

## Color image segmentation based on intensity and hue clustering – a comparison of LS and LAD approaches

Petar Taler<sup>1,\*</sup> and Kristian Sabo<sup>1</sup>

<sup>1</sup> Department of Mathematics, Josip Juraj Strossmayer University of Osijek  
Trg Lj. Gaja 6, 31 000 Osijek, Croatia  
E-mail: `{petar, ksabo}@mathos.hr`

**Abstract.** Motivated by the method for color image segmentation based on intensity and hue clustering proposed in [26] we give some theoretical explanations for this method that directly follows from the natural connection between the maximum likelihood approach and Least Squares or Least Absolute Deviations 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: October 26, 2014; accepted: December 18, 2014; available online: December 30, 2014

### 1. Introduction

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

Most attention with respect to image segmentation has been focused on gray scale images (see, e.g., [15]). 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 this does not coincide with the vision psychology of human eyes because of a 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, *I* component represents *intensity*, *H* component represents *hue*, and *S* component represents *saturation*. To convert RGB representation to HSI representation, first compute [1]:

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

\*Corresponding author.

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, a color image could be represented (see, e.g., [26]) by intensity-data set  $\mathcal{I} = \{I_i \in \mathbb{R} : i = 1, \dots, m_I\} \subset \mathbb{R}$ , with corresponding weights  $w_i^I > 0$ , where  $I_i \neq I_j, i \neq j$ , and hue-data set  $\mathcal{H} = \{H_i : i = 1, \dots, m_H\} \subset [0, 2\pi)$  with corresponding weights  $w_i^H > 0$ , where  $H_i \neq H_j, i \neq j$ , and  $m_I, m_H$  are numbers of possible different discrete values for intensity and hue components, respectively.

There is plenty of literature considering the image segmentation problem (see, e.g., [1], [2], [15], [26]). Our starting point is the method proposed in paper [26] which deals with segmentation of an image represented in HSI space based on the least squares criterium. According to this method, image segmentation tasks are performed separately in the intensity and hue spaces. The results obtained in that way are associated with the so-called fuzzy membership function, the data is mapped into a multidimensional space and the multi-dimensional clustering problem is solved. For the purpose of the hue-data segmentation process, it is necessary to determine the centers of clusters, which is reduced to solving the problem of nonlinear optimization with a large number of local extrema. To address this problem, the method which finds the stationary point of the corresponding objective function that does not necessarily have to be the point of the global minimum is proposed in [26]. Furthermore, a natural criterion for defining the fuzzy membership function is not given in [26]. Motivated by the method in the cited paper, in this paper: (i) we show that the objective function used for determining cluster centers in huespace is Lipschitz continuous, hence the known global DIRECT method can be used for its minimization; (ii) on the basis of the maximum likelihood approach we construct a specific natural fuzzy membership function; (iii) besides the Least Squares approach, we also observe the Least Absolute Deviations approach, which is particularly important in the case of the presence of so-called outliers among data; (iv) we illustrate our methods by means of several graphical examples.

The paper is organized as follows. In the next section, a brief 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 the corresponding connection with the maximum likelihood approach is given. On the basis of the relation between one-dimensional optimal partitions, an 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  $\mathcal{A} = \{a_i \in \mathbb{R}^n : i = 1, \dots, m\} \subset \mathbb{R}^n$  with the corresponding weights  $w_i > 0$  into  $k$  disjoint subsets  $\pi_1, \dots, \pi_k$ ,  $1 \leq k \leq m$ , such that

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

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

Any function  $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ ,  $\mathbb{R}_+ := [0, +\infty)$ , that has the following two properties

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \quad d(x, y) \geq 0 \quad d(x, y) = 0 \quad \text{iff} \quad x = y,$$

(see, e.g., [11, 24]) is called a distance-like function. Let  $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$  be some distance-like function by applying the *minimum distance condition* (see, e.g., [11, 23]); to each cluster  $\pi_j \in \Pi$  we can associate its center  $c_j$  defined by

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

If we define an objective function  $\mathcal{F}: \mathcal{P}(\mathcal{A}, k) \rightarrow [0, +\infty)$  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  $\Pi^*$ , such that  $\mathcal{F}(\Pi^*) = \min_{\Pi \in \mathcal{P}(\mathcal{A}, k)} \mathcal{F}(\Pi)$ .

Conversely, for a given set of centers  $c_1, \dots, c_k \in \mathbb{R}^n$ , by applying the minimal distance condition we can define the partition  $\Pi = \{\pi_1, \dots, \pi_k\}$  of the set  $\mathcal{A}$  in the following way:  $\pi_j = \{a \in \mathcal{A} : d(c_j, a) \leq d(c_s, a), \forall s = 1, \dots, k\}$ ,  $j = 1, \dots, k$ , 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

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

where  $F: \mathbb{R}^{kn} \rightarrow \mathbb{R}_+$ . 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 [5, 16, 20].

The most known algorithm for searching for a locally optimal partition is the  $k$ -median algorithm [4, 13], which can be described by two steps which are iteratively repeated.

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

$$\pi_j = \{a \in \mathcal{A} : d(c_j, a) \leq d(c_s, a), \forall s = 1, \dots, k\}, \quad j = 1, \dots, k. \quad (3)$$

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

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

Let us mention that for the least squares distance-like function  $d(x, y) = \|x - y\|_2^2$  the previous algorithm is also known as the  $k$ -means algorithm.

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

$$U_i = \{j \in J : d(c_j, a_i) \leq d(c_s, a_i), \forall s \in J\}, \quad J = \{1, \dots, k\}. \quad (5)$$

Note that the set  $U_i$  is not empty, 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  $\Pi = \{\pi(z_1), \dots, \pi(z_k)\}$  is said to be a *well-separated partition*.

### 3. Least squares and least absolute deviations one-dimensional weighted clustering in intensity space

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

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

It is generally well known that the LAD approach ignores outliers among the data [3, 19], while the LS approach stresses them.

In order to apply the  $k$ -median 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  $\pi_j^I(p), j = 1, \dots, k_I$  can be explicitly determined by the following formula:

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

where  $\text{med}_{I_i \in \pi_j^I(1)}(w_i^I, I_i)$  is a weighted median of the set  $\pi_j^I(1)$  (see, e.g., [17, 18, 25])

and  $\text{mean}_{I_i \in \pi_j^I(2)}(w_i^I, I_i) = \frac{\sum_{I_i \in \pi_j^I} w_i^I I_i}{\sum_{I_i \in \pi_j^I} w_i^I}$  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.

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

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

$\sum_{j=1}^{k_I} \lambda_j = 1$ ,  $\lambda_j \geq 0$ ,  $j = 1, \dots, k_I$ , where  $x \mapsto \frac{1}{\sigma_j} \varphi^{(p)}\left(\frac{x - c_j^I(p)}{\sigma_j}\right)$ ,  $p = 1, 2$  are 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

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

It can be shown that in the limit case for sufficiently small variances ( $\sigma_j^2$ ,  $j = 1, \dots, k_I$ ) the corresponding negative log-likelihood function can be approximated by the objective function  $F^{(p)}$  (see, e.g., [12]).

### 3.2. Data representation with the 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  $\pi_j^I(p)$ ,  $j = 1, \dots, k_I$ ,  $p = 1, 2$  have been determined. Motivated by paper [26] and the 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 the so-called fuzzy membership function.

Note that

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

where  $U_i^{(p)}$  is defined by (5) for the 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 the  $k_I$ -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, a  $k_I - 1$ -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  $\mathcal{H} = \{H_i : i = 1, \dots, m_H\} \subset [0, 2\pi)$  of hue-data, with corresponding weights  $w_i^H > 0$  has to be divided into  $k_H$  clusters  $\Pi^H = \{\pi_1^H, \dots, \pi_{k_H}^H\}$ ,  $1 \leq k_H \leq m_H$ . Note that without loss of generality we can suppose that  $H_i \neq H_j$ ,  $i \neq j$ . In this section, we consider a one-dimensional clustering problem using the Least Absolute Deviation (LAD) optimality criterion, based on the LAD distance function [14, 22]:

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

and the Least Squares (LS) optimality criterion the based on the LS distance-like function [14, 26]:  $D_2(x, y) = D_1^2(x, y) = (\pi - ||x - y| - \pi|)^2$ . The problem of finding an optimal partition of the set  $\mathcal{H}$  according to (2) reduces to the following nonconvex and nonsmooth optimization problem

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

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

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

In paper [26], a very useful formula for calculating the centers of clusters in hue space was given. Unfortunately, it is only a local solution of 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 [7, 8]. The DIRECT algorithm requires the objective function to be Lipschitz continuous. In this context, let us show the following proposition.

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

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

**Proof.** Since the proofs for  $p = 1$  and  $p = 2$  are very similar, we will prove only the case  $p = 1$ . Let  $\alpha, \beta \in [0, 2\pi)$ ; 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 that  $g_1(\beta) - g_1(\alpha) \leq \sum_{H_i \in \pi_j^H(1)} w_i^H |\alpha - \beta|$ , 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  $\pi_j^H(p)$ ,  $j = 1, \dots, k_H$  and  $\sigma_j > 0$ ,  $j = 1, \dots, k_H$  given positive numbers. Let us suppose that hue-data  $H_i$ ,  $i = 1, \dots, m_H$  are independent and come from mixing distribution with probability density function (see, e.g., [14]):

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

$\sum_{j=1}^{k_H} \lambda_j = 1$ ,  $\lambda_j \geq 0$ ,  $j = 1, \dots, k_H$ , where  $x \mapsto \frac{1}{\sigma_j} \sum_{l=-\infty}^{\infty} \varphi^{(p)} \left( \frac{x - c_j^H(p) + 2l\pi}{\sigma_j} \right)$ ,  $p = 1, 2$  are a probability density function of wrapped Laplace random variable  $\mathcal{LW}(c_j^H(p), \sigma_j^2)$ , i.e., wrapped normal random variable  $\mathcal{NW}(c_j^H(p), \sigma_j^2)$ , and  $\varphi^{(p)}(x) = \frac{1}{\sqrt{2}} e^{-|x|^p/2}$ . It can be shown that in the limit case for sufficiently small variances ( $\sigma_j^2$ ,  $j = 1, \dots, k_H$ ) the corresponding negative log-likelihood function can be approximated by the objective function  $G^{(p)}$  (see, e.g., [14]).

#### 4.2. Data representation with the 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  $\pi_j^H(p)$ ,  $j = 1, \dots, k_H$ ,  $p = 1, 2$  have been determined. Analogously to the intensity space case, any pixel with respect to partition  $\Pi^H(p)$  could be represented by the  $k_H - 1$ -tuple

$$H_i \equiv H_i(\sigma, \Pi^H(p)) = \left( v_\sigma^{(p)}(H_i, c_1^H(p)), \dots, v_\sigma^{(p)}(H_i, c_{k_H-1}^H(p)) \right) \in [0, 1]^{k_H-1}, \quad i = 1, \dots, m_H,$$

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  $a_i$ ,  $i = 1, \dots, m$  is uniquely determined by a 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 [13], a *k-median 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  $a_i$ ,  $i = 1, \dots, m$  could be represented by the  $(k_I + k_H - 2)$ -tuple

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

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

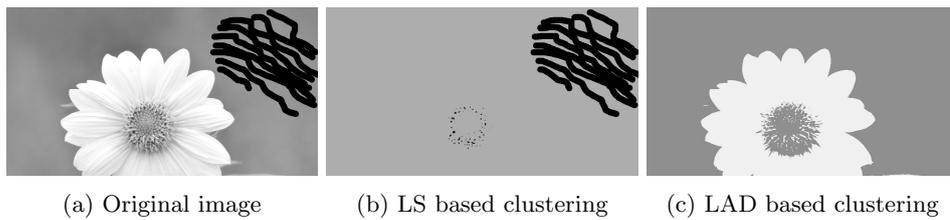
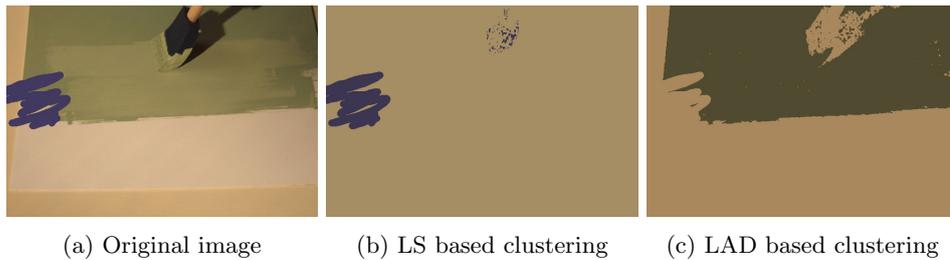
## 6. Numerical and illustrative examples

We have implemented all described methods in a program written in C++ programming language and tested it with different parameters on the following examples.

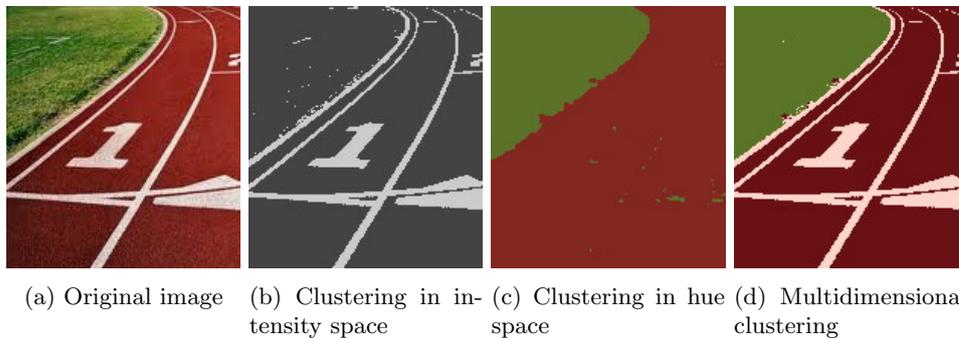
**Example 1.** *Figure 1 shows a greyscale image with “noise” (outliers) which is clustered in intensity space with  $k_I = 2$  clusters. It is noticeable that the LS – optimality criterion separates outliers in the individual cluster, while the rest of the image falls into another cluster. Nevertheless,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

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

**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 fall into the same cluster. A similar situation occurs when clustering in hue space with  $k_H = 2$  clusters: the grass and the track are separated, but the numbers*

Figure 1: *Weighted clustering in intensity space*Figure 2: *Weighted clustering in hue space*

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 the Silhouette width criterion (see [10]).

Figure 3: *Multidimensional weighted clustering*

## 7. Conclusions

Some theoretical improvements of results given in [26] are proposed in this paper. First of all, the global method DIRECT for determining the centers of clusters in hue space is proposed. For this purpose, theoretical assumptions on the corresponding objective function ensuring suitability of the DIRECT method are demonstrated. The

fuzzy membership function connecting hue and intensity spaces is constructed on the basis of motivation provided by the maximum likelihood approach. The method is extended on the Least Absolute Deviations criterion and several examples that illustrate the differences between the approaches are presented. It is shown that the Least Absolute Deviation approach is much less sensitive to the presence of outliers compared to the Least Squares approach.

In further research, instead of multiple launching of the  $k$ -median algorithm, we will endeavor to construct an efficient method for finding a nearly global optimal partition based on the the method described in [21], and a method which automatically determines the appropriate number of clusters in a partition.

## Acknowledgement

The authors would like to thank Professor Rudolf Scitovski (Department of Mathematics, Josip Juraj Strossmayer University of Osijek) for his useful comments and remarks. We are also thankful to anonymous referees and journal editors for their careful reading of the paper and insightful comments that helped us improve the paper. This work is supported by the Ministry of Science, Education and Sports, Republic of Croatia.

## References

- [1] Carron, T., Lambert, P.(1994). Color edge detector using jointly hue, Saturation and Intensity, International Conference on Image Processing, 977–981.
- [2] Cheng, H. D. Jiang, X. H., Sun Y.(1988) Survey on color image segmentation, The First International Workshop on CVPRIP, Triangle Park, North Carolina.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Finkel, D. E., Kelley, C. T.(2006). Additive scaling and the DIRECT algorithm, J. Glob. Optim. 36, 597–608.
- [7] 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.
- [8] Jones, D. R., Perttunen, C. D., Stuckman, B. E.(1993). Lipschitzian optimization without the Lipschitz constant, J. Optim. Theory Appl. 79, 157–181.
- [9] Jung, R.(2007). Unsupervised multiscale segmentation of color images, Pattern Recognition Letters, 28, 523–533.
- [10] Kaufman, L., Rousseeuw, P. (2005). Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Son, Hoboken, USA.

- [11] Kogan, J.(2007). Introduction to Clustering Large and High-Dimensional Data, Cambridge University Press.
- [12] Kung, S. Y.(2014). Kernel Methods and Machine Learning, Cambridge University Press.
- [13] Leisch, F.(2007). A toolbox for K-centroids cluster analysis, Computational Statistics & Data Analysis 51, 526–544.
- [14] Mardia, K. V., Jupp, P.(2000). Directional Statistics, John Wiley and Sons Ltd., 2nd edition.
- [15] Pal, N. R., Pal, S. K.(1993), A review on image segmentation techniques, Pattern Recognition, 26, 1277–1294.
- [16] Reese, J.(2006). Solution methods for the  $p$ -median problem: an annotated bibliography, Published online in Wiley InterScience, Wiley.
- [17] Sabo, K.(2014). Center-based  $l_1$ -clustering method, International Journal of Applied Mathematics and Computer Science, 24, 151–163.
- [18] Sabo, K., Scitovski, R., Vazler, I.(2013). One-dimensional center-based  $l_1$ -clustering method, Optimization Letters, 7, 5–22.
- [19] 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.
- [20] Schöbel, A., Scholz, D.(2010). The big cube small cube solution method for multidimensional facility location problems, Computers & Operations Research, 37, 115–122.
- [21] Scitovski, R., Scitovski, S. (2013). A fast partitioning algorithm and its application to earthquake investigation, Computers & Geosciences 59, 124–131.
- [22] 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.
- [23] Späth, H.(1983). Cluster-Formation und Analyse, R. Oldenburg Verlag, München.
- [24] Teboulle, M.(2007). A unified continuous optimization framework for center-based clustering methods, Journal of Machine Learning Research 8, 65–102.
- [25] 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.
- [26] 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.