Implementation of GF-HOG Technique for Effective Commercial and Industrial Load Clustering and Classification for Better Demand Response

Aniruddha Bhattacharya[1] and Madhusudan Singh[2] (iD)

[1]*Assistant Professor, Department of Electrical Engineering, Delhi Technological University, Delhi, India*
[2]*Professor, Department of Electrical Engineering, Delhi Technological University, Delhi, India*

*****Correspondence:** Aniruddha Bhattacharya; Email: a.b.bhattacharya@dtu.ac.in ; Tel.: +91-9910977405

ABSTRACT: With increased penetration of renewables since last decade has evolved measures from the regulator for robust distribution network mostly catering to residential load. With increasing future demand of commercial and industrial load (CIL) due to aspiring GDP growth and insistence of popular governments to encourage renewable use, large chunk of the CIL will be served by renewables. So increasing the robustness alone on the renewable supply side will be in vain unless effective Demand response with rationalized tariff system for CIL which is more profitable energy market than subsidized residential tariff in India .The present method employed for clustering and classifying Load profile of CIL loads with wild swings will impede effective demand response implementation due to tariff cartelization in favor of not so popular energy systems. The method of GF-HOG technique with SVM will adopt cluster and classification of pattern of Load profiles for better recognition generated out from large CIL data. This will help produce more vibrant tariff structure for effective CIL load management. The algorithms were tested on a local network which 70% -30 % of load in favor of CIL and found to generate better DR.

Keywords: Gradient function histogram of oriented gradients (GF-HOG), Commercial and Industrial Load (CIL), Targeted Load Profile (TLP), Support vector mechanics (SVM), Demand Response (DR), Probability density function (PDF).

1. INTRODUCTION

THE development of the electrical power distribution policy makers offer distribution companies fresh area of evaluating rationalized prices, for different types of consumers. For the effective tariff mechanism, the electricity buyer's categorization is based upon the type of customer (*viz.* residential, commercial and industrial) are sparsely correlated to the actual electrical consumption providing distorted image of load profiles distribution [1]. The load profiles of the customers and clubbing consumers exhibiting similar load profile is the hallmark of residential load distribution studies [2], [3]. The customer groups, is divided into major categories i.e. residential, commercial and industrial, Inside each category, fresh clustering and classification can be identified by their actual loading patterns of the consumers. A residential load customer, the load behavior for a specific loading condition refer to as Targeted Load profile(TLP) *They* can be produced by averaging the load data monitored from the Load dispatch center for a period 15 minutes to 30 minutes of measurements of load. The 24-hour load profile data are normalized and few numbers of clustering methods have been used in the previous literature building load profile, traditional

statistical clustering methods [1], [4][5][6][7], fuzzy logic and neural networks [6], [8][9][10][11], The clustering will create formation of the *load patterns* suitable for residential load. Each pattern is created with load patterns of optimal number of clusters with smallest variance during classification. However, since the Load pattern required for commercial and industrial load profiling offer large swing of load due to intermittent industrial and commercial load operations so proper classification and clustering with traditional methods will cause irrational tariff structure and weird DR, various types of clustering methods have been investigated for accurate classification of customer load profile of wild swinging CIL with shortlisted substation servicing load mix which is predominantly CIL type From the initial set of TLPs (Target Load Profiles) three clustering algorithms i.e., K-means, fuzzy hierarchical clustering, and Summon maps are adopted. This implies applications of classical statistical clustering techniques [1, 4, 5, 7, 6], fuzzy logic [8] [9] [10], and neural networks [10, 11]. Every profile is created on account of time averaged load (KW) profiles in the same class and each activity class of industrial and commercial load and may be Target Load Profile of the class. The TLPs used in clustering and classification methods are normalized, the load cluster centroids cannot provide two-dimensional clustering. The data point is always marked to the nearest center, the above said techniques were not able to find load clusters other than spherical clusters [7]. Which always remain typical for the clustering requirement of this type of commercial and industrial load distribution. Industrial load requires the load profile in 2D form with different class of customer and different classes of activities within it as Load profile. With wild swing in the above profiles for hard loads, the model load

profiles (MLP's) based on conventional clustering and classification approach will nonlinear quadratic with large statistical errors.

1.1 Contribution to clustering solution for effective demand response of commercial and industrial load

(a) Firstly For large distribution-based data vis load profile data points, clustering with strongly overlapping peaks like the case discussed above, The realization of TLP's load data points are the mix of kernel probability distribution functions (PDF) [8]; 2D distribution of Gradient density distribution subjected to Euclidian distance over the kernel .The precision of these techniques depends on the accuracy of probability of the template to represent data.. The cluster centers with a higher local density surrounded with lower local density in the neighborhood. The edge detection of these densities leads to Gradient distribution. Each data point, two quantities of density and gradients are computed: its density ρ_i and its distance δ_i from high density point and their gradients. They depend on the distances δ_{ij} between high density data points, which satisfies the equality.

(b) Density estimation techniques provide an accurate estimate of a observable PDF based on an illustrated collection of time averaged load data [1] , [2], [3] ,[4] . The gradient density functions obtained are known as (GF_HOG) Gradient Function of histogram of oriented gradients and are used to find, swing detection in Targeted Load profile (TLP) [10] [11]. The method developed in [12] [13] with effective application of GF-HOG [15]   supported by Kernel PCA(K-Principal Component Analysis) and Support Vector Mechanism(SVM) for detecting sketch based high swinging high density distribution

(c) The GF-HOG are primarily employed for pattern recognition techniques [6, 24]. The approaches provided in [19, 1] demonstrate an HOG application in association with support vector machines [20] for identifying probability density cluster image of 2D. In [10], Gradient Field HOG (GF-HOG) is applied to retrieval of sketch-based image the density of is taken as a probability function over a 2D domain, and the density distribution in 1D-PCA and gradients as 1D –PCA or edge detection PCA with SVM is used as the feature identifier to describe the wild swing in CIL load.

2. PAPER STRUCTURE

Section (3) will review of advance clustering and classification techniques to find suitable method for CIL load clustering in 2D.

Section (4) In K-medoids method, K value it is calculated as distance within data points. In the mean shift technique, is used to detect only spherical load clusters which seems unlikely in commercial and Industrial Load (CIL) load clustering in 2D and to find automatically correct number of clusters.

Section (5) will explore the popular concept used in the density estimation, these gradient density functions which are

referred as GF-HOG which are primarily used for pattern recognition likely to be used in this research work for detecting distributed density gradient images of CIL Load [12] [13]. The approaches adopted in [14] [15] demonstrate an application of HOG for detecting wild swings in non-circular CIL load density images in 2D. In the lately published article [16], These HOG's are called the GF-HOG which are used primarily for sketch-based image retrieval seems ideally suited for detection of CIL load clusters. The 2D GF-HOG will reconstructed to 1D Gradient PCA vector and 1D Density distribution PCA vector with non-vanishing Eigen values.

Section (6) we provide experimental result to validate the GF-HOG use in feature extraction of Gradient density distribution images obtained from CIL loads. Clusters. The present work was inspired by recent work on application of pattern recognition to quantum supremacy [17] [18] [19]. This method basically used for density determination and gradient feature extraction from the images created out from substation data.

3. ADVANCE CLUSTERING AND CLASSIFICATION TECHNIQUES

3.1 Substation load profile 2D data distribution has three major steps.

Step 1: The substation metered load profiles data are mined for two features (1) data density and (2) feature and shape extraction which uses normalization for shape data only. Substations are grouped as per the TLP's shape, producing load cluster images using various iterative algorithms. An average normalized template, load shape of TLP within load group, to represent the overall load profile shape from the substations metered data with each cluster for metered over days.

Step 2: The relationship between fixed retrieved data and these developed clusters of substations is characterized. A classification method uses the multinomial Logical Regression model. It can mark on the HOG plot undetected substations metered data to each cluster with a probability completely based on the fixed data. This is imitable by time periods and unknown areas, for which real-time metered load (KW) data.

Step 3: The normalized vectors will be scaled up by suitable Eigen value to load magnitude density distribution vectors and shape vectors. For unmonitored substations, the actual scaling process is sometimes used to estimate peak loads distribution as bright spots in the images meant for feature extraction.

4. CLOSE COMPARATIVE STUDIES

4.1 K-Means clustering technique

In Low Voltage networks servicing CIL loads cannot be classified into micro categories, the clustering handles large volume of intermittently swinging and sometimes wildly varying meter data through the study period [25]. It cannot do assessment of the optimal clusters number. Number of Cluster were investigated in terms of the dissimilarity within each group but cannot converge for largely fluctuating CIL. The

difference in the cluster is identified by the sum of Euclidian error between substation metered load profiles within CIL groups. With a smaller number of clusters, the targeted set will be at the centroid. The Euclidian distance of all clusters is calculated as the error and the primary disadvantage of this approach is its inability to scale and iteratively convergent solution for deciding precise number of clusters with large fix of CIL loads, So more effective and traditional techniques like histograms used for pattern recognition ,which can compute the densities and gradient of different shapes called as Histogram of Gradients (GF-HOG), by smoothing the image, and computing the gradient and binning the resulting gradients can be useful to cluster and classify CIL Loads.. For the discrete N samples of Set S, it is important to know the faster rate of convergence of Eigen values of the proposed density estimation vector for GF-HOG method. For the kernel density estimators and the histograms used for shape vector estimation [14, 15], the convergence rates are the expected value (N samples size) of the square of the error between the computed and the true PDF's and are equal to be $O(N^{-1/3})$ and $O(N^{-3/5})$ [17],[21]

4.2 Multinomial logistic regression (MLR) for classification within the cluster groups

Logistic regression is used to predict categorical dependent variable y, outcome is used here to produce classification tools [26]. The independent variables (template) and the dependent variable x (Load clusters), the regression coefficients will produce probability of outcomes by logistic function expressed in (5)

$$P(y = 1) = F(z) = \frac{1}{1 + e^{-(b_0 + b_1 \times z)}}.$$

(5)

Where $z = \{z_{i1}, z_{i2}, \ldots z_{im}\}$ is set of fixed independent variables $b_k = \{b_{1k} \ldots b_{mk}\}$ are constant of regressions $P(y=1)$: Probability of the substation to remain in the first cluster.

The result of equation (5) is between 0 and 1, For N number of clusters (MLR), multinomial logistic regression is used to define the relationship between z values and P(y) membership values. For N clusters, the model will give N-1 logistic regression equations and probability of metered data assigned to a cluster that can be determined by (6) and (7).

$$P(Y_i = n) = \frac{e^{b_n \times z_i}}{1 + \sum_{k=1}^{K-1} e^{b_k \times z_i}}, \quad n = 1, 2, \ldots K - 1$$

(6)

$$P(Y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} e^{b_k \times z_i}}$$

(7)

A substation meter data assigned to the cluster with highest probability. We can able to develop an estimate of simple mean-shift. The approach will provide observations about a point and related to the density gradient. A simple modification provides k-nearest neighbor estimate of mean-shift class. This is same as Loftsgaarden and Quesenberry's [11] estimate of k-nearest neighbor density, where within

number of observations do not change, but the other estimates the volume of region remain same

4.3 Density Gradient Estimation

In issues related to pattern recognition, probability density function (PDF) is known in greater details. Limited information about density function, nonparametric techniques is used to obtain density gradient estimates. The nonparametric density estimates as discussed in Cacoullos' [5] Kernels of multivariate problems of Parzen's [4] univariate kernel estimates.is is possible n-dimensional space.

Let X (n) are N independent n-dimensional random vectors identically. Kernel density estimators of multivariate type as

$P_N(X) = (Nh^n)^{-1} \sum k(h^{-1}(X - Xj))$ where K(x) is the scalar functions satisfying

Sup[k(Y)]<inf

$Y \in R^n$

$$\int_{R^n} [k(Y)] dy < infinity$$

(8)

$$\lim_{[[Y]]+=} [[Y]]^n k(Y) = 0$$

(9)

$$\int_{R^n} k(Y) \, dY = 1$$

(10)

5. CLUSTER REPRESENTED USING NON-PARAMETRIC 2D DENSITY DISTRIBUTION IMAGES

Clusters having irregular shapes detected based on densities of data points. In density-based clustering of noise [11], density threshold is chosen, which discards noise as the points in regions of densities where threshold is greater, it assigns the data point to different clusters as outliers This drawback is not present in the mean-shift method of clustering [10, 11] is choosing an appropriate threshold. For a chosen group of M customers, each targeted Load Profile (TLP) of CIL load is characterized by a vector $x^m = \{x_h^m, h = 1,2,3,4 \ldots H\}$, where H represents time-domain data time intervals of 15 min, 30 min, representing the Target Load Profile [1]. The whole set of data is X = { x^m with m =1,2,3,4 ….M} . The clustering process associates the initial M Target Load Profile (TLP) to customer classes, with each cluster represented as X^k C X contains n^k TLPs, for k= 1,2,3, 4.K. Let us further call c $^{k, as}$ the component vector having the centroid of the cluster There set of points representing density distribution function. That converge to a maximum this method is useful for finding non-spherical clusters. [10], it is useful to find non spherical clusters and to calculate number of clusters. The cluster centers are defined, as maxima of the density of data points. The algorithm converges to cluster centers which are surrounded by lower local density neighbors and have large distance from few points of higher local density. For every

data point, we find two quantities: *δi as its distance from points of higher density and ρi as its local density*. It depends on the distances δ_{ij} (the distance between data points), satisfies the Euclidian criterion. The density ρ_i of data point can be found as

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \tag{11}$$

δc is a cutoff distance and ρ_i is equal to density of point distribution. The algorithmic only take as input, relative magnitude of ρ_i in different points. The choice of critical distance is key.is key for large data sets for the results to be robust with respect to δ_i which is the distance between any point and other point of higher density the highest density points, we represent $\delta_i = \max(d_{ij})$. The δ_i is larger than nearest neighbor distance only for points which are global and are density local maxima. The cluster centers are identified as data points with large δ_{ij}. These observations, which, is illustrated by the simple example in *Figure 1A*. Solved using Euclidean distance framework described below illustrated diagrammatically by *Figure 1B* shows 29 points embedded

The Euclidean distance is used to measure different types of distances represented as follows

(1) Distance between the vectors with H components

$$d(\mathbf{y}, \mathbf{x}) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(y_h - x_h)^2} \tag{12}$$

*(2) The distance calculated between the vector **y** and each of **M** members of set **X** i.e. Vector-to-set distance,*

$$d(\mathbf{y}, \boldsymbol{X}) = \sqrt{\frac{1}{M}\sum_{\mathbf{x}\in\boldsymbol{X}} d^2(\mathbf{y}, \mathbf{x})} \tag{13}$$

the mean distance of members x_q the set (for Q members) and y_i of set Y (for J members) The algorithm can be run iteratively for unlimited number of clusters with all TLP data accommodated in single image of density distribution clusters
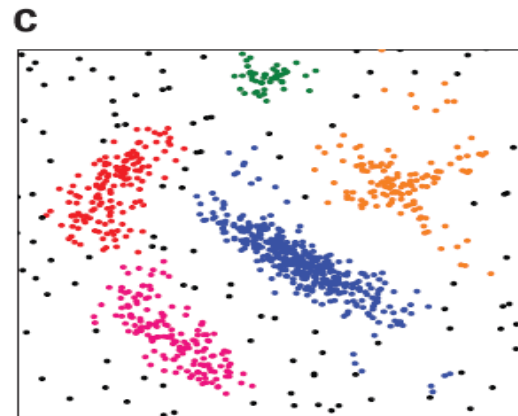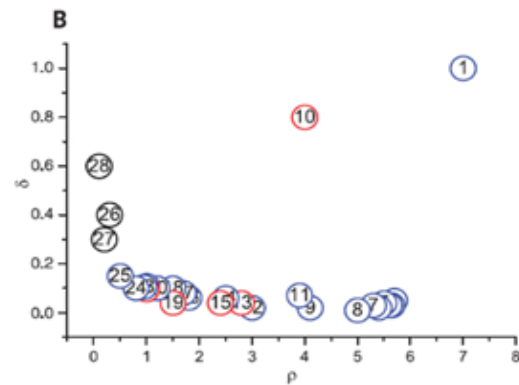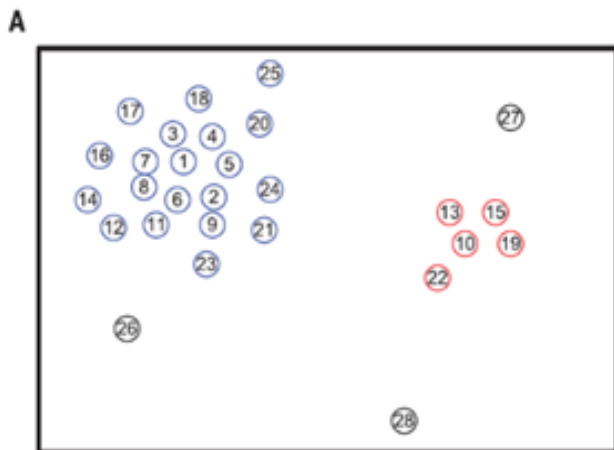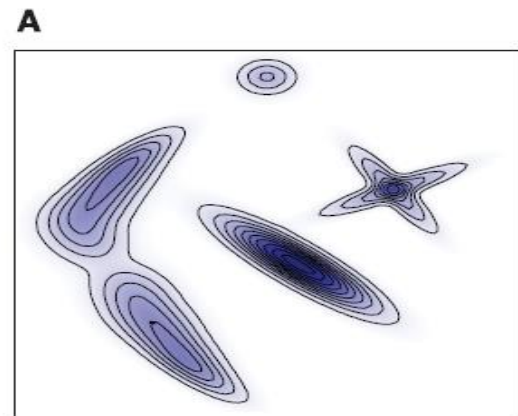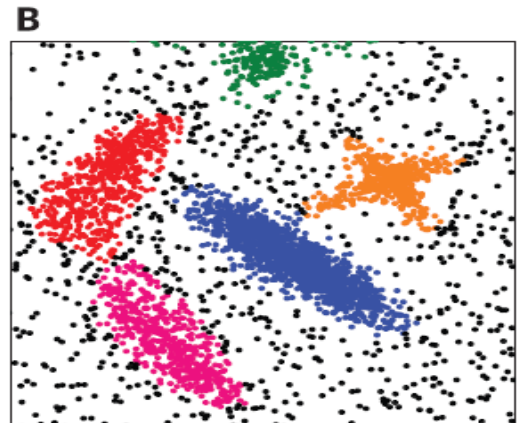

Figure 1: 2D representation. (A)Distribution of points... (B) Data Decision graph (C) different clusters represented with Different colors

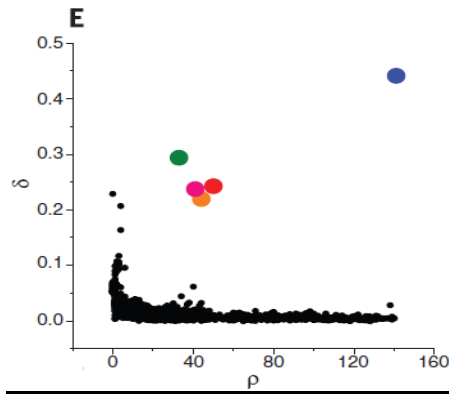*(1)* **X axis** represents **ρ** *(2)* **Y axis** represents **δ**

Figure 2: Images of point distributions. (A) The probability distribution. The regions with lowest density (B) Probability distributions for 5000 and 3000 sample points, respectively. Points are marked with different color according to the cluster. (E) The decision graphs, with colored centers of cluster.

5.2 Theoretical interpretation of the images obtained

The density distribution images represented in two-dimensional space distribution as shown in *Figure 2*. Then the maximum density is found at various points, which are centers of clusters. Figure 1B, the plot of ρi as a function of δi, mentioned as decision graph. The δ values of points 8 and 9, of same values of ρ, is different: Point 10 remains in cluster group of point 1, but higher ρ points are closest to cluster, but the neighborhood of high density point 9 belongs to different cluster. The only high δ and high ρ points are the centers of clusters. Points 25, 26, 27 are, outliers. When the cluster centers are found, other points are representing the same cluster as nearby points of highest cluster density. The cluster allocation is achieved in a single step, compared to other clustering and classification algorithms where objectivity is optimized iteratively [3, 7]. In clustering, it is also useful to asses reliability of an assignment. The approaches are obtained by function optimization [2, 8], and convergence is a natural defined measure.

5.3 Special kernel functions for identification non-regular density distribution

The 2D density distribution clusters are investigated using two components (1) Density (2) Density gradients by retaining Hessians functions at each point of the clusters and extracting one dimensional feature vector for easy identification used depending upon its prudence. The density gradient estimates do normally have kernel functions for gradient estimation. This result offers intuitive estimate for the density gradient mean shift. It can be expanded to a K nearest- neighbor approach in gradient estimation as discussed Loftsgaarden and Quesenberry's [11] extension of kernel density estimates.

5.4 Mean-Shift gradient estimates

The kernel Gaussian function is often the differentiable kernel function of multivariate type which satisfies the conditions for consistency of estimates of density gradient. The kernel stating probability Gaussian density function with zero mean with covariance matrix stated as

$$k(X) \equiv (2\pi)^{-n/2} \exp\left(-\tfrac{1}{2}X^T X\right). \tag{12}$$

$$\hat{\nabla}_x p_N(X) = N^{-1} \sum_{i=1}^{N} (X_i - X)(2\pi)^{-n/2}h^{-(n+2)}$$
$$\cdot \exp\left[-(X - X_i)^T \left(\frac{X - X_i}{2h^2}\right)\right]. \tag{13}$$

Intuitive interpretation of (13) is a mean shift of the measure of weighted the observations about the point X. The mean samples of the weighted shifts are taken as estimate of gradient. This will give the kernel function is of the form

$$k(X) = g(X^T X). \tag{14}$$

$$k(X) = \begin{cases} c(1 - X^T X), & X^T X \leq 1 \\ 0, & X^T X > 1 \end{cases} \tag{15}$$

Where

$$c = \pi^{-n/2}\left(\frac{n+2}{2}\right)\Gamma\left(\frac{n+2}{2}\right) \tag{16}$$

C is the constant for normalizing required to integrate the kernel function to 1. This kernel function can satisfy the conditions for consistency of the estimate of gradient and similar to the kernel stated in Epanechnikov [lo]. By putting *equations (15) in (17)*, we compute the gradient in (18)

$$\hat{\nabla}_x p_N(X) \equiv (Nh^n)^{-1} \sum_{j=1}^{N} \nabla_x k(h^{-1}(X - X_j)) \tag{17}$$

$$\hat{\nabla}_x p_N(X) = (Nh^{n+2})^{-1}2c \sum_{X_i \in S_h(X)} (X_i - X) \tag{18}$$

$$= \left(\frac{k}{Nv_h(X)}\right)\frac{n+2}{h^2}\left(\frac{1}{k}\sum_{X_i \in S_h(X)}(X_i - X)\right)$$

$$v_h(X) \equiv \int_{S_h(X)} dY = \frac{h^n \pi^{n/2}}{\Gamma(n + 2/2)} \tag{19}$$

The volume of the region

$$S_h(X) \equiv \{Y : (Y - X)^T(Y - X) \leq h^2\}$$

and k is the number of TLP data points within $S_h(X)$ and, Equation (18) provides with estimation of the gradient estimation process. The last term in (19) is a mean shift of data points of the observations in S(X) of X.

$$M_h(X) \equiv \frac{1}{k}\sum_{X_i \in S_h(X)}(X_i - X) \tag{20}$$

If the gradient values are zero, then it represents uniform density over cluster S(X), the average mean shift is zero with a nonzero density gradient heading to quick increase in PDF, the time average data points will align along the direction than anywhere else in S(X). The mean shift average will head along the direction and magnitude of the gradient is proportional to length [13]

5.5 Gradient Clustering Algorithm

The method of clustering a data points is to assign each data points to the nearest mode along the direction of the gradient at that point. One could move each observation with small step in the direction of gradient and iteratively repeat the process till denser clusters produces near the modes. The other approach is to shift each observation by similar amount proportional to the gradient of the observation point [14]

$$X_j^0 \equiv X_j, \qquad j = 1,2,\cdots,N \tag{21}$$

$$X_j^{i+1} = X_j^i + a\nabla_x \ln p(X_j^i). \tag{22}$$

Value of (a) is optimally chosen to obtain convergence. N-dimensional analog can be found putting into the roots of the equation

$$\nabla_x p(X) = 0 \tag{23}$$

Above equation represents the combined density p(X). The minima will be the roots of (23), the algorithm resulted from (22) will move data points away as the gradients moves away from them. From every iteration, points will move closer to center. The following two observations are

(1) The step size for the gradient which is same near the centroid will be lesser than that of farther point. This will cause data points move far from the centroid and near to the local minimum which move to the mode quicker.

(2) Gradient normalization is considered if a mean M Gaussian density and covariance matrix represented as

$$p(X) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}(X-M)^T(X-M)\right]. \tag{24}$$

$$X_j^1 = X_j + \nabla_x \ln p(X_j) = X_j - (X_j - M) = M. \tag{25}$$

Gaussian density, cluster merged to single point after one iteration. This will help mix of density problem of which each component densities which look Gaussian

(3) The normalized gradient, convergence can be achieved as mean-square sense. [23][29]

$$Y = X + a\nabla_x \ln p(X) \tag{26}$$

Minimize covariance of our samples with smaller (a) as shown in (27) & (28),

$$E\{Y - M_Y\}^T(Y - M_Y)\} \le E\{(X - M_X)^T(X - M_X)\} \tag{27}$$

$$(E\{Z\})_i \equiv \left(\int_{R^n} \frac{\nabla_x p(Y)}{p(Y)} p(Y) \, dY\right)_i \tag{28}$$

5.6 Complete Gradient Cluster Algorithm

1. Set of n-dimensional vectors as random sample for n-dimensional random variable X, which has density distribution of E as described in *Equations (24)-(28).*

2. Using the methodology described, he kernels estimator $\hat{E}^*$ can be computed.

3. Taking the clusters particularly to local maxima of $\hat{E}$ function, by transposing those elements in the gradient $\nabla \hat{E}$ direction described in *Equation (28),* with the appropriate step.

The Gradient Clustering Algorithm, are of classic Newton method, defined as

$$x_j^{k+1} = x_j^k + b\frac{\varLambda f(x_j^k)}{f(x_j^k)} \text{ For j=1\dots m and k=0, 1\dots\dots.k*} \tag{29}$$

4. Formulate the kernel estimator $\hat{E}$;

5. Identify condition to stop the program (find also number of steps represented as k^*);

6. Initiation of process to create cluster with particular elements are assigned to set after the last, k^* step;

7. Analyze the effect of the values of parameters on obtained results.

6. HOG FOR CLUSTER TYPE DETECTION

This section explains feature extraction chain of the 2D density model functions, which is summarized in *Figure 3*. Evaluation of local histograms normalized over image gradient in a dense grid as described in Section 2 and 3. Similar features is explained in pattern recognition methods [16, 18, 19, and 20].
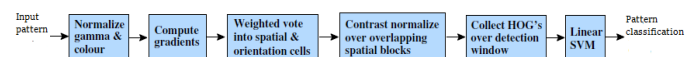

Figure 3: Object detection chain and feature extraction block chain

The local density distribution gradients or sometimes called as edge directions can be characterized by local object appearance and shape, without having corresponding edge positions or gradients. Then by splitting the image into proportional spatial cell and constructing a local 2-D histogram of density and direction of gradient. This is a measure of local histogram over larger distributed regions and to normalize all cells. This is known as 2 D Histogram of Oriented Gradient (HOG) descriptors. HOG descriptors are combined with feature vector for conventional SVM classifier provides detection chain the orientation histograms offer many precludes [22, 23, 24], but when combined with distributed histogram and normalized in Lowe's Scale Invariant Feature Transformation (SIFT) to precise image matching [30], either image patch descriptor for matching scale invariant points. SIFT perform well in this application [30, 31]. "The Shape Context studied in [32] provides alternative block shapes, although initially using only edge features without orientation histogram that makes the representation effective". The HOG representation has many advantages. It detects density and gradient structure that is very unique of local shape, and easily has controllable invariance. The 2-D histogram is represented, two parameters of image i.e. one is the image density, which is denoted and the other is the gradient of the image [6] [15].

6.1 Improved 2D Gradient Histography Image

The 2-D histogram is composed of density image, which is denoted by f(x, y), and gradient image. For the given function f(x, y), the gradient at the location (x, y) is defined as the 2-D column vector [6]

FOREX Publication
Open Access | Rapid and quality publishing

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]^T .$$
(48)

For a discrete image representing HOG-GF, This will generate the gradient which signifies running difference of density over the horizontal and vertical axis of the image, which is as follows:

$$\frac{\partial f}{\partial x} = f(x+1, y) - f(x-1, y)$$
(49)

$$\frac{\partial f}{\partial y} = f(x, y+1) - f(x, y-1).$$
(50)

Horizontal and vertical gradients effectively can be used to find combined gradient g(*x, y*) to find gradient of images as follows. This will remove noise and outliers from the space for better image analysis

$$g(x,y) = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2}.$$
(51)

Each distributed cluster image has its GF pair which consist of united gradient (G) and density (F) representing density. Traversing clusters of images, and computing the frequency of G-F pair is easily possible. If p_{ij} is the frequency of G-F pair (i, j), The cluster density will be represented by

$$p_{ij} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \delta_{ij}$$
(52)

6.2 Mathematical analysis for retrieving two 1D Principal Component Analysis (PCA) components using Unsupervised Learning Algorithm
The consolidate noise free 2 D images are sorted by unsupervised algorithm to retrieve two 1D density images of density and gradient. Following is a sample numerical exercise solved and can be replicated for large data set of running data from substations

Step 1: Data Set

Substation Nos	Mean data of 15 minutes (D1)	Mean data of the next 15 minutes (D2)	Mean data of the next 15 minutes (D3)
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

Step 2: Spread it in matrix and compute mean
The mean matrix A = [66 60 60]

Step 3: Compute covariance

	D1	D2	D3
D1	504	360	180
D2	360	360	0
D3	180	0	720

Step 4: Calculating one dimensional PCA using Eigen values and vector

Coefficients are values are λ= 44.81 λ= 629.11 λ=910.01
Eigen values = -3.75; -0.505; -1.05

Step 5: Sort the eigenvectors with decreasing eigenvalues and choose eigenvectors of largest eigenvalues to form a d × k dimensional matrix W.

Eigen values in decreasing order
910.06; 629.11; 44.81

Eigen vector corresponding to two highest eigenvalues
W=   1.05594   -0.594
          0.69108   -0.675
             0   0

6.3 HOG-GF image segmentation through 2D Principal Component Analysis (PCA)
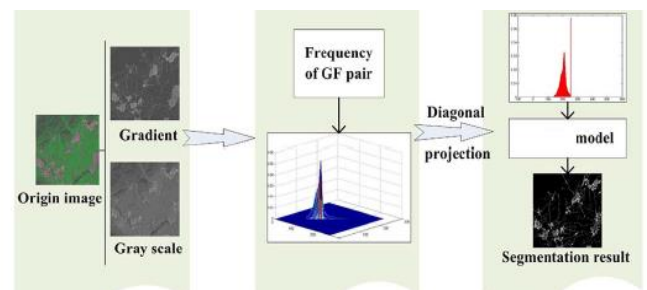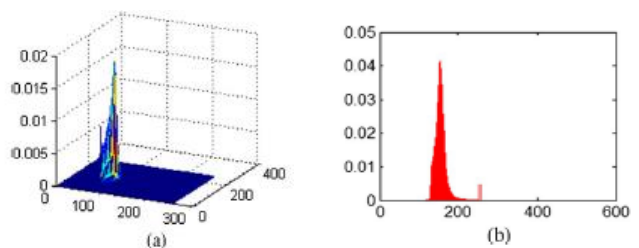

Figure 4(i)

Figure 4(ii)

Figure 4: (*i*) Improved gradient of histogram of 1D PCA histogram density image; (*ii*) (**a**) Improved 1D PCA gradient histogram. (**b**) Two dimensional PCA of gradient histogram

6.4 MAD (Mean Average Distance) Model
For 1D histogram, k is considered as threshold segment value for binary bits. The binary segments, density distribution and its gradient images are splitted into two classes i.e. of density and its gradient corresponding to r ∈ {0, 1, 2 . . . k} in the histogram of 1D size [23]

$$\begin{cases} \omega_0(k) = \sum_{r=0}^{k} p_r \\ \omega_1(k) = \sum_{r=k+1}^{2(L-1)} p_r \end{cases}$$
(54)

$$\begin{cases} \mu_0(k) = \sum_{r=0}^{k} r p_r \\ \mu_1(k) = \sum_{r=k+1}^{2(L-1)} r p_r. \end{cases}$$
(55)

Next, we can produce mean absolute deviation on the 1D histogram, where sample mean is μ, and sample variance is σ where MAD is Mean Average Distance models in 1D space

$$\begin{cases} \text{MAD}_0(k) = \sum_{r=0}^{k} |r - \mu_0(k)| \, p_r / \omega_0(k) \\ \text{MAD}_1(k) = \sum_{r=k+1}^{2(L-1)} |r - \mu_1(k)| \, p_r / \omega_1(k). \end{cases} \tag{56}$$

The residential loads, the gradients and densities are same within one type of image. Unlike in CIL load, 2D histogram images have density and its gradient, it appears as a Gaussian curve, and threshold will be the lowest descendent in between modes. The modes of two images are represented by $f(x/\mu 0, \sigma 0)$ and $f(x/\mu 1, \sigma 1)$, respectively. Each probability of $f(x/\mu 0, \sigma 0)$ and $f(x/\mu 1, \sigma 1)$ is $\omega 0(x)$ and $\omega 1(x)$, and the function of the 1-D histogram is shown as follows [23]:

$$h(x) = \omega_0(x) f(x|\mu_0, \sigma_0) + \omega_1(x) f(x|\mu_1, \sigma_1)$$

$$\text{where } 0 \le x \le 2(L-1). \tag{57}$$

For the optimal threshold is represented as $k*$, where $k*$ satisfies

$$k^* = \frac{\frac{\sigma_0 \sigma_1}{\sqrt{2}} \ln\left(\frac{\omega_0 \sigma_1}{\omega_1 \sigma_0}\right) + \mu_1 \sigma_0 + \mu_0 \sigma_1}{\sigma_0 + \sigma_1}. \tag{58}$$

The mean deviation in absolute of σ_1 *and* σ_0 is shown as follows:

$$\sum_{x=0}^{k} f(x|\mu_1, \sigma_1) |x - \mu_0(k)| + \sum_{x=k+1}^{2(L-1)} f(x|\mu_1, \sigma_1) |x - \mu_1(k)|$$

$$= f(x|\mu_1, \sigma_1) |\mu_1 - \mu_1(k)| = |\mu_1 - \mu_1(k)|. \tag{59}$$

7. EXPERIMENT AND RESULT ANALYSIS

A set of CIL load customers connected to the Medium Voltage distribution system with reasonable penetration of renewables having 162 Nos 33KV/411KV 3 Phase AUSGRID NSW Load dispatch center at Australian State of New South Wales has been considered [12]. The system was considered due to mix of residential and nonresidential loads having 75% in favor of CIL loads with the Target Load Profile (TLP) of a customer is found by calculating the mean of 15-min interval in a day data od Load. The original vector space with specific set of data will then be with magnitude and shape of the TLP which are closely monitored. The clustering algorithm assigns to each TLP of every customer to a particular cluster, providing predominantly without overlapping for all customers. The Gradient clustering algorithms used require an optimal number of clusters to be formed. The clusters numbers decrease with increase in the distance threshold [9], and the shape of the clusters considering CIL loads remain non-overlapping. Some indications optimum number of clusters from the view point of classification may be obtained from cluster density and gradient cluster function optimization. The best number of clusters has been found to be the knee of the curve principle

[27]. The cluster patterns are obtained from algorithms described in *Section 4.*

Sample images – 1 of TLP cluster image constructed and its ID conversion
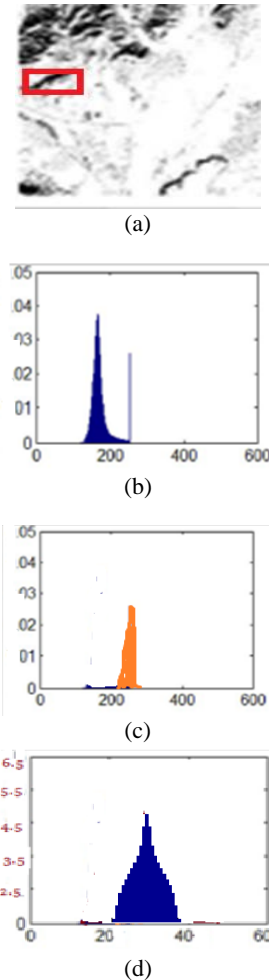

(a)

(b)

(c)

(d)

Figure 5: (a). Sample marked in Red (b). PCA (Eigen Value) of (c). PCA of Gradient Distribution Represents of typical CIL Load densities distribution of the sample marked in Red (d). Covariance distribution
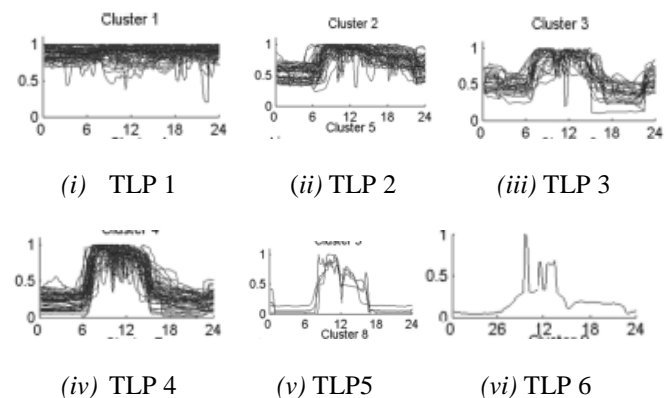

(i) TLP 1   *(ii)* TLP 2   *(iii)* TLP 3

(iv) TLP 4   *(v)* TLP5   *(vi)* TLP 6

Figure 6: Images of the clusters formed during 24-hour period of the day from the Sample images- 1 Clustering results for the HOG-GF METHOD with K = 6 clusters

Sample images – 2 extracted from the substation and its ID conversion
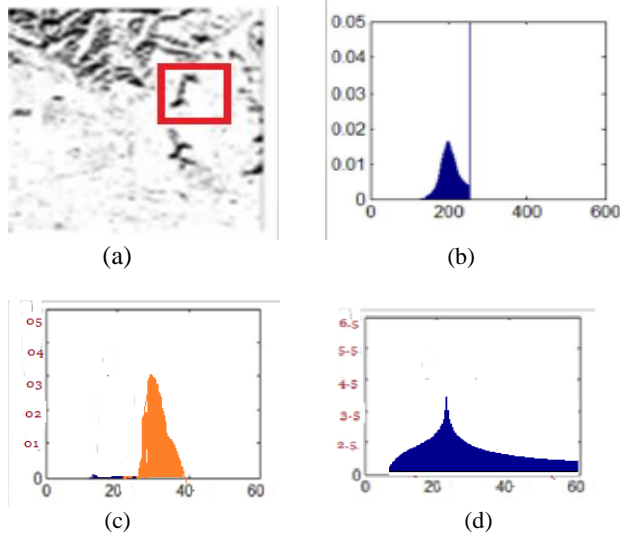

(a) (b)

(c) (d)

Figure 7: (a). Sample marked in Red (b) PCA of densities distribution (c). PCA of Gradient Represents of typical CIL Load of the sample marked in Red Distribution Cluster distribution (d) Covariance distribution

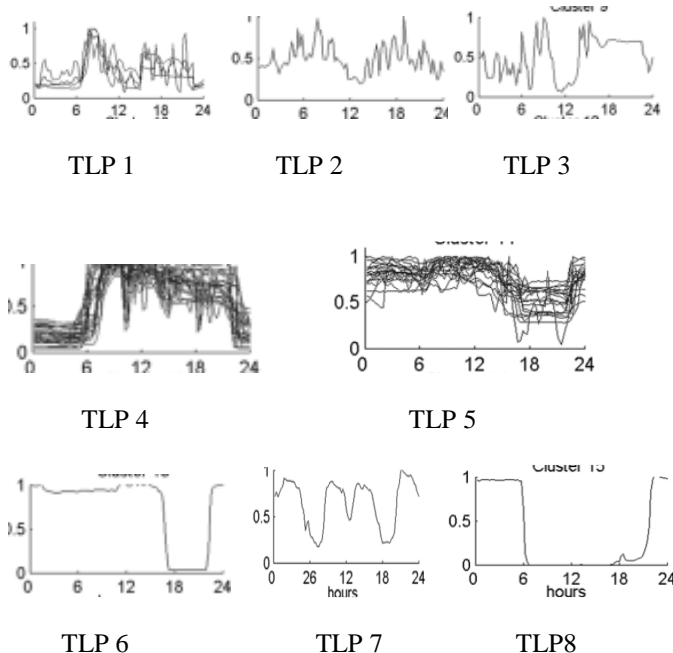TLP 1   TLP 2   TLP 3

TLP 4   TLP 5

TLP 6   TLP 7   TLP8

Figure 8: Images of the clusters formed during 24-hour period fo the day from the Sample image 2 Clustering results for the HOG-GF with K = 8 clusters

7.1 Result analysis

The results clearly indicate the importance of the HOG-GF method compare to other techniques to represent the density and its shape. This superiority also depends on techniques to build separate classes of TLP's and to discriminate load patterns for CIL Loads, as in Fig. 3 unlike hierarchical clustering. The various clustering techniques will be useful to find the number of TLP's, equal to number of clusters i.e., is a

cluster center. However, with increasing fluctuations in the load pattern with decrease in cluster numbers, every method adopts a different strategy for load pattern grouping, the customer partitioning resulting, compute the centroids from the set of data points. [27][28]

8. CONCLUSION

The distribution service providers use Load clustering techniques for density distribution of load pattern for electricity customer classification. The study carried here emphasize most important aspects for the distribution companies for the possible use of clustering and Pattern recognition techniques for gradient density distribution. The indicative number of CIL load customer classes using the GF-HOG method take into consideration the intermittently turbulent CIL load, even though criteria to find out the best number of clusters could result in higher values The results of the clustering quality assessment done in this paper show that one algorithms—HOG-GF, will give highly detailed isolation of the clusters, discriminating load patterns which seems perfectly suited to large swinging CIL Loads. These avenues provide algorithms will perform classification, by making customer groups into a number of customer classes particularly fit for a customer for tariff formation purposes for CIL loads. The various algorithms are used to distribute the TLP's within some clusters and, as such, are less effective for CIL Load.

▨ REFERENCES

[1] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," IEEE Trans. Power Syst., vol. 18, no. 1, pp. 381–387, Feb. 2003

[2] J. MacQueen, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, L. M. Le Cam,J. Neyman, Eds. (Univ. California Press, Berkeley, CA, 1967), vol. 1, pp. 281–297.

[3] E. Parzen, "On estimation of a probability density function and mode," Ann. Math. Statist. yol. 33, pp. 1065-1061, 1962

[4] T. Cacoullos. "Estimation of a multivariate densltv." Ann. Instkatist. Math:, vol. 18, pp. 179-189, 1966

[5] N. N. Cencov, Estimation of an unknown distribution density from observations, Soviet Math., 3 (1962), pp. 1559–1562.

[6] A. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 107-119, Jan. 2000

[7] C. S. Chen, J. C. Hwang, Y. M. Tzeng, C. W. Huang, and M. Y. Cho,"Determination of customer load characteristics by load survey system at Taipower," IEEE Trans. Power Del., vol. 11, pp. 1430–1436, 1996.

[8] G. J. McLachlan, T. Krishnan, The EM Algorithm and Extensions (Wiley Series in Probability and Statistics vol. 382, Wiley-Interscience, New York, 2007

[9] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, in Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, E. Simoudis, J. Han, U. Fayyad, Eds. (AAAI Press, Menlo Park, CA, 1996), pp. 226–231.

[10] G. Chicco, R. Napoli, F. Piglione, M. Scutariu, P. Postolache, and C. Toader, "Load pattern-based classification of electricity customers," IEEE Trans. Power Syst., vol. 19, no. 2, pp. 1232–1239, May 2004.

[11] D. 0. Loftsgaarden and C. P. Quesenberry, "A nonparametric estimate of a multivariate density function," Ann. Math. Statist.,vol. 36, pp. 1049-1051, 1965

[12] "Sample report of Northern India state load dispatch " provided by SLDC Delhi

[13] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review,"ACM Comput. Surv., vol. 31, no. 3, pp. 264–323, 1999.

[14] C. S. Chen, M. S. Kang, J. C. Hwang, and C. W. Huang, "Synthesis of power system load profiles by class load study," Elect. Power Energy Syst., vol. 22, pp. 325–330, 2000.

[15] C.M. Bishop, Pattern recognition and machine learning (Information science and statistics),Springer, New York, NY, 2006.

[16] W. T. Freeman and M. Roth. Orientation histograms for hand gesture recognition. Intl. Workshop on Automatic Faceand Faceand Gesture- Recognition, IEEE Computer Society, Zurich,Switzerland, pages 296–301, June 1995.

[17] D. Scott, On optimal and data-based histograms, Biometrika, 66 (1979), pp. 605–610.

[18] W. T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma. Computer vision for computer games. 2nd International Conference on Automatic Face and Gesture Recognition, Killington,VT, USA, pages 100–105, October 1996

[19] D. G. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60(2):91–110, 2004

[20] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. IJCV, 60(1):63–86, 2004

[21] G. Wahba, Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation, Ann. Stat., 3 (1975), pp. 15–29.

[22] R. K. McConnell. Method of and apparatus for pattern recognition,January 1986. U.S. Patent No. 4,567,610.

[23] N. N. Cencov, Estimation of an unknown distribution density from observations, Soviet Math., 3 (1962), pp. 1559–1562.

[24] W. T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma. Computer vision for computer games. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, pages 100–105, October 1996

[25] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression. New York, NY, USA: Wiley-Interscience, 2000, vol. 354.

[26] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to Trans.Power Syst., vol. 20, no. 2, pp. 548–555, May 2005.

[27] G. Schwarz, "Estimating the dimension of a model," Ann. Stat., vol. 6, pp. 461–464, 1978.

[28] D. M. Gavrila. The visual analysis of human movement: A survey. CVIU, 73(1):82–98, 1999.

[29] D. G. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60(2):91–110, 2004.

[30] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. PAMI, 2004. Accepted

[31] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. The 8th ICCV, Vancouver, Canada, pages 454–461, 2001