are the same as those in Section 4.1. Table 3 shows the *Error* results of different methods on the color images. For the purpose of comparing the performance of algorithms visually, only a portion of experiments are shown from Figs. 1, 2 and 3.

As we can see from Table 3 the feature weighting clustering algorithms can always achieve smaller *Error* values than other methods. It should be noted that the proposed methods are not sensitive to the initialization, leading to more stable results compared with IFWFCM. For example, WKHM outperforms the other methods for images 135069, 238011, 113016, 80099, 24063, 108073 and 134052. However, the segmentation

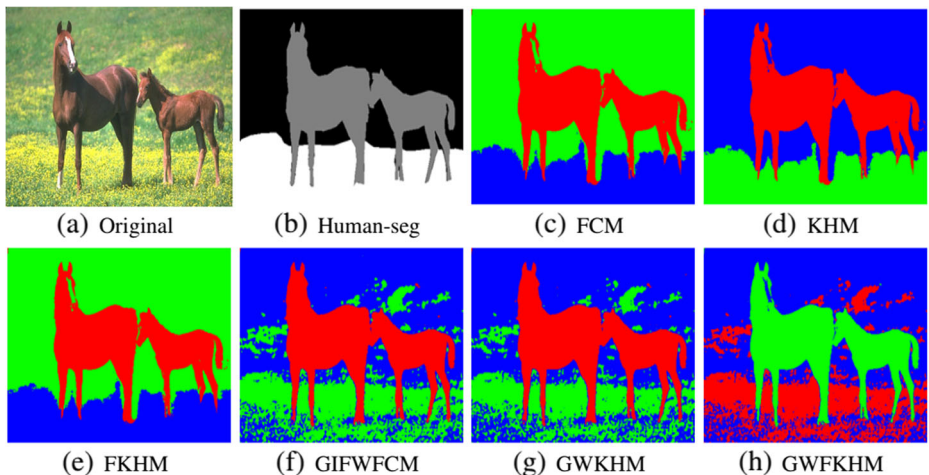


Fig. 8 The segmentation results according to different algorithms on image 113016

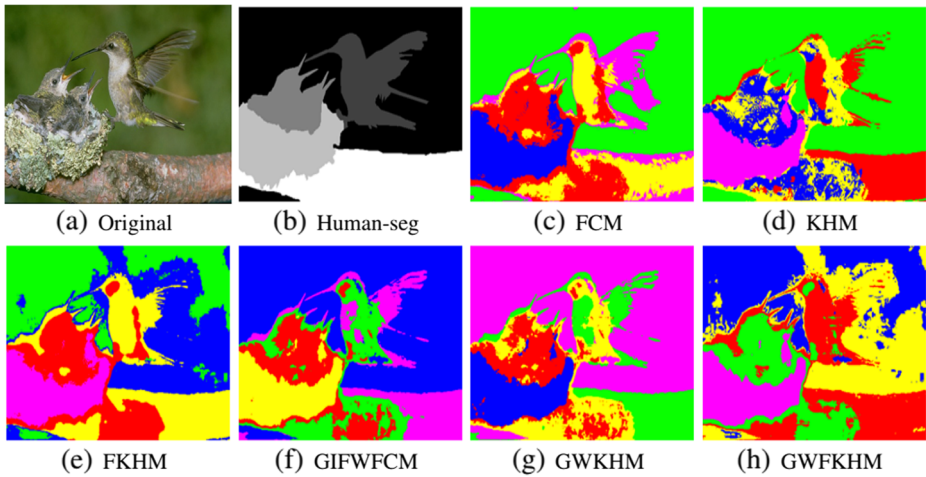


Fig. 9 The segmentation results according to different algorithms on image 163062

performances of images 299091, 163062 and 124084 are not desirable as observed from Figs. 1–3. This may be because the Lab color space is not very sufficient to describe these images, so the feature group weighting approaches with more types of features (i.e. Lab color space, local homogeneity, texture of the image) can be more effective and their segmentation performances are evaluated in the next subsection.

4.2.2 Experiments for feature group weighting based algorithms with the extracted features

In this section, the segmentation results of FCM, KHM, FKHM, GIFWFCM, GWKHM and GWFKHM with the distance $d_{ij}^{(gw)}$ described in the Section 3.4 are presented and the

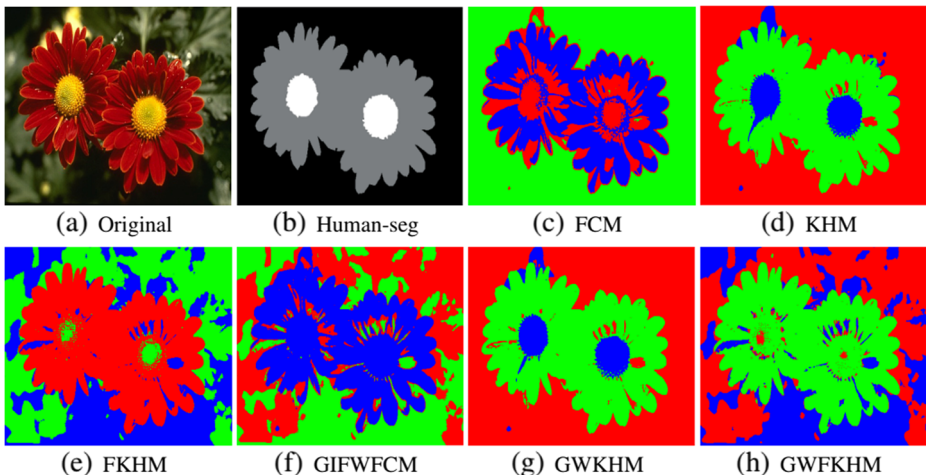


Fig. 10 The segmentation results according to different algorithms on image 124084

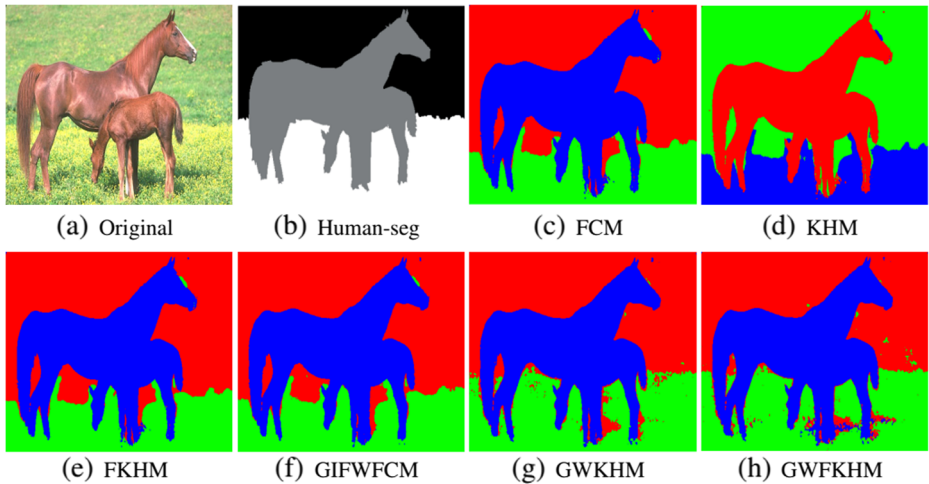


Fig. 11 The segmentation results according to different algorithms on image 113044

performances are also evaluated in terms of the ratio of misclassification *Error*. The parameters of the clustering algorithms are also the same as those in Section 4.1. We use traversal search to find the optimal feature group weights, for each feature group weight, the value ranges from 0.1 to 0.8, and the step size is set as 0.1. The experimental results demonstrate the weight of the Lab feature group is larger than that of the other two features for most images.

The *Error* results of different methods with 8 features are shown as Table 4. Only a portion of experiments are shown on Figs. 4, 5, 6, 7, 8, 9, 10, 11 and 12. For the images 3096, 135069, 238011, 299091, 113016, 124084, 80099, 113044, 24063, 108073 and 134052, the proposed algorithm GWKHM or GWFKHM outperforms the others, obtaining the best *Error* value respectively as 0.6664%, 0.4152%, 11.1580%, 6.3154%, 3.1807%, 3.5699%,

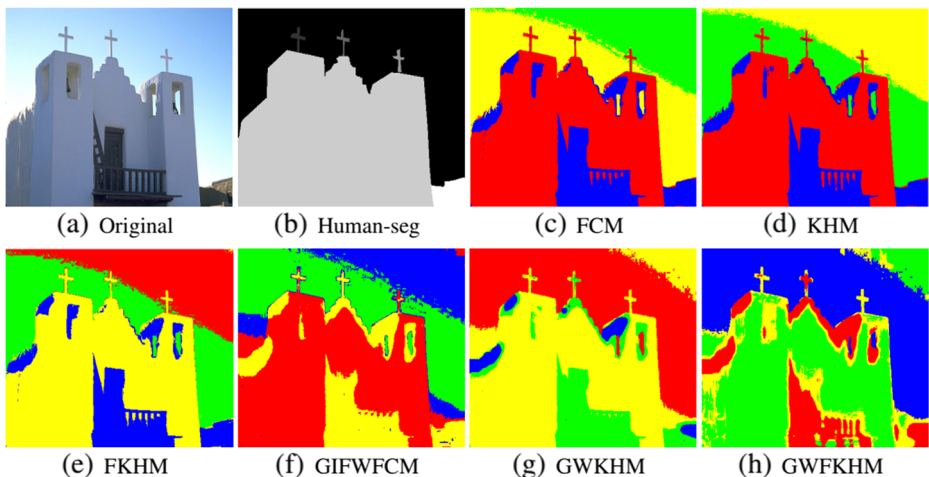


Fig. 12 The segmentation results according to different algorithms on image 24063

0.3258%, 4.1062%, 13.6612%, 34.2122%, 31.6177%. The three feature group weighting based methods can achieve smaller *Error* values than the standard ones in most cases. Note that, the *Error* values of the standard clustering algorithms are smaller than the proposed feature group weighting based methods for images 163062. Their segmentation results shown in Fig. 9c, d, e are closer to the human segmentation results. But we can observe from Fig. 9f, g, h that the segmentation results of the algorithms using feature group weighting can represent the detailed information better. To make a comparison between Tables 3 and 4, the results of the later are better than those of the former in most cases, hence we can conclude that the extracted feature vector with 8 dimensional components is superior to that merely in the Lab color space.

5 Conclusions

In this paper, we proposed novel KHM-type clustering algorithms with feature weighting for color image segmentation. To effectively compute the dissimilarity between different pixels and cluster centers, the Lab color space, local homogeneity and texture of the image are extracted to construct the feature vector. These features are divided into three groups during the segmentation process and a weight value is assigned to each group to measure its importance for clustering. Experimental results of both benchmark datasets and color images have shown the superiority of proposed methods, and the feature group weighting based clustering algorithms can achieve better performance for image segmentation than other related works. Adding more complicated features for more accurate image segmentation results in the proposed model will be further researched.

Acknowledgements This research is supported by the National Natural Science Foundation of China (Grant No. 61373126) and the Fundamental Research Funds for the Central Universities of China (Grant No. JUSRP51510).

Appendix A

In this Appendix, the detailed derivations for obtaining the update equation of \mathbf{c}_j and w_q are provided. At first, the partial derivation by \mathbf{c}_j of L is calculated as follows.

$$\frac{\partial L}{\partial \mathbf{c}_j} = -kp \sum_{i=1}^n \frac{[d_{ij}^{(w)}]^{-p-2} \text{diag}(\mathbf{w}^2)(\mathbf{x}_i - \mathbf{c}_j)}{\left(\sum_{j=1}^k [d_{ij}^{(w)}]^{-p}\right)^2} \quad (\text{A.1})$$

As we can see from the above equation, $\text{diag}(\mathbf{w}^2)$ is not related to variable i and $\frac{[d_{ij}^{(w)}]^{-p-2}}{\left(\sum_{j=1}^k [d_{ij}^{(w)}]^{-p}\right)^2} = m_{WKHM}(\mathbf{c}_j / \mathbf{x}_i) \cdot w_{WKHM}(\mathbf{x}_i)$, thus the equation of \mathbf{c}_j is obtained as

(2) by letting (A.1) be equal to 0. But it should be noted that the Euclidian distance d_{ij} is replaced by $d_{ij}^{(w)}$ in $m_{WKHM}(c_j / x_i)$ and $w_{WKHM}(x_i)$.

$$m_{WKHM}(c_j / x_i) = \frac{[d_{ij}^{(w)}]^{-p-2}}{\sum_{j=1}^k [d_{ij}^{(w)}]^{-p-2}} \tag{A.2}$$

$$w_{WKHM}(x_i) = \frac{\sum_{j=1}^k [d_{ij}^{(w)}]^{-p-2}}{\left(\sum_{j=1}^k [d_{ij}^{(w)}]^{-p}\right)^2} \tag{A.3}$$

Then, letting the partial deviation by w_q of L , which is denoted as (A.4), to be equal to 0, hence the update equation of w_q is obtained as (A.5), where the Lagrange multiplier λ should be eliminated.

$$\frac{\partial L}{\partial w_q} = kpw_q \sum_{i=1}^n \frac{\sum_{j=1}^k \left([d_{ij}^{(w)}]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k [d_{ij}^{(w)}]^{-p} \right)^2} - \lambda \tag{A.4}$$

$$w_q = \frac{\lambda}{kp \sum_{i=1}^n \frac{\sum_{j=1}^k \left([d_{ij}^{(w)}]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k [d_{ij}^{(w)}]^{-p} \right)^2}} \tag{A.5}$$

In terms of the constraint of feature weights $w_q \in [0, 1]$, $\sum_{q=1}^d w_q = 1$, in which the (A.5) is substituted and the calculation of λ is obtained as follows.

$$\lambda = \sum_{l=1}^d kp \sum_{i=1}^n \frac{\sum_{j=1}^k \left([d_{ij}^{(w)}]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k [d_{ij}^{(w)}]^{-p} \right)^2} \tag{A.6}$$

Therefore, the (A.6) is substituted in (A.5) to obtain the update equation of $w_q (q = 1, 2, \dots, d)$ shown as (9).

Appendix B

In this Appendix, the detailed derivation for obtaining the update equation (20) is provided. First, the partial derivation by $c_j (j = 1, 2, \dots, K)$ of L_2 is calculated and the result is set to be 0, then the update equation of cluster centers can also be obtained with the same form as (2), where $d_{ij}^{(g,w)}$ is utilized in $m_{WKHM}(c_j / x_i)$ and $w_{WKHM}(x_i)$. For the computation of feature weights, we firstly analyze the case of $q \in G(1)$, the partial deviation by w_q of

L_2 is calculated as follows.

$$\frac{\partial L_2}{\partial w_q} = kpw_q \sum_{i=1}^n v_1 \frac{\sum_{j=1}^k \left(\left[d_{ij}^{(gw)} \right]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k \left[d_{ij}^{(gw)} \right]^{-p} \right)^2} - \lambda_1 \quad (\text{B.1})$$

Then, letting the value of (B.1) to be 0 and the equation of w_q is obtained as (B.2), which is substituted in $\sum_{q \in G(1)} w_q = 1$, the constraint of feature group $G(1)$, then the calculation of λ_1 is obtained and substituted in (B.2) again, therefore the update equation of feature weights of $G(1)$ is shown as (20).

$$w_q = \frac{\lambda_1}{kpv_1 \sum_{i=1}^n \frac{\sum_{j=1}^k \left(\left[d_{ij}^{(gw)} \right]^{-p-2} \cdot (x_{iq} - c_{jq})^2 \right)}{\left(\sum_{j=1}^k \left[d_{ij}^{(gw)} \right]^{-p} \right)^2}} \quad (\text{B.2})$$

References

1. Ahmed MN, Yamany SM, Mohamed N et al (2002) A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Trans Med Imaging* 21(3):193–199
2. Alguwaizani A, Hansen P, Mladenovic N et al (2011) Variable neighborhood search for harmonic means clustering. *Appl Math Modell* 35(6):2688–2694
3. Arbelaez P, Maire M, Fowlkes CC et al (2011) Contour detection and hierarchical image segmentation. *IEEE Trans Pattern Anal Mach Intell* 33(5):898–916
4. Cai W, Chen S, Zhang D (2007) Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recogn* 40(3):825–838
5. Carrizosa E, Alguwaizani A, Hansen P et al (2015) New heuristic for harmonic means clustering. *J Glob Optim* 63(3):427–443
6. Chen X, Ye Y, Xu X et al (2012) A feature group weighting method for subspace clustering of high-dimensional data. *Pattern Recogn* 45(1):434–446
7. Cheng H-D, Sun Y (2000) A hierarchical approach to color image segmentation using homogeneity. *IEEE Trans Image Process* 9(12):2071–2082
8. Gan G, Ng MK-P (2015) Subspace clustering with automatic feature grouping. *Pattern Recogn* 48(11):3703–3713
9. Gong M, Liang Y, Shi J et al (2013) Fuzzy c-means clustering with local information and kernel metric for image segmentation. *IEEE Trans Image Process* 22(2):573–584
10. Hung C-H, Chiou H-M, Yang W-N (2013) Candidate groups search for K-harmonic means data clustering. *Appl Math Modell* 37(24):10123–10128
11. Huang JZ, Ng MK, Rong H et al (2005) Automated variable weighting in k-means type clustering. *IEEE Trans Pattern Anal Mach Intell* 27(5):657–668
12. Huang X, Ye Y, Zhang H (2014) Extensions of kmeans-type algorithms: a new clustering framework by integrating intracluster compactness and intercluster separation. *IEEE Trans Neural Netw Learn Syst* 25(8):1433–1446
13. Jiang H, Yi S, Li J et al (2010) Ant clustering algorithm with K-harmonic means clustering. *Expert Syst Appl* 37(12):8679–8684
14. Kumar S, Pant M, Kumar M et al (2015) Colour image segmentation with histogram and homogeneity histogram difference using evolutionary algorithms. *Int J Mach Learn Cybern* 6:1–21
15. Li Q, Mitianoudis N, Stathaki T (2007) Spatial kernel K-harmonic means clustering for multi-spectral image segmentation. *IET Image Process* 1(2):156–167
16. Liu N, Chen F, Lu M (2013) Spectral co-clustering documents and words using fuzzy K-harmonic means. *Int J Mach Learn Cybern* 4(1):75–83

17. Sag T, Cunkas M (2015) Color image segmentation based on multiobjective artificial bee colony optimization. *Appl Soft Comput* 34:389–401
18. Tan KS, Isa NA, Lim WH et al (2013) Color image segmentation using adaptive unsupervised clustering approach. *Appl Soft Comput* 13(4):2017–2036
19. Wu X, Wu B, Sun J et al (2015) A hybrid fuzzy K-harmonic means clustering algorithm. *Appl Math Modell* 39(12):3398–3409
20. Xing HJ, Ha MH (2014) Further improvements in Feature-Weighted Fuzzy C-Means. *Inf Sci* 267:1–15
21. Yang FQ, Sun TEL, Zhang CH (2009) An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization. *Expert Syst Appl* 36(6):9847–9852
22. Yeh W-C, Lai C-M, Chang K-H (2016) A novel hybrid clustering approach based on K-harmonic means using robust design. *Neurocomputing* 173:1720–1732
23. Zhang B (2000) Generalized k-harmonic means. Hewlett-Packard Laboratoris Technical Report
24. Zhou Z, Zhu S, Zhang D (2015) A novel K-harmonic means clustering based on enhanced firefly algorithm. In: Proceedings of the the 6th international conference on intelligence science and big data engineering. Springer, Suchow, pp 140–149



Zhiping Zhou received a PhD degree in detection technology and automation devices from East China University of Science and Technology, China, in 2007. He visited Purdue University as a visiting scholar in 2010. He is currently working as a professor in College of Internet of Things Engineering, Jiangnan University, China. His research interests include image processing, multimedia signal processing, Computer network and the control system.



Xiaoxiao Zhao received a B.S degree in Internet of Things Engineering from Jiangnan University, China, in 2015. She is currently doing her M.S research study in College of Internet of Things Engineering, Jiangnan University. Her research interests include data mining and digital image processing.



Shuwei Zhu received a M.S degree in Control science and Engineering from Jiangnan University, China, in 2016. He is currently getting his PhD degree in College of Electronics and Information Engineering, Tongji University, China. His research interests include data mining and multi-objective optimization.