Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Magenetic Tile Fault Detecion of High Voltage Electitric Machine: A consistent soft-label-based Multi-view feature selection Method

Junhang Chen[1,2] | **Zuanyuan Yang**[1,2] | **Mingyang Liu**[1,2] | **Daoyuan Li**[1,2]

[1]School of Automation, Guangdong Hong Kong Macao Joint Laboratory for Smart Discrete Manufacturing, Guangdong University of Technology, Guangdong, China

[2]School of Automation, Guangdong Key Laboratory of IoT Information Technology, Guangdong University of Technology, Guangdong, China

Correspondence
Corresponding author : Junhang Chen.
Email: junhanggdgy@aliyun.com

Present address
School of Automation, Guangdong University of Technology, Guangdong, China

Abstract

The detection of high voltage permanent magnet motors has always been a big problem due to the interference of high voltage and magnetic field on the diagnosis. Especially the magenetic tile of the motor, the failure of the magenetic tile will directly lead to the operation failure of the motor. We propose Multi-view Unsupervised Consistent Soft-label Feature Selection(MUCSFS). This method constructed consistent pseudo-labels through soft labels of clustering affinity of each view sample and constructed the model by integrating selection constraints into the mapping model. This model is used to filter the fault data set to get the feature subset, and the feature subset is used to cluster. We verify the effectiveness of the method by simulating multi-view data and through the fault clustering experiment of the magnetic tile fault data set of high voltage motor, it is confirmed that our method can effectively cluster the fault categories.

KEYWORDS
Unsupervised feature selection, multi-view learing, Fault clustering, soft-label

1 | INTRODUCTION

In high-voltage motor equipment, we need to pay close attention to the running state of the motor, which is determined by the harsh working environment and the high-precision structure of the equipment itself[1,2]. The general construction diagram of a high-voltage permanent magnet motor is shown in FIGURE. 1[†]. During its operation, the failure of the magnet will lead to low operation efficiency or even damage to the whole motor. Therefore, we need to detect the motor magnet to ensure the reliability of the motor, so fault detection technology is essential in the detection of high-voltage motor[3]. Standard fault diagnosis techniques can be divided into vibration signal analysis and image object detection. Because the fault feature information of the motor in the early stage of the fault is fragile, it is often submerged by interference signals such as environmental noise. Traditional methods based on signal analysis cannot make an accurate judgment. It takes a long time to accumulate the corresponding data of motor magnet faults and a lot of time and manpower to establish a sample database. Sometimes it is difficult to use deep learning methods for detection. The collected data will be of high dimensionality[4,5]. Therefore, it is necessary to use dimensionality reduction techniques to process data in fault detection.

Image dimensionality reduction can be divided into feature extraction and feature selection methods. Feature extraction methods use methods such as matrix decomposition to decompose data by changing the value of the data[6]. Feature selection method is to reduce the dimension by screening the features of the data and selecting the representative feature set in the data set to represent the features of the data set in the form of low rank without changing the value of the original data in the feature set[7]. Due to the diversity of data, it gives us a lot of challenges, but it also provides a new way of data processing mode. The same thing can be described in different forms by different sensors, and the data collected under such multi-form collection can be called multi-view data. Compared with single-view data, multi-view data can help us understand the object's related properties more deeply. The standard unsupervised multi-view feature selection method uses graph structure to learn the features of each

[†] Oleh Lillian D'Amore, November 16, 2022 Posting[https://warezism.blogspot.com/2022/11/23-how-does-permanent-magnet-motor-work.html]

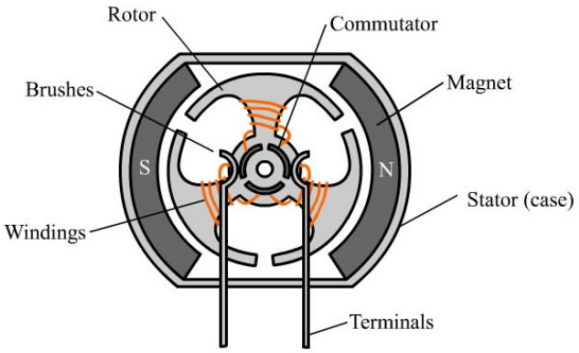


FIGURE 1 Internal construction diagram of high voltage permanent magnet motor

view, and then uses the consistency principle of multi-view to build a unified pseudo-label, and uses the pseudo-label to guide the feature selection process[8].

In unsupervised feature selection methods, there are three difficult problems. The first is how to build pseudo-label. He et al. used graph structure to guide the feature selection method[9]. Nie et al. closely combine the graph structure with the feature selection method to map each other to obtain a better low-rank representation[10]. Expanding to multi-view feature selection, Zhang et al. proposed a multi-view feature selection method based on a multistage projection of an adaptive neighbor graph model[11]. Shang et al. presented a joint sparse transform representation module and pseudo-label matrix module, using the pseudo-label matrix to guide the learning of sparse low-dimensional Spaces while obtaining the relationship between data points and mining the relationship between data, features, and pseudo-labels[12]. The second challenge is integrating pseudo-labels with the feature selection process. The original feature selection mapping model can be regarded as a special matrix factorization model. The feature selection model selects k features in the data set as the feature matrix by filtering rules and then calculates the corresponding coefficient matrix to ensure the minimum error. However, it is difficult to establish the filtering rules for selecting the model, and the selected features are only sometimes the best. Therefore, Li and Yi et al. used the linear regression model to change the nonlinear selection structure into a linear mapping structure so that the features correspond to the pseudo-labels, and it is convenient to obtain more representative features[13]. Li and Zhang et al. added an error term in the regression model and considered the uncorrelated nature of features. When the number of features required is small, differentiated but redundant features can be selected[14]. The third challenge is how to find the most appropriate subset of features. Due to the different correspondence between features and samples, we also hope to find features with more independent features for features. Hence, we need to consider the relationship between feature views. Qian et al. initially used the L2,1 norm to find the sparse features with the largest proportion as the feature subset[15]. However, the L2, 0 norm can only guarantee a certain sparsity. Therefore, Li et al. used the structural relationship of the feature graph in the original data set to preserve the feature related to the feature subset[16]. Lim et al. used mutual information constraints to distinguish the selected features and tries their best to find the features with strong independence[17]. Du et al. improve the L2,0 norm, utilizes the feature group effect, select feature groups at the same time, and replace the mode of searching for the best feature group one by one with the way of searching for the best feature group[18].

This paper uses a multi-view regression mapping model combined with soft labels related to the affinity of cluster categories to screen feature subsets with variations of the L2,0 norm. This method is Multi-view Unsupervised Consistent Soft-label Feature Selection(MUCSFS). It has the following three advantages:

First, the multi-view regression mapping model is used to add weights to each feature mapping relationship, which is conducive to maintaining the mapping relationship between the feature weight matrix and the pseudo-label and improving feature selection. The second is to obtain the same pseudo-label by using the clustering category affinity of each sample under each view and the consistency principle of multi-view. The affinity distance between each sample and the centroid of different categories is obtained, which helps to learn the relationship between samples and ensure the validity of pseudo-labels. Thirdly, the feature selection screening mode is integrated into the mapping model of feature selection, and a matrix is used to reflect the screening rules, which helps to improve the selection accuracy and select the feature subset that represents the result more strongly.

The rest of the paper is organized as follows. In Section 2, we describe in detail our research on feature selection methods. In Section 3, we introduce the experiment and analyze the results. The Section 4 is the summary of the thesis.

2 | THE METHODOLOGY

In this section, we introduce the details of the proposed method. First, we present the problem definition. Then, we formulate the overall objective function and offer an alternative optimization strategy. Finally, we discuss the time complexity of the proposed method.

2.1 | Notation

In this paper, we denote the matrices by boldface capital letters, the vectors by boldface lower-case letters, and the scalar values by italic letters. For an arbitrary matrix $\mathbf{V} \in \Re^{d\times n}$, $\mathbf{v}_{ij}$ denotes its (i,j)-th entry, $\mathbf{v}_i$ denotes its i-th row vector. $Tr(\mathbf{V})$ denotes the trace of a matrix $\mathbf{V}$ when $\mathbf{V}$ is square, and $\mathbf{V}^T$ denotes the transpose of $\mathbf{V}$. The l_2-norm of matrix $\mathbf{V}$ is defined as $\|\mathbf{V}\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$, where $\mathbf{v}_i$ represents the i-th entry of $\mathbf{v}$. $\|\mathbf{V}\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{d} v_{ij}^2} = \sqrt{Tr(VV^T)}$ denotes the Frobenius norm of $\mathbf{V}$. $\|\mathbf{V}\|_{2,1} = \sum_{i=1}^{n}\sqrt{\sum_{j=1}^{d}\mathbf{v}_{ij}^2}$ denotes $l_{2,1}$-norm of $\mathbf{V}$. $\mathbf{v} \geq 0$ or $\mathbf{V} \geq 0$ indicates all of their entries are larger than or equal to zero. $\mathbf{I}_n$ denotes the identity matrix of $n \times n$. $1_n = [1, 1, \cdots, 1]^T$, $1_n \in \Re^{n\times 1}$ is a column vector with the corresponding dimension. For clarity, TABLE 1 shows some frequently-used notations and their descriptions.

TABLE 1 Main notations used in this paper

Notation	Description
n	The number of samples
d^v	The number of features in the v-th view
$d = d^1 + d^2 + \cdots + d^v$	The sum of the number of all feature dimensions for all views
c	The number of clusters
V	The number of views
$\mathbf{X}^v \in \Re^{d^v\times n}$	Data matrix in the v-th view
$\mathbf{W} \in \Re^{d\times c}$	The feature weight projection matrix
$\mathbf{K}^v \in \Re^{d^v\times m}$	The subspace projection matrix
$\mathbf{G} \in \Re^{c\times n}$	The consistent feature pseudo-labels
$\mathbf{S}^v \in \Re^{n\times c}$	The learned affinity matrix

2.2 | Objective formulation

In the unsupervised feature selection method, the pseudo-label constructed by the structure information of the data set is needed to guide the selection. In the past unsupervised feature selection methods, we often used regression models to select features. The regression model is as follows:

$$\min \|\mathbf{W}^T\mathbf{X} + 1_n a^T - \mathbf{G}^T\|_F^2 \tag{1}$$

$\mathbf{W} \in \Re^{d\times c}$ is the feature weight matrix, $\mathbf{X} \in \Re^{d\times n}$ is the data set, $\mathbf{G} \in \Re^{c\times n}$ is the pseudo-label to guide the selection, and a is the deviation of the regression map. With the iteration of the model, $\mathbf{G}$ constantly updates the learned cluster structure and gradually records the mapping relationship between data and labels. Find the partial derivative of a on the expression, so that it is equal to 0, you can get the expression of a:

$$a = \frac{1}{n}(\mathbf{G}^T - \mathbf{W}^T\mathbf{X})1_n \tag{2}$$

Bringing this formula into the original model, we can get the new formula:

$$\begin{aligned} &\min \|(\mathbf{W}^T\mathbf{X} - \mathbf{G}^T)\mathbf{H}\|_F^2 \\ &s.t.\ \ \mathbf{H} = \mathbf{I}_n - \tfrac{1}{n}1_n 1_n^T, \mathbf{I}_n \in \Re^{n\times n}, 1_n \in \Re^{n\times 1} \end{aligned} \tag{3}$$

However, how to select features is another big problem in feature selection. Unsupervised feature selection methods often use L2,1 norm to make the $\mathbf{W}$ matrix sparse enough to score features, but indirect constraints cannot avoid feature redundancy. Therefore, the L2,0 norm is used to define the sparsity of $\mathbf{W}$:

$$\begin{aligned} &\min \|(\mathbf{W}^T\mathbf{X} - \mathbf{G}^T)\mathbf{H}\|_F^2 + \|\mathbf{W}\|_{2,0} \\ &s.t.\ \ \mathbf{H} = \mathbf{I}_n - \tfrac{1}{n}1_n 1_n^T, \|\mathbf{W}\|_{2,0} = k \end{aligned} \tag{4}$$

However, L2,0-norm is difficult to optimize sparse to only k rank, so other methods are often used to assist optimization. Due to the diversity of data, we must deal with multi-view data. Feature selection for multi-view data needs to consider the features from multi-view, so the features from all views are stacked together, which makes the optimization more difficult. Because of this situation, we can divide $\mathbf{W}$ into two matrix forms for optimization, one of which represents the matrix of feature scores of all views, and the other represents the screening matrix. The screening process and sparsity are transformed into another matrix: let $\mathbf{W} = \mathbf{UM}$. $\mathbf{M}$ represents a feature weight matrix of k dimensions, while $\mathbf{U}$ represents a screening matrix. At the same time, each view will have different structures and characteristics because of the different construction principles so that each view will have the most suitable for its pseudo-labels. When screening, we need to unify multiple pseudo-labels. Therefore, according to the above, the new multi-view data set model we obtained is:

$$\begin{aligned} &\min \textstyle\sum_v^V \|((\mathbf{UM})^T\mathbf{X} - \mathbf{F}^T)\mathbf{H}\|_F^2 + \|\mathbf{F} - \mathbf{G}^{\mathrm{v}}\|_F^2 \\ &s.t.\ \ \mathbf{H} = \mathbf{I}_n - \tfrac{1}{n}1_n 1_n^T. \end{aligned} \tag{5}$$

In the formula, $\mathbf{U} = [\mathrm{u}_1, \mathrm{u}_2, \cdots, \mathrm{u}_\mathrm{k}]$ can be represented as a matrix composed of k vector groups, and the elements of each vector are 0 or 1, and the expression is $\mathrm{u}_\mathrm{k} = [\underbrace{0...0}_{t-1}, 1, \underbrace{0...0}_{d-t}]$. Where, the value of t is related to the feature indicator vector, and the feature indicator vector is the matrix composed of the indicators of k features in the data set screened by the L2,0 norm of $\mathbf{W}$. However, obtaining accurate pseudo-labels from each view is another difficult problem, especially the feature selection problem based on surface detection targets, where various faults are very similar. Therefore, we consider using the affinity between samples from each view and each cluster category as a soft label to guide the final feature selection process. In this way, the learned soft labels are more consistent with the inherent fuzziness of data distribution and thus have stronger discriminative ability. However, since the data set may have noise interference, we use the pattern of finding the sample's affinity with each cluster category in the subspace to construct the pseudo-label. This helps to obtain the genuine sample affinity to the class attribute, then the new multi-view model is:

$$\begin{aligned} &\min \sum_{v=1}^{V}\sum_{i=1}^{n}\sum_{j=1}^{c} \mathrm{P_v}\|(\mathbf{K}^v)^T x_i^v - o_j^v\|_F^2 s_{ij}^v + \alpha\|\mathbf{G} - \mathbf{S}^v\|_F^2 + \beta\|\left((\mathbf{UV})^T\mathbf{X} - \mathbf{G}^T\right)\mathbf{H}\|_F^2 \\ &s.t. s_1^T 1 = 1, 0 \le \mathrm{s_{ij}} \le 1, \mathbf{H} = \mathbf{I_n} - \tfrac{1}{\mathrm{n}}1_\mathrm{n}1_\mathrm{n}^\mathrm{T}, \mathrm{K^v(K^v)^T} = \mathbf{I}_{\mathrm{d_v}}. \end{aligned} \tag{6}$$

In this model, $\mathbf{K}^v \in \Re^{d^v\times m}$ represents the subspace weight matrix. In order to reduce the noise interference in the data set, we convert the data to cluster in the subspace. At the same time, not every view is the most important, so we introduce the weight matrix $\mathrm{P_v}$ to unify the soft label matrix obtained from different views into a consistent pseudo-label matrix $\mathbf{G}$, and use $\mathbf{G}$ to guide feature selection.

2.3 | Optimization

$f(\mathbf{U}, \mathbf{M}, \mathbf{G}, \mathbf{K}^v, \mathbf{S}^v, \mathbf{O}_j, \mathrm{P}_v)$ is a function of six variables with respect to MUCSFS. The objective function is non-convex due to the presence of multiple variables. It isn't easy to obtain the optimal global solution. However, the objective function is convex to a single variable. The best solution is alternating optimization over a single variable. This paper uses an alternating iterative optimization method suitable for multivariate optimization to optimize the Objective Eq. (6).

1) 1) Solving $\mathbf{M}$ with fixed variables $\mathbf{U}, \mathbf{G}, \mathbf{K}^v, \mathbf{S}^v, \mathbf{O}_j, \mathrm{P}_v$. Take the function about the $\mathbf{M}$ part and transform it, then there is:

$$\begin{aligned} &\min_{\mathbf{M}} Tr[\left(\mathbf{G}^T - \mathbf{M}^T\mathbf{U}^T\mathbf{X}\right)^T\mathbf{H}\left(\mathbf{G}^T - \mathbf{M}^T\mathbf{U}^T\mathbf{X}\right)] \\ &= \min_{\mathbf{M}} Tr(\mathbf{GHF} + \mathbf{M}^T\mathbf{U}^T\mathbf{XHX}^T\mathbf{UM} - \mathbf{GHX}^T\mathbf{UM} - \mathbf{M}^T\mathbf{U}^T\mathbf{XHF}) \end{aligned} \tag{7}$$

Then, according to the above formula, $\mathbf{M}$ is derived, and the solution formula for $\mathbf{M}$ is finally obtained as follows:

$$\mathbf{M} = (\mathbf{U}^T\mathbf{XHX}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{XHG}^T \tag{8}$$

2) Solving $\mathbf{U}$ with fixed variables $\mathbf{M}, \mathbf{G}, \mathbf{K}^v, \mathbf{S}^v, \mathbf{O}_j, \mathrm{P}_v$. Similar to solving $\mathbf{M}$, we can also obtain the solution formula for $\mathbf{U}$ as follows:

$$\min_{\mathbf{U}} Tr(-\mathbf{GHX}^T\mathbf{U}(\mathbf{U}^T\mathbf{XHX}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{XHF} + \mathbf{GHG}^T) \tag{9}$$

Since $\mathbf{GHG}^T$ is a fixed value, we remove this value and reverse the sign to get a new solution matrix:

$$\max_{\mathbf{U}} Tr((\mathbf{U}^T\mathbf{XHX}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{XHG}^T\mathbf{GHX}^T\mathbf{U}) \tag{10}$$

For ease of solution, let's set $\mathbf{A} = \mathbf{XHX}^T$ and $\mathbf{B} = \mathbf{XHFF}^T\mathbf{HX}^T$, then the formula becomes:

$$\max_{\mathbf{U}} Tr((\mathbf{U}^T\mathbf{AU})^{-1}\mathbf{U}^T\mathbf{BU}) \tag{11}$$

After constructing such a model, we can follow the solution method of problem (7) in [19]. According to the solution of this problem, we can choose the maximum k vectors and their subscript vector group Q. Combined with the above methods, we can derive the composition of the selection matrix $\mathbf{U}$:

$$u_{ij} = \begin{cases} 1, if j \in \mathrm{Q} \\ 0, otherwise \end{cases} \quad j = 1, 2, \cdots, d; i = 1, 2, \cdots, k \tag{12}$$

3) Solving $\mathbf{O}_j$ with fixed variables $\mathbf{M}, \mathbf{G}, \mathbf{K}^v, \mathbf{S}^v, \mathbf{U}, \mathrm{P}_v$. According to the solution formula of the cluster centroid, we can obtain:

$$\mathbf{O}_j^v = \frac{\sum_{i=1}^n s_{ij}^v(\mathbf{K}^v)^T x_i^v}{\sum_{i=1}^n s_{ij}^v} \tag{13}$$

4) Solving $\mathbf{K}^v$ with fixed variables $\mathbf{M}, \mathbf{G}, \mathbf{O}_j, \mathbf{S}^v, \mathbf{U}, \mathrm{P}_v$. According to [20], we can know the solution model of $\mathbf{K}^v$as follows:

$$\begin{cases} \min_{\mathbf{K}^v} Tr((\mathbf{K}^v)^T\mathbf{Q}^v\mathbf{K}^v) \\ \mathbf{Q}^v = \mathbf{X}^v(\mathbf{L}^v - \mathbf{S}^v\mathbf{D}(\mathbf{S}^v)^T)(\mathbf{X}^v)^T \end{cases} \tag{14}$$

Where we have $\boldsymbol{L}^v \in \Re^{n\times n}$ and $\boldsymbol{D}^v \in \Re^{c\times c}$ are diagonal matrices, defined as $\boldsymbol{L}_{ii}^v = \sum_{j=1}^c s_{ij}^v$ and $\boldsymbol{D}_{jj}^v = diag(\frac{1}{\sum_{i=1}^n s_{ij}^v})$, and in this way, we can solve for K^v.

5) Solving $\mathbf{S}^v$ with fixed variables $\mathbf{M}, \mathbf{G}, \mathbf{O}_j, \mathbf{K}^v, \mathbf{U}, \mathrm{P}_v$. According to the solution formula related to $\mathbf{S}^v$ and the solution method of graph in [10], let $z_{ij}^v = \|(\boldsymbol{K}^v)^T x_i^v - o_j^v\|_2^2$, we can get the solution formula as follows:

$$\min_{s_i^v} \|s_i^v + \frac{z_i^v}{2\beta}\|_2^2 + \alpha\|\mathbf{G}_i - s_i^v\|_2^2 \tag{15}$$

6) Solving $\mathbf{G}$ with fixed variables $\mathbf{M}, \mathbf{S}^v, \mathbf{O}_j, \mathbf{K}^v, \mathbf{U}, \mathrm{P}_v$. According to the solution formula related to $\mathbf{G}$, the solution formula of the consistency matrix G can be obtained as follows:

$$\mathbf{G} = (\alpha\mathbf{I} + \beta\mathbf{H})^{-1}(\alpha\sum_{v=1}^V \mathrm{P_v}\mathbf{S}^v + \beta\mathbf{M}^T\mathbf{U}^T\mathbf{XH}) \tag{16}$$

7) Solving P_v with fixed variables $\mathbf{M}, \mathbf{S}^v, \mathbf{O}_j, \mathbf{K}^v, \mathbf{U}, \mathbf{G}$. To distinguish the weight of each view according to the relevant similarity component of each view, we set: $N_v = \sum_{i=1}^{n}\sum_{j=1}^{c} \|(\mathbf{K}^v)^T x_i^v - \mathbf{O}_j^v\|_F^2 s_{ij}^v$ then:

$$\begin{aligned} &\min \sum_{v=1}^{V} P_v^t N_v \\ &s.t. \sum_{v=1}^{V} P_v = 1, 0 \leq P_v \leq 1 \end{aligned} \tag{17}$$

The weight calculation formula can be obtained as follows:

$$P_v = \frac{(\frac{1}{N_v})}{\sum_{v=1}^{V}(\frac{1}{N_v})} \tag{18}$$

After synthesizing the calculation results of the above parameters, we use a table to represent the specific iterative process of our algorithm, as shown in the Algorithm 1:

Algorithm 1 Alternative optimization to solve the model (6)

Input: Multi-view feature matrix $\mathbf{X}$, the dimension of the feature subspace m; Number of selected features k; the number of iterations t; the number of clusters c and the parameters α, β, γ and η.
Output: k features for the data
1. Initialize the various parameters
2. **repeat**
3. Update $\mathbf{M}$ with Eq. (8).
4. Update $\mathbf{U}$ with Eq. (12).
5. Update $\mathbf{O}_j$ with Eq. (13).
6. Update $\mathbf{K}^v$ with Eq. (14).
7. Update $\mathbf{S}^v$ with Eq. (15).
8. Update $\mathbf{G}$ with Eq. (16).
9. Update P_v with Eq. (18).
10. **until Convergence**
11. We use **W=UM** to compute the feature weight matrix $\mathbf{W}$. Use the rank size of $\mathbf{W}$ to find the k largest Ordinal Numbers, and use the Ordinal Numbers to find the k features in the dataset.

2.4 | Complexity analysis

According to this algorithm, the computational complexity of the proposed MUCSFS method is analyzed. According to the setting, n represents the total number of data samples, V is the number of views, d^v represents the number of features in a single view, the number of features in all view datasets can be represented by $d = d^1 + d^2 + \cdots + d^v$, c is the number of potential clusters, r is the dimension of the projected subspace, and t is the total number of iterations. Update the matrix $\mathbf{U}$ by solving MUCSFS. The complexity of updating $\mathbf{U}$ is given by $O(k)$. Update the projection matrix $\mathbf{M}$ by solving MUCSFS. The complexity of updating $\mathbf{M}$ is given by $O(k^2d + k^2m + kdn + kmn)$. Update the $\mathbf{O}_j$ by solving MUCSFS. The complexity of updating $\mathbf{O}_j$ is given by $O(vnc)$. Update the $\mathbf{K}^v$ by solving MUCSFS. The complexity of updating $\mathbf{K}^v$ is given by $O(vdm)$. Update the $\mathbf{S}^v$ by (15). The complexity of updating $\mathbf{S}^v$ is given by $O(vnc)$. Update the matrix $\mathbf{G}$ by (16). The complexity of updating $\mathbf{G}$ is given by $O(nc)$. Update the P_v by solving MUCSFS. The complexity of updating P_v is given by $O(v)$. Since this is a multi-view operation, there is $O(d^3t)$, and the total complexity of algorithm. MUCSFS is combined with the complexity results of the above iteration parameters.

3 | EXPERIMENT

The experiment is divided into two parts, one is a simulation experiment on a data set similar to the surface detection data set, and the other is a fault detection experiment using the magnetic tile surface picture data set in a high-voltage motor.

3.1 | simulation experiment

In order to prove the reliability of our method, we will introduce two data sets to prove the superiority of our method, namely MSRC-v1 and Handwritten data set. MSRC-v1[21] † is by trees, buildings, plane, car, bicycle, cattle, face of seven class, There are 30 images for each refinement class. Handwritten[22]‡ is handwritten digital data set multiple points of view. the two data sets are based on shape feature based data set, It is suitable for testing analog surface fault data sets. At the same time, we also introduce five methods to compare with MUCSFS to illustrate the optimization of this method. The four methods are ASVW, ASCL, NSGL and MRUFS. We present the characteristics of the data set in Table 2.

TABLE 2 Statistics of two datasets

View	MSRC-v1	Handwritten
1	CM(24)	FCCS(76)
2	CF(254)	PC(216)
3	GIST(512)	KLC(64)
4	LBP(256)	PA(240)
5	-	ZM(76)
6	-	Morp(6)
#Size	210	2000
#Classes	7	20

We will describe and explain the comparison method to ensure the experiment's reliability. Adaptive Similarity and View Weight (ASVW): This approach uses a learning mechanism to adaptively represent common structures through an objective function defined by a standard graph Laplace operator across different views, and sparse L2,p norm constraints designed for feature selection[23]. Adaptive Collaborative Similarity Learning (ACSL): By considering the final feature selection performance, this method adaptively learns collaborative similarity structure and similarity combination weights[24]. Non-negative Structured Graph Learning (NGSL): Adaptive similarity graph learning and feature selection learn ideal graphs by imposing reasonable rank constraints, effectively capturing the relationship between multi-view samples[25]. MRUFS(Multi-level regularization based unsupervised multi-view feature selection): This method adaptively learns a suitable similarity matrix in the dimensionality reduction feature space based on the learned projection matrix[26]. In order to reduce the redundancy and noise information in multi-view data, the intellectual projection matrix is constrained by using the structural sparsity, dependence, and diversity information of multi-view data.

To prove the effectiveness of the experimental results, we use two evaluation indicators to verify and illustrate the experimental results, which are clustering accuracy (ACC) and normalized mutual information (NMI)[27,28]. ACC represents the proportion of the sample correctly predicted by the algorithm, resulting in accurate data. The calculation process is as follows:

$$ACC = \frac{1}{n}\sum_{i=1}^{n}\delta(v_i, map(t_i)) \tag{19}$$

In the equation, n represents the number of samples, v_i represents the predicted label of the ith sample, t_i represents the ith true label, and $map(t_i)$ represents the best mapping function of the real label cluster.

NMI represents the quality of cluster quality, the calculation process is as follows

$$NMI = \frac{MI(C, L)}{\max(H(C),\ H(L))} \tag{20}$$

† https://github.com/goodbai-sdnu/NSGL_code
‡ https://archive.ics.uci.edu/ml/datasets/Multiple+Feature

After describing the experimental data, we adopted matlab2018 under the win10 system as the experimental operating environment. Since the method needs parameter setting, we give the range of parameters here, and most parameter setting values are $[10^{-4}, 10^{4}]$. For the specific parameters in each method, m in the MUCSFS method is generally in the range of [5 100]. In contrast, the nearest neighbor search ranges k of the method with graph structure is set to 5. The optimal number of features will vary from one method to another for different datasets, so we use an increasing sequence of values. The number of features in MSRC-v1 datasets and Handwritten datasets will be [100,200,300,400,500] and [250,300,350,400,450], respectively. The results obtained by our experiment are shown in Table 3 to Table 6:

TABLE 3 ACC of all approaches on MSRC-v1 datasets (%).

Datasets	FeaDim	ASVW	ACSL	NSGL	MRUFS	MUCSFS
MSRC-v1	100	41.77	30.00	39.76	45.01	**58.09**
	200	46.10	31.24	44.29	50.52	**59.38**
	300	48.05	31.24	50.95	57.16	**63.00**
	400	50.01	32.19	52.76	53.00	**63.23**
	500	56.38	34.00	57.95	55.46	**65.95**

TABLE 4 NMI of all approaches on MSRC-v1 datasets (%).

Datasets	FeaDim	ASVW	ACSL	NSGL	MRUFS	MUCSFS
MSRC-v1	100	29.35	16.35	25.57	32.37	**49.08**
	200	31.30	18.75	30.46	36.74	**50.89**
	300	31.44	19.12	38.43	46.43	**53.12**
	400	40.55	19.05	41.25	42.26	**53.85**
	500	50.21	21.46	49.99	44.12	**57.35**

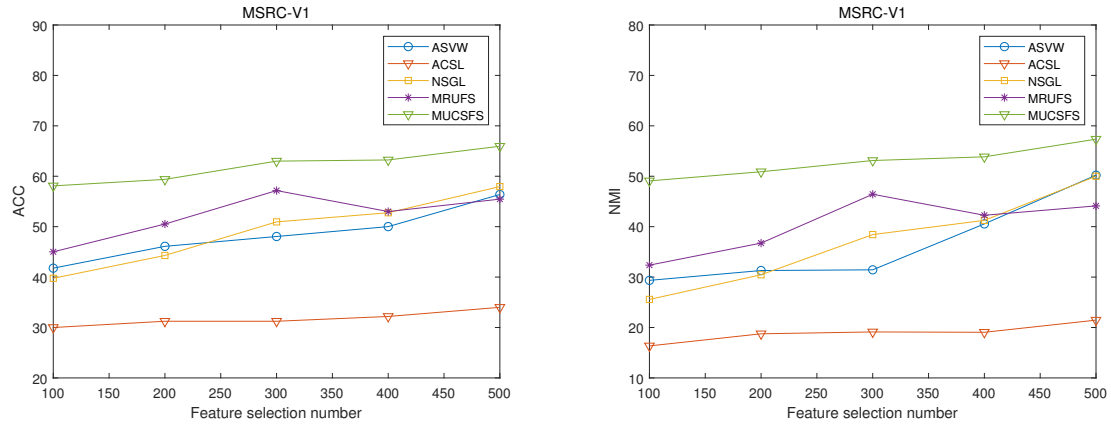
TABLE 5 ACC of all approaches on Handwritten datasets (%).

Datasets	FeaDim	ASVW	ACSL	NSGL	MRUFS	MUCSFS
Handwritten	250	70.01	76.30	67.28	75.24	**82.00**
	300	71.62	77.40	76.59	73.25	**81.60**
	350	63.27	45.10	77.04	68.93	**82.30**
	400	64.73	43.30	61.19	70.33	**81.30**
	450	63.03	51.90	65.31	73.51	**76.65**

TABLE 6 NMI of all approaches on Handwritten datasets (%).

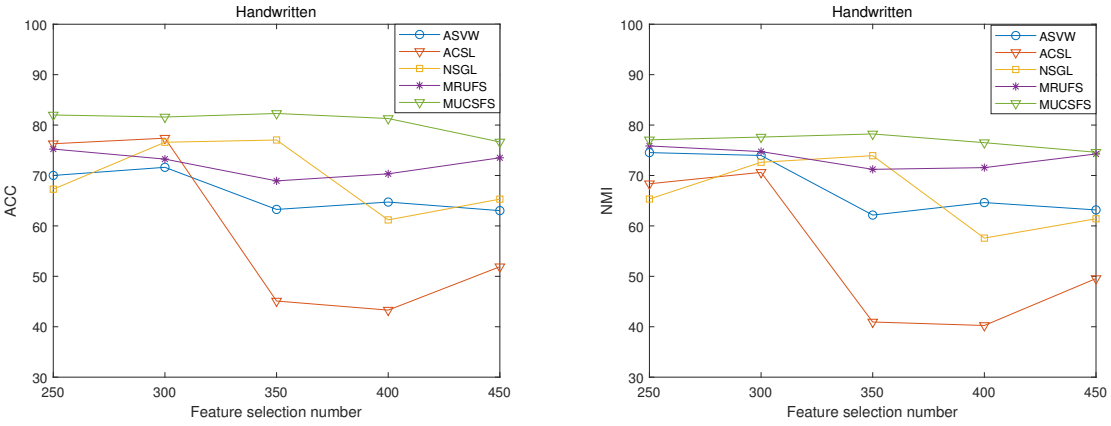
Datasets	FeaDim	ASVW	ACSL	NSGL	MRUFS	MUCSFS
Handwritten	250	74.54	68.36	65.33	75.86	**77.07**
	300	73.97	70.62	72.62	74.74	**77.62**
	350	62.15	37.94	73.93	71.22	**78.24**
	400	64.63	37.25	57.58	71.56	**76.51**
	450	63.17	49.56	61.42	74.27	**74.61**

In Table 3 to Table 6, the best are shown in bold, and the next best are underlined. As can be seen from the experimental results table, the results obtained by our method are better on average than those obtained by other methods in the MSRC-v1 and Handwritten data sets. However, the two results in the Handwritten data sets could be better. This is because the number of optimal features selected by each feature selection method differs, but the overall trend is gradually increasing and decreasing.



(a) (b)

FIGURE 2 On the MSRC-v1 datasets, ACC and NMI fluctuation status of all methods as the number of features changes.



(a) (b)

FIGURE 3 On the Handwritten datasets, ACC and NMI fluctuation status of all methods as the number of features changes.

For more intuitive reflection, we enumerate a decomposition result graph that changes with the number of features, as shown in the FIGURE 2 and FIGURE 3:

3.2 | Fault clustering experiment on magnetic tile surface of high voltage motor

We use the magnetic tile surface image data collected by the Chinese Academy of Sciences for fault clustering[29]. Magnetic tile is a kind of tile magnet mainly used in permanent magnet motors, which can make the simple motor structure easy maintenance, lightweight, small volume, reliable use, less copper, low copper consumption, low energy consumption, and other advantages. However, the actual use of the process due to high voltage caused by a high magnetic environment will make the adoption of the target detection results not accurate, and because of the diversity of magnetic tile faults, it is difficult to distinguish the naked eye. The fault types of magnetic tiles can be divided into Blowhole, Break, Crack, Fray, and Uneven. The specific fault picture is shown in FIGure 4:

However, the original data set image will have corresponding interference during the acquisition process, such as light and electromagnetic field interference. Therefore, we use the LBP features and HOG features of the original image as the second and third views. LBP(Local Binary Pattern) is an operation used to describe image local features. LBP features have significant advantages, such as gray invariance and rotation invariance. It was proposed by T. Ojala, M. Pitikainen, and D. Harwood in

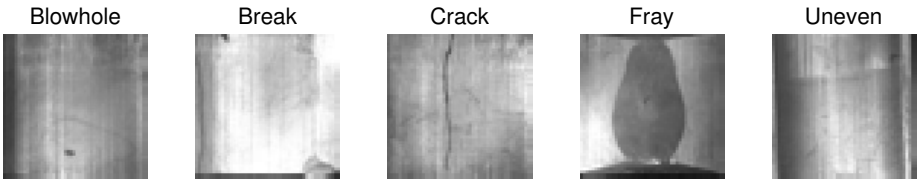


FIGURE 4 In the figure, We show pictures of each class of faults in the Magnet tile fault dataset

1994[30]. LBP feature has been widely used in many fields of computer vision due to its simple computation and good effect. The LBP feature operator can effectively reflect the feature texture of the magnetic tile fault image, and the feature vector group is used to represent the feature of the image. It constructs features by calculating and counting the gradient direction histogram of the local area of the image. HOG(Histogram of Oriented Gradient) is operated on the local grid unit of the image so that it can maintain a good invariance to the geometric and optical deformation of the image[31].

After determining the contents of each perspective, we randomly selected 15 images from the five types of fault feature images, totaling 75 images, to construct the magnetic vado perspective data set. The magnetic vado perspective data set we obtained is shown in the Table 7:

TABLE 7 Magnet tile fault dataset

View	Magnet tile fault
1	Gray(2500)
2	LBP(90)
3	HOG (900)
#Size	75
#Classes	5

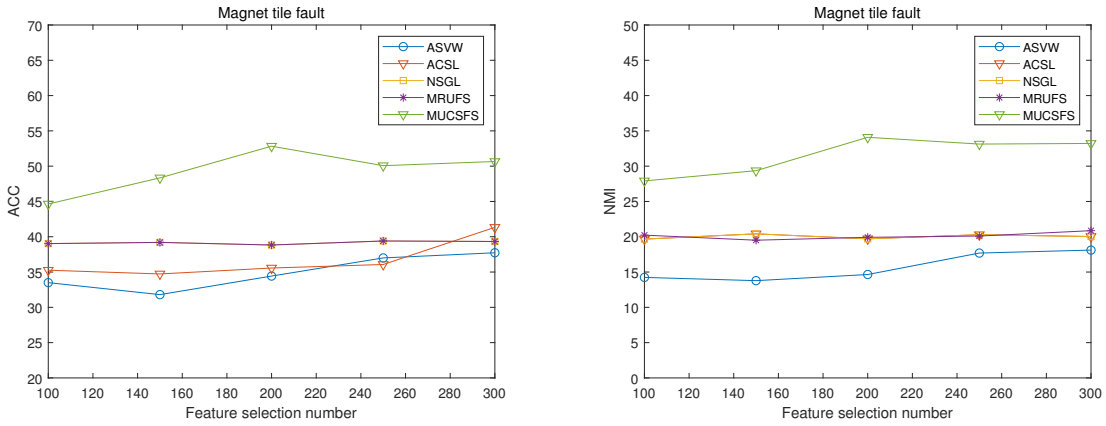
After processing the magnetic tile data set, we decomposed it. In order to further prove the effectiveness of the proposed method in the fault clustering experiment, the method of simulation experiment in the previous paper is also introduced in the experiment, and the results of all methods are compared. The parameter setting is the same as that of the simulation experiment above, and the number of feature selections was set as [100,150,200,250,300]. The experimental results are shown in Table 8 and Table 9:

TABLE 8 ACC of all approaches on four datasets (%).

Datasets	FeaDim	ASVW	ACSL	NSGL	MRUFS	MUCSFS
Magnet tile fault	100	33.50	35.26	39.03	39.04	**44.64**
	150	31.80	34.73	39.10	39.20	**48.34**
	200	34.43	35.60	40.13	38.83	**52.83**
	250	37.00	36.07	38.93	39.40	**50.07**
	300	37.73	41.33	39.56	39.33	**50.66**

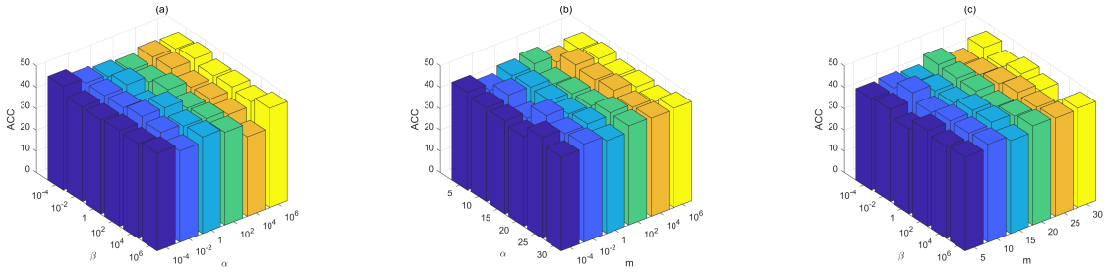
TABLE 9 NMI of all approaches on four datasets (%).

Datasets	FeaDim	ASVW	ACSL	NSGL	MRUFS	MUCSFS
Magnet tile fault	100	14.24	14.46	19.67	20.23	**27.91**
	150	13.78	14.81	20.41	19.51	**29.36**
	200	14.65	15.48	19.71	19.92	**34.09**
	250	17.69	16.62	20.29	20.10	**33.13**
	300	18.11	22.45	20.03	20.86	**33.22**



(a) (b)

FIGURE 5 On the magnet tile fault datasets, ACC and NMI fluctuation status of all methods as the number of features changes.



(a) (b) (c)

FIGURE 6 The parameter variation of this method on the the magnet tile fault datasets, each graph shows the influence between two parameters, and the remaining parameters are fixed.

In these tables, the result of our method is the best, and it has a good effect on the clustering accuracy of faults. Our method adopts the L2,0 norm and obtains sufficiently sparse features compared to other models. At the same time, the model has the bias structure and the subspace soft label to avoid noise interference in the image data set of the magnetic tile fault. These help distinguish different fault pictures of magnetic tiles. To further demonstrate the effectiveness of our method, we use a graph to show the trend of all methods in the number of feature selections.

3.3 | Parameter analysis

The model is mainly affected by three parameters, α, β, and m. To explain the influence between α and β, considering the close relationship between the value of m and the characteristics of each view, the influence of multi-view will be more obvious. Therefore, in order to intuitively explain the impact of parameters on the result. We take the Magnet tile fault data set as an example to illustrate the relationship between the three parameters, and the results are shown as FIGURE.6:

We compare the three parameters separately and fix the other parameters when compare the two parameters. For α and β, the fixed value is 10^4, and for the feature dimension of the feature subspace the fixed value is 30. The number of feature selections taken is fixed to 250. Then, as shown in the FIGURE. 6, (a), (b) and (c) show the model accuracy changes under the combination of α and β, α and m, and β and m respectively. It can be seen that the parameters α and β fluctuate in the range $[10^{-4}, 10^6]$ and have little impact on the accuracy, while the m value only has a large increase in a fixed range, and there is little difference in other ranges. There is no direct interaction between the parameters.

4 | CONCLUSIONS

This paper proposes a multi-view feature selection method called MUCSFS, which constructs consistent pseudo-labels by soft labels of sample cluster affinity in each view, integrates the selection constraints into the mapping model, reduces the screening steps, and finally obtains the feature subset. This method can be used for target surface detection. We verify the effectiveness of this method by using handwritten and MSRC-v1 data sets. Then, through the fault clustering experiment of the magnetic tile fault data set of high voltage motors, it is confirmed that our method can effectively enumerate the fault categories. In the future, we will study to improve the clustering accuracy of fault categories.

ACKNOWLEDGMENTS

This work was supported by the Guangdong Basic and Applied Basic Research Foundation, grant number 2022A1515010688.

AUTHOR CONTRIBUTIONS

All authors have participated in the design and approval of the final manuscript version.

FINANCIAL DISCLOSURE

There are no financial conflicts of interest to disclose.

CONFLICT OF INTEREST

All authors declares that he/she has no conflict of interest.

REFERENCES

1. Abbasi AR. Fault detection and diagnosis in power transformers: a comprehensive review and classification of publications and methods. *Electric Power Systems Research.* 2022;209:107990.
2. Yang W, Zimroz R, Papaelias M. Advances in Machine Condition Monitoring and Fault Diagnosis. *Electronics.* 2022;11(10):1563.
3. Hoang DT, Kang HJ. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research.* 2019;53:42–50.
4. Kumar PH, Lakhimsetty S, Somasekhar VT. A low-cost fault-tolerant permanent magnet brush-less direct current motor drive for low-power electric vehicle applications. *International Journal of Circuit Theory and Applications.* 2023.
5. Hsu HY, Lo TY, Hsieh MF. A simple signal extraction-based online real-time diagnosis approach for interturn short-circuit fault of permanent magnet motor. *International Journal of Circuit Theory and Applications.* 2023.
6. Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends.* 2020;1(2):56–70.
7. Alelyani S, Tang J, Liu H. Feature selection for clustering: a review.. *Data clustering: algorithms and applications.* 2013;29(1).
8. Zhang R, Nie F, Li X, Wei X. Feature selection with multi-view data: A survey. *Information Fusion.* 2019;50:158–167.
9. He X, Cai D, Niyogi P. Laplacian score for feature selection. *Advances in neural information processing systems.* 2005;pp:507–514.
10. Nie F, Zhu W, Li X. Unsupervised feature selection with structured graph optimization. *Proceedings of the AAAI conference on artificial intelligence.* 2016;30(1):1302–1308.
11. Zhang H, Wu D, Nie F, Wang R, Li X. Multilevel projections with adaptive neighbor graph for unsupervised multi-view feature selection. *Information Fusion.* 2021;70:129–140.
12. Shang R, Zhang X, Feng J, Li Y, Jiao L. Sparse and low-dimensional representation with maximum entropy adaptive graph for feature selection. *Neurocomputing.* 2022;485:57–73.
13. Li Z, Yang Y, Liu J, Zhou X, Lu H. Unsupervised feature selection using nonnegative spectral analysis. *Proceedings of the AAAI conference on artificial intelligence.* 2012;26(1):1026–1032.
14. Li X, Zhang H, Zhang R, Nie F. Discriminative and uncorrelated feature selection with constrained spectral analysis in unsupervised learning. *IEEE Transactions on Image Processing.* 2019;29:2139–2149.
15. Qian M, Zhai C. Robust unsupervised feature selection. *Twenty-third international joint conference on artificial intelligence.* 2013:1621–1627.
16. Li S, Tang C, Liu X, Liu Y, Chen J. Dual graph regularized compact feature representation for unsupervised feature selection. *Neurocomputing.* 2019;331:77–96.
17. Lim H, Kim DW. Pairwise dependence-based unsupervised feature selection. *Pattern Recognition.* 2021;111:107663.
18. Du X, Nie F, Wang W, Yang Y, Zhou X. Exploiting Combination Effect for Unsupervised Feature Selection by L2,0-Norm Norm. *IEEE transactions on neural networks and learning systems.* 2018;30(1):201–214.
19. Nie F, Dong X, Tian L, Wang R, Li X. Unsupervised Feature Selection With Constrained L2,0-Norm and Optimized Graph. *IEEE transactions on neural networks and learning systems.* 2020;33(4):1702–1713.
20. Wang F, Zhu L, Li J, Chen H, Zhang H. Unsupervised soft-label feature selection. *Knowledge-Based Systems.* 2021;219:106847.
21. Winn J, Jojic N. Locus: Learning object classes with unsupervised segmentation. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1.* 2005;1:756–763.
22. Breukelen vM, Duin RP, Tax DM, Den Hartog J. Handwritten digit recognition by combined classifiers. *Kybernetika.* 1998;34(4):381–386.
23. Hou C, Nie F, Tao H, Yi D. Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Transactions on Knowledge and Data Engineering.* 2017;29(9):1998–2011.
24. Wang Z, Feng Y, Qi T, Yang X, Zhang JJ. Adaptive multi-view feature selection for human motion retrieval. *Signal Processing.* 2016;120:691–701.

25. Bai X, Zhu L, Liang C, Li J, Nie X, Chang X. Multi-view feature selection via nonnegative structured graph learning. *Neurocomputing.* 2020;387:110–122.
26. Chen T, Zeng Y, Yuan H, Zhong G, Lai LL, Tang YY. Multi-level regularization-based unsupervised multi-view feature selection with adaptive graph learning. *International Journal of Machine Learning and Cybernetics.* 2023;14(5):1695–1709.
27. Knops ZF, Maintz JA, Viergever MA, Pluim JP. Normalized mutual information based registration using k-means clustering and shading correction. *Medical image analysis.* 2006;10(3):432–439.
28. Zhou X, Wang X, Dougherty ER, Russ D, Suh E. Gene clustering based on clusterwide mutual information. *Journal of Computational Biology.* 2004;11(1):147–161.
29. Huang Y, Qiu C, Yuan K. Surface defect saliency of magnetic tile. *The Visual Computer.* 2020;36:85–96.
30. Hsiang SM, Brogmus GE, Courtney TK. Low back pain (LBP) and lifting technique—a review. *International Journal of Industrial Ergonomics.* 1997;19(1):59–74.
31. Pang Y, Yuan Y, Li X, Pan J. Efficient HOG human detection. *Signal processing.* 2011;91(4):773–781.