Adaptive Non-Homogeneous Granulation-Aided Density-Based Deep Feature Clustering for Far Infrared Sign Language Images

Pritam Paral, *Member, IEEE*, Saibal Ghosh, *Graduate Student Member, IEEE*, Sankar K. Pal, *Life Fellow, IEEE*, and Amitava Chatterjee, *Senior Member, IEEE*

***Abstract*—In image clustering applications, deep feature clustering has recently demonstrated impressive performance, which employs deep neural networks for feature learning that favors clustering exercises. In this context, density-based methods have emerged as the preferred choice for the clustering mechanism within the framework of deep feature clustering. However, as the performance of these clustering algorithms is primarily effective on the low-dimensional feature data, deep feature learning models play a crucial role here. With far infrared (FIR) thermal imaging systems working in real-world scenarios, the images captured are largely affected by blurred edges, background noise, thermal irregularities, few details, etc. In this work, we demonstrate the effectiveness of granular computing-based techniques in such scenarios, where the input data contains indiscernible image regions and vague boundary regions. We propose a novel adaptive non-homogeneous granulation (ANHG) technique here that can adaptively select the smallest possible size of granules within a purview of unequally-sized granulation, based on a segmentation assessment index. Proposed ANHG in combination with deep feature learning helps in extracting complex, indiscernible information from the image data and capturing the local intensity variation of the data. Experimental results show significant performance improvement of the density-based deep feature clustering method after the incorporation of the proposed granulation scheme.**

***Index Terms*—Adaptive non-homogeneous granulation, American sign language, deep feature clustering, density-based clustering, far infrared thermal imaging, granular computing.**

I. Introduction

In RECENT times, collaborative robotics has gained much prominence, and, from industrial sectors to domestic utilities, assistive robots and human-robot interaction (HRI) systems have emerged as more useful options [1]. In such systems, vision-based, audio-based, and wearable sensor-based interfaces offer commonly used interaction modalities [2]. In the applications of speech detection, speaker identification, source localization, aural emotions study, musical stimulation, etc., auditory signal-based modalities are the most commonly used ones for interaction [2]. Non-contact, wear-free sensors are largely favored over the contact-based sensors in e.g. environmental monitoring, human detection and following, human-centered navigation, etc. [3], [4], [5]. Non-contact, vision sensing systems include RGB imaging cameras, depth cameras, thermal/infrared (IR) imagers [6], [7], [8], [9], [10], [11], [12]. Among vision-based systems, hand gestures from human are popular input stimuli because of their hassle-free acquisition and intuitive interpretation [13]. However, human hand gestures can often be open to various expositions and ambiguous meaning across different geographical and socio-cultural boundaries.

To uniformly standardize the interpretation of hand gestures in a more distinct and unambiguous way, several sign languages (SLs) have been formed [14]. As of today, there exist several SL dictionaries worldwide, among which American sign language (ASL) is one of the most popularly used ones [13]. The American manual alphabet (AMA) library contains 36 hand signs i.e. 26 letters from the English language and 10 numerals. In the present research work, we have considered hand signs corresponding to those 10 numerals from the AMA. 10 distinct hand gesture signs according to the standard finger-spelling library of AMA have been captured from human volunteers using a far infrared thermal imaging camera (FIR-TIC). Despite user comfort and other advantages of color vision sensing, they impose some challenges in real-world scenarios, e.g., improper ambient illumination, poor night vision, sensor noise, etc. To combat issues like insufficient lighting or varying illumination, thermal imaging systems offer excellent alternatives to the regular RGB systems. Thermal cameras can work efficiently in both day and night conditions, under adverse weather conditions, significant background variations, etc. However, infrared images still suffer from challenges, such as blurred edges, background noise, thermal irregularities, low signal-to-noise ratios, as well as few details.

In this study, we propose a novel, robust framework of image clustering to partition the dataset of challenging FIR sign language images into clusters, such that images that are similar to one another are placed together in the same cluster, while dissimilar images are grouped into different clusters.

Received 10 February 2024; revised 8 September 2024; accepted 6 October 2024. Date of publication 11 December 2024; date of current version 27 March 2025. *(Pritam Paral and Saibal Ghosh are co-first authors.) (Corresponding author: Pritam Paral.)*

Pritam Paral is with the Department of Electrical Engineering, IIEST, Shibpur, Howrah 711103, India (e-mail: callinpritam@gmail.com).

Saibal Ghosh and Amitava Chatterjee are with the Department of Electrical Engineering, Jadavpur University, Kolkata 700032, India (e-mail: saibal436ghosh@gmail.com; amitava.chatterjee@ieee.org).

Sankar K. Pal is with the Center for Soft Computing Research, ISI, Kolkata 700108, India (e-mail: sankar@isical.ac.in).

This article has supplementary downloadable material available at https://doi.org/10.1109/TETCI.2024.3510292, provided by the authors.

Recommended for acceptance by J. Qiang.

Digital Object Identifier 10.1109/TETCI.2024.3510292

Being an unsupervised approach of learning, clustering some unlabeled irregular real-world data becomes more challenging than classifying the same with labeled inputs. Image clustering is a technique used in computer vision and machine learning to group similar images together into clusters based on their visual features. However, for higher-dimensional image data, clustering becomes a very cumbersome task, primarily because of difficulty in obtaining reliable similarity measures in the high-dimensional space [15]. Therefore, in general, the image clustering approaches first lower the dimensionality of the data using feature extraction or feature selection methods, and then carry out clustering in reduced dimensional space. The features could include color histograms, texture patterns, shapes, edges, or even higher-level features extracted from pre-trained deep neural networks (DNNs).

Recent development of deep clustering techniques in this regard, have shown significant improvement in the clustering performance in the latent learned feature space [16]. Deep clustering methods combine the power of deep learning techniques with traditional clustering algorithms to learn feature representations from data and then perform clustering based on those learned representations. Deep embedded clustering (DEC) [17], deep clustering network (DCN) [18], deep convolutional clustering method (DCCM) [19], etc. are a few state-of-the-art deep clustering techniques to name here. Despite being able to utilize deep features for cluster representations, most of these methods essentially apply a partitional clustering mechanism in the latent feature space and usually require the prior information of number of clusters, which is not very practical in real-world situations.

Among popular traditional clustering techniques, such as partitional clustering, hierarchical clustering, density-based clustering, spectral clustering, etc. [20], density-based methods [20] have the advantages, over other clustering algorithms, of being able to detect clusters that are irregularly shaped and sized, and they do not require the number of clusters to be specified prior, which is particularly useful for practical clustering tasks. In addition, these methods are less sensitive to initialization conditions and are also very robust to outliers. Density-based spatial clustering of applications with noise (DBSCAN) [21], DENsity-based CLUstEring (DENCLUE) [22], ordering points to identify cluster structure (OPTICS) [23] are some of the popular density-based clustering methods. However, these techniques are primarily implemented in the original feature space and they do not perform well when grouping images with high dimensionality, owing to the limited representation capability.

Very recently, several deep clustering techniques, namely, deep density clustering of unconstrained faces (DDC-UF) [24], deep embedding determination (DED) [25], etc., have been proposed to deal with the issue of estimating the number of clusters $\mathcal{K}$, which is a major issue in clustering problems and a proper estimation of $\mathcal{K}$ is non-trivial. Nevertheless, these techniques do not take into account the local structures inside each cluster and prevent points from having distinct roles based on their densities. Addressing the same, Ren et al. has proposed a two-stage deep density-based clustering (DDC) [15] method for images that significantly improves clustering performance by taking into consideration both the importance of points and local cluster structures. However, the performance of deep clustering techniques is largely influenced by the aptness of features extracted by DNN models, whereas the performance of DNNs are dependent on the nature of the input data to a great extent.

In this work, we have incorporated a computing framework known as granular computing (GrC) [26] to broadly abstract useful knowledge and information from the image data before feeding into the deep networks. In recent years, GrC has rapidly gained popularity, which helps in extracting complex, indiscernible information from data in form of information granules. The GrC has been an integral part of a rough entropy thresholding (RET) method applied in various computer vision applications. Pal et al. first developed the *maximum RET* (MRET) method [27] in which crisp granulation is performed on the image, i.e., non-overlapping square blocks of fixed size are used to granulate the image. The granule size is chosen to be half of the smaller peak of the image histogram. Later, the MRET method was extended by Dariusz and Jaroslaw [28] to include multilevel thresholding in combination with evolutionary algorithms (EAs). In this study, overlapping rectangular blocks were employed to perform image granulation. In [29], a novel spatio-temporal segmentation method based on the MRET algorithm was proposed for moving object recognition. This study considered *quad-tree decomposition* (QtD) for image granulation. In recent times, Lei and Fan [30] defined a novel form of *square rough entropy* to quantify the image roughness, and proposed the corresponding image thresholding algorithm. In this work, the granule size was selected to be one half of the minimum peak width in the homogeneity histogram, taking noise into account. However, a common drawback of all these granulation techniques is that none of them can automatically determine the granule size in conjunction with image information, which is especially crucial in the context of real-world images subjected to photometric challenges, e.g., non-uniform ambient illumination, poor night vision, background clutter, sensor irregularities, etc.

To address this limitation, very recently, a novel granulation method for adaptively selecting the granule size, called *adaptive granulation Renyi rough entropy thresholding* (ARRET) algorithm [31], has been proposed, which works on the principle of maximizing the uniformity of the segmented regions of image. In this method, the size of the granule can be chosen adaptively in accordance with the input image. The study also introduces a parametric rough entropy on the basis of Renyi entropy form. More flexibility is provided by the new Renyi rough entropy measure with the inclusion of parameter. Homogeneous crisp granulation serves as the foundation for the ARRET algorithm, where the size of the even granules stays constant throughout the image and the algorithm adaptively determines the optimal granule size based on the image information.

However, our experimental study reveals that the image segmentation performance of the homogeneous granulation based ARRET method deteriorates when, instead of standard datasets, raw high-dimensional vision data with challenging attributes acquired by real sensors is taken into consideration (for instance, visual color images or infrared images having local

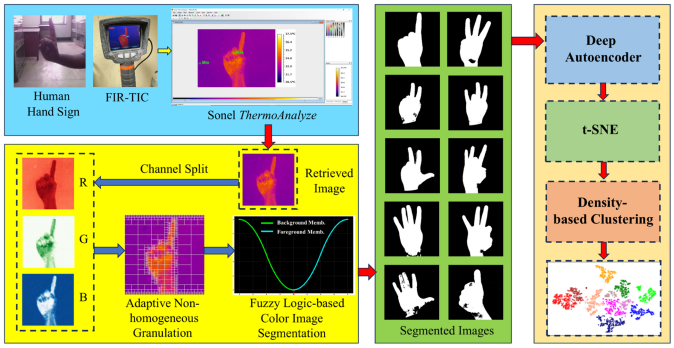

Fig. 1. Overall architecture of the proposed thermal sign language recognition scheme ANHG-DDC.

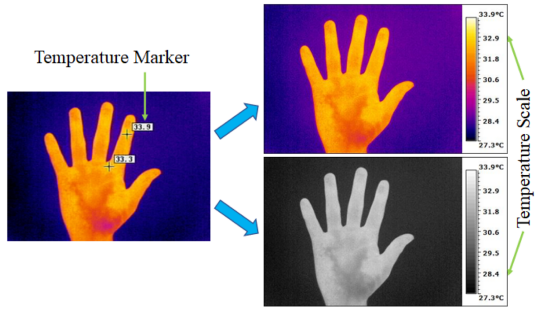

Fig. 2. Originally captured raw FIR image with Sonel $KT-384$ and its processed versions with Sonel *ThermoAnalyze* software v1.7.0.10.

intensity variations and varying clumping profiles captured by a real camera or thermal imager). As already mentioned, there are several techniques of granule formation for dealing with the ambiguity in images. The shape of the granule could be even or uneven with fixed or variable size. Uneven granules with varying size are more natural and effective in dealing with real-world circumstances, particularly for image areas with varied levels of distinguishability. The quad-tree decomposition (QtD) [29] is a popular non-homogeneous granulation approach in this context, which generates different sized uneven granules of regular shape based on the spatial proximity and gray level homogeneity of the image pixels. However, the extent of decomposition, i.e., the smallest possible granule size is not uniform for all images and directly influences the segmentation outcomes. Therefore, selecting an appropriate minimum granule size is a crucial task in the context of image thresholding [31]. In this work, a novel adaptive non-homogeneous granulation (ANHG) method is proposed to adaptively select the minimum granule size for the QtD algorithm based on a segmentation assessment index, so that an optimal segmentation threshold is obtained for a single-channel image. The concept of threshold determination can easily be extended for three-channel RGB images.

In the present study, a new density-based deep feature clustering model integrating our proposed ANHG technique is developed, which is referred to as *ANHG-aided deep density-based clustering* (ANHG-DDC). The functional diagram of the ANHG-DDC scheme is given in Fig. 1.

The contributions of the work are summarized as follows:

i) To the best of our knowledge and belief, this study is first of its kind to demonstrates how the photometrically challenging FIR image data of AMA finger-spelling numerals can be grouped into an appropriate number of disjoint clusters, not knowing anything about the label information or the number of clusters beforehand.

ii) Considering real-world photometric challenges, such as non-uniform ambient illumination, poor night vision, background clutter, sensor irregularities, etc., this study proposes a novel adaptive non-homogeneous granulation technique to determine the threshold for segmentation of a single-channel image. The smallest possible granule size for the non-homogeneous granulation is selected adaptively in accordance with the input image.

iii) The proposed work is also unique in formulating a robust two-tier feature extraction and learning framework, where an adaptive non-homogeneous granulation scheme is cascaded to a deep feature learning mechanism with a view to extracting the complex, indiscernible information from the image data as well as capturing the varying local intensity distribution of the data.

The rest of this article is organized as follows: Section II demonstrates the FIR-TIC-based data acquisition system and the dataset created. Section III first presents the proposed ANHG based threshold determination (ANHG-TD) algorithm for a single-channeled image, and then demonstrates how an advanced fuzzy logic-based color image segmentation scheme can subsequently be employed to perform segmentation of the three-channel FIR images. Section IV describes the deep density clustering framework involved in this work, which incorporates a deep feature learning mechanism and a density-based clustering approach. The obtained results with extensive experimental studies are reported in Section V along with its analyses and inference studied. Finally, Section VI concludes the article.

II. Experimental Bench and Data Acquisition

Thermal cameras have advantages against illumination variation and are suitable especially under dark environments and unstructured environments in presence of background clutter [9]. For capturing the hand sign images, we have utilized a far infrared thermal imaging camera. FIR thermal images demonstrating the fingerspelled AMA numerals have been captured from 20 volunteers aged $18-45$ years. During data acquisition, the plane of the volunteer's hands in the air and the plane of the camera were kept parallel. An originally captured raw image using *IRONBOW* color palette with temperature marker and its processed equivalents are shown in Fig. 2 [33]. The colormap with temperature gradient demonstrates how the variation in temperature contributes in different color tones across the images. For detailed experimental platform and configuration, please refer to Section V-A.

For each volunteer, 10 sets of images have been taken corresponding to individual classes i.e., $1-10$, giving a total of 2000 images. Fig. 3 presents representative hand sign images from individual classes, acquired from various participants.

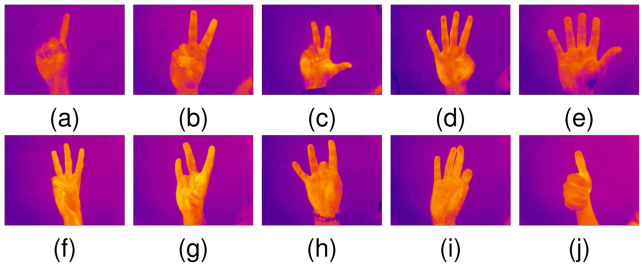

Fig. 3. Sample FIR thermal fingerspelling images depicted in *IRONBOW* palettes representing the AMA numerals (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g) 7, (h) 8, (i) 9, and (j) 10.

III. Color Image Segmentation Based on Adaptive Non-Homogeneous Granulation

A. Concept of Adaptive Non-Homogeneous Granulation Based on the Uniformity Measure (UM)

In QtD method [29], a rectangular/square window is picked from the image and then QtD is performed over the pixels inside the corresponding window, depending on the differences in their gray levels. The image is continued to be decomposed as long as the difference between the minimum and maximum gray levels within a granule does not drop below a certain threshold. The decomposition is carried out by recursively partitioning a chosen image segment into four equal quadrants. The size and extent of decomposed granules are governed by a pre-specified granule detection threshold, which is roughly approximated based on the 25th and 75th percentile information from the image gray-level distribution.

The experimental analysis, however, demonstrates that the empirical threshold may give rise to improper granule formation, especially when non-bimodal FIR sign-language images with non-uniform local densities and differing clumping tendencies are considered. Moreover, if the smallest possible granule size (denoted as min$GrSz$) for the QtD is reasonably large, it can potentially result in the loss of desirable image information. On the other hand, relatively smaller min$GrSz$ may lead to the detection of spurious unwanted regions, as well as a failure to capture distinctive features of the images. Therefore, variation in the min$GrSz$ can have a direct impact on the segmentation results, which is demonstrated with a representative example in Fig. 4. In this work, we present a new approach for selecting the most favorable min$GrSz$ by optimizing the segmentation evaluation index corresponding to each region of the segmented image, with a view to automatically granulating the image using the image information.

A popular index for evaluating the quality of image segmentation is *uniformity measure* (UM) [31], denoted herein by α. The uniformity of an image region has an inverse relationship with the variance of the region. A greater region uniformity indicates that the grayscale distribution is more concentrated. Assuming that τ is the threshold used to segment the image, the UM α can be calculated by using the following equation:

$$\alpha = 1 - \frac{1}{\beta}\sum_{k=0}^{1}\vartheta_k^2 \tag{1}$$

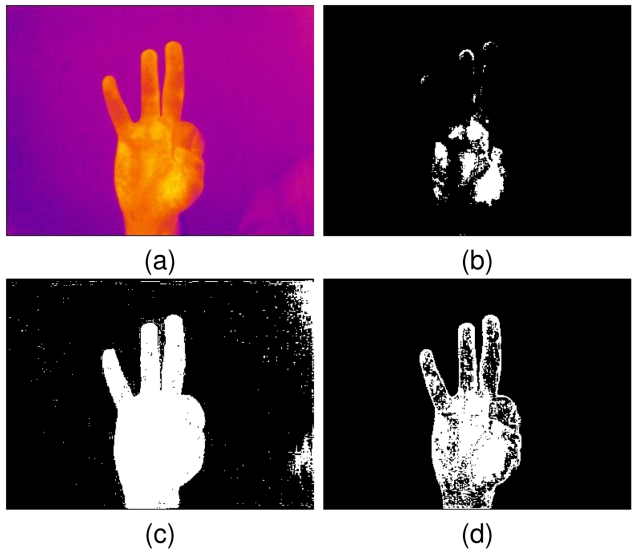

Fig. 4. (a) A representative FIR sign language image; Segmentation results: (b) A relatively higher min$GrSz$ (min$GrSz = 35$) leads to the loss of several desirable regions, (c) The min$GrSz = 2$ results in detection of different spurious unwanted regions, as well as missing crucial image features, and (d) An intermediate min$GrSz$ (min$GrSz = 17$) demonstrates comparatively better granulation effect.

with ϑ_k^2 denoting the variance within the k-th region in the image undergoing segmentation ($k = 0, 1$), the normalized constant, β, makes α fall within the interval $[0, 1]$. The value of β is in general taken as greater than the maximum image variance. The larger the α value, the greater the uniformity of the regions in the segmented image.

Now, as discussed earlier, results of segmentation may vary depending on the smallest possible granule sizes min$GrSz$s. The optimal min$GrSz$, denoted by min$GrSz^*$, should be such that it leads the α to approach its maximum value

$$\text{min}GrSz^* = \arg\max_{\text{min}GrSz}\ \alpha \tag{2}$$

To maximize the UM α, the choice of the parameter min$GrSz$ is varied. The α corresponding to each threshold τ can be calculated using (1). For each min$GrSz$ within a defined range, an optimal threshold $\tilde{\tau}$ will be obtained by the thresholding algorithm. The optimal min$GrSz$, i.e. min$GrSz^*$, can be determined using (2) while the min$GrSz$ travels through its value range. Hence, this reflects a *nested* optimization framework [31]. The value interval of min$GrSz$ is selected as: $[2, \lfloor\hat{U}/2\rfloor]$, $\hat{U} = \min\{U, V\}$, with $\hat{U}$ being the minimum between the width V and height U of the image.

B. Proposed Adaptive Non-Homogeneous Granulation Based Threshold Determination (ANHG-TD) Algorithm

The work proposes a segmentation threshold determination algorithm for a single-channel image based on a novel adaptive non-homogeneous granulation technique, where the smallest possible size of the granules formed by quad-tree decomposition technique is adaptively selected based on the segmentation assessment index (UM) α. The value of the index α depends on the threshold τ, while τ itself varies in accordance with

the minimum granule size. Hence, the said algorithm, referred to as ANHG-TD, is executed through a nested optimization mechanism. The proposed ANHG-TD algorithm is detailed in Algorithm 1.

The ANHG-TD algorithm contains two loops. The first loop sees the min$GrSz$ rise from its minimum to maximum values. For each individual min$GrSz$, quad-tree decomposition technique is applied on the input image and the image is subdivided into non-overlapping square blocks with the smallest possible size of min$GrSz$, which are more homogeneous compared to the image itself. Inside the second loop, the threshold τ rises from the minimum to the maximum gray levels present in the input image. For each τ, based on the knowledge of the non-homogeneous granules formed, the rough set representations of the foreground F_τ and background B_τ are given as follows [27]:

The inner and outer approximations of the foreground (denoted as $\underline{F}_\tau$ and $\overline{F}_\tau$, respectively):

$$\underline{F}_\tau = \left\{ \bigcup_k \Phi_k | q_i > \tau \forall i = 1, \ldots, u_k v_k, \text{ with } q_i \in \Phi_k \right\} \quad (3)$$

$$\overline{F}_\tau = \left\{ \bigcup_k \Phi_k, \exists i, i = 1, \ldots, u_k v_k \text{ s.t. } q_i > \tau, \text{ with } q_i \in \Phi_k \right\} \quad (4)$$

The inner and outer approximations of the background (denoted as $\underline{B}_\tau$ and $\overline{B}_\tau$, respectively):

$$\underline{B}_\tau = \left\{ \bigcup_k \Phi_k | q_i \leq \tau \forall i = 1, \ldots, u_k v_k, \text{ with } q_i \in \Phi_k \right\} \quad (5)$$

$$\overline{B}_\tau = \left\{ \bigcup_k \Phi_k, \exists i, i = 1, \ldots, u_k v_k \text{ s.t. } q_i \leq \tau, \text{ with } q_i \in \Phi_k \right\} \quad (6)$$

where a non-overlapping window with size $u_k \times v_k$ is considered as the granule Φ_k and q_i is a pixel in Φ_k. Then, the foreground and background roughnesses (denoted as ρ_{F_τ} and ρ_{B_τ}, respectively) are calculated as follows [27]:

$$\rho_{F_\tau} = 1 - \frac{|\underline{F}_\tau|}{|\overline{F}_\tau|} = \frac{|\overline{F}_\tau| - |\underline{F}_\tau|}{|\overline{F}_\tau|} \quad (7)$$

$$\rho_{B_\tau} = 1 - \frac{|\underline{B}_\tau|}{|\overline{B}_\tau|} = \frac{|\overline{B}_\tau| - |\underline{B}_\tau|}{|\overline{B}_\tau|} \quad (8)$$

where $|\cdot|$ is the cardinality of a set. Given the roughnesses ρ_{F_τ} and ρ_{B_τ}, the rough entropy of the image (denoted as σ_τ) can be computed by using the following equation [27]:

$$\sigma_\tau = -\frac{e}{2} \left[\rho_{F_\tau} \log_e (\rho_{F_\tau}) + \rho_{B_\tau} \log_e (\rho_{B_\tau}) \right] \quad (9)$$

The optimal threshold $\tilde{\tau}$ is obtained by performing the maximization of the rough entropy, as indicated in (9)

$$\tilde{\tau} = \arg\max_\tau \sigma_\tau \quad (10)$$

The threshold $\tilde{\tau}$ categorizes the pixels in the input image into two classes. Then, with the help of (1), the UM α is computed according to $\tilde{\tau}$. So, an α value and an optimal threshold $\tilde{\tau}$ will be obtained corresponding to each min$GrSz$. A certain min$GrSz$ for which the UM α achieves the maximum value is selected as the optimal min$GrSz$ (denoted as min$GrSz^*$). The optimal $\tilde{\tau}$, which corresponds to the optimal min$GrSz$, i.e. min$GrSz^*$, is considered the segmentation threshold $\tilde{\tau}_s$ for the input image.

Algorithm 1: The Proposed ANHG-TD Algorithm.

Input: ***Single-channel image***;
Output: ***Segmentation threshold for the image*** $\tilde{\tau}_s$;

1: **Initialization:** m: lower limit of the smallest possible granule size min$GrSz$, M: upper limit value of the min$GrSz$, $gray\_min$: minimum gray level of the image, and $gray\_max$: maximum gray level of the image.
2: **for** min$GrSz = m$ to M **do**
3: Perform QtD on the image with min$GrSz$;
4: **for** threshold $\tau = gray\_min$ to $gray\_max$ **do**
5: Compute the inner and outer approximations of the foreground and background by the threshold τ in accordance with (3)–(6);
6: Compute the roughness values of the foreground and background by using (7) and (8);
7: Compute the rough entropy by (9);
8: **end for**
9: Determine the optimal threshold $\tilde{\tau}$(min$GrSz$) with the help of (10);
10: Calculate the uniformity measure α(min$GrSz$) corresponding to $\tilde{\tau}$(min$GrSz$) based on (1);
11: **end for**
12: Determine the optimal minimum granule size min$GrSz^*$ according to (2);
13: The optimal threshold corresponding to the min$GrSz^*$ is the segmentation threshold $\tilde{\tau}_s$ for the image.

C. Fuzzy Logic-Based Color Image Segmentation

Considering the aspects, such as loss of information in RGB to grayscale conversions [8] or lack of useful information in single-channel images compared to multi-channel images, an advanced fuzzy logic-based color image segmentation scheme has been adopted in this work. In this approach, instead of a single-channel or grayscale FIR image, a more informative RGB (three-channel) FIR image is considered, whose individual channels are granulated with our proposed ANHG technique and the respective thresholds for segmentation are determined. Let the segmentation threshold values in R, G, and B channels be respectively denoted as $\tilde{\tau}_{sr}$, $\tilde{\tau}_{sg}$, and $\tilde{\tau}_{sb}$. Based on these threshold values, pixels in an image are assigned as foreground or background pixels. As the intensity value $I(p)$ of an individual pixel p approaches the segmentation threshold $\tilde{\tau}$ from the lower side, its likelihood to be categorized as a foreground pixel increases. Similarly, pixel intensity values lesser than the segmentation threshold $\tilde{\tau}$ make the pixel's likelihood tend towards becoming a background pixel. This foreground-background likelihood of the image pixels can be better represented by

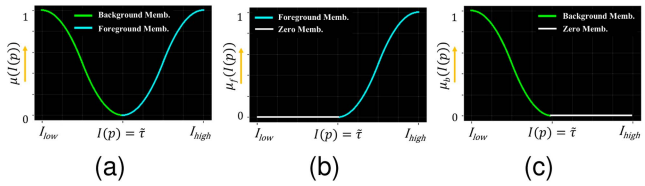

Fig. 5. Reverse order pi fuzzy membership function for granular foreground-background segmentation: (a) Overall membership, (b) Foreground membership, and (c) Background membership.

fuzzy membership functions. Considering the nature of variation, we have considered a *reverse order pi fuzzy membership* (ROPFM) function here. Another equally fitting fuzzy rule-base in this regard is offered by a *reverse order triangular fuzzy membership* (ROTFM) function. Let the minimum and maximum values of pixel intensities be defined as I_{low} and I_{high} which are usually set to 0 and 255, respectively. For an individual pixel p, the foreground membership $\mu_f(I(p))$, background membership $\mu_b(I(p))$, and the overall membership $\mu(I(p))$, with respect to the pixel intensity variation following ROPFM function is demonstrated in Fig. 5. Consequently, the foreground fuzzy membership values in R, G, and B color channels are obtained as $\{\mu_f(I(p))\}_r$, $\{\mu_f(I(p))\}_g$, and $\{\mu_f(I(p))\}_b$, respectively. Similarly, the respective background fuzzy membership values are calculated as $\{\mu_b(I(p))\}_r$, $\{\mu_b(I(p))\}_g$, and $\{\mu_b(I(p))\}_b$. Hence, the combined foreground and background memberships are respectively obtained as:

$$\begin{aligned}\{\mu_f(I(p))\}_c &= \{\mu_f(I(p))\}_r + \{\mu_f(I(p))\}_g + \{\mu_f(I(p))\}_b \\ \{\mu_b(I(p))\}_c &= \{\mu_b(I(p))\}_r + \{\mu_b(I(p))\}_g + \{\mu_b(I(p))\}_b\end{aligned} \tag{11}$$

For $\{\mu_f(I(p))\}_c \geq \{\mu_b(I(p))\}_c$, the pixel is assigned as a foreground pixel and for $\{\mu_f(I(p))\}_c < \{\mu_b(I(p))\}_c$, the pixel is assigned as a background pixel. Now, to obtain the final segmented image, foreground pixels are thresholded up to a value of 1, whereas the background pixels are thresholded down to 0. Segmented images thus obtained based on ANHG technique are next fed to the deep feature learning module for the purpose of extracting effective low-dimensional feature representations from high-dimensional image data, as demonstrated in the following section.

The flowchart of the overall color image segmentation scheme based on our proposed ANHG technique with reverse order pi fuzzy rule-base is presented in Fig. 6. Moreover, representative RGB-channel hand sign images and their corresponding ANHG-segmented images are shown in Fig. 7.

IV. Density-Based Deep Feature Clustering

Let a given dataset comprising n number of raw RGB FIR images be denoted as $\underline{\mathcal{X}} = \{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n\}$, with each $\underline{x}_i \in \Re^D$ lying in a D-dimensional space. Also, we assume that the dataset of the corresponding segmented images obtained with the proposed ANHG technique is represented as $\mathcal{X} = \{x_i \in \Re^{D'}\}_{i=1}^n$. The overall framework of DDC is divided into two stages: *(i)* deep granulated feature learning and low-dimensional mapping, and *(ii)* density-based clustering. It does not require any predefined cluster count or class labels, as it inherently estimates an suitable number of clusters and shapes in the reduced space.

A. Granulated Feature Learning Via Deep Autoencoder

A deep autoencoder or simply autoencoder is a typical artificial neural network for encoding and representing the input data in a dimensionality reduced form. Such representations become very useful in unsupervised data learning problems such as image clustering. An autoencoder is made of three components: *encoder*, *code*, and *decoder*. The encoder $z = f_\theta(x)$ projects an input data point x to its low-dimensional representation z. This encoded representation can be retrieved from the part defined as *code*. Finally, a decoder reconstructs the learned data in its original space as $x' = g_\phi(z)$. Here, θ and ϕ define the set of network parameters for the encoder and decoder, respectively. In this work, deep convolutional autoencoder (CAE) is adopted for clustering the thermal sign images. CAE solves the optimization problem defined as [34]:

$$\arg\min_{\theta,\phi} \frac{1}{n}\sum_{i=1}^{n} \|x_i - g_\phi(f_\theta(\tilde{x_i}))\|_2^2 \tag{12}$$

Here, corrupted samples are denoted by $\tilde{x}$, which can be infiltrated by random noises e.g., Gaussian noise. (12) uses the generalized representation of a denoising autoencoder. After the deep autoencoder is trained (solving (12)), a set of encoded feature representations is observed, which is defined as $\mathcal{Z} = \{z_i = f_\theta(x_i) \in \Re^d\}_{i=1}^n$. Later, for better visual representation and optimal fitting of the density-based clustering, this d-dimensional representation of the data is further transformed into a 2-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) [36], which has excellent capability to preserve pairwise similarity.

B. Density-Based Clustering

The density-based clustering approach DDC approximates suitable clusters from the data in the 2-dimensional feature space without any previously defined cluster number. Following t-SNE, let the data in the 2D feature space be defined as $\mathcal{Y} = \{y_i \in \Re^2\}_{i=1}^n$.

1) Generation of Local Clusters: Similar to 'clustering by fast search and find of density peaks' (DenPeak) [37], in DDC, two quantities are computed first for each point y_i: the local density ρ_i and its distance δ_i from the higher density points. In DDC, the density parameter ρ_i is defined as [15]

$$\rho_i = \sum_{y_j \in \mathcal{Y}\setminus\{y_i\}} \exp\left(-\left(\frac{\Delta_{ij}}{\Delta_c}\right)^2\right) \tag{13}$$

The Euclidean distance between data points y_i and y_j is defined as Δ_{ij} with a predefined distance cutoff Δ_c. With a higher density of data point y_i, the value of ρ_i increases. The distance parameter δ_i of y_i refers to the minimum Euclidean distance

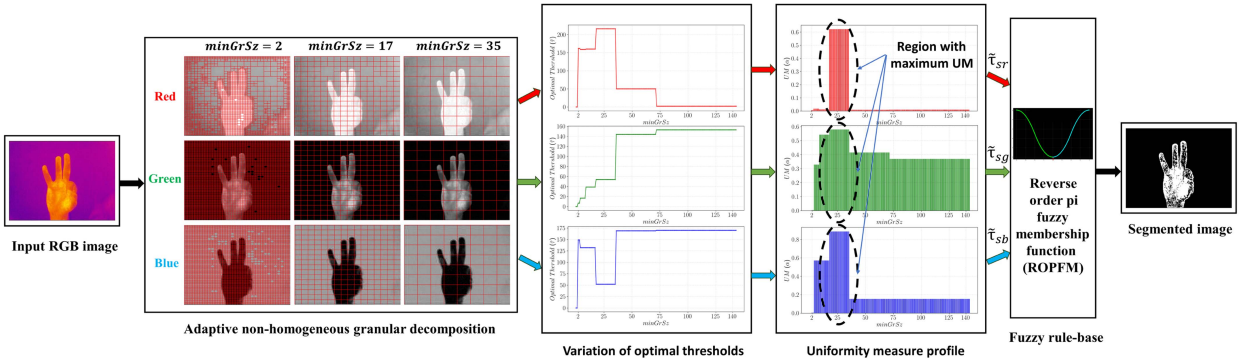

Fig. 6. Flow diagram of the overall color image segmentation scheme based on adaptive non-homogeneous granulation.

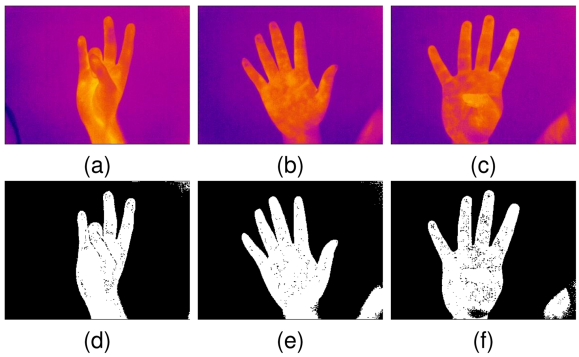

Fig. 7. (a)–(c) Representative FIR hand sign images in *IRONBOW* color palettes, and (d)–(f) their corresponding ANHG-segmented versions.

between y_i and points with larger densities than y_i.

$$\delta_i = \min_{j:\rho_j > \rho_i} (\Delta_{ij}) \tag{14}$$

For the data points with highest density, its density parameter ρ is considered as the maximum of pairwise distances. In DDC, the points with comparatively larger ρ and δ values are considered as *local cluster centers*. Mathematically, the criteria can be expressed as

$$\delta_i > \Delta_c \;\; and \;\; \rho_j > \tilde{\rho} \tag{15}$$

where $\tilde{\rho}$ represents the average of all the density values $\{\rho_i\}_{i=1}^{n}$ corresponding to the data points $\{y_i\}_{i=1}^{n}$.

It can easily be verified that a local cluster center y_i possesses the highest density in its Δ_c-neighborhood represented by a circular region with center y_i and radius Δ_c. After obtaining all of the local cluster centers, each remaining data point is assigned to the cluster as its nearest neighbor with a greater density. A group of clusters are then identified, which will be utilized to produce the final clustering outcome.

2) Combining Local Clusters: Assuming N local clusters $\{\mathcal{LC}^{(1)}, \mathcal{LC}^{(2)}, \ldots, \mathcal{LC}^{(N)}\}$ are found, the final clustering output will be achieved by combining them. In this context, the definitions of *core* and *border* points need to be realized. A data point y_i, assumed to be from local cluster $\mathcal{LC}^{(m)}$, is considered a core point if the following condition is met:

$$\rho_i > \tilde{\rho}^{(m)} \tag{16}$$

with $\tilde{\rho}^{(m)} = \frac{1}{c_m} \sum_{y_j \in \mathcal{LC}^{(m)}} \rho_j$ being the average density of the data points belonging to $\mathcal{LC}^{(m)}$ and c_m being the total number of data points in $\mathcal{LC}^{(m)}$. If not, y_i is taken into account as a border point.

The *connectivity* of clusters is also need to be accounted for attaining the final clustering output, which is defined as [15]:

I. A cluster $\mathcal{LC}^{(m)}$ is considered to be *directly density-connectable* from another cluster $\mathcal{LC}^{(p)}$ when

$$\exists \text{ core points } y_i \in \mathcal{LC}^{(m)} \text{ and } y_j \in \mathcal{LC}^{(p)}, \text{ such that } \Delta_{ij} < \Delta_c \tag{17}$$

II. A cluster $\mathcal{LC}^{(m)}$ is considered to be *density-connectable* to another cluster $\mathcal{LC}^{(p)}$ when

$$\exists \text{ a path } \mathcal{LC}^{(m)} = \mathcal{LC}_1, \mathcal{LC}_2, \ldots, \mathcal{LC}_u = \mathcal{LC}^{(p)} \tag{18}$$

where local cluster $\mathcal{LC}_l$ is directly density-connectable from local cluster $\mathcal{LC}_{l-1}$ $(l = 2, \ldots, u)$ and u is length of the path.

The final clustering result is eventually obtained after merging all the density-connectable local clusters. Following the merger of two local clusters, the higher-density cluster center becomes the center of the newly combined cluster. Further details of the clustering algorithm DDC can be found in [15].

For ease of comprehension, the overall ANHG-DDC model is summarized in Algorithm 2. Also, due to space limitations, computational complexity analysis of the ANHG-DDC model is presented in Section S-I of the "Supplementary File".

V. Results and Discussions

A. Experimental Platform and Configuration

In our experiments, we have utilized a completely radiometric thermal imaging camera $KT-384$ manufactured by Sonel, Poland, as shown in Fig. 8 [32]. The microbolometric, non-cooled, matrix-type detector of $KT-384$ offers a thermal pixel resolution of 384×288 and thermal sensitivity of $< 0.08°C$ [32]. The human volunteered image data acquisition process has been carried out inside an indoor laboratory with an approximate heat index of $22-24°C$ temperature and 77% humidity. After acquisition, raw images are preprocessed through the Sonel *ThermoAnalyze* software v1.7.0.10. The acquired hand sign images have been taken into the RGB-channel *IRONBOW* color palette before feeding into the ANHG-DDC module. For

Algorithm 2: The Proposed ANHG-DDC Approach.

Input: 1) ***Dataset of RGB FIR sign language images*** $\underline{\mathcal{X}}$,
2) ***Distance cutoff*** Δ_c;
Output: ***The final clustering distribution***;$\tilde{\tau}_s$;

Stage 1 $\rightarrow$ **RGB color image segmentation based on ANHG-TD algorithm**
for each image $\underline{x}_i \in \underline{\mathcal{X}}$ **do**
  Apply **Algorithm 1** on individual channels of $\underline{x}_i$ and determine the respective segmentation thresholds $\tilde{\tau}_{sr}^{(i)}$, $\tilde{\tau}_{sg}^{(i)}$, and $\tilde{\tau}_{sb}^{(i)}$.
  for each pixel $p \in \underline{x}_i$ **do**
    Compute the *combined* foreground and background fuzzy memberships, $\{\mu_f(\underline{x}_i(p))\}_c$ and $\{\mu_b(\underline{x}_i(p))\}_c$, by using ROPFM function, in accordance with (11).
    if $\{\mu_f(\underline{x}_i(p))\}_c \geq \{\mu_b(\underline{x}_i(p))\}_c$
      Classify p as a *foreground* pixel.
    else
      Classify p as a *background* pixel.
  end for
  Generate the segmented image for $\underline{x}_i$, denoted as x_i.
end for
Form the dataset of ANHG-based segmented images $\mathcal{X}$ corresponding to $\underline{\mathcal{X}}$.
Stage 2 $\rightarrow$ **Deep granulated feature learning**
Train a deep convolutional autoencoder (CAE) using (12).
Use the encoder $f_\theta(\cdot)$ to transform $\mathcal{X}$ into lower-dimensional feature representations $\mathcal{Z}$.
Perform mapping of $\mathcal{Z}$ into a two-dimensional space $\mathcal{Y}$ by using t-SNE.
Stage 3 $\rightarrow$ **Density-based clustering**
$\backslash*$ **Generation of local clusters** $*\backslash$
for each data point $y_i \in \mathcal{Y}$ **do**
Calculate ρ_i and δ_i using (13) and (14), respectively.
end for
Select local cluster centers using (15).
Assign all the remaining data points to the clusters and realize a set of local clusters $\{\mathcal{LC}^{(1)}, \mathcal{LC}^{(2)}, \ldots, \mathcal{LC}^{(N)}\}$.
$\backslash*$ **Producing the ultimate clustering outcome** $*\backslash$
Specify the data points from the local clusters as *core* or *border* points based on (16).
Combine all the density connectable (whether directly or not) local clusters using (17) and (18).
Return the final clustering output achieved.

all the experimental run, we have used our FIR-TIC-captured fingerspelling image database described in Section II, originally containing a total 2000 images corresponding to 10 different types of AMA numerals.

B. Parameter Settings of the Deep Neural Network (DNN)

The network of the convolutional autoencoder (CAE) implemented in this work is: $CONV_{32}^5 \rightarrow CONV_{64}^5 \rightarrow CONV_{128}^3 \rightarrow FC_{10} \rightarrow CONV_{128}^3 \rightarrow CONV_{64}^5 \rightarrow CONV_{32}^5$ [15]. Here, for example, $CONV_{64}^5$ indicates a convolutional layer characterized by a 5×5 kernel and 64 filters. The value of stride is selected as 2. Moreover, FC_{10} represents a fully connected layer comprising 10 neurons. Except for the input, embedding, and output layers, all internal layers in the CAE are activated by the ReLU function. The number of neurons in the FC layer matches the number of output classes, which is also manifested by the dimensionality of learned feature representations $\mathcal{Z}$ (i.e., 10).

Fig. 8. Sonel $KT-384$ FIR-TIC used in the data acquisition process.

The *loss function* is the error function which is assessed during the neural network training phase and the training attempts to minimize the losses. In this work, the 'mean squared error' loss function is tested. The *optimizer* that is used to minimize the loss is 'adam'. The *batch size*, which is the number of samples per gradient update, is set to 256.

In order to avoid intractable computation time during the hyperparameter optimization process, the *number of epochs* is kept at a reasonably low value for all parameter combinations. It has been observed that 500 epochs are sufficient for the CAE to converge. Furthermore, the initialization scheme adopted for the CAE is 'Xavier initialization' or 'Glorot uniform Initialization'. Xavier Initialization aims to initialize the weights in a way that maintains the same variance of activations across every layer. The constant variance helps prevent the gradient from disappearing or exploding.

C. Experimental Results and Performance Evaluations

Originally the dataset contains 2000 granulated binary images of hand signs uniformly distributed across classes $1-10$. To improve the performance of the DNN model, the dataset has been further augmented and enhanced by increasing the image samples count to 10000. We have selected a width shift of 5%, height shift of 5%, and a rotational range of $\pm 10^\circ$ for data augmentation. Before feeding into the DNN, individual granulated images have been resized into a pixel dimension of 32×32. With the deep features obtained from the DNN model of deep autoencoder, t-SNE has been performed further to realize the final two-dimensional features of the data. These two-dimensional features have been utilized to obtain the final clusters in the latent feature space by using the DDC technique.

TABLE I
CLUSTERING PERFORMANCE OF DIFFERENT GRANULATION TECHNIQUES-AIDED DEEP DENSITY-BASED CLUSTERING APPROACHES ON THE ORIGINALLY ACQUIRED DATASET

Method	ACC	NMI	ARI
XG-DDC	0.419	0.404	0.252
HG-DDC	0.570	0.681	0.454
NHG-DDC	0.629	0.690	0.461
ARRET-DDC	0.646	0.710	0.489
ANHG-DDC	**0.698**	**0.725**	**0.633**

Bold entities represent the best results obtained with the competing techniques corresponding to three performance metrics.

TABLE II
CLUSTERING PERFORMANCE OF DIFFERENT GRANULATION TECHNIQUES-AIDED DEEP DENSITY-BASED CLUSTERING APPROACHES ON THE AUGMENTED DATASET

Method	ACC	NMI	ARI
XG-DDC	0.564	0.667	0.391
HG-DDC	0.661	0.705	0.562
NHG-DDC	0.709	0.710	0.574
ARRET-DDC	0.733	0.741	0.639
ANHG-DDC	**0.788**	**0.764**	**0.667**

Bold entities represent the best results obtained with the competing techniques corresponding to three performance metrics.

TABLE III
PERFORMANCE COMPARISON AMONG THE PROPOSED ANHG-DDC AND THE INTEGRATED APPROACHES COMBINING VARIOUS DR TECHNIQUES AND k-MEANS ALGORITHM

Approach	ACC	NMI	ARI
LE + k-means	0.208	0.092	0.044
ISOMAP + k-means	0.274	0.211	0.101
t-SNE + k-means	0.286	0.209	0.104
UMAP + k-means	0.341	0.259	0.147
ANHG-DDC	**0.788**	**0.764**	**0.667**

Bold entities represent the best results obtained with the competing techniques corresponding to three performance metrics.

For the comparative analyses, we have considered *no granulation* (XG), *homogeneous granulation* (HG), *non-homogeneous granulation* (NHG), ARRET, and the proposed approach of ANHG in conjunction with DDC approaches (denoted as XG-DDC, HG-DDC, NHG-DDC, ARRET-DDC, and ANHG-DDC, respectively) in the experimental studies. The *average* clustering performance of the originally acquired dataset with 2000 samples without data augmentation has been reported in Table I, for all of the four above-mentioned competing techniques. Three clustering evaluation metrics have been considered here, namely clustering accuracy (ACC),[1] normalized mutual information (NMI), and adjusted rand index (ARI) [15]. Bold entities in Table I represent the best results obtained with the competing techniques corresponding to three performance metrics. The same convention is followed for the remaining tables presented in this paper and the tables in the associated Supplementary File. Without any form of granulation the system achieves a clustering accuracy of 41.85%. While, even with the most primitive form of granulation, i.e., HG-aided DDC, the accuracy jumps up to a value of 57.03%. This clearly demonstrates a significant improvement in the clustering performance on the FIR images after the incorporation of image granulation. The accuracy has been further improved to 62.86% and 64.61% with the unevenly shaped granulation technique (i.e., NHG)-aided DDC and adaptive crisp granulation technique (i.e., ARRET)-aided DDC models, respectively. Finally, with our proposed granulation technique ANHG-aided DDC, a substantially enhanced clustering accuracy of 69.77% has been achieved, outperforming the previously competing methods. The corresponding NMI and ARI values as reported in Table I also develop a similar understanding here. The clustering results in this case is visually depicted in Fig. 9(a1)–(a5).

In a similar manner, experiments have been carried out with the augmented FIR hand sign dataset containing 10000 image samples. The average results for all the granulation techniques-aided DDC methods have been reported in Table II. With data augmentation, significant performance improvement can be seen corresponding to all of the five granulation techniques, owing to the architecture of deep feature extraction models. Without any granulation, clustering accuracy has been improved to 51.46%. The ARRET-DDC model produces an accuracy of 73.28%, which is a significant improvement of clustering performance, especially for a database with highly variational features. However, our proposed ANHG technique based DDC outperforms the remaining four models with an accuracy of 78.77%. The corresponding NMI and ARI values have also been mentioned in the table, and a similar performance improvement in terms of those can be seen with the ANHG-DDC. In this case, the representative clustering results of all the competing models are pictorially demonstrated in Fig. 9(b1)–(b5), which reveals a more coherent and ordered clustering behavior for the ANHG-DDC.

After that, to demonstrate the effectiveness of our proposed ANHG-DDC, the method is compared with a number of integrated approaches combining various non-granulation techniques and non-density-based clustering methods. In this context, various popular classical *dimensionality reduction* (DR) techniques, namely *Laplacian eigenmap (LE)*, *isometric mapping (ISOMAP)*, *t-SNE*, and *uniform manifold approximation and projection (UMAP)* [39], [40], have been combined with a widely used partitional clustering method, i.e., k-means clustering [20] (herein referred to as LE + k-means, ISOMAP + k-means, t-SNE + k-means, and UMAP + k-means, respectively). The average clustering performances of these integrated approaches, as well as the ANHG-DDC are presented in Table III, which evidences that the ANHG-DDC significantly outperforms all four models in terms of clustering accuracy, NMI and ARI results.

Also, for the sake of fair and unbiased comparison, the clustering performance of the proposed ANHG-DDC is compared to that of three other well-established density-based clustering

[1]Computation of *clustering accuracy* is based on figuring out the optimal setting that would maximize the metric, where "setting" indicates what labels in predicted clusters correspond to what labels in ground-truth clusters. The label mapping, essentially a *linear sum assignment problem*, is also referred to as minimum weight matching in bipartite graphs. Here *modified Jonker-Volgenant* algorithm [38] with no initialization is used to solve the graphs.

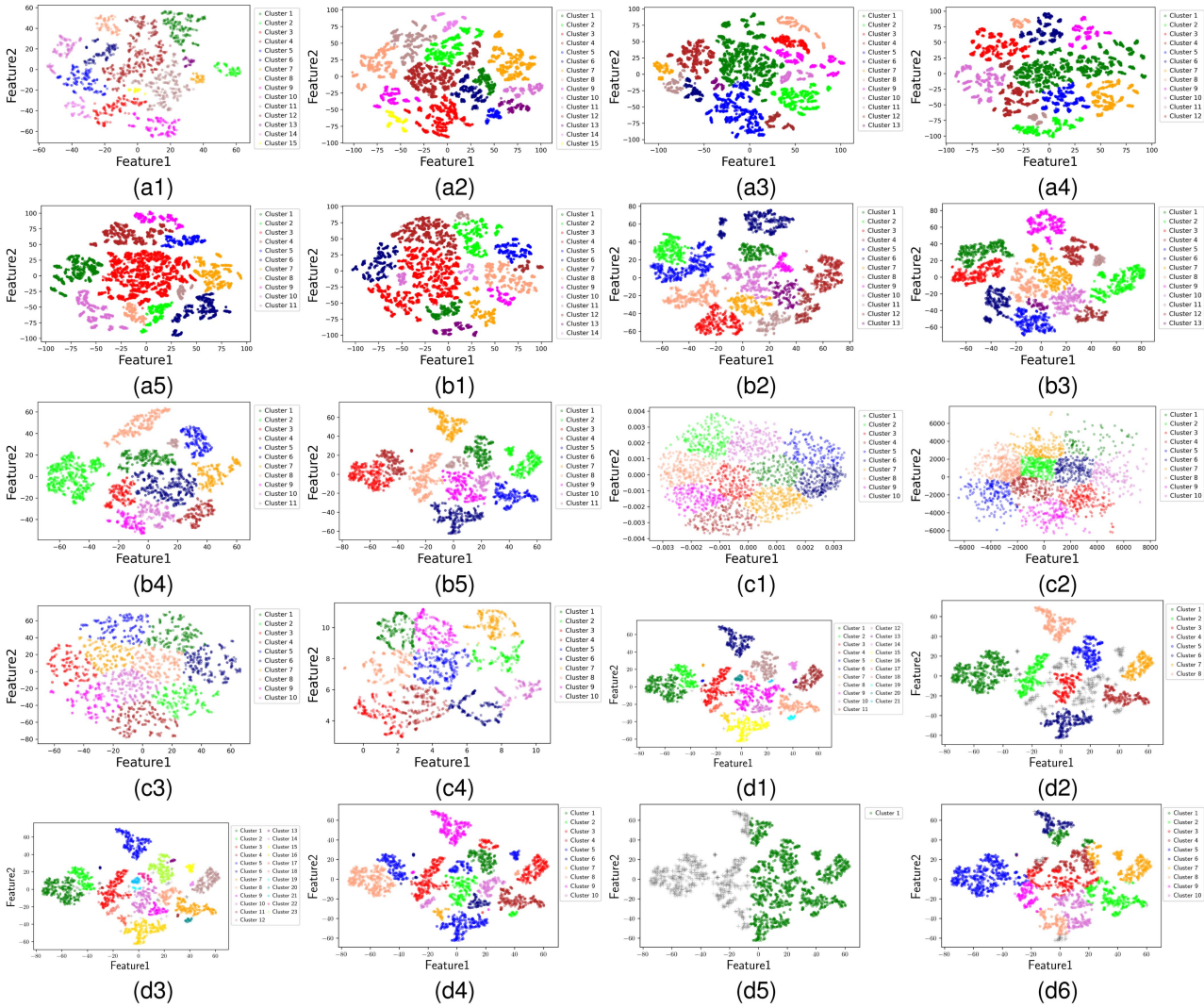

Fig. 9. Clustering performances of different granulation technique-aided deep density-based clustering approaches on the FIR image dataset without augmentation: (a1) XG-DDC, (a2) HG-DDC, (a3) NHG-DDC, (a4) ARRET-DDC, and (a5) ANHG-DDC. Clustering performances of different granulation technique-aided deep density-based clustering approaches on the FIR image dataset with augmentation: (b1) XG-DDC, (b2) HG-DDC, (b3) NHG-DDC, (b4) ARRET-DDC, and (b5) ANHG-DDC. Clustering performances of the integrated approaches combining various DR techniques and k-means algorithm on the augmented FIR image dataset: (c1) LE + k-means, (c2) ISOMAP + k-means, (c3) t-SNE + k-means, and (c4) UMAP + k-means. Clustering performances of various ANHG-aided deep density-based, deep spectral, and deep subspace clustering models: (d1) ANHG-DDC-DN, (d2) ANHG-DDC-OP, (d3) ANHG-DDC-DB, (d4) ANHG-DSPC, (d5) ANHG-DSSC(bu), and (d6) ANHG-DSSC(td) on the augmented FIR image dataset.

algorithms, viz. DENCLUE, OPTICS, and DBSCAN, embedded variants of ANHG-DDC, formulated in this study itself (herein referred to as ANHG-DDC-DN, ANHG-DDC-OP, and ANHG-DDC-DB, respectively). The corresponding results are reported in Table IV, which shows that the ANHG-DDC consistently produces better mean clustering results than its competing variants embedding density-based mechanisms.

Furthermore, performance comparison is made between the ANHG-DDC and several other deep granulated feature clustering approaches. In this context, a deep spectral clustering (DSPC) method based on popular *Ng–Jordan–Weiss (NJW)* algorithm [41], and two deep subspace clustering (DSSC) methods respectively based on well-established *CLIQUE (CLustering In QUEst)* [42], a bottom-up subspace approach, and *PROCLUS (PROjected CLUStering)* [43], a top-down subspace approach, have been introduced in the proposed ANHG-DDC framework replacing the DDC method. The resulting models, formulated in this work itself, are referred to as ANHG-DSPC,

TABLE IV
COMPARISON OF CLUSTERING PERFORMANCE OF THE PROPOSED ANHG-DDC AND VARIOUS ANHG-AIDED DEEP DENSITY-BASED, DEEP SPECTRAL, AND DEEP SUBSPACE CLUSTERING MODELS

Method	ACC	NMI	ARI
ANHG-DDC-DN	0.731	0.762	0.639
ANHG-DDC-OP	0.738	0.738	0.598
ANHG-DDC-DB	0.742	0.763	0.659
ANHG-DSPC	0.602	0.623	0.455
ANHG-DSSC(bu)	0.217	0.322	0.136
ANHG-DSSC(td)	0.613	0.631	0.465
ANHG-DDC	**0.788**	**0.764**	**0.667**

Bold entities represent the best results obtained with the competing techniques corresponding to three performance metrics.

ANHG-DSSC(bu), and ANHG-DSSC(td), respectively. Similar to ANHG-DDC, ANHG-DSPC or ANHG-DSSCs first extracts low-dimensional feature representations from granulated image data using a deep CAE and then adopt t-SNE to map the learned features into a 2-D space, while maintaining the data instances' pairwise similarity. Finally, a spectral or subspace clustering method, i.e., NJW, CLIQUE, or PROCLUS is applied on the 2-D embedded data to produce the final clustering result. Considering a k clustering problem, the NJW algorithm divides up a dataset by employing the largest k eigenvectors of the normalized affinity matrix obtained from the dataset. CLIQUE is a grid-based subspace approach that locates density-based clusters in subspaces. PROCLUS is a k-medoid-like approach, which, using a sample of a high-dimensional data set, first forms k candidate cluster centers for the dataset and then iteratively refines the subspace clusters. Table IV presents the comparison of average clustering results obtained by ANHG-DSPC, ANHG-DSSC(bu), ANHG-DSSC(td), and ANHG-DDC, which reveals the supremacy of the proposed ANHG-DDC over other three competing models. As evidenced by Table IV, overall, the class of deep density-based clustering algorithms outperforms the deep spectral or subspace clustering methods, indicating that they are better suited for the thermal image dataset under investigation.

The representative visual demonstrations of the clustering results obtained by the above 10 approaches, i.e., *(i)* LE + k-means, *(ii)* ISOMAP + k-means, *(iii)* t-SNE + k-means, *(iv)* UMAP + k-means, *(v)* ANHG-DDC-DN, *(vi)* ANHG-DDC-OP, *(vii)* ANHG-DDC-DB, *(viii)* ANHG-DSPC, *(ix)* ANHG-DSSC(bu), and *(x)* ANHG-DSSC(td) are presented in Figs. 9(c1)–(d6), respectively. The selection of parameter values for different models follows the guidelines provided in the respective seminal works. From the comparison of Figs. 9(d1)–(d6) and Fig. 9(b5), it is also evident that the ANHG-DDC manifests more regular and definite distribution of clustering in comparison to the competing methods.

Additionally, we conduct a comprehensive experimental investigation to evaluate the clustering performance of the proposed ANHG-DDC in various challenging settings, which, due to space limitations, are presented in different sections of the "Supplementary File". Firstly, we assess the clustering efficacy of the ANHG-DDC in diverse thermally challenging environments, which is presented in Section S-II. Then, to demonstrate the generalizability and universality of the ANHG-DDC on other relevant potential datasets and datasets of other application fields, performance evaluations are conducted in Section S-III. Finally, robustness of the ANHG-DDC at different noise levels is studied in Section S-IV.

VI. Conclusion and Future Works

The present study demonstrates how the photometrically affected far infrared image data of AMA finger-spelling numerals can be categorized into a suitable number of disjoint clusters, without having any prior knowledge about the label information or the number of clusters. For this purpose, a robust granular computing-aided deep feature learning framework has been designed to derive effective low-dimensional feature representations from high-dimensional image data. In this context, the work has proposed a novel granulation approach ANHG, where the smallest possible granule size for quad-tree decomposition is obtained adaptively in alignment with the characteristics of the input image. Once the feature extraction and learning process is over, the low-dimensional representations are further reduced to a 2-D space with the application of t-SNE. Then density-based clustering is exercised on the 2-D embedded data to automatically identify a suitable number of arbitrarily shaped clusters. Experimental results and performance evaluations aptly demonstrate the superiority of our proposed ANHG-aided DDC over the original DDC, as well as its variants integrating different rough entropy based granulation techniques, formulated in this work itself. An intriguing area of future research is to embed the concept of transfer learning or semi-supervised learning into the framework of DDC.

Acknowledgment

The authors want to express their sincere thankfulness to Mr. Aninda Sundar Mondal, Jadavpur University, India, regarding the process of data acquisition with FIR-TIC. The authors also acknowledge the kind cooperation of all the volunteers in the Instrumentation & Cyber Physical System laboratory, Jadavpur University, India, during the process of data acquisition. Prof. S.K. Pal acknowledges the National Science Chair, SERB-DST, Government of India.

References

[1] K. -H. Park, H. -E. Lee, Y. Kim, and Z. Z. Bien, "A steward robot for human-friendly human-machine interaction in a smart house environment," *IEEE Trans. Autom. Sci. Eng.*, vol. 5, no. 1, pp. 21–25, Jan. 2008.

[2] F. Karray, M. Alemzadeh, J. A. Saleh, and M. N. Arab, "Human-computer interaction: Overview on state of the art," *Int. J. Smart Sens. Intell. Syst.*, vol. 1, no. 1, pp. 137–159, 2008.

[3] R. C. Luo and O. Chen, "Wireless and pyroelectric sensory fusion system for indoor human/robot localization and monitoring," *IEEE/ASME Trans. Mechatron.*, vol. 18, no. 3, pp. 845–853, Jun. 2013.

[4] P. Paral, A. Chatterjee, A. Rakshit, and S. K. Pal, "Extended target tracking in human–robot coexisting environments via multisensor information fusion: A heteroscedastic Gaussian process regression-based approach," *IEEE Trans. Ind. Inform.*, vol. 19, no. 9, pp. 9877–9886, Sep. 2023.

[5] C.-P. Lam, C.-T. Chou, K.-H. Chiang, and L.-C. Fu, "Human-centered robot navigation–towards a harmoniously human-robot coexisting environment," *IEEE Trans. Robot.*, vol. 27, no. 1, pp. 99–112, Feb. 2011.

[6] S. Ghosh, P. Paral, A. Chatterjee, and S. Munshi, "Rough entropy-based fused granular features in 2D locality preserving projections for high-dimensional vision sensor data," *IEEE Sensors J.*, vol. 23, no. 16, pp. 18374–18383, Aug. 2023.

[7] P. Paral, A. Chatterjee, and A. Rakshit, "OPTICS-based template matching for vision sensor-based shoe detection in human–robot coexisting environments," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 11, pp. 4276–4284, Nov. 2019.

[8] P. Paral, A. Chatterjee, and A. Rakshit, "Vision sensor-based shoe detection for human tracking in a human–robot coexisting environment: A. photometric invariant approach using DBSCAN algorithm," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4549–4559, Jun. 2019.

[9] S. Joardar, A. Chatterjee, S. Bandyopadhyay, and U. Maulik, "Multi-size patch based collaborative representation for palm dorsa vein pattern recognition by enhanced ensemble learning with modified interactive artificial bee colony algorithm," *Eng. Appl. Artif. Intell.*, vol. 60, pp. 151–163, 2017.

[10] J. Bin, Z. Bahrami, C. A. Rahman, S. Du, S. Rogers, and Z. Liu, "Foreground fusion-based liquefied natural gas leak detection framework from surveillance thermal imaging," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 4, pp. 1151–1162, Aug. 2023.

[11] C.-H. Kuo, P.-C. Chang, and S.-W. Sun, "Behavior recognition using multiple depth cameras based on a time-variant skeleton vector projection,"

IEEE Trans. Emerg. Top. Comput. Intell., vol. 1, no. 4, pp. 294–304, Aug. 2017.
[12] C. Nie, Z. Ju, Z. Sun, and H. Zhang, "3D object detection and tracking based on LiDAR-camera fusion and IMM-UKF algorithm towards highway driving," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 7, no. 4, pp. 1242–1252, Aug. 2023.
[13] S. Z. Gurbuz et al., "American sign language recognition using RF sensing," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3763–3775, Feb. 2021.
[14] J. Wu, L. Sun, and R. Jafari, "A wearable system for recognizing American sign language in real-time using IMU and surface EMG sensors," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 5, pp. 1281–1290, Sep. 2016.
[15] Y. Ren, N. Wang, M. Li, and Z. Xu, "Deep density-based image clustering," *Knowl.-Based Syst.*, vol. 197, 2020, Art. no. 105841.
[16] X. Peng, S. Xiao, J. Feng, W. Y. Yau, and Z. Yi, "Deep subspace clustering with sparsity prior," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 1925–1931.
[17] J. Xie, R. B. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *Proc. Int. Conf. Mach. Learn.*, 2016, vol. 48, pp. 478–487.
[18] B. Yang, X. Fu, N. D. Sidiropoulos, and M. Hong, "Towards K-means-friendly spaces: Simultaneous deep learning and clustering," in *Proc. Int. Conf. Mach. Learn.*, 2017, vol. 70, pp. 3861–3870.
[19] J. Wu et al., "Deep comprehensive correlation mining for image clustering," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8128–8137.
[20] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 3rd ed. Waltham, MA, USA: Morgan Kaufmann Publishers, 2011, pp. 471–479.
[21] M. Ester, H. -P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Int. Conf. Knowl. Disc. Data Mining*, 1996, pp. 226–231.
[22] A. Hinneburg and D. A. Keim, "An efficient approach to clustering in large multimedia databases with noise," in *Proc. Int. Conf. Knowl. Discov. Data Mining*, vol. 98, 1998, pp. 58–65.
[23] M. Ankerst, M. M. Breunig, H. -P. Kriegel, and J. Sander, "OPTICS: Ordering points to identify the clustering structure," *ACM SIGMOD Rec.*, vol. 28, no. 2, pp. 49–60, 1999.
[24] W.-A. Lin, J.-C. Chen, C. D. Castillo, and R. Chellappa, "Deep density clustering of unconstrained faces," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8128–8137.
[25] Y. Wang, E. Zhu, Q. Liu, Y. Chen, and J. Yin, "Exploration of human activities using sensing data via deep embedded determination," in *Proc. Int. Conf. Wireless Algorithms Syst. Appl.*, 2018, pp. 473–484.
[26] S. K. Pal, S. K. Meher, and A. Skowron, "Data science, Big Data and granular mining," *Pattern Recognit. Lett.*, vol. 67, part 2, pp. 109–112, Dec. 2015.
[27] S. K. Pal, B. U. Shankar, and P. Mitra, "Granular computing, rough entropy and object extraction," *Pattern Recognit. Lett.*, vol. 26, no. 16, pp. 2509–2517, Dec. 2005.
[28] D. Małyszko and J. Stepaniuk, "Adaptive multilevel rough entropy evolutionary thresholding," *Inf. Sci.*, vol. 180, no. 7, pp. 1138–1158, Apr. 2010.
[29] D. Chakraborty, B. U. Shankar, and S. K. Pal, "Granulation, rough entropy and spatiotemporal moving object detection," *Appl. Soft Comput.*, vol. 13, no. 9, pp. 4001–4009, Sep. 2013.
[30] B. Lei and J. Fan, "Image thresholding segmentation method based on minimum square rough entropy," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105687.
[31] B. Lei and J. Fan, "Adaptive granulation Renyi rough entropy image thresholding method with nested optimization," *Exp. Syst. Appl.*, vol. 203, 2022, Art. no. 117378.
[32] "Sonel KT-384 operating manual," Accessed: Oct. 24, 2023. [Online]. Available: https://www.manualslib.com/manual/759604/Sonel-Kt-384.html
[33] A. S. Mondal, "Hybrid-uniform mixture model based iterative robust coding for image recognition problems using thermal imaging," M. E. Thesis, Dept. Elect. Eng., Jadavpur Univ., Kolkata, India, 2023.
[34] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
[35] X. Guo, E. Zhu, X. Liu, and J. Yin, "Deep embedded clustering with data augmentation," in *Proc. Int. Asian Conf. Mach. Learn.*, 2018, pp. 550–565.
[36] L.v.d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.
[37] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
[38] D. F. Crouse, "On implementing 2D rectangular assignment algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 4, pp. 1679–1696, Aug. 2016.
[39] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2020, *arXiv:1802.03426*.
[40] S. Ghosh, A. Chatterjee, and S. Munshi, "Visual cue-aided human supervised robot navigation guidance in photometrically challenging environments using adaptive spatial-feature kernel-guided bilateral LPP," *Meas. Sci. Technol.*, vol. 34, 2023, Art. no. 105404.
[41] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 849–856.
[42] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," in *Proc. 1998 ACM-SIGMOD Int. Conf. Manage. Data*, Seattle, WA, USA, Jun. 1998, pp. 94–105.
[43] C. C. Aggarwal, C. Procopiuc, J. Wolf, P. S. Yu, and J.-S. Park, "Fast algorithms for projected clustering," in *Proc. 1999 ACM-SIGMOD Int. Conf. Manage. Data*, Philadelphia, PA, USA, Jun. 1999, pp. 61–72.

Pritam Paral (Member, IEEE) received the B.Tech. degree in electronics and instrumentation engineering from the West Bengal University of Technology, Kolkata, India, in 2012, the M. E. degree in electrical engineering from Jadavpur University, Kolkata, India in 2014, and the Ph.D. (Engg.) degree in instrumentation and measurement from Electrical Engineering Department, Jadavpur University, in 2022. He is currently a Faculty with the Electrical Engineering Department, IIEST, Shibpur, Howrah, India. His research interests include human-centered robotics, instrumentation and signal processing, computer vision and machine learning, granular computing, evolutionary computation, and fractional-order chaotic systems.

Saibal Ghosh (Graduate Student Member, IEEE) received the B.E. and M.E. degrees in electrical engineering from Jadavpur University, Kolkata, India, in 2016 and 2020, respectively. He is currently working toward the Ph. D. degree with the Department of Electrical Engineering, Jadavpur University. His research interests include machine learning, computer vision, deep learning, artificial intelligence, signal processing, instrumentation, and robotics.

Sankar K. Pal (Life Fellow, IEEE) received the first Ph.D. degree in radio physics and electronics from the University of Calcutta, Kolkata, India, in 1979, and the second Ph.D. degree in electrical engineering along with DIC from Imperial College, University of London, London, U.K., in 1982. He is currently a National Science Chair, Government of India, and the President with Indian Statistical Institute (ISI), Kolkata. He is also a Distinguished Scientist and former Director with ISI, a former Distinguished Professor of Indian National Science Academy, and a former Chair Professor of Indian National Academy of Engineering. He was with the University of California at Berkeley, Berkeley, CA, USA, and was also with the University of Maryland at College Park, College Park, MD, USA, NASA JSC, Houston, Texas, and US Naval Research Laboratory, Washington, DC. He has co-authored 21 books and more than 500 research publications in his research interests which include the areas of pattern recognition, machine learning, image/video processing, data mining, web intelligence, soft computing, bioinformatics, and cognitive machines. He is also on the Editorial Board of 30 internationally well-known scientific journals in computer science and engineering, including several IEEE transactions. He is a Fellow of TWAS, IAPR, IFSA, and all four National Academies for Science/Engg. in India.

Amitava Chatterjee (Senior Member, IEEE) visited Saga University, Saga, Japan, on a Monbukagakusho Scholarship, in 2003. From 2004 to 2005, he was with the University of Electro-Communications, Tokyo, Japan, as a JSPS Postdoctoral Fellow. He visited Université Paris XII and Université Paris-Est, Champs-sur-Marne, France, as an Invited Teacher, in 2004, 2009, and 2017, respectively. He is currently a Professor with the Electrical Engineering Department, Jadavpur University, Kolkata, India. His research interests include nonlinear control, intelligent instrumentation, signal processing, image processing and pattern recognition, and robotics.