

Received March 9, 2022, accepted March 28, 2022, date of publication April 1, 2022, date of current version April 7, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3164210

Modeling a Local Dissimilarity Map With Weibull Distribution—Application to 2-Class and Multi-Class Image Classification

MOUSTAPHA DIAW¹, AGNÈS DELAHAIES¹, JÉRÔME LANDRÉ¹, FLORENT RETRAINT²,
AND FRÉDÉRIC MORAIN-NICOLIER¹

¹CRéSTIC, EA 3804, University of Reims Champagne-Ardenne, IUT Troyes, 10026 Troyes, France

²LIST3N, University of Technology of Troyes (UTT), 10300 Troyes, France

Corresponding authors: Moustapha Diaw (moustapha.diaw@univ-reims.fr) and Frédéric Morain-Nicolier (frederic.nicolier@univ-reims.fr)

This work was supported in part by Troyes Champagne Métropole.

ABSTRACT Due to the considerable increase of images in everyday life, many applications require a study on their similarity. The main challenge is to find a simple and efficient method to compare and classify image pairs into similar and dissimilar classes. This study presents a new method to image pairs comparison and classification based on the modeling of the Local Dissimilarity Map (LDM). The LDM is a tool for locally measuring the dissimilarity between two binary or grayscale images. It is a measure of dissimilarities based on a modified version of the Hausdorff distance, which allows quantifying locally the dissimilarities between images. This measure is completely without parameters and generic. The image pairs classification (2-class classification) method is structured as follows. First, a statistical model for the LDM is proposed. The model parameters, used as descriptors, are relevant to discriminate similar and dissimilar image pairs. Second, classifiers are applied to compute the classification scores (2-class classification problem). In addition, this approach is robust with respect to geometric transformations such as translation compared to the state-of-the-art similarity measures. Although the main objective of this paper is to apply our approach to image pairs classification, it is also performed on a classification with more than two classes (multi-class classification). Experiments on the well-known image data sets *NIST and on old print data set prove that the proposed method produces comparable, even better results than the state-of-the-art methods in terms of accuracy and F_1 score.

INDEX TERMS Local dissimilarity map, Euclidean distance transform, Weibull distribution, supervised classification.

I. INTRODUCTION AND STATE-OF-THE-ART

Image comparison and image classification have received a great attention from researchers in the recent years, due to the considerable increase of available images in our digital world. Many analysis and processing techniques have been studied to compare and classify images (mainly image pairs).

Usually, image similarity is being assessed by measures such as Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR). MSE and PSNR are widely applicable because they are simple to calculate and also convenient to implement mathematically in an optimization context. But they are very poorly adapted to perceive visual quality [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao¹.

However, researchers have developed two measures of structural and feature similarity to enhance the quality of visual perception: Structural Similarity Index (SSIM) and Feature Similarity Index (FSIM). In [2], Wang proposed a Structural Similarity Index for quality assessment based on the degradation of structural information. The computation of the SSIM index depends on three sub-indices: luminance, contrast and structure. After calculating the SSIM through a local sliding window, the global image quality is evaluated by calculating the average: Mean SSIM (MSSIM [3]). The great success of SSIM is due to the fact that the human visual system is adapted to the structural information in images. But it does not detect information on images with low-level features. However, the visual information in an image is redundant, while the human visual system understands

an image mainly based on its low-level features, such as edges and zero crossings. In [4], Zhang proposed the feature similarity index (FSIM) to detect low-level features similarity between images. The computation of this measure is based on two main parameters which are the phase congruence and the gradient magnitude. Gradient magnitude uses convolution masks to express the gradient operators [4], [5]. To our knowledge, no researcher has developed another similarity or dissimilarity measure more recent and better than these two above-mentioned measures for image pairs comparison and classification. However, these two similarity measures are not constructed from a local distance. This makes their statistical modeling difficult. To our best knowledge, no researcher has been able to model a similarity or dissimilarity measure map by a statistical distribution, neither MSSIM, nor FSIM nor the other existing similarity measures. In each case, similarity indices are used as descriptors to classify image pairs.

Generally for multi-class image classification the process can be divided into two steps: feature extraction and classification. The first step aims to extract descriptors vectors from the local distinguishing features of images. In the second step, a classifier is applied on all the resulting vectors to recognize the class of the unknown images. The performance of the classification depends on the quality of descriptors. Currently, machine learning and deep learning algorithms (convolutional neural networks) [6]–[9] are mostly used and give very good results for image recognition and classification. But convolutional Neural Networks architectures are sometimes complex and hard to interpret.

This paper proposes a new efficient and low complexity technique for image comparison, image pairs classification (2-class classification), and essentially multi-class image classification. In [10], Baudrier proposed a measure to characterize the local differences between binary images. It is based on the Hausdorff distance and has produced good results in several research works [11], [12]. This measure is called the Local Dissimilarity Map (LDM). In this work, an extension of this measure is adapted to grayscale images. In [13], Molchanov and Teran have extended the Euclidean distance transform to grayscale images. This work has enabled us to adapt our approach to structural non-binary images. In contrast to other measures, the LDM computes the local distance between the pixels of images. The LDM will lead to detect low-level features between images.

In the literature many researchers have proposed statistical distributions to model images and classify them using model parameters as descriptors. In [14], Nguyen introduced the gamma distribution to model images noise. The parameters of this model are relevant to distinguish authentic face images and high quality presentation attack images. In [15], Huang proposed a statistical model to distinguish the natural images acquired by digital cameras and images created by computer graphics rendering software. For classification,

using the estimated parameters, the authors proposed to establish a generalized likelihood ratio test. In [16], Qiao used a parametric model to expose the traces of resampling forgery, which is described with the distribution of residual noise. The authors proposed a statistical model describing the residual noise from a resampled image. In [17], Doan introduced a noise model which is relevant to describe a natural image acquired by a digital camera. The parametric model is characterized by two fingerprints which are used for falsification identification.

In this paper, a statistical two-parameter Weibull distribution is proposed to model the LDM of binary and grayscale images regardless of their size. The model is applied to old print images and is validated by the Kolmogorov-Smirnov ($K-S$) test. The LDMs of images are characterized by the two parameters of the model. In addition, image pairs classification (from *NIST and old print data sets) is proposed applying supervised classifiers and taking the two parameters of Weibull distribution as descriptors. To extend the effectiveness of our approach, we applied it on a classification with more than two classes (multi-class classification). Results are compared with state-of-the-art methods [6].

This paper is structured as follows. Section II presents the principle of the Local Dissimilarity Map for binary and grayscale images. Since this measure depends on the calculation of the Euclidean distance transform, we will first introduce the notion of distance transform. Section III presents the robustness of the Local Dissimilarity Map to geometric transformations. We will test the robustness on geometric transformations such as pixels translation. It is also true for rotation, barrel deformation and noise addition. Section IV introduces the statistical model and its theoretical validation. We will show theoretically that the LDM follows the proposed statistical distribution. Section V reports the numerical experimentation. We will first validate the theoretical model on old print data set and evaluate the performances of our classifications on *NIST and old print data sets. Section VI addresses the comparison of our approach to state-of-the-art methods. Finally, section VII concludes the paper with a few perspectives.

II. LOCAL DISSIMILARITY MAP

The Local Dissimilarity Map (LDM) [10] is an image processing tool allowing characterizing local differences between two binary images. Let's consider two images A and B of the same size (w, h) , the LDM between A and B , $LDM(A, B)$, is a 2D array of size (w, h) . It can easily be transformed by normalization of its values into an image in order to display it. In this section, the aim is to compare two grayscale images with a local dissimilarity measure and to model it statistically. The LDM of grayscale images are constructed from the LDM of binary images which are themselves constructed from Euclidean distance transforms. The proposed dissimilarity measure is completely without parameters and generic.

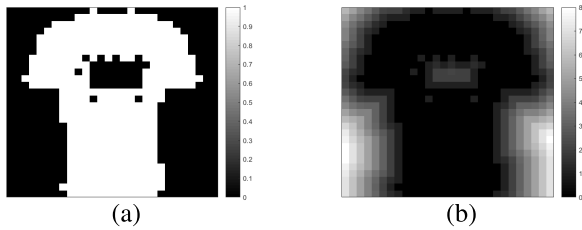


FIGURE 1. A binary image X (a) and its euclidean distance transform (b).

A. EUCLIDEAN DISTANCE TRANSFORM

The distance transform is a measuring tool which plays a crucial role in computer vision [18], in pattern recognition [19], [20], in robotics [21]. The calculation of the distance transform depends on the chosen underlying distance d . The classic choices for d are: the Euclidean distance from the L_2 norm, the Manhattan distance from the L_1 norm, which produces the 4-neighborhood, the Chebyshev distance from the L_∞ norm, which produces the 8-neighborhood. In this work, the Euclidean distance is used to compute the distance transforms on binary images.

1) FOR BINARY IMAGES

The distance transform for a binary image X is the application that associates to each point $x \in X$ the distance to the nearest non-zero point of X . Consider a binary image X and $p = (p_1, p_2)$, $q = (q_1, q_2)$, two pixels in X , the Euclidean distance transform of X is therefore written as follows:

$$DT_X(p) = \min_{q|X(q) \neq 0} \{ \|p - q\|_2 \} \quad (1)$$

where $\|p - q\|_2 = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$ is the Euclidean distance between $p, q \in X$. Figure 1 presents a binary image and its Euclidean distance transform. The pixel values become weak if they are close to the object.

2) FOR GRAYSCALE IMAGES

Consider F as an upper semi-continuous function ($F \subset \mathbb{R}^2$) that takes values in $\{0, 1, \dots, 255\}$. This function is called a grayscale image.

Since the computation of the Euclidean distance transform only works for binary images, Molchanov and Teran [13] have extended the Euclidean distance transform to grayscale images. A grayscale image is divided into N binary sub-images obtained using N thresholds τ_i , $i = 1 \dots N$. The Euclidean distance transform of a grayscale image is obtained by summing the distance transform of every computed sub-image. For an upper semi-continuous function F , the set F_τ of the sub-images is given by:

$$F_\tau = \{p \in F : F(p) \geq \tau\}. \quad (2)$$

At any level τ , F_τ has values in $\{0, 1\}$ (binary image). So, the Euclidean distance transform of a grayscale image F (RVDT $_F$) is the sum of the Euclidean distance transform of

the binary sub-images $F_{\tau_i}, DT_{F_\tau}(x)$:

$$RVDT_F(p) = \frac{1}{N} \sum_{\tau} DT_{F_\tau}(p). \quad (3)$$

Figure 2 shows a grayscale image and its RVDTs with the number of slices 2, 5, 15 and 25. It also shows the graph of standard L_2 squared of the difference of RVDT for N and $N+1$ slices as a function of the number of slices N . We started with $N = 2$ and we observe that the Euclidean distance $D = \|DT_{N+1} - DT_N\|_2^2$ is low in the interval [7, 9], and increases again in [10, 13]. From $N = 15$ until the end, D becomes low again and thus stable. Hence, from $N = 15$ the Euclidean distance transforms are nearly identical. The same behavior can be observed for other images. In any case $D \simeq 0$ for $N \geq 15$. In this paper, we could have chosen $N = 7$, $N = 8$ or $N = 9$, but we risked losing information when we calculated the RVDT. To ensure the opposite, we choose $N = 15$, in the rest of this paper, which is sufficient to obtain good results with less computing time compared to the $N = 255$ thresholds of the original paper [13].

B. LDM FOR BINARY IMAGES

For binary images, A and B take their values in the set $\{0, 1\}$ and $LDM(A, B)$ takes its values in \mathbb{R}^+ . It corresponds to a map of distances between A and B . Let $p = (x, y)$ be a pixel at coordinates (x, y) . The LDM is defined from $\mathbb{R}^2 \times \mathbb{R}^2$ to \mathbb{R}^2 by [10]:

$$\begin{aligned} LDM(A, B)(p) &= 1_A \Delta_B(p) \max(DT_A(p), DT_B(p)) \\ &= |A(p) - B(p)| \max(DT_A(p), DT_B(p)), \end{aligned} \quad (4)$$

where

$$1_A \Delta_B(p) = \begin{cases} 1 & \text{if } p \in A \Delta B \\ 0 & \text{otherwise,} \end{cases}$$

with $A \Delta B = (A \setminus B) \cup (B \setminus A)$ and $DT_A(p)$ is the Euclidean distance transform of A between the pixel p and the nearest non-zero pixel of A . Eq. (5) can be simplified for binary images [12]:

$$LDM(A, B)(p) = B(p) DT_A(p) + A(p) DT_B(p). \quad (6)$$

Eq. (6) removes the max operator and the absolute value. It has a great interest in the modeling of the Local Dissimilarity Map.

To illustrate the LDM behavior, two examples of LDM between three binary images are given in Figure 3. As the LDM locates and quantifies image pairs difference, Figure 3 shows, for similar image pairs, the LDM values will be lower in intensity and quantity than for dissimilar image pairs. For similar image pairs (see Figure 3d), the LDM appears really darker because on the two images A and B , the pixel values are locally close to each other. For dissimilar image pairs (see Figure 3e), the LDM appears clearer than for similar image pairs because only a small number of pixels are equal. Images A, B and C are taken from the MNIST data set, initially in grayscale and they have been binarized using the Sauvola

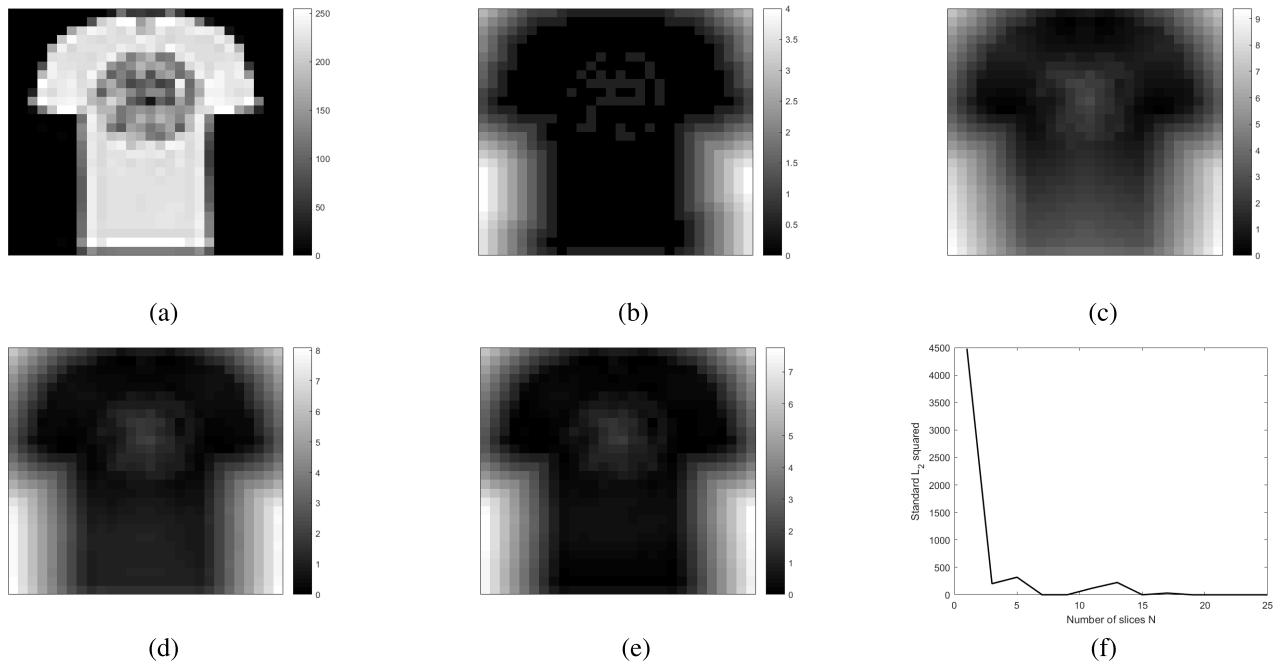


FIGURE 2. A grayscale image from the F-MNIST data set (a), RVDT_F with $N = 2$ (b), RVDT_F with $N = 5$ (c), RVDT_F with $N = 15$ (d), RVDT_F with $N = 25$ (e), standard L_2 squared of the difference of distance transform as a function of N (f). The distance transform with 15 slices is almost equal to the distance transform with 25 slices. In addition, the standard L_2 squared of the difference of distance transform for N slices and $N + 1$ slices ($D = \|\text{DT}_{N+1} - \text{DT}_N\|_2^2$) remains close to zero from the 15 slices ($D \approx 0$ for $N \geq 15$). Thus, for the computation of the distance transform, using only 15 slices are sufficient to obtain good results.

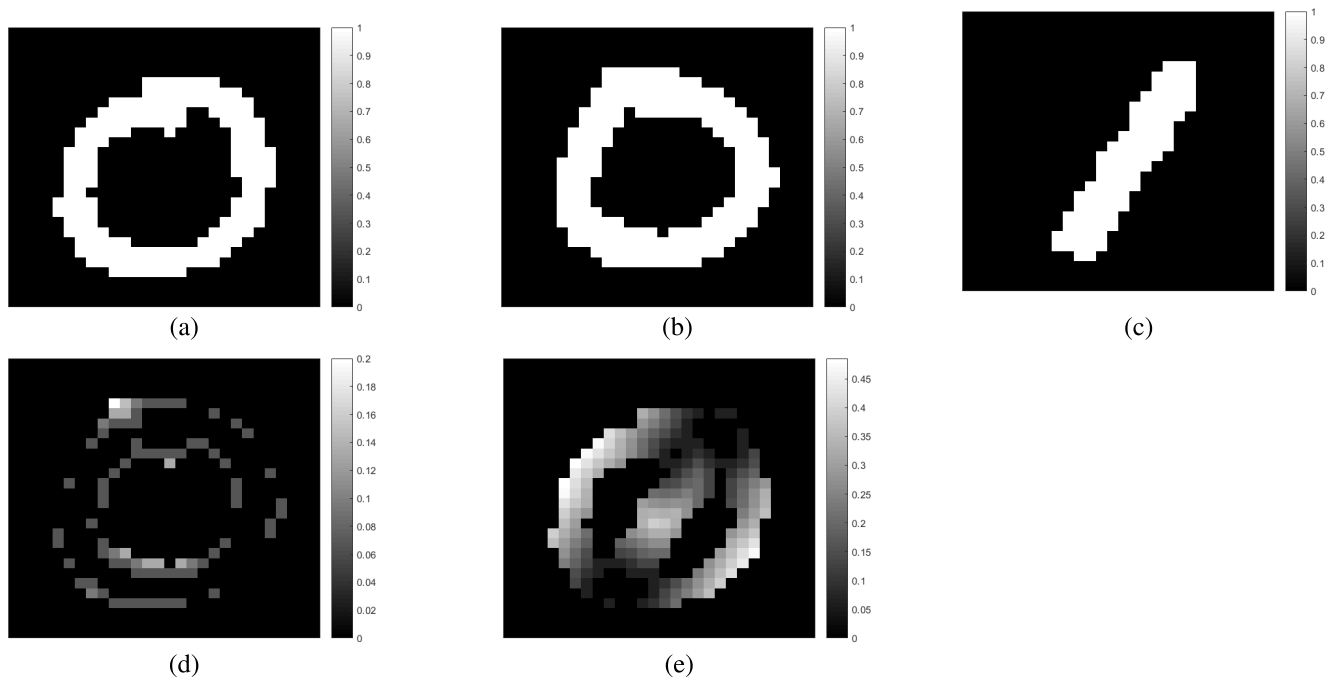


FIGURE 3. Binary image A (a), binary image B (b), binary image C (c), LDM(A, B) (d) and LDM(A, C) (e).

threshold [22]. The object is the pixel values equal to 1 and the background is the pixel values equal to 0. For both LDM, the areas where the pixels of the two images have a very large gap are in gray and where they are locally close are 0 (black).

The LDM of dissimilar images has a maximum grayscale value equal to 0.2 while for similar images the maximum value is close to 0.5, so the grayscale of the LDM of dissimilar images appears clearer than those of similar images.

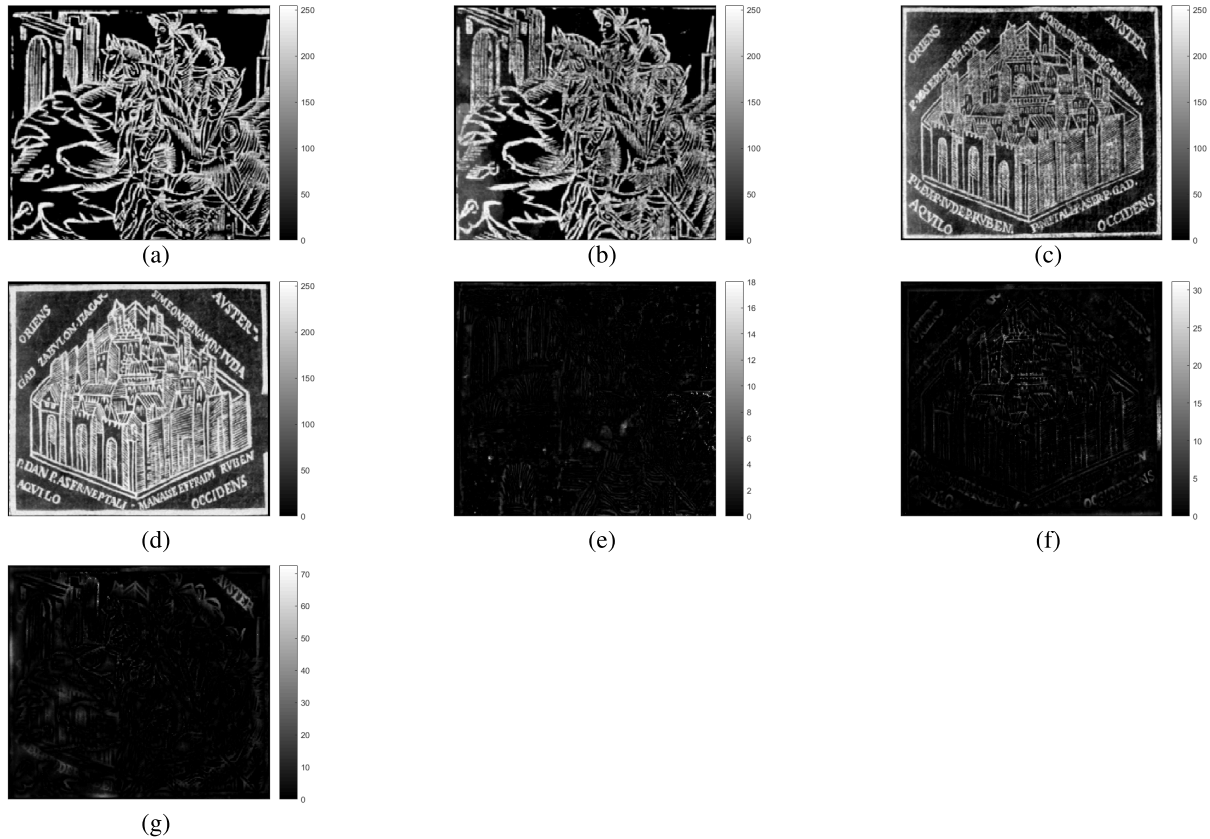


FIGURE 4. Grayscale images *A* (a), grayscale image *B* (b), grayscale image *C* (c), grayscale image *D* (d), LDM(*A*, *B*) (e), LDM(*C*, *D*) (f) and LDM(*A*, *C*) (g).

C. LDM FOR GRAYSCALE IMAGES

In order to extend the LDM to grayscale images, in eq. (6), the RVDT from Molchanov and Teran [13] can be used instead of the classical distance transform in eq. (5) or (6).

But in this case, distances between positions (x, y) and pixel values (luminance) are mixed together leading to interpretation and dynamics problems. Indeed, by mixing pure distances $(\max(DT_A(p), DT_B(p)))$ and luminance differences $(|A(p) - B(p)|)$, the distances obtained are no longer pure positional deviations between two pixels (p, q) . We clearly want in this paper to keep true distances for the LDM, which will be important in the section IV. One way to avoid this problem is to use the thresholding techniques of Molchanov by slicing grayscale images into several binary images and then compute simple multiple LDM between these binary images.

So the Real Value Local Dissimilarity Map (RVLDM) is defined as the sum of thresholded images:

$$RVLDM(A, B)(p) = \frac{1}{N} \sum_{i=1}^N LDM(A_i, B_i)(p), \quad (7)$$

where N is the number of thresholds used in the sum, τ_i is a threshold, $A_i = \{p \in A : A(p) \geq \tau_i\}$. Each threshold τ_i is chosen according to a regular spacing between

$m = \max(\min(A), \min(B))$ and $M = \min(\max(A), \max(B))$. So this regular spacing has a step of $s = (M - m)/N$.

Three examples of RVLDM between four grayscale images *A*, *B*, *C* and *D* are given in Figure 4. These are images extracted from old print data set where two images are considered similar if they represent the same scene. For similar images (see Figures (4e) and (4f)), the map appears much darker than for dissimilar images (see Figure (4g)). There are small pixel differences (in gray) on images from the same print. The areas where the two images being compared are not identical and clearly highlighted. The map of the dissimilar images (not coming from the same print) shows a very large difference between the two images with higher pixel values. It can also be seen, as in the case of binary images, because the maximum value of the gray levels of the RVLDM for dissimilar images is much higher than for similar images. The maximum RVLDM value for dissimilar images is 70 while it is only 18 and 30 for similar images. For grayscale images the same behavior of the RVLDM is observed as with two binary images (see Figure 3).

III. ROBUSTNESS OF THE LDM TO GEOMETRIC TRANSFORMATIONS

In this section, we present a comparison of the robustness between RVLDM, SSIM and FSIM measures. Using an input image from the F-MNIST data set, we applied deformations

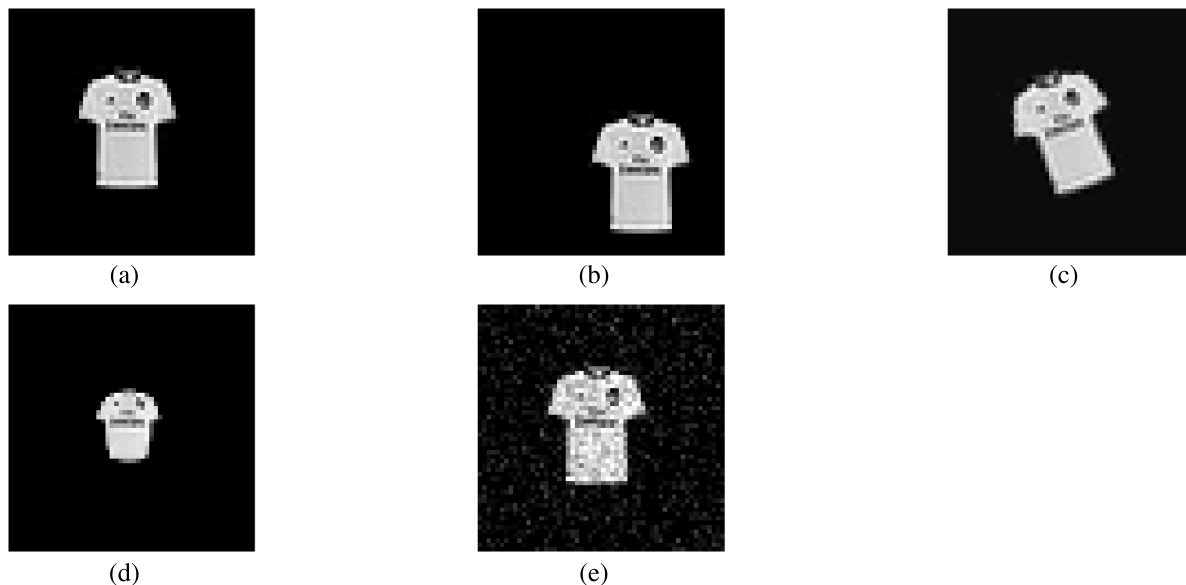


FIGURE 5. An original image (a) and examples of its variations (b), (c), (d) and (e). Translation with $t = 10$ (b), rotation with $\theta = 20^\circ$ (c), barrel deformation with $a = 10^{-4}$ (d) and Gaussian noise with PSNR = 22.7 dB (e).

which are translation, rotation, barrel deformation and noise addition. For each deformation, RVLDM, SSIM and FSIM are computed with respect to the input image. Subsequently, each measure is normalized to obtain values between 0 and 1. Figure 5 shows the input image and examples of its deformations. FSIM and SSIM are initially similarity measures and are here transformed into dissimilarity measures.

$$SSIM'(p) = 1 - SSIM(p), \tag{8}$$

$$FSIM'(p) = 1 - FSIM(p), \tag{9}$$

in order to have the same behavior as the RVLDM. In addition, normalizations are:

$$RVLDM_n = \frac{RVLDM}{\max(RVLDM)}, \tag{10}$$

$$SSIM_n = \frac{SSIM'}{\max(SSIM')}, \tag{11}$$

$$FSIM_n = \frac{FSIM'}{\max(FSIM')}, \tag{12}$$

where $RVLDM_n$, $SSIM_n$ and $FSIM_n$ are array containing the dissimilarity indices between the original image and the distorted images.

In Figure 6 we plot the maximum value of the normalized measures $RVLDM_n$, $SSIM_n$ and $FSIM_n$ with respect to the amount of deformation. The maximum value is taken into account, we do not measure a global similarity from the map, but we measure the local dissimilarity in order to localize the changes. Thus, the particular values of the maps are important. Taking into account the maximum provides some upper bound of the measures. We can observe in each graph (Figure 6) that the RVLDM and FSIM is monotonic with respect to all deformations. It can be seen in the graphs on Figure 6 that the RVLDM and FSIM gradually reach

their maximum. This ensures a good distribution of the RVLDM values. Since we cannot compute a map with FSIM, we took only the dissimilarity measure index. This shows the robustness of the measure of all deformations (except translation). On the other hand, SSIM gave erratic evolution for all deformations.

The next section will go deeply into the theory and will define a statistical model of the behavior of RVLDMs.

IV. STATISTICAL MODEL AND THEORETICAL VALIDATION

A. THE WEIBULL DISTRIBUTION

The Weibull distribution is a continuous probability distribution, discovered by the Swedish mathematician Waloddi Weibull [23]. Generally, the Weibull probability density function is defined as follows:

$$f(t) = \frac{\alpha}{\beta} \left(\frac{t - \gamma}{\beta} \right)^{\alpha - 1} e^{-\left(\frac{t - \gamma}{\beta} \right)^\alpha} \tag{13}$$

where $f(t) \geq 0$, $t > 0$, $\alpha > 0$, $\gamma \geq 0$ and $\beta > 0$. α is the shape parameter, β is the scale parameter and γ is the location parameter. The probability density f and the cumulative distribution function F of a two-parameter Weibull distribution ($\gamma = 0$) are:

$$f(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha - 1} e^{-\left(\frac{t}{\beta} \right)^\alpha}, \tag{14}$$

and

$$F(t) = 1 - e^{-\left(\frac{t}{\beta} \right)^\alpha}, \tag{15}$$

respectively, for $t > 0$, $\alpha > 0$ and $\beta > 0$. t is a random variable that represents in this paper the pixel values of RVLDM. Figure 7 shows the densities and the cumulative

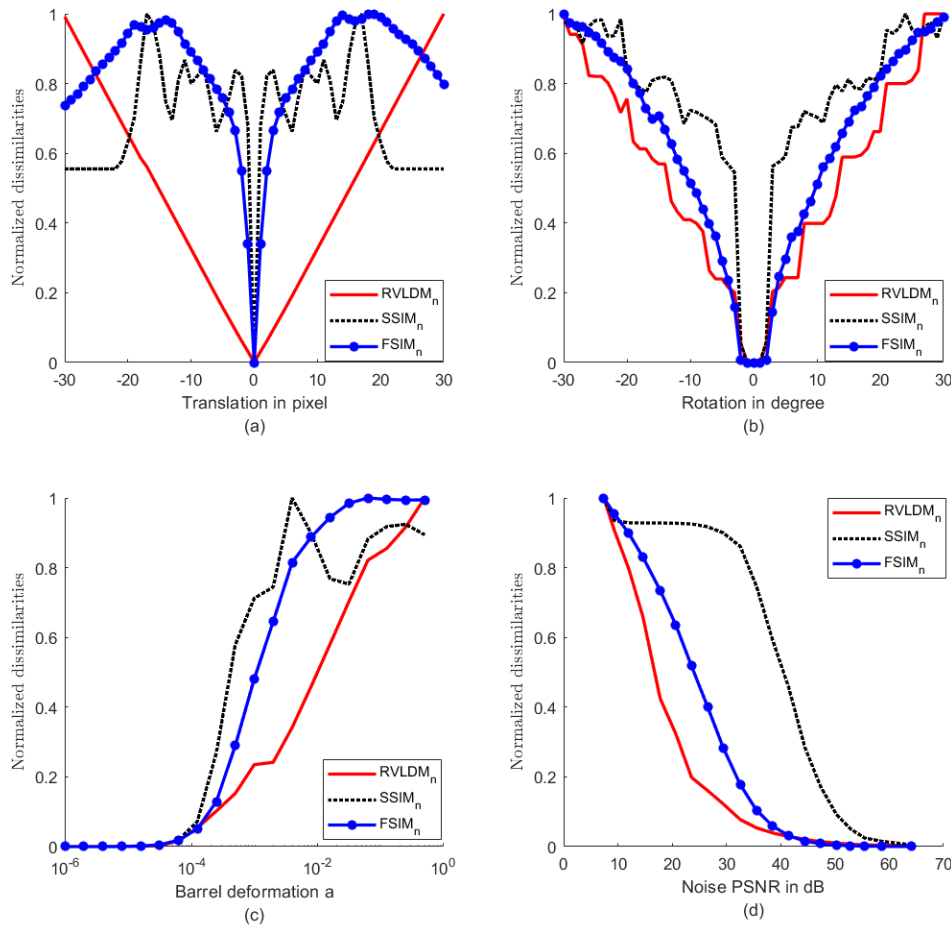


FIGURE 6. A comparison of RVLDL, SSIM and FSIM with respect to translation (a), rotation (b), barrel deformation (c) and noise addition (d). The maximum value of the measure map between a test image and its translation is plotted with respect to the deformation amount. For SSIM and FSIM we took the dissimilarity index. Note that for the barrel deformation, the bigger is the parameter a the smaller is the deformed image. In (d) the noise is a Gaussian white noise.

Weibull distribution functions for different shapes and scale parameters. In the next section, thanks to the *theorem 1*, we will show theoretically that the RVLDL follows a two-parameter Weibull distribution.

B. THEORETICAL MODEL VALIDATION

The modeling of a Local Dissimilarity Map is based on the modeling of the L_2 norm. The following theorem [24] gives us the distribution of the L_r ($r \geq 1$) norm under certain conditions.

Theorem 1: For non-identical, correlated and upper-bounded random variables $X_i = |S_i - T_i|^r$, the random variable $Z = \sum_{i=1}^N X_i$, with finite N , is Weibull-distributed. The density of Z is given by the eq. (14).

With S_i and T_i are feature vectors. In our case, S_i and T_i represent the pixel vector of image 1 or image 2.

For the proof, see [25]–[27].

Property 1: If the random variable Z is Weibull-distributed with shape (resp. scale) parameter α

(respectively β), then the variable $Z^{1/r} = (\sum_{i=1}^N X_i)^{\frac{1}{r}}$ is a Weibull distribution with parameters $r\alpha$ and $\beta^{\frac{1}{r}}$.

For the proof it is sufficient to calculate the distribution function of Z .

When $r = 2$, the L_2 norm between two features vectors follows a Weibull distribution $Wbl(2\alpha, \beta^{\frac{1}{2}})$.

1) FOR THE LDM OF BINARY IMAGES

The pixel values of a distance transform are non-identical and upper-bounded by $d = \sqrt{w^2 + h^2}$, where w and h are the dimensions of the image. Since the distance transform is constructed by the spatial propagation of a distance, then the neighboring pixel values of the distance transform are correlated. As an example, we have retrieved the values of the neighboring pixels of the distance transform obtained in Figure 1, plotted them against each other (see Figure 8) and computed the correlation coefficient which is equal to 0.9911. The conditions of the theorem 1 are checked, so the distance transform on a binary image is a Weibull distribution.

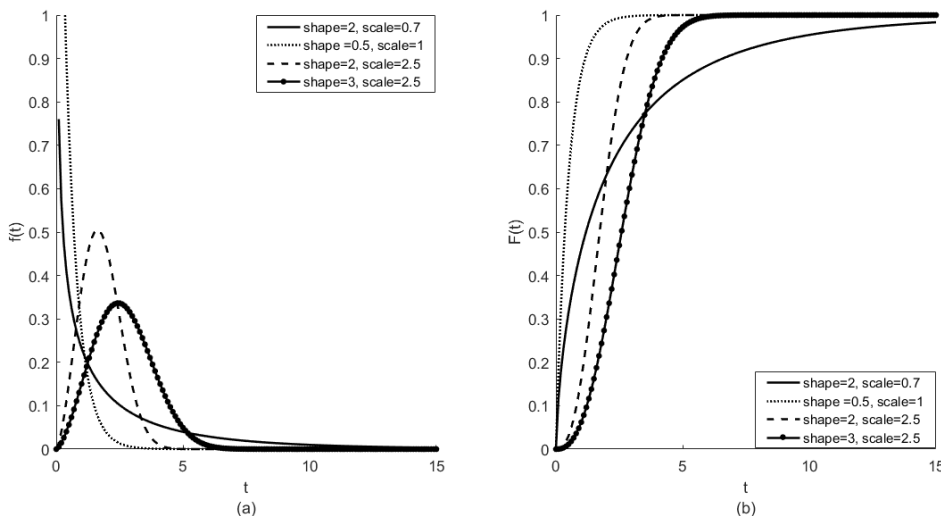


FIGURE 7. Densities of weibull distribution for different shapes and scale parameters (a). Cumulative distribution functions of weibull for different shapes and scale parameters (b).

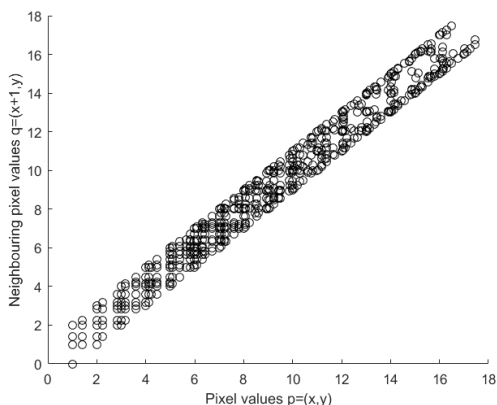


FIGURE 8. Neighboring pixel values graph of the distance transform map.

For any pixel p , $\min_q \|p - q\|_2$ is still a Euclidean distance. The Euclidean distance transform (dt_X) of a binary image X is then Weibull-distributed.

For A, B two binary images and $p \in A \triangle B$. From eq. (6) we have:

- if $A(p) = 1$ and $B(p) = 0$, then $DT_A(p) = 0$ and $LDM(A, B)(p) = DT_B(p)$,
- if $B(p) = 1$ and $A(p) = 0$, then $DT_B(p) = 0$ and $LDM(A, B)(p) = DT_A(p)$.

Previously we showed that the Euclidean distance transform of a binary image follows a Weibull distribution. So we can conclude that the Local dissimilarity Map of two binary images is Weibull-distributed. \square

2) FOR THE RVLDM OF GRAYSCALE IMAGES

As discussed in the previous section, in order to compute the Euclidean distance transform and the RVLDM between grayscale images, the images are divided into sub-images obtained using τ thresholds. The sub-images obtained after thresholding are binary images. From eq. (6) and (7),

we have:

$$\begin{aligned} RVLDM(p) &= \frac{1}{N} \sum_{i=1}^N (B_i(p)DT_{A_i}(p) + A_i(p)DT_{B_i}(p)) \\ &= \frac{1}{N} \sum_{i=1}^N B_i(p)DT_{A_i}(p) \\ &\quad + \frac{1}{N} \sum_{i=1}^N A_i(p)DT_{B_i}(p). \end{aligned} \tag{16}$$

Since RVLDM is a sum of Euclidean distance transforms, it is still a distance transform. As seen above, the distance transform follows a Weibull distribution. Hence, RVLDM is Weibull-distributed. \square

V. NUMERICAL EXPERIMENTATION

A. DATA DESCRIPTION

To evaluate the proposed method we used a local old print data set [10], and the *NIST data sets which are: MNIST [6], [9], F-MNIST [6], [9], E-MNIST [30], [31], K-MNIST [32]. The MNIST, F-MNIST, E-MNIST, K-MNIST data sets contain a large number of images of shape $28 * 28$ pixels. They are composed of 10 classes (0-9), except the E-MNIST data set. The E-MNIST data set contains uppercase letters (A-Z) and lowercase letters (a-z) which are each composed of 26 classes. F-MNIST data set contains a lot of contrasting images (see Figure 9(c)). All these data sets, except the old print data set (64 images), contain 60, 000 grayscale images for training and 10, 000 images grayscale for testing a model.

In [6], Xiao proposed different types of machine learning algorithms for solving MNIST and F-MNIST data set. In [9], Kadam applied convolutional neural networks (CNN) for image classification. Authors used five different architectures with varying convolutional layers, filter size and fully connected layers are proposed. They used MNIST and F-

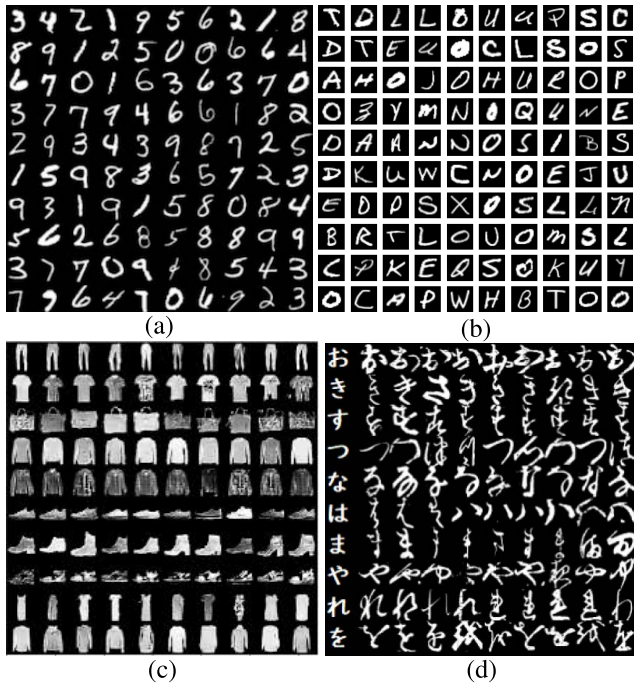


FIGURE 9. Example of *NIST images; MNIST (a), E-MNIST (b), F-MNIST (c) and K-MNIST (d).

MNIST to test the performances of CNN. In [31], Vaia proposed a deep unsupervised feature learning spiking neural networks with binarized classification layers for the E-MNIST classification. Authors have used binary activations to extract features from spiking input data, and a gradient descent on the output layer to perform training for classification. The E-MNIST data set is intended to represent a more challenging classification task for neural networks and learning systems. In [32], Clanuwat introduced the K-MNIST data set to engage the machine learning community to the field of Japanese literature.

In this paper, we randomly selected 130 images (to avoid bias in the results) from each class on MNIST, F-MNIST and K-MNIST data sets and 100 images from each class on E-MNIST data set. In addition, to compare images, each image must be compared to the rest of images in the data set in order to provide the RVLDMs of image pairs. We then obtained a huge number of RVLDMs. It can be seen in Table 1 that the comparison of 1,300 images provided 844,350 RVLDMs of distinct image pairs on F-MNIST data set, and 100 images on E-MNIST data set supplied 3,378,700 RVLDMs distinct image pairs. In practice, a storage constraint limits the number of usable images in the * NIST data sets.

B. NUMERICAL MODEL VALIDATION WITH THE K-S TEST

The Kolomorov-Smirnov test is an adjustment test based on distribution function F rather than density. It is based on these two assumptions:

- H_0 : the empirical distribution function \hat{F} is close to the distribution function F of a continuous law.
- H_1 : otherwise.

TABLE 1. Number of images, classes and RVLDM in each data set.

Data sets	# of images used	# of classes	# of RVLDMs
MNIST	1 300	10	844 350
F-MNIST	1 300	10	844 350
E-MNIST	2 600	26	3 378 700
K-MNIST	1 300	10	844 350
Old print	64	10	2 016

To perform this test, we first look for an estimate of the distribution function from an observed sample in order to compare it with the theoretical distribution function. The measurement of the fit from F to \hat{F} is done using the Kolmogorov-Smirnov statistic, K-S statistic [28]. To know whether to accept or reject the null hypothesis, we compare this Kolmogorov Smirnov statistic with a critical value $D_{\sigma}(n)$. This critical value depends on the σ risk of being wrong and the number of samples n .

In the previous sections, we have shown that the RVLDM is a two parameter (α, β) Weibull distribution. Parameters are estimated by using the maximum likelihood method [29], which not only allows us to model the RVLDMs but will also be used later to discriminate images (separating RVLDMs of similar image pairs and dissimilar image pairs, section V-C). To illustrate, we have modeled the RVLDMs of grayscale images given in Figure 4. Their histograms and their empirical distribution functions, fitted with a two-parameter Weibull distribution, are shown in Figure 10. The fits of the Weibull distribution on histograms and empirical distribution functions of RVLDMs in Figure 4 are perfect. So, the Weibull distribution fits well with the grayscale values of the RVLDMs. This fit is evaluated by Kolmogorov Smirnov’s statistical test at a confidence level of $\sigma = 0.05$. This experiment is repeated on more 1000 RVLDMs of old print images and every time the distance $D_{\sigma(n)}$ is less than the critical value. This experience remains true for any Local Dissimilarity Maps of binary or grayscale images.

C. BEHAVIOR OF DISTRIBUTION PARAMETERS FOR DATA SETS

The RVLDMs in each data set are labeled into similar and dissimilar classes. A RVLDM of image pairs is considered as similar when they belong to the same class. As mentioned above, the RVLDMs of images follow a two-parameter Weibull distribution. To illustrate RVLDMs discrimination, we took 6 classes and 15 images in each class of MNIST and F-MNIST data sets. The RVLDMs of all images are computed and the two parameters are then extracted and represented (scale parameter with respect to the shape parameter, see Figure 11). For these two data sets, we obtained 4005 RVLDMs, 630 of similar image pairs and 3,375 of dissimilar image pairs.

With our method, any image differences (whatever their size) can be summarized with only two values (α, β) extracted from the Weibull distribution of their RVLDMs. To classify image pairs, we only need to give a vector composed of two values into the classifiers to compute the results.

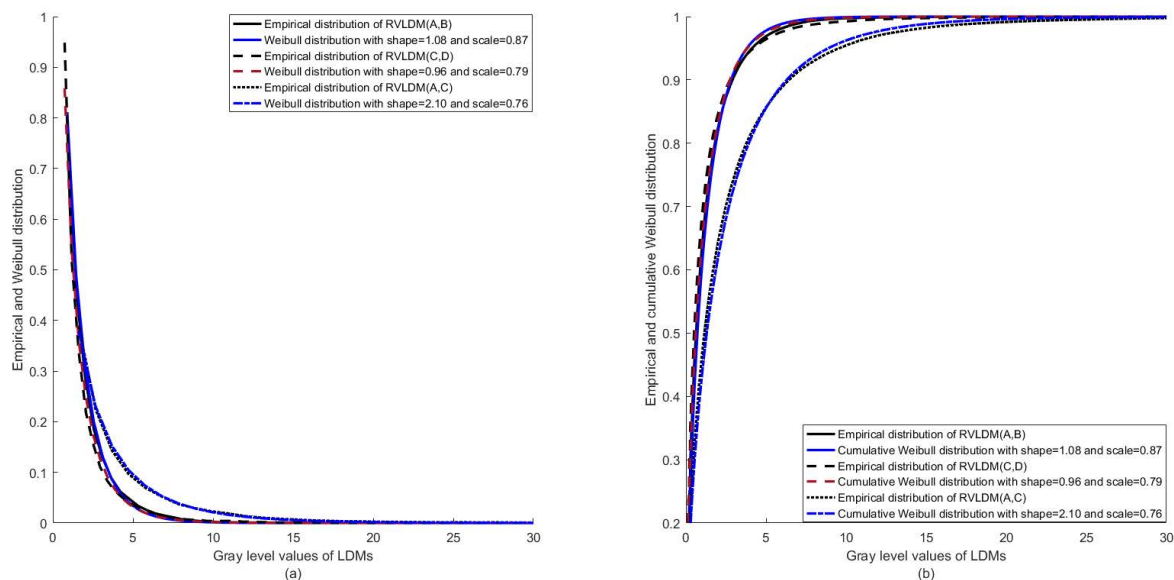


FIGURE 10. Empirical densities and empirical distribution functions of the RVLDMs in Figure 4 fitted with a two-parameter Weibull distribution. In (a), we merge on the same graph the histograms of RVLDM(A, B), RVLDM(A, C) and RVLDM(C, D) fitted by the weibull distribution with parameters $(\alpha_1 = 1.08, \beta_1 = 0.87)$, $(\alpha_2 = 0.96, \beta_2 = 0.79)$ and $(\alpha_3 = 2.10, \beta_3 = 0.76)$ respectively. In (b), we merge on the same graph the empirical distribution functions and the cumulative distribution functions of Weibull for the three RVLDM with the parameters (α_1, β_1) , (α_2, β_2) and (α_3, β_3) . Note that the curves of the two RVLDMs of similar images overlap and are detached from those of dissimilar images.

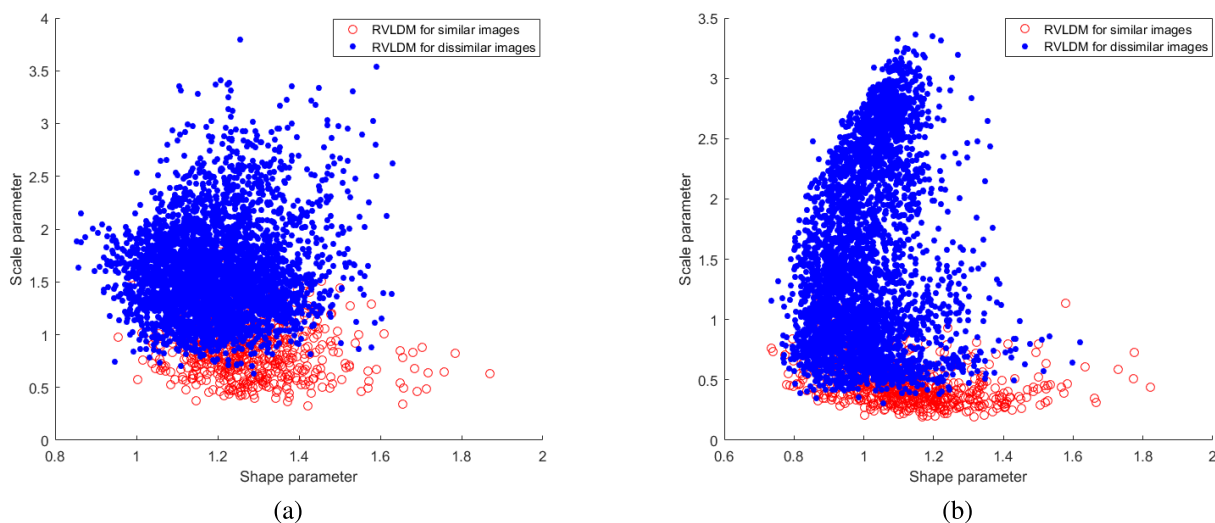


FIGURE 11. Representation of the scale and shape parameters of the RVLDM of images from MNIST data set (a) and F-MNIST data set (b).

It is clear that in the scale-shape parameter space, similar and dissimilar image pairs belong to distinct clusters. Figure 11 shows the relevance of the two parameters of Weibull to distinguish the RVLDMs of similar and dissimilar image pairs. So the next step is to use supervised binary classifiers to be able to discriminate similar and dissimilar image pairs. A multi-class classification will also be tested to show the effectiveness of our method with limited data available. The following sections present the results of binary and multi-class classification obtained from our proposed approach.

D. 2-CLASS CLASSIFICATION

In this section, supervised algorithms are used to classify similar image pairs and dissimilar image pairs. In this work, the following three supervised classifiers are used: k -Nearest Neighbors (k -NN) [33], [34], Artificial Neural Networks (ANN) [35], [36] and the Logistic Regression (LOGR) [36]. The k -NN is one of the easiest supervised learning algorithms to implement. It can be used to solve both classification and regression problems. The ANN are widely used in image processing. They do more in-depth learning and give better performance depending on the number of hidden layers and

TABLE 2. Classifiers and their implementation details.

Classifiers	Scikit-learn class	Hyperparameter setting
k -NN	sklearn.neighbors.KNeighbors	weights=distance, n_neighbors=9, p=1
ANN	sklearn.neural_network.MLPClassifier	activation='relu', hidden_layer_sizes=(100, 10)
LOGR	sklearn.linear_model.LogisticRegression	C = 1, multi_class = ovr, penalty=l ₁

TABLE 3. Accuracy, recall, precision, and F_1 score obtained by the classifiers k -nearest neighbors, artificial neural network and logistic regression. Application on MNIST, F-MNIST, K-MNIST, E-MNIST and old print data sets. It is a binary (two classes) classification: class of similar image pairs and class of dissimilar image pairs using the RVLDM, MSSIM and FSIM approaches. In bold is where our approach gives best performances for the different data sets.

Metrics / Classifiers & measures	k -Nearest Neighbors			Artificial Neural Network			Logistic Regression			
	RVLDM	MSSIM	FSIM	RVLDM	MSSIM	FSIM	RVLDM	MSSIM	FSIM	
MNIST	Accuracy	0.904	0.872	0.880	0.920	0.910	0.913	0.919	0.905	0.912
	Recall	0.34	0.20	0.27	0.30	0.16	0.20	0.24	0.07	0.16
	Precision	0.61	0.33	0.39	0.75	0.72	0.73	0.82	0.79	0.78
	F_1 score	0.43	0.25	0.32	0.43	0.26	0.32	0.37	0.12	0.27
F-MNIST	Accuracy	0.891	0.891	0.892	0.906	0.913	0.905	0.905	0.914	0.914
	Recall	0.36	0.33	0.41	0.40	0.33	0.37	0.29	0.37	0.40
	Precision	0.49	0.58	0.55	0.55	0.50	0.52	0.55	0.62	0.60
	F_1 score	0.42	0.42	0.47	0.46	0.39	0.43	0.38	0.46	0.48
E-MNIST (upper case letter)	Accuracy	0.965	0.961	0.962	0.966	0.962	0.963	0.966	0.962	0.963
	Recall	0.22	0.09	0.09	0.21	0.08	0.08	0.13	0.06	0.04
	Precision	0.53	0.35	0.35	0.57	0.36	0.36	0.61	0.51	0.77
	F_1 score	0.31	0.14	0.14	0.30	0.13	0.13	0.21	0.10	0.08
K-MNIST	Accuracy	0.883	0.858	0.860	0.905	0.901	0.900	0.901	0.900	0.901
	Recall	0.13	0.08	0.11	0.10	0.02	0.05	0.01	0.00	0.01
	Precision	0.44	0.25	0.27	0.64	0.51	0.56	0.89	0.35	0.65
	F_1 score	0.20	0.12	0.15	0.17	0.05	0.09	0.02	0.00	0.02
Old print	Accuracy	0.953	0.935	0.930	0.952	0.952	0.949	0.947	0.952	0.951
	Recall	0.37	0.20	0.09	0.23	0.21	0.04	0.19	0.10	0.06
	Precision	0.53	0.45	0.46	0.57	0.55	0.69	0.43	0.33	0.40
	F_1 score	0.43	0.27	0.15	0.32	0.30	0.08	0.26	0.15	0.10

the number of neurons in each layer. Finally, LOGR is one of the statistical approaches that can be used to evaluate and characterize the relationships between a binary response variable and exploratory variables that can be categorical

or numerical. In our case, we used a supervised learning algorithm.

In the rest of this paper, we use these three algorithms proposed by Xiao [6]. First, the k -NN using the number of

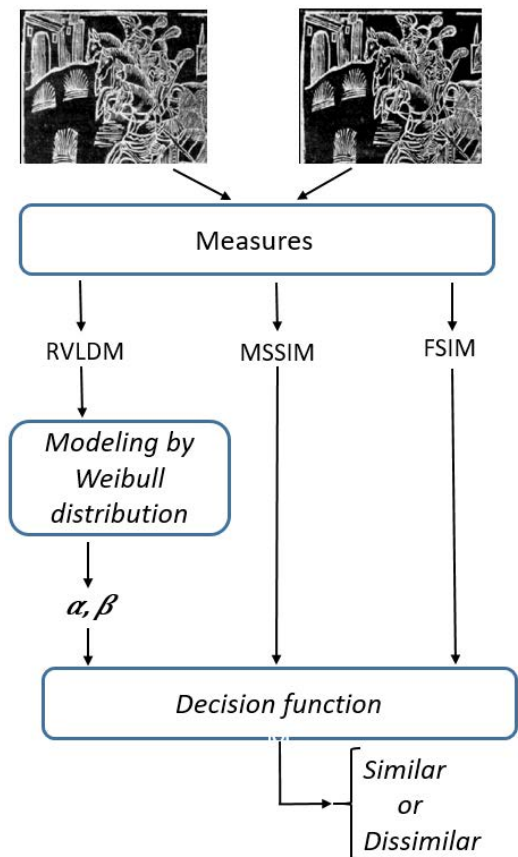


FIGURE 12. The whole process for classifying an image pair into similar and dissimilar. Our approach uses the two weibull parameters as descriptors and the others use the similarity indices.

neighbors $k = 9$ and Manhattan distance (l_1). Second, the ANN using 'relu' as activation function, and 2 hidden layers with 100 neurons on the first hidden layer and 10 neurons on the second hidden layer. Third, the LOGR using the inverse of regularization strength $C = 1$, and (l_1) penalty term as hyperparameters. Their implementation details are illustrated in Table 2. For each classifier, hyperparameters chosen provide good performances. In this section, the purpose is to classify image pairs into similar and dissimilar using the Weibull parameters as descriptors. Figure 12 shows the whole process for classifying an image pair into similar and dissimilar classes. The performance of our proposed method is compared with the state-of-the-art similarity measures such as MSSIM and FSIM.

In the MNIST, F-MNIST and K-MNIST data sets, 83, 850 RVLDMs of similar image pairs and 760, 500 RVLDMs of dissimilar image pairs were obtained. For the E-MNIST data set, we obtained 128, 700 similar image pairs and 3, 250, 000 dissimilar image pairs. Finally, for the old print data set we have 96 RVLDMs of similar image pairs and 1, 795 RVLDMs of dissimilar image pairs. In each case, 10-cross validation is used for all data to compute performances. Since these two classes are very unbalanced for each data set (the class of dissimilar image pairs contains

twice or three times as much RVLDMs as that of similar image pairs), we evaluate the performance of our proposed method by calculating the accuracy, the Recall, the Precision, and the F_1 score with respect to the similar image pairs class. The accuracy is the percentage of correct predictions. In our case, it is the percentage of correctly classified image pairs. The F_1 score is the harmonic mean of Precision and Recall. The Precision refers to positive predictive value and the Recall refers to true positive rate.

Algorithm 1 2-Class Classification Procedure

Input: Image pairs (A_i, A_j) , $i = 1, \dots, N-1$ and $j = i + 1, \dots, N$, where N is the number of images
Output: The prediction vector for 2-class classification ℓ

- 1: **for** $i = 1$ to N step 1 **do**
- 2: **for** $j = i+1$ to $N-1$ step 1 **do**
- 3: Compute $RVLDM(A_i, A_j)$, refer to eq.(7).
- 4: **end for**
- 5: **end for**
- 6: **for** $i = 1$ to Number of RVLDMs **do**
- 7: Extract (α_i, β_i) using the maximum likelihood method.
- 8: **end for**
- 9: Apply a 10-cross validation in the data (α, β) using classifiers k -NN, ANN and LOGR.

- 1) Accuracy = $\frac{TP+TN}{TP+FP+TN+FN}$
- 2) Recall = $\frac{TP}{TP+FN}$
- 3) Precision = $\frac{TP}{TP+FP}$
- 4) F_1 score = $\frac{2 \times (Recall \times Precision)}{(Recall + Precision)}$
 - TP = Number of similar image pairs correctly classified.
 - TN = Number of dissimilar image pairs correctly classified.
 - FP = Number of dissimilar image pairs classified as similar.
 - FN = Number of similar image pairs classified as dissimilar.

E. APPLICATIONS OF GEOMETRIC TRANSFORMATIONS ON DATA SETS

In this section, we will see the robustness of the RVLDM on geometric transformations such as pixel translations. Figure 13 shows the evolution of the F_1 score as a function of the number of translated pixels in the image to be compared in the MNIST, FMNIST, KMNIST and old print data sets. Since k -NN with $k = 9$ gave the best performances for classifying image pairs into similar and dissimilar without geometric transformations (see Table 3), we took the same classifier to compute the F_1 score of similar image pair classes. We also tested for the ANN classifier and the same behavior was observed. Even before calculating the F_1 score, we had tested for both classifiers the recall and the precision. For each metric and regardless of the classifier, the RVLDM is more robust than the state-of-the-art similarity measures to pixel

Algorithm 2 Multi-Class Image Classification Procedure

Input: Images $A_i, i = 1, \dots, N$ where N is the number of images

Output: The prediction vector for the multi-class image classification v_1

- 1: **for** $i = 1$ to N step 1 **do**
- 2: **for** $j = i + 1$ to $N - 1$ step 1 **do**
- 3: Compute RVLDM(A_i, A_j), refer to eq.(7).
- 4: **end for**
- 5: **end for**
- 6: **for** $= 1$ to Number of RVLDMs **do**
- 7: Extract (α_i, β_i) using the maximum likelihood method.
- 8: **end for**
- 9: Apply a 10-cross validation in the data (α, β) using classifiers k -NN, ANN and retrieve the prediction label ℓ . The vector ℓ contains, in our case, 0 and 1. 1 corresponds the class of similar image pairs and 0 the class of dissimilar image pairs.
- 10: Initialize a square matrix M of size N .
- 11: Initialize $x = 0$
- 12: **for** $i = 1$ to N step 1 **do**
- 13: **for** $j = i + 1$ to N step 1 **do**
- 14: $M[i, j] = \ell[x]$
- 15: $x = x + 1$
- 16: **end for**
- 17: **end for**
- 18: $R = M + M^t$. M^t is the transpose matrix of M
- 19: Initialize a matrix C of size $N * m$ where m is the number of classes in the data set.
- 20: $N_1 = N/m$ is the number of images in each class
- 21: **for** $k = 1$ to m step 1 **do**
- 22: $s_1 = (k - 1) * N_1$ and $s_2 = k * N_1$
- 23: **for** $i = 1$ to N step 1 **do**
- 24: **for** $j = s_1$ to s_2 step 1 **do**
- 25: $C[i, k] = \text{sum}(R[i, s_1 : s_2] == 1)$, count the number of similar RVLDM correctly classified in each line in the matrix R .
- 26: **end for**
- 27: **end for**
- 28: **end for**
- 29: Initialize a vector v_1 of size N and a vector v_2 ranging from 0 to 9 with step 1
- 30: **for** $i = 1$ to N step 1 **do**
- 31: $S = C[i, :]$
- 32: **for** $j = 1$ to m step 1 **do**
- 33: **if** $\max(S) = S[j]$ **then**
- 34: $v_1[i] = v_2[j]$
- 35: **end if**
- 36: **end for**
- 37: **end for**

translations. The same behavior was also observed for other transformations such as rotation, noise addition and barrel deformation.

TABLE 4. Accuracy obtained by the classifiers K-nearest neighbors, artificial neural networks. Application on MNIST, F-MNIST, K-MNIST, E-MNIST data sets. It is a multi-class classification using our approach based on RVLDM and those proposed by Xiao [6]. The latter is a well known and widely used method. Values in bold are where our approach gives the best performances for the different data sets.

	<i>k</i> -NN		ANN	
	RVLDM	Xiao [6]	RVLDM	Xiao [6]
MNIST	0.910	0.891	0.910	0.896
F-MNIST	0.75	0.909	0.762	0.93
E-MNIST	0.892	0.812	0.896	0.745
K-MNIST	0.792	0.782	0.836	0.731

After evaluating the effectiveness of the proposed approach on similar and dissimilar image pairs classification (2-class classification), we will test it on multi-class classification in the next section.

F. MULTI-CLASS CLASSIFICATION

In this part, the proposed method is tested on 10-class image classification for MNIST, F-MNIST, and K-MNIST data sets and 26-class image classification for E-MNIST data set. This approach consists of extracting the Weibull's parameters (α, β) in each RVLDM and to use them as input data in the classifiers to be able to give the class of unknown images. We have 10 classes in each data set (except E-MNIST data set which has 26 classes). The classes are given from 0-9 for MNIST, F-MNIST, K-MNIST data sets and A-Z for E-MNIST data set. In this case, we make a multi-class classification instead of a 2-class classification. To know the class of an unknown image, the RVLDM of this image is calculated with all images in each class and the proposed class of the image is the one that contains the most similar RVLDMs. The method is compared with the widely used state-of-the-art machine learning methods proposed by Xiao [6] (see Table 4). There are methods that consist of injecting images directly into a classifier. It is the role of the classifier to look directly at the image features and to decide their class. We used the same classifiers quoted in the section V-D with the same hyper parameters. Table 3 shows that the K-Nearest Neighbors (*k*-NN) and Artificial Neural Network (ANN) are given the best F_1 score. Hence, in the rest of this paper, these two classifiers and 10-cross validation are used to perform the classification.

G. INFERENCE TIMES IN BINARY AND MULTI-CLASS CLASSIFICATION

In this section, we present the inference times of our proposed method and the state-of-the-art methods on *NIST and old print data sets. In this paper, it is the time of extracting input parameters and classification. For our approach the inputs represent (α, β) parameters of the Weibull distribution,

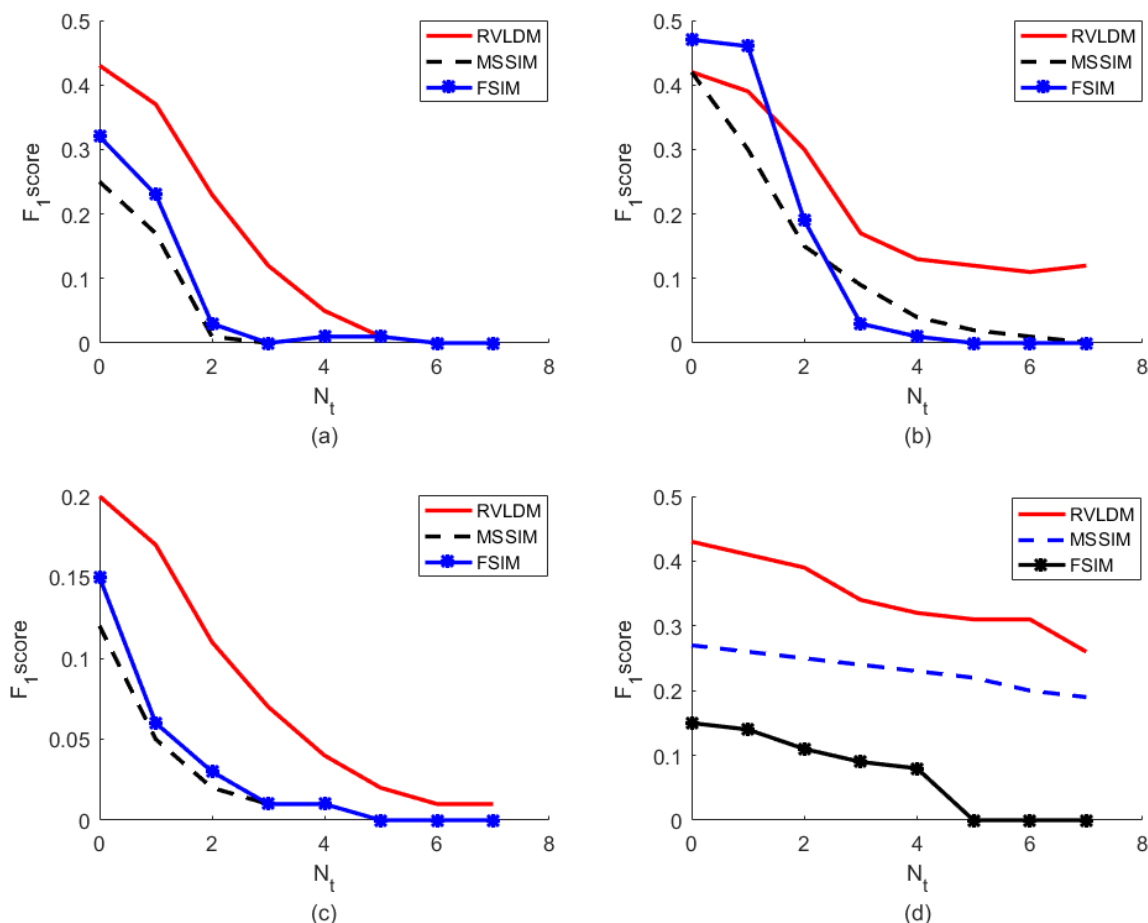


FIGURE 13. The F_1 score curves for each measure as a function of the number of translated pixels N_t of the image to be compared in the MNIST (b), FMNIST (c), KMNIST (d) and old print (e) data sets. In each instance, k -NN with $k = 9$ is used to compute the F_1 score. We also tested for the ANN classifier and the same behavior has been observed. Even before calculating the F_1 score, we had tested for both classifiers the recall and the precision and in each instance our method is more robust the state-of-the-art methods.

extracted from the RVLDMs of image pairs. For MSSIM and FSIM, the inputs are the similarity indices. Table 5 shows the inference times in seconds of our approach and those of the state-of-the-art methods for image pairs classification (2-class classification). Steps of the approach are illustrated in Algorithm 1. Table 6 presents the inference times of our method and methods presented in [6] for multi-class image classification. Algorithm 2 present the steps of this approach.

VI. COMPARISON TO STATE-OF-THE-ART METHODS

First, we evaluate our approach against the two state-of-the-art similarity measures: the MSSIM and the FSIM. For both measures, the similarity index of image pairs is calculated and it is used as descriptors for classifiers. Table 3 present image pairs classification (2-class classification) results based on similarity measures. We can see that the proposed method based on the RVLDM outperforms the state-of-the-art methods in terms of accuracy and F_1 score when we took k -NN and ANN as classifiers. However, our method does not outperform the state-of-the-art methods in the F-MNIST data set when we used the Logistic Regression.

Indeed, the F-MNIST data set contains contrasted images and the RVLDM is sensitive to these kinds of images. Despite the fact that it is below the other methods, it gives good performances in terms of accuracy, and even in terms of F_1 score.

Second, geometric transformations such as pixel translation is applied to show the robustness of our proposed method against the state-of-the-art similarity measures. In Figure 13, we can see that our method is more robust than others to pixel translation. Whatever the number of translated pixels, our approach gives the highest F_1 score compared to other methods. This behavior is also true for other transformations such as rotation, noise addition and barrel deformation.

Finally, to further investigate the effectiveness of our method, we applied it to a multi-class image classification problem. It is compared against the state-of-the-art classification methods based on machine learning proposed by Xiao [6]. In this part, the classifiers used in [6] are the same ones cited in section V-D. Table 4 presents the results of our proposed method and the state-of-the-art image classification methods. The results show that our approach

TABLE 5. *NIST and old print data sets: Inference times in seconds for extracting input parameters (α, β) for our measure RVLDM and similarity indices for the state-of-the-art measures MSSIM and FSIM), and image pairs classification for each classifier.

Inference time (s)/ Classifiers and measures		k -Nearest Neighbors			Artificial Neural Network			Logistic Regression		
		RVLDM	MSSIM	FSIM	RVLDM	MSSIM	FSIM	RVLDM	MSSIM	FSIM
*NIST data sets	Inference time (s)	597.7	208.6	3599.8	1764.7	1377.6	4768.8	612.7	225.6	3617.8
Old print data set	Inference time (s)	105.2	27.02	721.9	343.7	194.1	948.9	107.5	29.3	723.0

TABLE 6. *NIST data sets: Inference times in seconds for extracting input parameters (α, β) for our measure RVLDM and multi-class image classification for k -NN and ANN classifiers. We also provide inference in seconds of the state-of-the-art methods proposed by Xiao [6] in the same data sets.

Inference time (s)/ Classifiers and measures		k -NN		ANN	
		RVLDM	Xiao [6]	RVLDM	Xiao [6]
*NIST data sets		612.7	744.1	1779.9	1818.2

based on the RVLDM outperforms the other methods in terms of accuracy in data sets, except for the F-MNIST data set.

In recent years, researchers have used deep learning methods based on Convolutional Neural Network (CNN), which is a branch of machine learning, to classify images in these different data sets. CNN methods are currently very successful to image classification. Kadam proposed in [9] five architectures of CNN to classify images from the F-MNIST and the MNIST data sets. Authors used architectures with varying convolutional layers, filter size and fully connected layers. For the two data sets the best accuracy is given by the architecture 3. For MNIST data set, the accuracy obtained is 99.3% while the F-MNIST, the accuracy is 93.5%. Results are obtained with 128 batch size, softmax activation function, adam optimizer, 0.25 dropout after each pooling layer, 50 epochs and 2×2 kernel size. CNN and Long Short-Term Memory (LSTM) algorithms are introduced by [30] to classify images from the MNIST and E-MNIST data sets. For the MNIST data set, the accuracies obtained from CNN and LSTM are 98.2% and 98.3% respectively. For the E-MNIST data set, the accuracies obtained from CNN and LSTM are 85.1% and 85.7% respectively. To our knowledge, they are the most recent deep learning methods applied to these data sets. These models are all trained with 60,000 images and tested with 10,000 images from each data set. However, only a few of the images are needed by our approach to perform well. The proposed method is very efficient when we don't have a large number of image data sets at our disposal. In this case, deep learning models will not be well trained to make good predictions.

Although our approach gives less performance for the F-MNIST data set, it shows good accuracies for the other data sets. For the MNIST data set, there is no significant difference in performance between our approach and the state-of-the-art deep learning methods. However, we can see that our method

outperforms the methods presented in [30] for the E-MNIST data set. Machine learning methods require an input vector of size (w, h) (because it takes the entire image as an input) while our method needs only a vector with two values (α, β) as an input for any classifier. So, we obtained comparable, even better results compared to the state-of-the-art.

Figure 14 shows confusion matrices achieved when using both classifiers formed on the MNIST data set. For both classifiers, our method outperforms the others in terms of class accuracy. These results were also tested on the other data sets (except the F-MNIST data set) and the same results were obtained.

In terms of inference times, we can see that the proposed method uses more computing time than the MSSIM method and less computing time than the FSIM method (see Table 5). For the MSSIM, although it is the least expensive in inference times, it gives worse performances compared to our method and FSIM. However, the Table 6 shows that the proposed method is the least expensive in inference times compared to the method proposed in [6] for multi-class image classification. Deep learning methods are known to require a huge amount of resources for model learning. In contrast, our image classification method is simpler and more efficient and does not require this huge amount of resources and image data sets.

We also observed that the complexities of algorithms in time and memory used in the proposed method with different classifiers were $O(n^2)$. The complexity of algorithms increases quadratically when the image sizes or the number of images increases.

The main limitation of the proposed method is that it fails on images with mixed structures and textures. Contrasted images contain a lot of structures. Although the state-of-the-art methods outperform ours in terms of accuracy for multi-class image classification, our approach works correctly and has given good performances for these types of images.

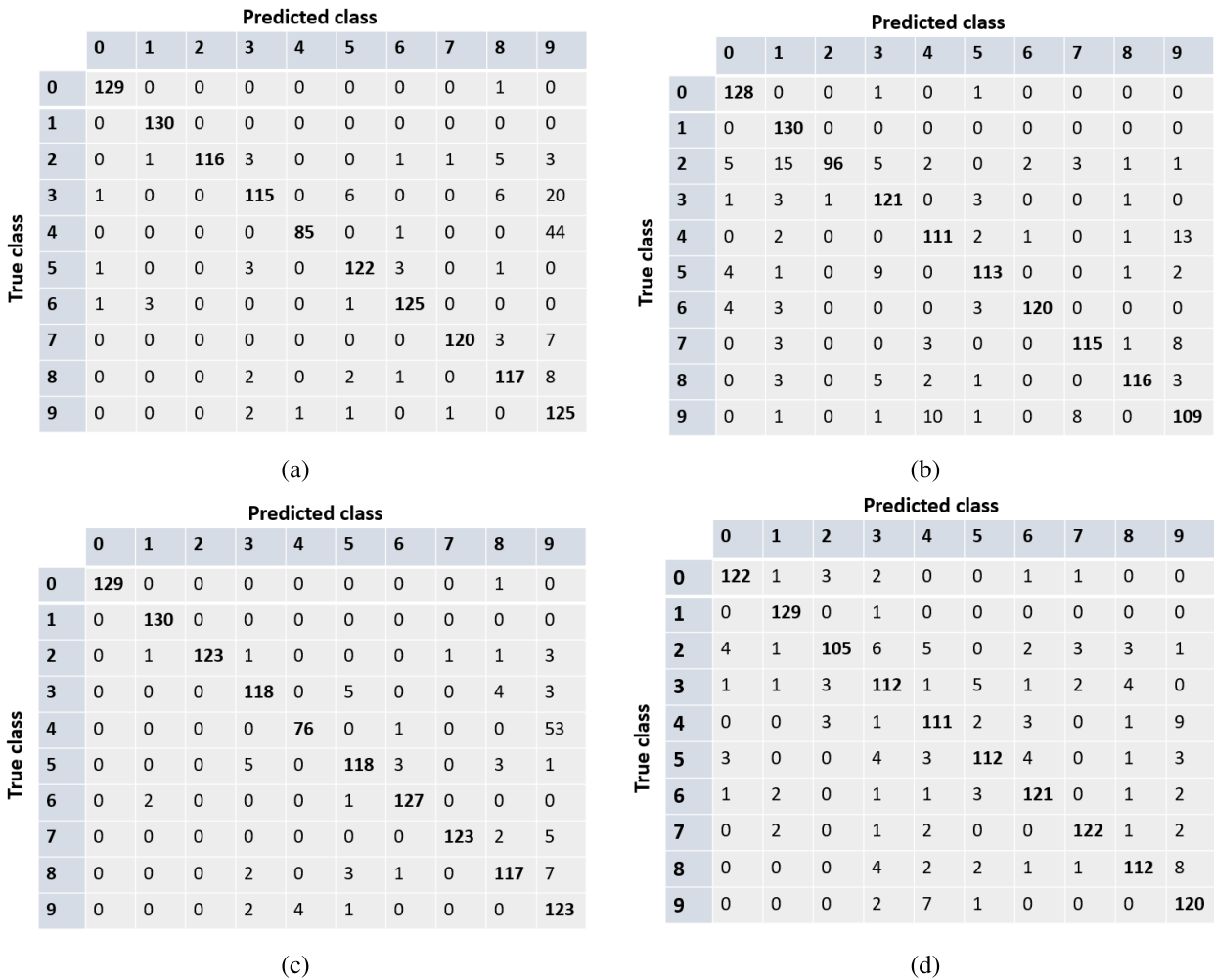


FIGURE 14. At the top, we have the confusion matrix obtained using *k*-NN of our proposed method (a) against the method proposed by Xiao [6] (b) on MNIST data set. At the bottom, we have the confusion matrix obtained using ANN of our proposed method (c) against the method proposed by Xiao [6] (d) on the same data set.

We decided not to do any preprocessing on the images to show the robustness of our proposed method.

VII. CONCLUSION AND PERSPECTIVES

Image comparison and image classification have become one of the research concerns in recent years, due to the considerable increase of images in our digital world. This paper presents a new method to image pairs classification and especially multi-class image classification based on the modeling of the Local Dissimilarity Map. We proposed a two-parameter Weibull distribution to model the Local Dissimilarity Map between two binary and grayscale images. A Kolmogorov-Smirnov test was used to validate the model. The two parameters characterizing the model are exploited to distinguish the RVLDMs of image pairs into similar and dissimilar classes. We showed that the proposed method is robust to geometric transformations such as pixel translation. The method is also applied to multi-class image classification using the parameters (α, β) as descriptors. In each case,

our proposed approach is comparable, even better than the state-of-the-art methods in terms of accuracy and F_1 score, except for the F-MNIST data set with contains contrasted images. The approach is also the least expensive in inference times compared to the state-of-the-art images classification methods. Unlike deep learning methods, our proposed approach is even very efficient when we don't have a large number of image data sets at our disposal. It is a method that could be used in the medical field or other fields where there is not a huge available data set.

In future works, the two following points will be addressed to overcome the problem of the RVLDM on images with mixed structures and textures and to improve the robustness of the method on contrasted images. They will enable us to improve the results of our proposed method in different data sets.

- We will inject in the RVLDM one of dissimilarity measures to detect differences on images images containing contrasts or mixed structures and textures with very

small windows. Among these measures, we have mutual information, disjoint information, Minkowski distance and Kullback-Leibler's divergence. Mutual information is a similarity measure that has shown good results in the field of medical image registration [37]. It allows us to analyze, interpret the data judiciously and quantify the information measured in terms of entropy. The disjoint information is the opposite of mutual information. The difference between these two measures is the role of joint entropy. For this reason, we want to exploit them, in the future, as local measures for the RVLDM. It could allow us to improve the results of our proposed method for each data set.

- As most of image classification methods, preprocessing on input images is required to have good performances. In the future, we will then use a Local Contrast Normalization method [38] to overcome the problem of the RVLDM on contrasted images. This preprocessing will make our measure more efficient at detecting dissimilarities and it will improve the results of our proposed method on each data set.

ACRONYMS

- LDM Local Dissimilarity Map
- PSNR Peak Signal to Noise Ratio
- MSE Mean Square Error
- MSSIM Mean Structural Similarity Index
- FSIM Feature Similarity Index
- DT Distance Transform
- RVDT Real Value Distance Transform (distance transform for grayscale images)
- RVLDM Real Value Local Dissimilarity Map (LDM for grayscale images)
- k -NN k -Nearest Neighbors
- ANN Artificial Neural Networks
- LOGR Logistic Regression

REFERENCES

- [1] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Orlando, FL, USA, vol. 4, May 2002, pp. 3313–3316.
- [2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 600–612, Apr. 2004.
- [3] M.-J. Chen and A. C. Bovik, "Fast structural similarity index algorithm," *J. Real-Time Image Process.*, vol. 6, no. 4, pp. 281–287, Dec. 2011.
- [4] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [5] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study," *J. Comput. Commun.*, vol. 7, no. 3, pp. 8–18, 2019.
- [6] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*. [Online]. Available: <https://github.com/zalandoresearch/fashion-mnist>
- [7] W. Gao, M. Wu, S.-K. Lam, Q. Xia, and J. Zou, "Decoupled self-supervised label augmentation for fully-supervised image classification," *Knowl.-Based Syst.*, vol. 235, Jan. 2022, Art. no. 107605.
- [8] A. Sellami and S. Tabbone, "Deep neural networks-based relevant latent representation learning for hyperspectral image classification," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108224.
- [9] S. S. Kadam, A. C. Adamuthe, and A. B. Patil, "CNN model for image classification on MNIST and fashion-MNIST dataset," *J. Sci. Res.*, vol. 64, no. 2, pp. 374–384, Jan. 2020. [Online]. Available: <http://yann.lecun.com/exdb/mnist/>
- [10] É. Baudrier, F. Nicolier, G. Millon, and S. Ruan, "Binary-image comparison with local-dissimilarity quantification," *Pattern Recognit.*, vol. 41, no. 5, pp. 1461–1478, May 2008.
- [11] A. Kharbach, M. El Omari, A. Mardani, B. Bellach, and M. Rahmoun, "Medical images registration based on normalized dissimilarity index," *J. Comput. Sci.*, vol. 13, no. 9, pp. 380–392, Sep. 2017.
- [12] A. Mardani, A. Kharbach, M. Rahmoun, B. Bellach, M. Elayachi, and M. Elhitmy, "Local and global measure of dissimilarity between two segmentations," *Int. J. Innov. Technol. Exploring Eng.*, vol. 4, no. 6, pp. 2278–3075, 2014.
- [13] I. S. Molchanov and P. Terán, "Distance transforms for real-valued functions," *J. Math. Anal. Appl.*, vol. 278, no. 2, pp. 472–484, Feb. 2003.
- [14] H. P. Nguyen, A. Delahaies, F. Retraint, and F. Morain-Nicolier, "Face presentation attack detection based on a statistical model of image noise," *IEEE Access*, vol. 7, pp. 175429–175442, 2019.
- [15] M. Huang, M. Xu, T. Qiao, T. Wu, and N. Zheng, "Designing statistical model-based discriminator for identifying computer-generated graphics from natural," *J. Universal Comput. Sci.*, vol. 25, no. 9, p. 1151, 2019.
- [16] T. Qiao, R. Shi, X. Luo, M. Xu, N. Zheng, and Y. Wu, "Statistical model-based detector via texture weight map: Application in re-sampling authentication," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1077–1092, May 2019.
- [17] T. N. C. Doan, F. Retraint, and C. Zitzmann, "Image tampering detection based on a statistical model," *Multimedia Tools Appl.*, vol. 80, nos. 21–23, pp. 32905–32924, Sep. 2021, doi: [10.1007/s11042-021-11213-3](https://doi.org/10.1007/s11042-021-11213-3).
- [18] T. Strutz, "The distance transform and its computation," *Comput. Vis. Pattern Recognit.*, vol. 2106, 2021.
- [19] J. Wang and Y. Tan, "Efficient Euclidean distance transform algorithm of binary images in arbitrary dimensions," *Pattern Recognit.*, vol. 46, no. 1, pp. 230–242, 2013.
- [20] P. Ramya and R. Rajeswari, "Human action recognition using distance transform and entropy based features," *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 8147–8173, Mar. 2021.
- [21] D. Zhu, C. Wang, W. Wang, R. Garg, S. Scherer, and M. Q.-H. Meng, "VDB-EDT: An efficient Euclidean distance transform algorithm based on VDB data structure," 2021, *arXiv:2105.04419*.
- [22] E. A. Pambudi, "Improved Sauvola threshold for background subtraction on moving object detection," *Int. J. Softw. Eng. Comput. Syst.*, vol. 5, no. 2, pp. 78–89, Aug. 2019.
- [23] W. Wang and T. Okaze, "Statistical analysis of low-occurrence strong wind speeds at the pedestrian level around a simplified building based on the Weibull distribution," *Building Environ.*, vol. 209, Feb. 2022, Art. no. 108644.
- [24] G. J. Burghouts, A. W. M. Smeulders, and J. M. Geusebroek, "The distribution family of similarity distances," in *Proc. 20th Int. Conf. Neural Inf. Process.*, 2007, pp. 201–208.
- [25] E. Bertin, "Global fluctuations and Gumbel statistics," *Phys. Rev. Lett.*, vol. 95, no. 17, pp. 170–601, Oct. 2005.
- [26] E. Bertin and M. Clusel, "Generalized extreme value statistics and sum of correlated variables," *J. Phys. A, Math. Gen.*, vol. 39, pp. 7607–7619, May 2006.
- [27] E. J. Gumbel, *Statistics of Extremes*. New York, NY, USA: Columbia Univ. Press, 1958.
- [28] Z. Cong, L. Chu, Y. Yang, and J. Pei, "Comprehensible counterfactual explanation on Kolmogorov–Smirnov test," 2020, *arXiv:2011.01223*.
- [29] N. Balakrishnan and M. Kateri, "On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data," *Statist. Probab. Lett.*, vol. 78, no. 17, pp. 2971–2975, Dec. 2008.
- [30] A. B. Shetty, N. N. Ail, M. Sahana, Sushmitha, and V. P. Bhat, "Recognition of handwritten digits and English texts using MNIST and EMNIST data sets," *Int. J. Res. Eng., Sci. Manage.*, vol. 4, pp. 240–243, Jul. 2021.
- [31] R. Vaia, J. Chiasson, and V. Saxena, "A deep unsupervised feature learning spiking neural network with binarized classification layers for the EMNIST classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 1, pp. 124–135, Feb. 2022. [Online]. Available: <https://paperswithcode.com/dataset/emnist>, doi: [10.1109/TETCI.2020.3035164](https://doi.org/10.1109/TETCI.2020.3035164).

[32] T. Clanuwat, M. Bober-Irizar, A. Kitamoto, A. Lamb, K. Yamamoto, and D. Ha, "Deep learning for classical Japanese literature," 2018, *arXiv:1812.01718*. [Online]. Available: <https://github.com/rois-codh/kmnist>

[33] H. Singh, V. Sharma, and D. Singh, "Comparative analysis of proficiencies of various textures and geometric features in breast mass classification using k-nearest neighbor," *Vis. Comput. Ind., Biomed., Art*, vol. 5, no. 1, pp. 1–9, Dec. 2022, doi: [10.1186/s42492-021-00100-1](https://doi.org/10.1186/s42492-021-00100-1).

[34] Y. M. Wazery, E. Saber, E. H. Houssein, A. A. Ali, and E. Amer, "An efficient slime mould algorithm combined with K-nearest neighbor for medical classification tasks," *IEEE Access*, vol. 9, pp. 113666–113682, 2021, doi: [10.1109/ACCESS.2021.3105485](https://doi.org/10.1109/ACCESS.2021.3105485).

[35] O. I. Abiodun, M. U. Kiru, A. Jantan, A. E. Omolara, K. V. Dada, A. M. Umar, O. U. Linus, H. Arshad, A. A. Kazaure, and U. Gana, "Comprehensive review of artificial neural network applications to pattern recognition," *IEEE Access*, vol. 7, pp. 158820–158846, 2019, doi: [10.1109/ACCESS.2019.2945545](https://doi.org/10.1109/ACCESS.2019.2945545).

[36] S. Kang, "Using binary classifiers for one-class classification," *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115920.

[37] D. Sengupta, P. Gupta, and A. Biswas, "A survey on mutual information based medical image registration algorithms," *Neurocomputing*, vol. 486, pp. 174–188, May 2022.

[38] M. Rad, P. M. Roth, and V. Lepetit, "ALCN: Adaptive local contrast normalization," *Comput. Vis. Image Understand.*, vol. 194, May 2020, Art. no. 102947.



AGNÈS DELAHAIES received the Ph.D. degree in image processing from the University of Angers, France, in 2011. She is currently an Associate Professor with the University of Reims Champagne-Ardenne. Her research interests include image comparison, similarity measures, and dissimilarity detection.



JÉRÔME LANDRÉ received the B.S., M.S., and Ph.D. degrees in computer science from the University of Burgundy, Dijon, France, in 1993, 1995, and 2005, respectively. From 1996 to 2005, he was a Software Engineer with the University of Burgundy. He joined the University of Reims Champagne-Ardenne, in 2006, where he began working in image processing. His research interests include content-based image retrieval, medical imaging, and change detection.



FLORENT REIRAUNT received the M.Sc. degree in applied mathematics and the Ph.D. degree in image processing from the National Institute of Applied Sciences, Lyon, France, in 1994 and 1998, respectively. He is currently a Full Professor with the University of Technology of Troyes. His research interests include image modeling, statistical image processing, hypothesis testing theory, and anomaly detection and localization, with a main application to digital image forensics.



MOUSTAPHA DIAW received the M.Sc. degree in applied mathematics from the University of Aix-Marseille, France, in 2019. He is currently pursuing the Ph.D. degree in image processing with the University of Reims Champagne-Ardenne, Reims, France. His research interests include statistical image processing, similarity and dissimilarity measures, satellite image change detection, and digital image forensics.



FRÉDÉRIC MORAIN-NICOLIER received the M.Sc. degree in applied physics and the Ph.D. degree in image processing from the University of Bourgogne, France, in 2000. He is currently a Full Professor with the University of Reims Champagne-Ardenne, where he is integrated with the CREsTIC Laboratory. His research interests include image processing and analysis, similarity measures and dissimilarity detection, and biomedical imaging.

...