

Frequency-Based Deformable Models as a Distance Metric

Stelios Krinidis, Michail Krinidis and Vassilios Chatzis

Information Management Department
Technological Institute of Kavala
Ag. Loukas, 65404, Kavala, Greece
stelios.krinidis@mycosmos.gr; mkrinidi@gmail.com; chatzis@teikav.edu.gr

Received June 2012; revised September 2012

ABSTRACT. *This paper introduces a distance metric between two distributions that we call the Deformation Distance (DeD). The DeD is based on the “energy” that must be paid to deform one distribution into the other, presenting a perceptual similarity match better than other distribution distances. The DeD relies on the distribution’s frequency-based features. The frequency-based features are extracted by a physics-based deformable model that parameterizes the distribution. The DeD was evaluated on a variety of random as well as real distributions. Also, DeD was evaluated on image clustering and compared to other distances. The experimental results demonstrate the efficiency of the proposed distance metric.*

Keywords: Deformable model, modal analysis, finite element method, deformable curves, distance metric, multidimensional distribution.

1. **Introduction.** Image segmentation is one of the first and most important tasks in image analysis and computer vision. In the literature, various methods have been proposed for object segmentation and feature extraction, described in [1, 5, 7, 8, 11, 15, 18, 26, 27, 31]. However, the design of robust and efficient segmentation algorithms is still a very challenging research topic, due to the variety and complexity of images. Image segmentation is defined as the partitioning of an image into non-overlapped, consistent regions which are homogeneous in respect to some characteristics such as intensity, color, tone, texture, etc. The image segmentation can be divided into four categories: thresholding, clustering, edge detection and region extraction. In this paper, a clustering method for image segmentation will be considered.

Clustering is a process for classifying objects or patterns in such a way that samples of the same cluster are more similar to one another than samples belonging to different clusters [6, 12]. There are two main clustering strategies: the hard clustering scheme and the fuzzy clustering scheme. The conventional hard clustering methods classify each point of the data set just to one cluster [12]. As a consequence, the results are often very crisp, i.e., in image clustering each pixel of the image belongs just to one cluster. On the other hand, fuzzy set theory [32] has introduced the idea of partial membership, described by a membership function. Fuzzy c-means (FCM) algorithm [4] is the most popular method used in image segmentation because it has robust characteristics for ambiguity and can retain much more information than hard segmentation methods [9, 10, 25].

All clustering methods share a common problem, how “close” or “far” is one distribution from the other in a consistent manner. Hence, it is necessary to measure distances between

two distribution using a “distance metric”. A distance metric is a more than something that starts at zero and gets bigger as it get farther away. It must be defined in such a way that the shortest distance between any two distributions is a straight line. The effectiveness of a clustering method is tied to the exploited distance metric.

In this paper, we introduce a distance metric between two distributions, that we call *Deformation Distance* (DeD). DeD is a useful and flexible distance metric, based on the “energy” that must be paid to deform one distribution into the other. The DeD relies on the distributions’ frequency-based features. The frequency-based features are extracted by the free vibrations of a physics-based deformable model [13, 20, 21] that parameterizes the distributions under consideration. When used to compare distributions that have the same size, the DeD is a true metric.

The remainder of the paper is organized as follows. In Section 2, we review and survey some of the existing measures of similarity and their drawbacks. The physics-based deformable model [13] used as the frequency-based feature generator is presented in Section 3. In Section 4, the deformation distance (DeD) is introduced. Experimental results are presented in Section 5 and conclusions are drawn in Section 6.

2. Previous Work. Image clustering systems usually represent image features by multi-dimensional histograms. For example, the color content of an image is defined by the distribution of its pixels in some color space. These distributions are used to separate image clusters of an image. Thus, a measure of similarity between distributions must be defined. In this Section, distributions are formally defined, and the most common distribution similarity measures for image clustering are presented.

A distribution P is a function that maps an image and a set of reference colors and other characteristics into a vector p_i of nonnegative numbers. These vectors typically represent bins (or their centers) in a fixed partitioning of the relevant region of the underlying feature space, and the associated reals are a measure of the mass of the distribution that falls into the corresponding bin.

Several measures have been proposed for the similarity between two distributions $P = \{p_i\}$ and $Q = \{q_i\}$. They can be divided into two categories. The *bin-by-bin* similarity measures only compare contents of corresponding distribution bins, that is, they compare p_i and q_i for all i , considering no relation among p_i and q_j for $i \neq j$. On the other hand, the *cross-bin* measures contain terms that also compare non-corresponding bins.

2.1. Bin-by-Bin Similarity Measures. In this category, only pairs of bins in the two distributions that have the same index are matched. A combination of all the pairwise differences defines this kind of similarity.

Minkowski-form distance::

$$d_r(P, Q) = \left(\sum_i \|p_i - q_i\|^r \right)^{1/r}. \quad (1)$$

The L_1 distance has been proposed for computing the similarity scores between color images [30], and the L_∞ norm was used to measure texture similarity.

Kullback-Leibler divergence and Jeffrey divergence::

$$d_{KL}(P, Q) = \sum_i p_i \log \frac{p_i}{q_i}. \quad (2)$$

The Kullback-Leibler (K-L) divergence has been suggested in [22] as an image similarity measure and measures how inefficient on average it would be to code one

histogram distribution using the other as the code-book. Symmetric and numerically stable variants like the *Jeffrey* divergence (JD) [28] is also another similarity measure:

$$d_J(P, Q) = \sum_i \left(p_i \log \frac{p_i}{m_i} + q_i \log \frac{q_i}{m_i} \right), \quad (3)$$

where $m_i = \frac{p_i + q_i}{2}$.

χ^2 statistics::

$$d_{\chi^2}(P, Q) = \sum_i \frac{(p_i - m_i)^2}{m_i}, \quad (4)$$

where $m_i = \frac{p_i + q_i}{2}$. This distance measures how unlikely it is that one distribution was drawn from the population represented by the other.

The above mentioned similarity definitions can be appropriate in different areas. For example, the Kullback-Leibler divergence is justified by information theory and the χ^2 statistics by statistics. However, these measures do not necessarily match perceptual similarity well. The major drawback of these measures is that they account only the information of the bins with the same index, and do not use information across bins. Another drawback of these kind of similarity measures is their sensitivity to bin size. A binning that is too coarse will not have sufficient discriminative power, while a binning that is too fine will place similar features in different bins which will never be matched. On the other hand, cross-bin similarity measures, always produce better results with smaller bins.

2.2. Cross-Bin Similarity Measures.

Quadratic-form distance::

$$d(P, Q) = \sqrt{(\mathbf{p} - \mathbf{q})^T \mathbf{A} (\mathbf{p} - \mathbf{q})}, \quad (5)$$

where \mathbf{p} and \mathbf{q} are vectors that list all the entries in P and Q . Cross-bin information is incorporated via a similarity matrix $\mathbf{A} = [a_{ij}]$, where a_{ij} denote the similarity between bins i and j .

Weighted-Mean-Variance (WMV)::

$$d(P, Q) = \frac{|\mu_P - \mu_Q|}{\sigma(\mu)} + \frac{|\sigma_P - \sigma_Q|}{\sigma(\sigma)}, \quad (6)$$

where $\sigma(\cdot)$ denotes an estimate of the standard deviation of the respective entity.

Kolmogorov-Smirnov distance::

$$d(P, Q) = \max_i (|\hat{p}_i - \hat{q}_i|), \quad (7)$$

where \hat{p}_i and \hat{q}_i are cumulative distributions. The Kolmogorov-Smirnov distance is a common statistical measure for unbinned distributions and it is defined only for one distance.

Earth mover's distance (EMD)::

$$d(P, Q) = \frac{\sum_{i=1}^M \sum_{j=1}^N d_{ij} f_{ij}}{\sum_{i=1}^M \sum_{j=1}^N f_{ij}}, \quad (8)$$

where d_{ij} denotes the ground distance and f_{ij} the flow between p_i and q_i . The EMD [29] is based on the solution of a transportation problem which is a linear optimization problem. If the cost for moving a single feature unit in the feature space is defined based on the ground distance, then the distance between two distributions is given

as the minimal cost to transform one distribution to the other, where the total cost is the sum of the costs needed to move the individual features. As a key advantage of the EMD, each image may be represented by an individual binning that is adapted to its specific distribution.

Weighted-mean-variance (WMV) is a parametric measure relying on the means and variances of the marginal distributions. Kolmogorov-Smirnov distance is defined only for cumulative distributions, and therefore can be used only with marginal distributions, while the others are applicable to multidimensional histograms. The EMD has the additional advantage to be applicable to histograms with individual binning.

3. Physics-Based Deformable Modelling. In this Section, the physically based deformable model [20, 21] exploiting modal analysis, which is used to parameterize the distribution transformation, is presented. The model uses only elastic deformations, assuming that the distribution recovers its original configuration as soon as all applied forces causing the deformation are removed.

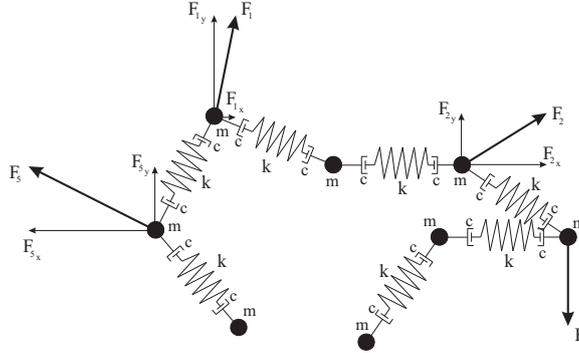


FIGURE 1. 2D model example of 8 nodes of mass m connected with identical springs of stiffness k . Four forces are acting on the model, that produce model deformation.

Modelling a M -dimensional distribution can be achieved by a chain topology of N virtual masses on the distribution. In Figure 1 is depicted an example of a 2D model of 8 nodes. Each model node has a mass m and is connected to its two nearest neighbors with identical springs of stiffness k . The ratio $a = \frac{k}{m}$ constitutes the so-called *characteristic value* of the model, which is a constant value that describes its physical characteristics and determines its physical behavior. When a increases, the deformable model tends to behave as a rigid one, which means in practice, that the model can be spatially moved without any deformation. On the other hand, when a decreases, model tends to be treated as a fully deformable one, which means that each force affects only the node (mass) it is applied to. Furthermore, these model nodes are points on the distribution at equilibrium. The node coordinates of the model under examination are stacked in vector:

$$\mathbf{v}_t = \{ (v_i^t(1), v_i^t(2), \dots, v_i^t(M)) \}_{i=1, \dots, N}, \quad (9)$$

where N is the number of vertices (masses) of the model, M is the dimension of the distribution and t denotes the t -th deformation time instance. In the following, \mathbf{v}_i^t denotes the i -th component of vector \mathbf{v}_t . The model under study, is a physics-based system governed by the fundamental equation of dynamics:

$$f_e(\mathbf{v}_i^t) + f_d(\mathbf{v}_i^t) + f_{ext}(\mathbf{v}_i^t) = m_i \ddot{\mathbf{v}}_i^t, \quad (10)$$

where m_i is the mass of the point under study and $\ddot{\mathbf{v}}_i^t$ its acceleration under total load of forces (i -th component of vector $\ddot{\mathbf{v}}_t$). $f_d(\cdot)$ is a damping force, $f_{ext}(\cdot)$ the external load on node under study, and $f_e(\cdot)$ is the elastic force due to node neighbors. The above governing equation is expressed for all model nodes, leading to a nonlinear system of coupled differential equations, since the displacement of a node depends on the displacement of its neighbors.

In order to solve this system of equations, we propose to set the natural length l_0 of the springs equal to zero. The length l_0 of the springs is included in the elastic force $f_e(\cdot)$ equation. This assumption does not import any restriction to the initial configuration of the model, if we add an equilibrium force $f_{eq}(\mathbf{v}_i^t) = -f_e(\mathbf{v}_i^t)$ in (10). This force keeps the model inflated, so that it does not shrink to a point. We assume that, at any future time, this equilibrium force is constant. Hence, the natural state of the model is its initial configuration. This assumption has a main advantage, that the model can be considered within the framework of linear elasticity. As a consequence, our solution lies in a set of linear differential equations with node displacements decoupled in each coordinate, regardless of the magnitude of the displacements. To enforce the assumption of constant equilibrium force $f_{eq}(\cdot)$, the angular variations of the springs orientation in any dimension should be sufficient small ($< 15^\circ$), in which case the aforementioned approximation is valid [23, 24].

The governing equation can now be written in a matrix form [3]:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f}_t, \quad (11)$$

where $\mathbf{u} = \mathbf{v}_t - \mathbf{v}_{t_0}$ is the nodal displacements vector. \mathbf{M} , \mathbf{C} , and \mathbf{K} [3, 19, 20, 21] are the mass, damping, and stiffness matrices of the model, respectively, and \mathbf{f}_t is the external force vector, usually resulting from the distance (Euclidean distance or any other distance metric) from one distribution to the other. Note, that the above formulation provides a simplification between forces $f_{eq}(\cdot)$ and $f_e(\cdot)$, so that they do not appear in the governing equation. These forces can be viewed as *internal* forces which do not need to be computed. Furthermore, equation (11) is a finite element formulation of the deformation process.

Instead of solving directly the equilibrium equation (11), one can transform it by a change of basis:

$$\mathbf{u} = \mathbf{\Psi}\tilde{\mathbf{u}}, \quad (12)$$

where $\mathbf{\Psi}$ is the square nonsingular transformation matrix of order N to be determined, and $\tilde{\mathbf{u}}$ is referred to as the *generalized displacement* vector. One effective way of choosing $\mathbf{\Psi}$ is setting it equal to $\mathbf{\Phi}$, a matrix whose entries are the eigenvectors of the generalized eigenproblem:

$$\mathbf{K}\phi_i = \omega_i^2\mathbf{M}\phi_i. \quad (13)$$

Thus, equation (12) is transformed to:

$$\mathbf{u} = \mathbf{\Phi}\tilde{\mathbf{u}} = \sum_{i=1}^N \tilde{u}_i\phi_i. \quad (14)$$

Equation (14) is referred to as the modal superposition equation. The i -th eigenvector, i.e. the i -th column of $\mathbf{\Phi}$, denoted by ϕ_i , is also called the i -th *vibration mode*. \tilde{u}_i (the i -th scalar component of $\tilde{\mathbf{u}}$) is its *amplitude*, and ω_i is the corresponding eigenvalue (also called *frequency*). Using the standard Rayleigh hypothesis [20], matrices \mathbf{K} , \mathbf{M} and \mathbf{C} are simultaneously diagonalized:

$$\begin{cases} \mathbf{\Phi}^T\mathbf{M}\mathbf{\Phi} = \mathbf{I} \\ \mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi} = \mathbf{\Omega}^2 \end{cases}, \quad (15)$$

where $\mathbf{\Omega}^2$ is a diagonal matrix whose elements are the eigenvalues ω_i^2 and \mathbf{I} is the identity matrix. Thus, in the modal space the governing matrix-form, equations decoupled into N scalar equations, by substituting (14) into (11) and premultiplying by $\mathbf{\Phi}^T$:

$$\ddot{\tilde{\mathbf{u}}} + \tilde{\mathbf{C}}\dot{\tilde{\mathbf{u}}} + \mathbf{\Omega}^2\tilde{\mathbf{u}} = \tilde{\mathbf{f}}, \quad (16)$$

where $\tilde{\mathbf{C}} = \mathbf{\Phi}^T \mathbf{C} \mathbf{\Phi}$ and $\tilde{\mathbf{f}} = \mathbf{\Phi}^T \mathbf{f}$. Solving these equations at time t leads to $\tilde{\mathbf{u}}$, and the displacement \mathbf{u} of the model nodes is obtained by the modal superposition equation (14).

In practice, we wish to approximate nodal displacements \mathbf{u} by $\hat{\mathbf{u}}$, which is the truncated sum of the N' low-frequency vibration modes, where $N' \ll N$:

$$\mathbf{u} \approx \hat{\mathbf{u}} = \sum_{i=1}^{N'} \tilde{u}_i \phi_i \quad (17)$$

Eigenvectors $(\phi_i)_{i=1, \dots, N'}$ form the *reduced modal basis* of the system. This is the major advantage of modal analysis: it is solved in a subspace corresponding to the N' truncated low-frequency vibration modes of the deformable structure [20, 21, 24]. The number of vibration modes retained in the distributions' displacement description, is chosen so as to obtain a compact but adequately accurate representation. A typical *a priori* value for N' , covering many types of standard deformations is equal to one quarter of the number of the vibration modes.

An important advantage of the formulations described so far, in the full as well as the truncated modal space, is that the vibration modes ϕ_i and the frequencies ω_i of a chain topology have an explicit expression [20] and they do not have to be computed using eigen-decomposition techniques (due to the dimensions of matrices \mathbf{K} and \mathbf{M}). The eigenvalues (frequencies) are given by:

$$\omega_i^2 = 4a \sin^2 \left(\frac{\pi i}{2N} \right), \quad (18)$$

and the eigenvectors (vibration modes) are obtained by:

$$\phi_i = \left[\dots, \cos \frac{\pi i (2j - 1)}{2N}, \dots \right]^T, \quad (19)$$

where $i \in \{0, 1, \dots, N - 1\}$ and $j \in \{1, 2, \dots, N\}$. This is the main reason we have chosen and used the so far described model topology to parameterize our distributions' displacements.

In many computer vision applications [21], when the initial and the final deformable states are known, it is assumed that a constant force load \mathbf{f} is applied to the initial model state. Thus, equation (11) is called the equilibrium governing equation and corresponds to the static problem:

$$\mathbf{K}\mathbf{u} = \mathbf{f}. \quad (20)$$

In the new basis, equation (20) is simplified to $2N$ scalar equations:

$$\omega_i^2 \tilde{u}_i = \tilde{f}_i. \quad (21)$$

In equation (21), ω_i designates the i -th eigenvalue and the scalar \tilde{u}_i is the amplitude of the corresponding vibration mode (corresponding to eigenvector ϕ_i). Equation (21), indicates that, instead of computing the displacements vector \mathbf{u} from equation (20), we can compute its decomposition in terms of the vibration modes of the original (initial) model. The physical representation $\mathbf{v}(\tilde{\mathbf{u}})$ is finally given by applying the deformations to the initial model:

$$\mathbf{v}(\tilde{\mathbf{u}}) = \mathbf{v}_0 + \mathbf{\Phi}\tilde{\mathbf{u}}. \quad (22)$$

External forces f_i denote the x and y components of the forces acting on node i in a 2D case:

$$\mathbf{f} = [f_{x,1}, f_{y,1}, f_{x,2}, f_{y,2}, \dots, f_{x,N}, f_{y,N}]^T, \quad (23)$$

where N is the number of model nodes.

4. The Deformation Distance (DeD). In this Section we will introduce the *Deformation Distance* (DeD) metric among two distributions, based on the frequencies of the physics-based deformable model (eq. 22) described in the previous Section.

Intuitively, given two distributions, one can be considered as the initial state of the above described deformable model, while the other as the final state. Then, the DeD measures the “energy” that must be paid to deform one distribution into the other.

In a more formulated way, we can say that having in hand two distributions, let $P = \{(p_i^1, p_i^2, \dots, p_i^M)\}_{i=1, \dots, N}$ one distribution with N points in M dimensions and let $Q = \{(q_i^1, q_i^2, \dots, q_i^M)\}_{i=1, \dots, N}$ a second distribution with equal number of points defined in the same dimension space; and $\mathbf{F}_{P,Q} = [f_i^j(P, Q)]$ the ground distance matrix, where $f_i^j(P, Q)$ is the ground distance between points p_i^j and q_i^j . This distance matrix is considered as the external forces acting on the deformable model used, deforming one distribution (initial state) to the other (final state).

To estimate the desired deformation energy, the generalized displacement vector $\tilde{\mathbf{u}}$, derived by (14) will be used, which expressed as:

$$\tilde{\mathbf{u}} = \{(\tilde{u}_i^1, \tilde{u}_i^2, \tilde{u}_i^3, \dots, \tilde{u}_i^M)\}_{i=1, \dots, N}. \quad (24)$$

$\delta_i = [\tilde{u}_i^1, \tilde{u}_i^2, \dots, \tilde{u}_i^M]^T$ describes the i^{th} coordinate of the vector $\tilde{\mathbf{u}}$, which describes the frequency-based properties of the deformation [13, 20, 24].

Having computed the generalized displacement vector $\tilde{\mathbf{u}}$ (24), the deformation energy or in other words the DeD is defined as:

$$DeD(P, Q) = \frac{1}{N * M} \sqrt{\sum_{i=1}^N \sum_{j=1}^M (\tilde{u}_i^j)^2}. \quad (25)$$

The DeD is a true distance metric if the ground distance is a metric, fact that allows endowing image spaces with a metric structure. A proof of this is given in Appendix.

In the case, where the distributions under examination have not equal number of points, let us say that P has N points and Q has N' points ($N \neq N'$), then the DeD can be defined as:

$$DeD(P, Q) = f(DeD(P, Z), DeD(Z, Q)), \quad (26)$$

where $f(\cdot, \cdot)$ is the ground distance metric (e.g. the L_1 norm) and Z is the distribution with all its points located in point zero. The physical explanation of above defined (26) equation is that the distance among the two distributions is equal to the deformation energy that must be paid to deform the first distribution (initial state) to the second (final state) through the zero distribution – the distribution which has all its points located in point zero.

Figure 2 illustrates a very simple 1D example of the DeD calculation. Two random histograms (1D distributions) have been selected (Figure 2a) and the external forces deform the model to adapted on the second distribution. The generalized displacement vector $\tilde{\mathbf{u}}$ of the deformation process is depicted in Figure 2b, while their deformation distance is $DeD = 2,55$.

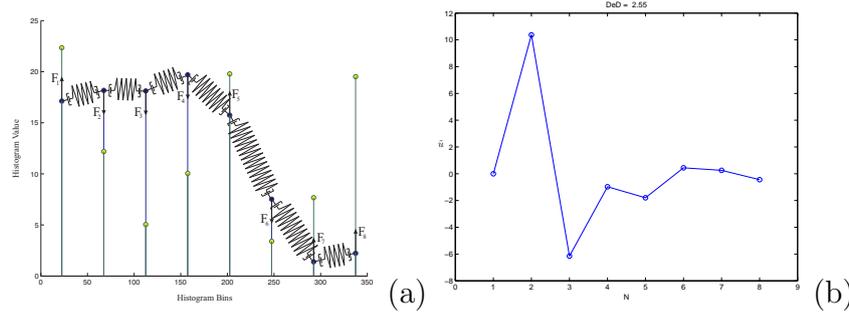


FIGURE 2. A graphic representation of the DeD estimation. **(a)** The deformable model adapted on the first of the two histogram distributions and the external forces deform it to the second, and **(b)** the \tilde{u} of the corresponding deformation process ($DeD = 2.55$).

5. Experimental Results. To evaluate the DeD distance metric, we applied it on image clustering problem. Our image clustering tests were performed on Berkeley Segmentation Dataset [16] and various other images. We compare the efficiency of the DeD with two other distances, L_1 norm and Earth Mover's Distance (EMD). We applied them on clustering using the well-known k-means [14] clustering method, and the performances of the three algorithms (distances) were compared with respect to the optimal segmentation accuracy (SA), where SA is defined as the sum of the correctly classified bins divided by the sum of the total number of bins of a distribution [2]:

$$SA = \sum_{i=1}^c \frac{A_i \cap C_i}{\sum_{j=1}^c C_j}, \quad (27)$$

where c is the number of clusters, A_i represents the set of bins belonging to the i -th class found by the algorithm, while C_i represents the set of bins belonging to the i -th class in the reference distribution.

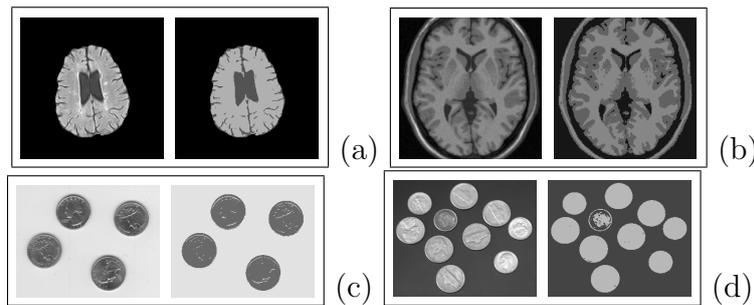


FIGURE 3. Four real gray-scale images and their clustering results after applying k-means algorithm exploiting DeD metric.

In our first set of experiments, we apply the k-means [14] clustering algorithm to some real gray-scale images exploiting the DeD metric (Figure 3). As distribution bin in that experiments has been thought the simple case of gray values of each pixel, e.g. each distribution bin was equal to the corresponding gray pixel value. Furthermore, four color real images, depicted in Figure 4, were also clustered using k-means exploiting DeD metric. In this case, each distribution bin was comprised by the RGB values of the corresponding pixel.

Furthermore, we apply the k-means clustering algorithm exploiting DeD metric and the images transformed to various color spaces. Because of how the human vision system

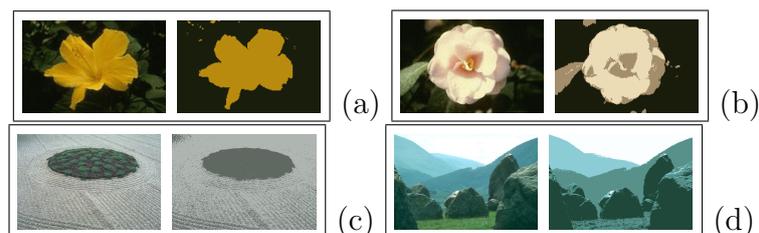


FIGURE 4. *Four real color images and their clustering results after applying k -means algorithm exploiting DeD metric.*

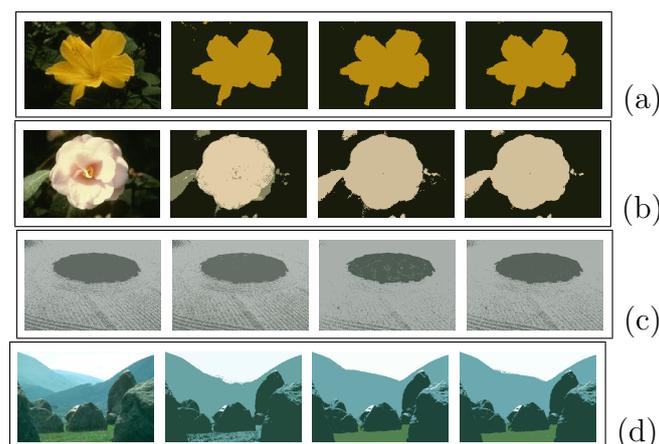


FIGURE 5. *Four real color images and their clustering results after applying k -means algorithm exploiting DeD metric. The first column depicts the initial image, while the second the clustering results in HSV color space. The third column shows the clustering results in CIE $L^*a^*b^*$ color space, and the last column the results combining all used color spaces.*

is built, color lives naturally in a three dimensional space. Color distributions, then, can describe the color contents of entire images. Due to the fact that RGB color space is very sensitive to illumination changes, we transform the images under examination into the HSV color space ignoring the luminance (value) information. Thus, instead of using HSV color information, the method uses HS vector colors. Luminance conveys information only about illumination intensity changes, while all color information is found in the hue and saturation domain. Also, we transform the images under examination into the CIE $L^*a^*b^*$ color space ignoring again the luminance information. Figure 5 illustrates a clustering example exploiting the above color spaces. The first column of Figure 5 shows the initial four images of Figure 4, while the second one depicts the clustering results exploiting DeD metric in HSV color space. That is, each distribution bin is comprised by the HS values of the corresponding pixel. The third column illustrates the clustering results in CIE $L^*a^*b^*$ color space. In this case, each distribution bin is comprised by the a^*b^* values of the corresponding pixel. Also, the last column of Figure 5 shows the clustering results exploiting DeD metric in a combination color space, i.e. each distribution bin is comprised by the RGB, HS and a^*b^* values of the corresponding pixel.

The deformable model used in the proposed approach adapted on the derived distributions. Although, adjacent distribution bins in general represent similar colors, they are some discontinuous bins in its body. These discontinuities are frequently appeared in HSV and CIE $L^*a^*b^*$ color space, namely, when the hue or a^* bin proceeds to next value. The deformable model adapted on the distribution bins “connects” similar color values, while

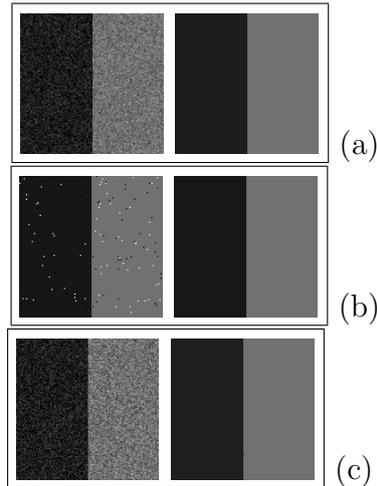


FIGURE 6. Clustering of a synthetic image. First column depicts the initial image, while the second the clustering result after applying the proposed algorithm. **(a)** The synthetic image with Gaussian noise (8%). **(b)** The same image with Salt & Pepper noise (8%). **(c)** The synthetic image with Uniform noise (25%).

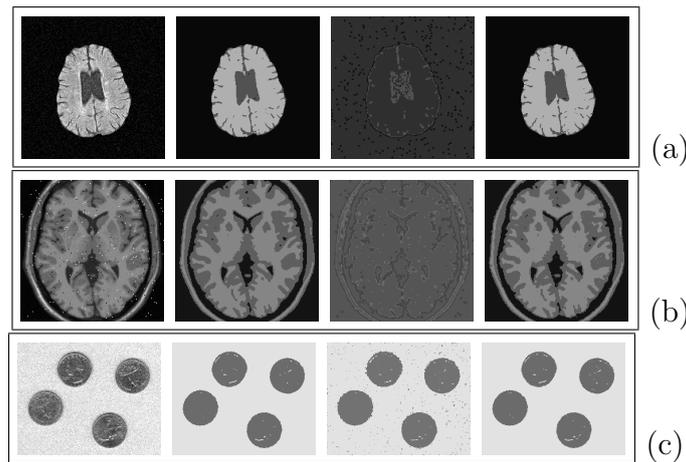


FIGURE 7. Clustering of real images. The first column depicts the initial image, while the second the clustering result after applying k -means algorithms using the L_1 norm. The third column shows the clustering results using the Earth Mover's Distance (EMD), while the final column depicts the results using the proposed distance metric. **(a)** A real image with Gaussian noise (8%). **(b)** An image with Salt & Pepper noise (8%). **(c)** An image with Uniform noise (15%).

the specific discontinuous parts are ignored by the proposed method and they are thought as continuous without losing the generality as it is shown. The connection of the adjacent distribution bins by the deformable model, influences the distribution representation in the way that changes in a color value usually convey changes to similar color values and as a consequence to adjacent distribution bin values.

Furthermore, we apply the proposed distance metric (DeD) to noisy images. We apply the proposed algorithm to a synthetic test image (Figure 6: 128×128 pixels, two classes with two gray level values taken as 20 and 120) corrupted by different levels of Gaussian

(8%), Salt & Pepper (8%) and Uniform (25%) noise, respectively. As a distribution for each pixel has been used noise resistance characteristics, e.g., the characteristics of the eight-connected neighborhood pixels, that is, the gray values of the corresponding pixel as well as the gray values of its eight-connected neighborhood pixels. The proposed distance metric (Figure 6) removes almost all the added noise achieving satisfactory results and proving a robustness to noise.

Furthermore, we apply three distance metrics (L_1 norm, Earth Mover’s Distance and DeD) with the k-means clustering algorithm to some real images [17] (Figure 7), contaminated with various kind of noise (gaussian, uniform and salt & pepper noise). The clustering results are shown in Figure 7. Figure 7(a) shows an MRI image contaminated with gaussian noise (8%), while Figures 7(b) and 7(c) illustrates other real images contaminated with salt & pepper (8%) and uniform (15%) noise respectively. The first column of Figure 7 shows the noisy image under consideration, while the second column illustrates the clustering results after applying k-means algorithm exploiting L_1 norm. The third and fourth columns depict the clustering results after applying the same clustering algorithm (k-means) using the EMD and DeD distance metrics respectively. It is clearly illustrated in Figure 7 that clustering results using EMD metric are influenced by the noise to different extents, which indicates that this distance metric lack enough robustness to noise, while L_1 norm exhibits enough resistance to all kind of noises. The proposed distance metric (DeD) eliminates almost all the effect of the noise. This remark is also enhanced by the results of the segmentation accuracy (SA) (27) as shown in Table 1. The ground truth clustered results has been manually extracted. Finally, as a distribution for each pixel has been used noise resistance characteristics, e.g., the characteristics of the eight-connected neighborhood pixels, that is, the gray values of the corresponding pixel as well as the gray values of its eight-connected neighborhood pixels.

TABLE 1. Segmentation accuracy (SA%) of three distance metrics on real images.

	L_1	EMD	DeD
Gaussian	98.61	63.14	99.27
Uniform	98.59	67.69	99.23
Salt & Pepper	98.66	70.44	99.33
Average	98.62	67.09	99.28

6. Conclusion. The deformation distance (DeD) is a general and flexible metric for distributions. It allows partial matches and it can be applied to variable-length representations of distributions. Because of these advantages, we believe that the DeD can be of use both for understanding (study, conclusion education, etc) distributions related to vision problems. Comparisons with other similarity measures show that the DeD matches perceptual similarity better as a fundamental element of image clustering systems.

The DeD is based on the “energy” that must be paid to deform one distribution into the other, relying on the distribution’s frequency-based features. The frequency-based features are extracted by a physics-based deformable model that parameterizes the distributions. Furthermore, the low frequency modal analysis of the distributions makes the metric robust to missing data or outliers.

Finally, the good quality of the results and the efficiency of DeD are rendered it a promising tool not only for the computer vision problems but also in pattern recognition systems and for problems outside the realm of computer vision.

Appendix. In this Appendix we prove that when the distributions under examination have equal number of points N and defined in the same space dimension M , and the ground distance $d(\cdot, \cdot)$ is metric, the DeD is a true metric. Non-negativity and symmetry hold trivially in all cases, so we only need to prove that the triangle inequality holds. Let $\mathbf{F}_{P,Q}$ the ground distance matrix from distribution P to Q , $\mathbf{F}_{Q,R}$ the distance matrix from Q to R and $\mathbf{F}_{P,R}$ the corresponding matrix from P to R . Consider, now, the flow $P \mapsto Q \mapsto R$. Since, the ground distance $d(\cdot, \cdot)$ is a metric, the triangle inequality is a fact:

$$\mathbf{F}_{P,R} \leq \mathbf{F}_{P,Q} + \mathbf{F}_{Q,R}. \quad (\text{A.1})$$

The DeD, as defined in eq. (25), between distributions P and R and the definition of \tilde{u}_i (14) leads us to:

$$DeD_{P,R} = \sqrt{\sum_{i=1}^N \sum_{j=1}^M \left(\frac{\sum_k [f_k^j(P,R)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2}. \quad (\text{A.2})$$

The factor $\frac{1}{N*M}$ is omitted because it is the same in all cases since all distributions under examinations have the same number of points N and defined in the same dimension space M , and does not infect at all the proof. Thus, the DeD between distributions P and R described in the above equation (A.2) can be transformed to:

$$\begin{aligned} DeD(P,R) &= \sqrt{\sum_i \sum_j \left(\frac{\sum_k [f_k^j(P,R)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2} \leq \sqrt{\sum_i \sum_j \left(\frac{\sum_k [(f_k^j(P,Q)+f_k^j(Q,R))\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2} \\ &= \sqrt{\sum_i \sum_j \left(\frac{\sum_k [f_k^j(P,Q)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} + \frac{\sum_k [f_k^j(Q,R)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2} \\ &\leq \sqrt{\sum_i \sum_j \left(\frac{\sum_k [f_k^j(P,Q)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2} + \sum_i \sum_j \left(\frac{\sum_k [f_k^j(Q,R)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2 \quad (\text{A.3}) \\ &\leq \sqrt{\sum_i \sum_j \left(\frac{\sum_k [f_k^j(P,Q)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2} + \sqrt{\sum_i \sum_j \left(\frac{\sum_k [f_k^j(Q,R)\phi_k(i)]}{(1+\omega_i^2) \sum_k \phi_k^2(i)} \right)^2} \\ &= DeD(P,Q) + DeD(Q,R). \end{aligned}$$

REFERENCES

- [1] M. Aghagolzadeh, H. Soltanian-Zadeh, B. Araabi, and A. Aghagolzadeh, A hierarchical clustering based on mutual information maximization, *Proc. of IEEE International Conference on Image Processing*, vol. 1, pp. 277-280, 2007.
- [2] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data, *IEEE Trans. Medical Imaging*, vol. 21, no. 3, pp. 193-199, 2002.
- [3] K. J. Bathe. *Finite Element Procedure*. Prentice Hall, Englewood Cliffs, New Jersey, 1996.
- [4] J. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New York, 1981.
- [5] C. H. Chen, T. Y. Chen, D. J. Wang, and T. J. Chen, A cost-effective people-counter for a crowd of moving people based on two-stage segmentation, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, no. 1, pp. 12-23, 2012.
- [6] T. P. Hong and C. H. Wu, An improved weighted clustering algorithm for determination of application nodes in heterogeneous sensor networks, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 173-184, 2011.
- [7] E. R. Hruschka, R. J. G. B. Campello, A. A. Freitas, and A. C. Ponce L. F. de Carvalho, A survey of evolutionary algorithms for clusterin, *IEEE Trans. Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 39, no. 2, pp. 133-155, 2009.
- [8] M. Krinidis and I. Pitas, Color texture segmentation based on the modal energy of deformable surfaces, *IEEE Trans. Image Processing*, vol. 18, no. 7, pp. 1613-1622, 2009.

- [9] S. Krinidis and V. Chatzis, Fuzzy energy-based active contours, *IEEE Trans. Image Processing*, vol. 18, no. 12, pp. 2747-2755, 2009.
- [10] S. Krinidis and V. Chatzis, A robust fuzzy local information c-means clustering algorithm, *IEEE Trans. Image Processing*, vol. 19, no. 5, pp. 1328-1337, 2010.
- [11] S. Krinidis and V. Chatzis, A physics-based dual deformable model, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, no. 1, pp. 100-121, 2012.
- [12] S. Krinidis, M. Krinidis, and V. Chatzis, An unsupervised image clustering method based on eemd image histogram, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, no. 2, pp. 151-163, 2012.
- [13] S. Krinidis and I. Pitas, Fast free-vibration modal analysis of 2-D physics-based deformable objects, *IEEE Trans. Image Processing*, vol. 14, no. 3, pp. 281-293, 2005.
- [14] J. MacQueen, Some methods for classification and analysis of multivariate observations, In *Proc. of 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281-297, 1967.
- [15] G. Malathi and V. Shanthi, Statistical measurement of ultrasound placenta images complicated by gestational diabetes mellitus using segmentation approach, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 4, pp. 332-343, 2011.
- [16] D. Martin, C. Fowlkes, D. Tal, and J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, In *Proc. of the 8th International Conference of Computer Vision*, vol. 2, pp. 416-423, 2001.
- [17] MA. Mathworks, Natick, Image processing toolbox, <http://www.mathworks.com>.
- [18] X. Munoz, J. Freixenet, X. Cufi, and J. Marti, Strategies for image segmentation combining region and boundary information, *Journal of Pattern Recognition Letters*, vol. 24, no. 1, pp. 375-392, 2003.
- [19] C. Nastar, *Models physiques deformables et modes vibratoires pour l'analyse du mouvement non-rigide dans les images multidimensionnelles*, Ph.D. Thesis, Ecole Nationale des Ponts et Chaussees, 1994.
- [20] C. Nastar and N. Ayache, Frequency-based nonrigid motion analysis: Application to four dimensional medical images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 11, pp. 1069-1079, 1996.
- [21] C. Nikou, G. Bueno, F. Heitz, and J. P. Armspach, A joint physics-based statistical deformable model for multimodal brain image analysis. *IEEE Trans. Medical Imaging*, vol. 20, no. 10, pp. 1026-1037, 2001.
- [22] T. Ojala, M. Pietikainen, and D. Harwood, A comparative study of texture measures with classification based feature distributions, *Journal of Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.
- [23] A. Pentland and B. Horowitz, Recovery of non-rigid motion and structure, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 730-742, 1991.
- [24] A. Pentland and S. Sclaroff, Closed-form solutions for physically-based shape modeling and recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 715-729, 1991.
- [25] D. L. Pham, and J. L. Prince, Adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities, *Journal of Pattern Recognition Letters*, vol. 20, pp. 57-68, 1999.
- [26] D. L. Pham, C. Xu and J. L. Prince, Current methods in medical image segmentation, *Annual Review of Biomedical Engineering*, vol. 2, pp. 315-337, 2000.
- [27] P. Puranik, P. Bajaj, A. Abraham, P. Palsodkar, and A. Deshmukh, Human perception-based color image segmentation using comprehensive learning particle swarm optimization, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 3, pp. 227-235, 2011.
- [28] J. Puzicha, T. Hofmann, and J. Buhmann, Non-parametric similarity measures for unsupervised texture segmentation and image retrieval, In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 267-272, 1997.
- [29] Y. Rubner, C. Tomasi, and L. J. Guibas, The earth movers distance as a metric for image retrieval, *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99-121, 2000.
- [30] M. Swain and D. Ballard, Color indexing, *International Journal of Computer Vision*, pp. 134-143, 1991.
- [31] T. H. Tsai and C. Y. Lin, Visual hand gesture segmentation using three-phase model tracking technique for real-time gesture interpretation system, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, no. 2, pp. 122-134, 2012.
- [32] L. A. Zadeh, Fuzzy sets, *Information and Control*, vol. 8, pp. 338-353, 1965.