Color Image Segmentation Based on Intensity and Hue Clustering – LS and LAD Approaches Comparison

Abstract. This paper addresses the color image segmentation problem. Motivated by the method for color image segmentation based on intensity and hue clustering proposed in the paper [22] we give some theoretical explanations for this method that directly follows from the natural connection between Maximum Likelihood approach and Least Square or Least Absolute Deviation clustering optimality criteria. The method is tested and illustrated on a few typical situations, such as the presence of outliers among the data.

Key words: data clustering, color image segmentation, least square, least absolute deviation

Received: xx xx, 201x; accepted: yy yy, 201x; available online: zz zz, 201x

1. Introduction

The term *Image Segmentation* refers to partitioning an image into two or more different regions that are "similar" in some image characteristic. It is an important task in image analysis process because all subsequent tasks, such as object recognition, depends on the quality of the segmentation. For this reason, methods for successful image segmentation are constantly being improved.

Most attention on image segmentation has been focused on gray scale images (see e.g. [12]). However, there are situations where this approach is not appropriate, and color components of the image have to be taken into account. Computers mostly use RGB (Red, Green, Blue) color space for image storage and manipulation, but it does not coincide with the vision psychology of human eyes because of high correlation among its three components (see e.g. [1]). For this reason, we will use HSI color space which is more compatible with the human vision.

In HSI color representation the I component represents *intensity*, H component represents *hue*, and S component represents *saturation*. To convert RGB representation to HSI representation, first compute [1]:

$$\begin{bmatrix} Y \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

Thereafter, HSI values can be computed as:

$$I = Y, \quad S = \sqrt{C_1^2 + C_2^2}, \quad H = \begin{cases} \arccos(\frac{C_2}{S}) & C_1 \geq 0 \\ 2\pi - \arccos(\frac{C_2}{S}) & C_1 < 0. \end{cases}$$

Among the three components of HSI representation, the most important ones are H and I, therefore, they will be used in the segmentation process. In that sense color image could be represented (see e.g. [22]) by intensity-data set $\mathcal{I} = \{I_i \in \mathbb{R} : i = 1, \ldots, m_I\} \subset \mathbb{R}$, with corresponding weights $w_i^I > 0$, where $I_i \neq I_j, i \neq j$, and hue-data set $\mathcal{H} = \{H_i : i = 1, \ldots, m_H\} \subset [0, 2\pi\rangle$ with corresponding weights $w_i^H > 0$, where $H_i \neq H_j$, $i \neq j$.

Our starting point is the method proposed in the paper [22], which utilize Least Squares optimality criterion for image clustering. We will extend that method by using the Least Absolute Deviation criterion and give some illustrative examples.

The paper is organized as follows. In the next section briefly introduction to the weighted clustering problem is given. In Sections 3 and 4, the one-dimensional clustering problem in intensity and hue space is considered and corresponding connection with Maximum Likelihood approach is given. On the basis of the relation between one-dimensional optimal partitions, image is represented in high-dimensional space. In Section 5 clustering in high-dimensional space is considered. Section 6 gives several illustrative numerical examples.

2. Weighted clustering problem

A partition of the set $\mathcal{A} = \{a_i \in \mathbb{R}^n : i = 1, \ldots, m\} \subset \mathbb{R}^n$ with the corresponding weights $w_i > 0$ into k disjoint subsets $\pi_1, \ldots, \pi_k$, $1 \leq k \leq m$, such that

$$\bigcup_{j=1}^{k} \pi_j = \mathcal{A}, \qquad \pi_r \cap \pi_s = \emptyset, \quad r \neq s, \qquad |\pi_j| \geq 1, \quad j = 1, \ldots, k,$$

will be denoted by $\Pi(\mathcal{A}) = \{\pi_1, \ldots, \pi_k\}$, and the elements $\pi_1, \ldots, \pi_k$ of such partition are called *clusters in* $\mathbb{R}^n$.

If $d\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty\rangle$ is some distance-like function (see e.g. [9, 20]), then, by applying the *minimum distance condition* (see e.g. [9, 19]), with each cluster $\pi_j \in \Pi$ we can associate its center c_j, defined by

$$c_j = c(\pi_j) := \underset{x \in \mathrm{conv}(\pi_j)}{\operatorname{argmin}} \sum_{a_i \in \pi_j} w_i d(x, a_i). \tag{1}$$

If we define an objective function $\mathcal{F}\colon \mathcal{P}(\mathcal{A}, k) \to [0, +\infty\rangle$ on the set of all partitions $\mathcal{P}(\mathcal{A}, k)$ of the set $\mathcal{A}$ containing k clusters by

$$\mathcal{F}(\Pi) = \sum_{j=1}^{k} \sum_{a_i \in \pi_j} w_i d(c_j, a_i),$$

then we define an optimal partition $\Pi^\star$, such that $\mathcal{F}(\Pi^\star) = \min\limits_{\Pi \in \mathcal{P}(\mathcal{A}, k)} \mathcal{F}(\Pi)$.

Conversely, for a given set of centers $c_1, \ldots, c_k \in \mathbb{R}^n$ applying the minimal distance condition we can define the partition $\Pi = \{\pi_1, \ldots, \pi_k\}$ of the set $\mathcal{A}$ in the following way: $\pi_j = \{a \in \mathcal{A} : d(c_j, a) \leq d(c_s, a), \forall s = 1, \ldots, k\}$, $j = 1, \ldots, k$, where one has to take care that every element of the set $\mathcal{A}$ occurs in one and only one cluster. Therefore the problem of finding an optimal partition of the set $\mathcal{A}$ can be reduced to the following optimization problem

$$\min_{c_1,\ldots,c_k\in\mathbb{R}^n} F(c_1,\ldots,c_k), \qquad F(c_1,\ldots,c_k) = \sum_{i=1}^{m} \min_{j=1,\ldots,k} w_i d(c_j, a_i), \tag{2}$$

where $F\colon \mathbb{R}^{kn} \to \mathbb{R}_+$, and $\mathbb{R}_+$ is the set of all vectors in $\mathbb{R}^n$ with nonnegative components. In general, this functional is not differentiable and it may have several local minima. Optimization problem (2) can also be found in literature as a k-median problem and it is most frequently solved by various metaheuristic methods [4] or by applying integer programming [13, 17].

The most known algorithm for searching for a locally optimal partition is the k-means algorithm [3, 10], which can be described by two steps which are iteratively repeated.

Step 1 For each set of mutually different assignment points $c_1,\ldots,c_k \in \mathbb{R}^n$ the set $\mathcal{A}$ should be divided into k disjoint unempty clusters $\pi_1,\ldots,\pi_k$ by using the minimal distance principle

$$\pi_j = \{a \in \mathcal{A} : d(c_j, a) \le d(c_s, a),\, \forall s = 1,\ldots,k\}, \qquad j = 1,\ldots,k. \tag{3}$$

Step 2 Given a partition $\Pi = \{\pi_1,\ldots,\pi_k\}$ of the set $\mathcal{A}$, one can define the corresponding centroids by

$$c_j = \operatorname*{argmin}_{x\in \operatorname{conv} \pi_j} \sum_{a_i\in\pi_j} w_i d(x, a_i), \quad j = 1,\ldots,k. \tag{4}$$

Let c_j be locally optimal centers of clusters π_j, $j = 1,\ldots,k$. For every $a_i \in \mathcal{A}$ define a set of indexes of the nearest assignment points

$$U_i = \{j \in J\colon d(c_j, a_i) \le d(c_s, a_i),\, \forall s \in J\},\ J = \{1,\ldots,k\}. \tag{5}$$

Note that the set U_i is unempty, and that it can be a single member set or a multi-member set. If for every $a_i \in \mathcal{A}$ the set U_i is a single member set, then a corresponding partition $\Pi = \{\pi(z_1),\ldots,\pi(z_k)\}$ is said to be a *well-separated partition*, i.e. the partition Π is said to be a well-separated partition if and only if the following holds

$$(\forall a_i \in \mathcal{A})(\exists j \in J) \quad d(c_j, a_i) < d(c_s, a_i), \quad \forall\, s \in J \setminus \{j\}.$$

3. Least square and least absolute deviation one dimensional weighted clustering in intensity space

The set $\mathcal{I} = \{I_i \in \mathbb{R} : i = 1,\ldots,m_I\} \subset \mathbb{R}$ of intensity-data, with corresponding weights $w_i^I > 0$ has to be divided into k_I clusters $\Pi^I = \{\pi_1^I,\ldots,\pi_{k_I}^I\}$, $1 \le k_I \le m_I$. Note that without loss of generality we can suppose that $I_i \ne I_j$, $i \ne j$. In this section, we consider a one-dimensional clustering problem using the Least Squares (LS), based on LS distance–like function $d_2(x,y) = (x-y)^2$ and the Least Absolute

Deviations (LAD) – optimality criterion, based on LAD distance function $d_1(x,y) = |x-y|$. The problem of finding an optimal partition of the set $\mathcal{I}$ according to (2) reduces to the following nonconvex and nonsmooth optimization problem

$$\min_{c_1,\ldots,c_k\in\mathbb{R}} F^{(p)}(c_1,\ldots,c_k), \qquad F^{(p)}(c_1,\ldots,c_k) = \sum_{i=1}^{m_I} \min_{j=1,\ldots,k_I} w_i^I d_p(c_j, I_i),\ p=1,2.$$

It is well known that generally, the LAD approach ignores outliers among the data [2, 16], while the LS approach stresses them.

In order to apply the $k-$means algorithm it is necessary to efficiently determine the centers of clusters in accordance with (1). It can be shown that for $p=1,2$ the centers $c_j^I(p)$ of clusters $\pi_j^I(p)$, $j=1,\ldots,k_I$ can be explicitly determined by the following formula:

$$c_j^I(p) = \underset{x\in\operatorname{conv}\pi_j^I(p)}{\operatorname{argmin}} \sum_{I_i\in\pi_j^I(p)} w_i^I d_p(x, I_i) = \begin{cases} \underset{I_i\in\pi_j^I(1)}{\operatorname{med}} (w_i^I, I_i),\ p=1 \\ \underset{I_i\in\pi_j^I(2)}{\operatorname{mean}} (w_i^I, I_i),\ p=2, \end{cases}$$

where $\underset{I_i\in\pi_j^I(1)}{\operatorname{med}} (w_i^I, I_i)$ is a weighted median of the set $\pi_j^I(1)$ (see e.g. [14, 15, 21]) and $\underset{I_i\in\pi_j^I(2)}{\operatorname{mean}} (w_i^I, I_i) = \frac{\sum_{I_i\in\pi_j^I} w_i^I I_i}{\sum_{I_i\in\pi_j^I} w_i^I}$ is a weighted mean of the set $\pi_j^I(2)$.

3.1. Connection with the Maximum likelihood approach

In this subsection we are going to illustrate the connection between minimization of the function $F^{(p)}$ and Maximum Likelihood approach. It will be shown that in some special situation the limit case of the Maximum Likelihood approach is equivalent to the problem of minimization of the objective function $F^{(p)}$. In order to do this, let us first suppose that $w_i^I = 1$, $i=1,\ldots,m_I$.

Let $c_j^I(p)$ be the centers of clusters $\pi_j^I(p)$, $j=1,\ldots,k_I$ and $\sigma_j>0$, $j=1,\ldots,k_I$ given positive numbers. Let us suppose that intensity-data I_i, $i=1,\ldots,m_I$ are independent and come from mixing distribution with probability density function

$$f^{(p)}(x;\lambda_1,\ldots,\lambda_{k_I}, c_1^I(p),\ldots,c_{k_I}^I(p),\sigma_1,\ldots,\sigma_{k_I}) = \sum_{i=1}^{k_I} \frac{\lambda_j}{\sigma_j}\varphi^{(p)}\left(\frac{x-c_j^I(p)}{\sigma_j}\right),$$

$\sum_{j=1}^{k_I}\lambda_j = 1$, $\lambda_j \geq 0$, $j=1,\ldots,k_I$, where $x\mapsto \frac{1}{\sigma_j}\varphi^{(p)}\left(\frac{x-c_j^I(p)}{\sigma_j}\right)$, $p=1,2$ are respectively a probability density function of Laplace random variable $\mathcal{L}(c_j^I(1),\sigma_j^2)$ i.e. Gaussian normal random variable $\mathcal{N}(c_j^I(2),\sigma_j^2)$, and $\varphi^{(p)}(x) = \frac{1}{\sqrt{2}}e^{-|x|^p/2}$.

The corresponding likelihood function reads

$$L^{(p)}(\gamma_1,\ldots,\gamma_{k_I}, c_1,\ldots,c_{k_I},\sigma_1,\ldots,\sigma_{k_I}) = \prod_{i=1}^{m}\sum_{j=1}^{k_I}\frac{\gamma_j}{\sigma_j}\varphi^{(p)}\left(\frac{I_i-c_j}{\sigma_j}\right).$$

Particulary, if we suppose $\gamma_j = \frac{1}{k_I}$ and $\sigma_j = \sigma > 0$, $j = 1, \ldots, k_I$ are given constants then the problem of centers estimation could be reduced to the following maximization problems

$$\max_{c_1,\ldots,c_k \in \mathbb{R}} \ln L^{(p)} \left(\frac{1}{k_I}, \ldots, \frac{1}{k_I}, c_1, \ldots, c_{k_I}, \sigma, \ldots, \sigma \right), \quad p = 1, 2$$

which is obviously equivalent to the following minimization problems

$$\min_{c_1,\ldots,c_{k_I} \in \mathbb{R}} F_\sigma^{(p)}(c_1, \ldots, c_{k_I}), \quad F_\sigma^{(p)}(c_1, \ldots, c_{k_I}) = -2\sigma^p \sum_{i=1}^{m} \ln \sum_{j=1}^{k_I} e^{-\frac{1}{2}\left|\frac{I_i - c_j}{\sigma}\right|^p}, \quad p = 1, 2.$$

Since for every vector $r = (r_1, \ldots, r_n) \in \mathbb{R}^n$ holds (see [9]) $\max\limits_{1 \le j \le n} r_j = \lim\limits_{\epsilon \to 0+} \epsilon \ln \left(\sum\limits_{j=1}^{n} \exp\left(\frac{r_j}{\epsilon}\right) \right)$, and $\min\limits_{1 \le j \le n} r_j = -\max\limits_{1 \le j \le n} (-r_j)$ it follows immediately

$$\lim_{\sigma \to 0^+} F_\sigma^{(p)}(c_1, \ldots, c_{k_I}) = F^{(p)}(c_1, \ldots, c_{k_I}), \quad p = 1, 2.$$

Generally, in the weighted case it is reasonable to consider the following weighted–likelihood function

$$L^{(p)} \left(\frac{1}{k_I}, \ldots, \frac{1}{k_I}, c_1, \ldots, c_{k_I}, \sigma, \ldots, \sigma \right) = \prod_{i=1}^{m} \left(\sum_{j=1}^{k_I} \frac{1}{k_I \sigma} \varphi^{(p)} \left(\frac{I_i - c_j}{\sigma} \right) \right)^{w_i^I},$$

which maximization is equivalent to the following minimization problem

$$\min_{c_1,\ldots,c_{k_I}} F_\sigma^{(p)}(c_1, \ldots, c_{k_I}),\ F_\sigma^{(p)}(c_1, \ldots, c_{k_I}) = -2\sigma^p \sum_{i=1}^{m_I} w_i^I \ln \sum_{j=1}^{k_I} e^{-\frac{1}{2}\left|\frac{I_i - c_j}{\sigma}\right|^p}, \quad p = 1, 2,$$

that in the limit case $\sigma \to 0^+$ converges to the corresponding weighted objective function $F^{(p)}$f.

3.2. Data representation with fuzzy membership function

Let $\sigma > 0$ be a given positive number and let us suppose that the optimal centers $c_j^I(p)$ of clusters $\pi_j^I(p)$, $j = 1, \ldots, k_I$, $p = 1, 2$ have been determined. Motivated by the paper [22] and Maximum likelihood approach, to every intensity-data I_i with respect to center $c_j^I(p)$ we can assign the value $\omega_\sigma^{(p)}(I_i, \pi_j^I(p))$, where

$$x \mapsto \omega_\sigma^{(p)}(x, c_j^I(p)) := \frac{\varphi^{(p)}\left(\frac{x - c_j^I(p)}{\sigma}\right)}{\sum_{l=1}^{k_I} \varphi^{(p)}\left(\frac{x - c_l^I(p)}{\sigma}\right)} = \frac{e^{-\frac{1}{2}\left|\frac{x - c_j^I(p)}{\sigma}\right|^p}}{\sum_{l=1}^{k_I} e^{-\frac{1}{2}\left|\frac{x - c_l^I(p)}{\sigma}\right|^p}},$$

is so called fuzzy membership function.

Note that

$$\sum_{j=1}^{k_I} \omega_\sigma^{(p)}(I_i, c_j^I(p)) = 1 \text{ and } \lim_{\sigma\to 0^+} \omega_\sigma^{(p)}(I_i, c_j^I(p)) = \begin{cases} \frac{1}{\mu_i^{(p)}}, & \text{if } j \in U_i^{(p)} \\ 0, & \text{if } j \in \{1,\dots,k\} \setminus U_i^{(p)}, \end{cases}$$

where $U_i^{(p)}$ is defined by (5) for corresponding distance-like function $d(x,y) = d_p(x,y)$, $p = 1,2$, and $\mu_i^{(p)} = |U_i^{(p)}|$. In this context every intensity-data I_i with respect to partition $\Pi^I(p)$ could be represented by k_I-tuple

$$\left(\omega_\sigma^{(p)}(I_i, c_1^I(p)), \dots, \omega_\sigma^{(p)}(I_i, c_{k_I-1}^I(p)), \omega_\sigma^{(p)}(I_i, c_{k_I}^I(p))\right) \in [0,1]^{k_I}.$$

Since $\sum_{j=1}^{k_I} \omega_\sigma^{(p)}(I_i, c_j^I(p)) = 1$ it is sufficient to consider $k_I - 1-$tuple representation with respect to partition $\Pi^I(p)$:

$$I_i \equiv I_i(\sigma, \Pi^I(p)) = \left(\omega_\sigma^{(p)}(I_i, c_1^I(p)), \dots, \omega_\sigma^{(p)}(I_i, c_{k_I-1}^I(p))\right) \in [0,1]^{k_I-1},\ i = 1,\dots,m_I.$$

4. Least square and least absolute deviation one dimensional weighted clustering in hue space

The set $\mathcal{H} = \{H_i : i = 1,\dots,m_H\} \subset [0, 2\pi\rangle$ of hue-data, with corresponding weights $w_i^H > 0$ has to be divided into k_H clusters $\Pi^H = \{\pi_1^H, \dots, \pi_{k_H}^H\}$, $1 \le k_H \le m_H$. Note that without loss of generality we can suppose that $H_i \ne H_j$, $i \ne j$. In this section, we consider a one-dimensional clustering problem using the Least Absolute Deviation (LAD) – optimality criterion, based on LAD distance function [11, 18]:

$$D_1(x,y) = \begin{cases} |x-y|, & |x-y| \le \pi \\ 2\pi - |x-y|, & \text{else} \end{cases} = \min\{|x-y|, 2\pi - |x-y|\} = \pi - ||x-y| - \pi|.$$

and Least Squares (LS) – optimality criterion, based on LS distance–like function [11, 22]: $D_2(x,y) = D_1^2(x,y) = (\pi - ||x-y| - \pi|)^2$. The problem of finding an optimal partition of the set $\mathcal{H}$ according to (2) reduces to the following nonconvex and nonsmooth optimization problem

$$\min_{c_1,\dots,c_{k_H} \in [0,2\pi\rangle} G^{(p)}(c_1,\dots,c_{k_H}), \quad G^{(p)}(c_1,\dots,c_{k_H}) = \sum_{i=1}^{m_H} \min_{j=1,\dots,k_H} w_i^H D_p(c_j, H_i),\ p = 1,2.$$

In order to apply the $k-$means algorithm it is necessary to efficiently determine the centers of clusters

$$c_j^H(p) = \operatorname*{argmin}_{x\in[0,2\pi\rangle} \sum_{H_i \in \pi_j^H(p)} w_i^H D_p(x, H_i),\ j = 1,\dots,k_H,\ p = 1,2. \tag{6}$$

In the paper [22] very useful formula for calculating the centers of clusters in hue space was given. Unfortunately, it is only a local solution of the problem (6). Instead of this, some efficient numerical method for a global optimization should be used.

One of the most popular algorithms for solving a global optimization problem for the Lipschitz continuous function is the `DIRECT` (DIvidingRECTangles) algorithm [6, 7]. The `DIRECT` algorithm requires the objective function to be Lipschitz continuous. In this context let us show the following proposition.

Proposition 1. *Functions* $g_p : [0, 2\pi\rangle \to \mathbb{R}_+$, $g_p(x) = \sum_{H_i \in \pi_j^H(p)} D_p(x, H_i)$, $p = 1, 2$ *are Lipschitz continuous on* $[0, 2\pi\rangle$, *i.e. there exists* $L > 0$ *such that*

$$g_p(\alpha) - g_p(\beta) \leq L|\alpha - \beta|, \quad \forall \alpha, \beta \in [0, 2\pi\rangle$$

Proof. Since the proofs for $p = 1$ and $p = 2$ are very similar, we will prove only the case $p = 1$. Let $\alpha, \beta \in [0, 2\pi\rangle$, then it holds

$$\begin{aligned} g_1(\alpha) - g_1(\beta) &= \sum_{H_i \in \pi_j^H(1)}^{m} w_i^H ||\alpha - H_i| - \pi| - \sum_{H_i \in \pi_j^H(1)} w_i^H ||\beta - H_i| - \pi| \\ &\leq \sum_{H_i \in \pi_j^H(1)} w_i^H ||\alpha - H_i| - |\beta - H_i|| \leq \sum_{i=1}^{m} w_i^H |\alpha - \beta|. \end{aligned}$$

Analogously it can be shown $g_1(\beta) - g_1(\alpha) \leq \sum\limits_{H_i \in \pi_j^H(1)} w_i^H |\alpha - \beta|$, and finally

$$g_1(\alpha) - g_1(\beta) \leq L|\alpha - \beta|, \quad L = \sum_{H_i \in \pi_j^H(1)} w_i^H.$$

□

4.1. Connection with the Maximum likelihood approach

Let $c_j^H(p)$ be the centers of clusters $\pi_j^H(p)$, $j = 1, \ldots, k_H$ and $\sigma_j > 0$, $j = 1, \ldots, k_H$ given positive numbers. Let us suppose that hue-data H_i, $i = 1, \ldots, m_H$ are independent and comes from mixing distribution with probability density function (see e.g. [11]):

$$g^{(p)}(x; \lambda_1, \ldots, \lambda_{k_H}, c_1^H(p), \ldots, c_{k_H}^H(p), \sigma_1, \ldots, \sigma_{k_H}) = \sum_{j=1}^{k_H} \frac{\lambda_j}{\sigma_j} \sum_{l=-\infty}^{\infty} \varphi^{(p)} \left(\frac{x - c_j^I(p) + 2l\pi}{\sigma_j} \right),$$

$\sum_{j=1}^{k_H} \lambda_j = 1$, $\lambda_j \geq 0$, $j = 1, \ldots, k_H$, where $x \mapsto \frac{1}{\sigma_j} \sum_{l=-\infty}^{\infty} \varphi^{(p)} \left(\frac{x - c_j^I(p) + 2l\pi}{\sigma_j} \right)$, $p = 1, 2$ are respectively a probability density function of Wrapped Laplace random variable $\mathcal{LW}(c_j^I(1), \sigma_j^2)$ i.e. wrapped normal random variable $\mathcal{NW}(c_j^I(2), \sigma_j^2)$, and $\varphi^{(p)}(x) = \frac{1}{\sqrt{2}} e^{-|x|^p/2}$. Completely analogously as in the intensity space it can be shown that in the limit case for sufficiently small variances (σ_j^2, $j = 1, \ldots, k_H$) the corresponding negative log-likelihood function can be approximated by the objective function $G^{(p)}$.

4.2. Data representation with fuzzy membership function

Let $\sigma > 0$ be a given positive number and let us suppose that the optimal centers $c_j^H(p)$ of clusters $\pi_j^H(p)$, $j = 1, \ldots, k_H$, $p = 1, 2$ have been determined. Analogously to the intensity space case, any pixel with respect to partition $\Pi^H(p)$ could be represented by $k_H - 1-$ tuple

$$H_i \equiv H_i(\sigma, \Pi^H(p)) = \Big(v_\sigma^{(p)}(H_i, c_1^H(p)), \ldots, v_\sigma^{(p)}(H_i, c_{s-1}^H(p))\Big) \in [0,1]^{k_H - 1},\ i = 1, \ldots, m_H,$$

where the corresponding fuzzy membership function is given by the following formula:

$$v_\sigma^{(p)}(x, c_j^H(p)) := \frac{\varphi^{(p)}\left(\frac{x - c_j^H(p) - 2\pi}{\sigma}\right) + \varphi^{(p)}\left(\frac{x - c_j^H(p)}{\sigma}\right) + \varphi^{(p)}\left(\frac{x - c_j^H(p) + 2\pi}{\sigma}\right)}{\sum_{l=1}^{k_H} \varphi^{(p)}\left(\frac{x - c_l^H(p) - 2\pi}{\sigma}\right) + \varphi^{(p)}\left(\frac{x - c_l^H(p)}{\sigma}\right) + \varphi^{(p)}\left(\frac{x - c_l^H(p) + 2\pi}{\sigma}\right)}.$$

5. Multidimensional weighted clustering

Any pixel a_i, $i = 1, \ldots, m$ is uniquely determined with pair (I_i, H_i), where I_i and H_i are corresponding values of intensity and hue. Here we suppose that $m_I, m_H \leq m$. Let $\Pi^I(p)$ and $\Pi^H(p)$, be optimal partitions obtained by clustering in intensity and hue space. According to [10], a *k-means algorithm* is initiated with e.g. 100 different randomly generated initial centers, and the one that gives the smallest value of the objective function (2) is taken as a solution.

With respect to partitions $\Pi^I(p)$ and $\Pi^H(p)$ every pixel a_i, $i = 1, \ldots, m$ could be represented by $(k_I + k_H - 2)-$tuple

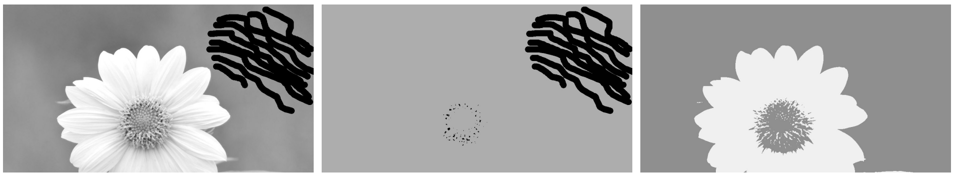
$$\begin{aligned} a_i &\equiv a_i(\sigma, \Pi^I(p), \Pi^H(p)) \\ &= \Big(\omega_\sigma(I_i, c_1^I(p)), \ldots, \omega_\sigma(I_i, c_{k_I - 1}^I(p)), v_\sigma^{(p)}(H_i, c_1^H(p)), \ldots, v_\sigma^{(p)}(H_i, c_{k_H - 1}^H(p))\Big) \in [0,1]^{k_I + k_H - 2}. \end{aligned}$$

Finally, the problem of image segmentation could be considered as a clustering in $(k_I + k_H - 2)$-dimensional space. For this purpose *k-means algorithm* based on distance-like function $d_p : [0,1]^{k_I + k_H - 2} \times [0,1]^{k_I + k_H - 2} \to [0, \infty\rangle$, defined by $d_p(x, y) = \|x - y\|_p^p$, $p = 1, 2$ with e.g. 100 randomly initialization should be used.

6. Numerical and illustrative examples

Example 1. *Figure 1 shows a greyscale image with "noise" (outliers) which is clustered in intensity space with $k_I = 2$ clusters. It is noticeable that LS – optimality criterion separates outliers in the individual cluster, while the rest of the image falls into another cluster. Nevertheless, LAD – optimality criterion ignores those outliers and segments the image more accurately.*

Example 2. *Similar example of clustering in hue space using $k_H = 2$ clusters is shown in Figure 2. LS optimality criterion stresses the outliers, while LAD optimality criterion ignores them.*



(a) Original image (b) LS based clustering (c) LAD based clustering

Figure 1: Weighted clustering in intensity space

(a) Original image (b) LS based clustering (c) LAD based clustering

Figure 2: Weighted clustering in hue space

Example 3. *Figure 3 shows the benefits of multidimensional weighted clustering. When the test image of an athletic track is segmented in intensity space with* $k_I = 2$ *clusters, numbers on the track are well separated, but pixels showing the grass and the track falls into the same cluster. Similar situation occurs when clustering in hue space with* $k_H = 2$ *clusters: the grass and the track are separated, but the numbers fall into the same cluster as the track. However, when we perform multidimensional weighted clustering (Section 5) in two-dimensional space* $(k_I + k_H - 2 = 2)$ *into 3 clusters, we obtain an accurate and well separated representation of the original image. Let us note that the appropriate number of clusters has been determined in accordance with Silhouette width criterion (see [8]).*

(a) Original image (b) Clustering in intensity space (c) Clustering in hue space (d) Multidimensional clustering

Figure 3: Multidimensional weighted clustering

Acknowledgement

The authors would like to thank Professor Rudolf Scitovski (Department of Mathematics, University of Osijek) for his useful comments and remarks.

References

[1] Carron, T., Lambert, P.(1994). Color Edge Detector Using Jointly Hue, Saturation and Intensity, International Conference on Image Processing, 977–981

[2] Cupec, R., Grbić, R., Sabo, K., Scitovski R.(2009). Three points method for searching the best least absolute deviations plane, Applied Mathematics and Computation, 215, 983–994

[3] Dhillon, I. S., Guan, Y., Kulis, B.(2004). Kernel k-means, spectral clustering and normalized cuts, Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), August 22–25, Seattle, Washington, USA, 551–556,

[4] Domínguez, E., Muñoz, J.(2005). Applying bio-inspired techniques to the p-median problem, Computational Intelligence Bioinspired Syst., 8th Int. Workshop Artificial Neural Networks, Springer-Verlag, Berlin, 67–74

[5] Finkel, D. E., Kelley, C. T.(2006). Additive scaling and the DIRECT algorithm, J. Glob. Optim. 36, 597–608

[6] Grbić, R., Nyarko, K. E., Scitovski, R.(2013). A modification of the DIRECT method for Lipschitz global optimization for a symmetric function, Journal of Global Optimization, 57, 1193–1212.

[7] Jones, D. R., Perttunen, C. D., Stuckman, B. E.(1993). Lipschitzian optimization without the Lipschitz constant, J. Optim. Theory Appl. 79, 157–181

[8] Kaufman, L., Rousseeuw, P. (2005). Finding groups in data: An introduction to cluster analysis, John Wiley & Son, Hoboken, USA

[9] Kogan, J.(2007). Introduction to Clustering Large and High-Dimensional Data, Cambridge University Press

[10] Leisch, F.(2007). A toolbox for K-centroids cluster analysis, Computational Statistics & Data Analysis 51, 526–544

[11] Mardia, K. V., Jupp, P.(2000). Directional Statistics, John Wiley and Sons Ltd., 2nd edition

[12] Pal, N. R., Pal, S. K.(1993), A review on image segmentation techniques, Pattern Recognition, 26, 1277–1294

[13] Reese, J.(2006). Solution methods for the p-median problem: an annotated bibliography, Published online in Wiley InterScience, Wiley

[14] Sabo, K.(2014). Center-based l_1-clustering method, International Journal of Applied Mathematics and Computer Science, 24, 151–163

[15] Sabo, K., Scitovski, R., Vazler, I.(2013). One-dimensional center-based l_1-clustering method, Optimization Letters, 7, 5–22

[16] Sabo, K., Scitovski, R., Vazler, I.(2011). Searching for a best LAD-solution of an overdetermined system of linear equations motivated by searching for a best LAD-hyperplane on the basis of given data, J. Optim. Theory Appl. 149, 293–314

[17] Schöbel, A., Scholz, D.(2010). The big cube small cube solution method for multidimensional facility location problems, Computers & Operations Research, 37, 115–122

[18] Scitovski, S., Scitovski, R.(2012). Cluster analysis of the data on unit circle, In Proceedings of The 1st Virtual International Conference in Advanced Research in Scientific Areas (ARSA-2012) Slovakia, December 3 - 7, 1574–1577

[19] Späth, H.(1983). Cluster-Formation und Analyse, R. Oldenburg Verlag, München
[20] Teboulle, M.(2007). A unified continuous optimization framework for center-based clustering methods, Journal of Machine Learning Research 8, 65–102
[21] Vazler, I., Sabo, K., Scitovski, R.(2012). Weighted median of the data in solving least absolute deviations problems, Communications in Statistics - Theory and Methods, 41, 1455-1465
[22] Zhang, C., Wang, P.(2000), A new method for color image segmentation based on intensity and hue clustering, In Proceedings of the 15th ICPR, Barcelona, 3, 613–616