JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 27, 227-242 (2011)

Characterization of WOSF by Equivalent Classes through Support Vector Machine

WEI-CHIH CHEN AND JYH-HORNG JENG*
Department of Information Engineering
I-Shou University
Kaohsiung, 804 Taiwan

It is known that a weighted order statistic filter (WOSF) generates a linearly separable Boolean function and 2 different WOSF may generate the same Boolean function. Therefore a natural question "how to characterize WOSF which correspond the same Boolean function" arises. In this paper, we propose a different representation of WOSF induced from SVM. Also, we construct equivalent classes for WOSF based on the maximal margin classification of SVM. Two types of equivalent classes are proposed. The first one is called BF equivalent class. The parameters representing the hyperplane are adopted as the representative of the class, which is unique. The second class is the global equivalent class which is derived by additional sign change and permutation on the components of the BF class representatives. Therefore we can efficiently characterize all of the WOSF through only few representatives of equivalent classes and save computation cost when searching for various WOSF. Finally, we provide 3 formulas to directly generate the corresponding outputs of each WOSF.

Keywords: linearly separable Boolean function, maximal margin classifier, hyperplane, WOSF, SVM, equivalent class

1. INTRODUCTION

Weighted order statistic filters (WOSF) make use of statistical and deterministic properties and have good performance for edge preservation and noise suppression in the field of image processing. Typical applications include noise cancellation, image reconstruction, edge enhancement, and texture analysis in [1, 2]. WOSF have a large of variations according to the practical implementation. Some commonly used technique such as median filters, weighted median filters (WMF), and order statistic filters (OSF) are special cases of WOSF [3].

Some scholars have presented several papers about the representation of WOSF. Koivisto, and Huttunen [4] designed WOS filters by the characterization of training-based optimization, Ropert, Saint-Martin, and Pele [5] represented WOS filters based on the statistical analysis of WOS filters, Savin, Ahmad, and Swamy [6] used linearly separable stack-like architecture to design WOS filters, and Hirata and Hirata Jr. [7] designed WOS filters from the examples applied on image. Marshall [8] directly designed weighted order statistic filters by threshold decomposition. Lukac *et al*. [9] generalized selection of weighted vector filters for multichannel signal processing. Yu and Liao [10] described the classification and convergence properties of WOSF, and proposed a conversion algorithm of positive Boolean function. Arce, Hall, and Barner [11] used latticed concept to propose a permutation of weighted order statistic filter. These representations of WOSF are based on the pair of weight vector and threshold value.

Received January 19, 2009; revised February 23, 2009; accepted March 30, 2009.
Communicated by H. Y. Mark Liao.
* Corresponding author.

Support vector machine (SVM) is a learning machine which was first introduced by [12]. A unique feature of SVM is that the discriminant function for classification problems or the predictive function for regression problems can be expanded on a small subset of training data as support vector [13]. The SVM theory has the characterization of maximal margin classification, so it is a well suited method for studying linearly separable Boolean functions. Lin, *et al*. [14] made the use of SVM to implement linearly separable Boolean functions. Yao and Yu [15] used dichotomous approach to design weighted order statistic filters by SVM. Chen and Jeng [16] deduced discriminant function of SVM as linearly separable Boolean function to represent WOS filters.

In this paper, we utilize SVM technique to represent WOSF. This method is based on the property of maximal margin classification of SVM. From a truth table generated by linearly separable Boolean function, we can construct a training set. By SVM training, we generate a maximal margin hyperplane which is formulated by an optimal normal vector and optimal bias. The hyperplane defines a discriminant function and this function has the same outputs as those of the WOSF. In other words, we only use a normal vector and bias to represent a lot of WOSF. Compared to the other representation of WOSF, the proposed method is obviously efficient and inexpensive.

Also, we propose an alternative method to characterize WOSF. The characterization consists of two stages of equivalent classes. The first class, referred to as BF class. The BF class is constructed by many WOSF with same outputs. From these WOSF, we can find one that the weight vector is just the normal vector and the threshold is just the bias. Because the normal vector and bias is unique, therefore they are used as the unique representative of BF equivalent class. The second class, referred to as global class. It is composed by sign change and permutation on the components of representative of BF class. The underlying idea is the sign change and permutation operations will form another pair of optimal parameters. This pair of parameters formulates a maximal margin hyperplane separating another linearly separable Boolean function. Based on this concept, we can rearrange a given BF representative such that all the components of the parameters are positive and decreasing component-wise. And this new BF representative is the unique representative of the global class.

This paper is organized as eight sections. The section 1 is an introduction. The WOSF and linearly separable Boolean function is described in section 2. In section 3, the SVM technique is induced to represent the WOSF. In section 4, the BF equivalent class is proposed to represent all WOSF. In section 5, we construct a global equivalent class by sign change and permutation. In section 6, we propose 3 translated formulas to generate the corresponding outputs. In section 7, we provide an experiment and illustration to compare the other method. Finally, a conclusion is given in section 8.

2. WOSF AND LINEARLY SEPARABLE BOOLEAN FUNCTION

Let $x \in \{0, 1\}^n$ be an n-dimension vector. The Hamming distance of x, denoted by $|x|$, is the number of components of x having value 1. Given an integer r, $0 \le r \le n + 1$, we define the function $F_r(x)$: $\{0, 1\}^n \to \{0, 1\}$ by

$$F_r(x) = \begin{cases} 1, & \text{if } |x| \ge r \\ 0, & \text{else} \end{cases}. \tag{1}$$

Such a function $F_r(x)$ is called an order statistic filter (OSF) and the parameter r is referred to as the order of the filter. An OSF $F_r(x)$ can be extended to a WOSF, a weighted OSF $F_{\Omega,r}(x)$. Given the weight vector $\Omega = (\omega_1, \omega_2, \ldots, \omega_n)$ with $\omega_i \in R$, let n' be the sum of the weights ω_i, *i.e.*, $n' = \lceil \sum \omega_i \rceil$. Given the integer r, $0 \le r \le n' + 1$, the function $F_{\Omega,r}(x)$ can be defined as

$$F_{\Omega,\,r}(x) = \begin{cases} 1, & \text{if } <\Omega,\ x> \ \ge r \\ 0, & \text{else} \end{cases}. \tag{2}$$

Here the parameter r is referred to as the threshold of the filter. For each fixed r in Eq. (2), varying the vector x in $\{0, 1\}^n$, the function $F_{\Omega,r}(x)$ will produce a sequence of outputs in $\{0, 1\}$. Altering r, the function will produce another sequence of outputs. Note that, when the weight vector is $\Omega = (1, 1, \ldots, 1)$, the WOSF is just an OSF. A Boolean function $B(x)$ is a function having inputs $x \in \{0, 1\}^n$ and outputs $y_i \in \{0, 1\}$. Each Boolean function can generate a truth table. A Boolean function is linearly separable if there exists a vector $\alpha = (\alpha_1, \ldots, \alpha_n) \in R^n$ and a threshold value $\theta \in R$, such that

$$B(x) = \begin{cases} 1, & \text{if } <\alpha,\ x> \ \ge \theta \\ 0, & \text{else} \end{cases}. \tag{3}$$

Compared to Eq. (2), every WOSF obviously generates a linearly separable Boolean function. For the following sections, we rewrite Eq. (2) in the form of linearly separable Boolean function as

$$F_{\Omega,\,r}(x) = \begin{cases} 1, & \text{if } <\Omega,\ x> - r \ge 0 \\ 0, & \text{else} \end{cases}, \text{ where } <\Omega,\ x> = \sum_{i=1}^{n} \omega_i x_i .$$

When r is replaced by $-\,r'$, we have

$$F_{\Omega,\,r'}(x) = \begin{cases} 1, & \text{if } <\Omega,\ x> + r' \ge 0 \\ 0, & \text{else} \end{cases}. \tag{4}$$

For the outputs of $F_{\Omega,r'}(x)$, we use the symbol "$-$ 1" to substitute "0", Eq. (4) is analogous to

$$F_{\Omega,r'}(x) = \text{sgn}(<\Omega, x> + r'). \tag{5}$$

For simplicity, we use the notation (Ω, r) to represent a WOSF. Every WOSF (Ω, r) defines a linearly separable Boolean function, different WOSF may define the same Boolean function. Since, let $[(\Omega_0, r_0)]_{BF}$ denote an equivalent class of WOSF generating the same Boolean function, *i.e.*, $[(\Omega_0, r_0)]_{BF} = \{(\Omega, r): F_{\Omega,r}(x) = F_{\Omega_0,r_0}(x)\}$.

3. SVM AND WOSF

The support vector machine (SVM) is a supervised learning machine and can be used as a classifier or regressor. When it is a classifier, the discriminant function is used to classify the training data. If the data is linearly separable, the most important character of SVM is that the separable hyperplane has maximum margin property. It means that the distance from the training data to the hyperplane can be maximized. Moreover, only few of the training data are required to represent the discriminant function forming the hyperplane. These representative data are referred to as support vectors.

Let $S = \{(x_i, y_i)\}_{i=1}^{l} \subseteq X \times Y$ is a training set, where $X \subseteq R^n$, $Y = \{1, -1\}$. The vector x_i is ith n-dimensional vector and called the data information, y_i is integer "1" or "– 1" and called the 2-class category. If the training set S is linearly separable, then there exists a hyperplane $H_{w,b}$.

$$H_{w,b} = \{x \in \Re^n: f_{w,b}(x) = \langle w, x\rangle + b = 0\},$$

such that this hyperplane can correctly classify the training data. Given a pair (w, b), $w \in R^n$ and $b \in R$ to define a function $f_{w,b}(x) = \langle w, x\rangle + b$. The normalized function of $f_{w,b}$ is $g_{w,b}$ and given as

$$g_{w,b}(x) = \| w \|^{-1} f_{w,b}(x) = \langle \| w \|^{-1} w, x\rangle + \| w \|^{-1} b,\ x \in R^n.$$

Also, the functional margin $\mu_s(w, b)$ and geometric margin $\eta_s(w, b)$ are defined as

$$\mu_s(w, b) = \min_{i=1}^{l}\{y_i \cdot (\langle w,\ x\rangle + b)\} = \min_{i=1}^{l}\{y_i \cdot f_{w,b}(x)\},$$
$$\eta_s(w, b) = \min_{i=1}^{l}\left\{y_i \cdot (\langle \|w\|^{-1} w,\ \bar{x}\rangle + \|w\|^{-1} b)\right\} = \min_{i=1}^{l}\{y_i \cdot g_{w,b}(x)\}.$$

The margin γ_s is the maximum geometric margin and is defined as

$$\gamma_s = \max_{w,b} \min_{i=1}^{l}\{y_i \cdot g_{w,b}(x)\} = \max_{w,b} \min_{i=1}^{l}\left\{y_i \cdot (\langle \|w\|^{-1} w,\ x\rangle + \|w\|^{-1} b)\right\}.$$

Such hyperplane with the maximum geometric margin is called maximal margin hyperplane. For a linearly separable training set S, the margin is equal to the inverse of Euclidean norm of w. Hence one considers the following primal optimization problem

$$\begin{aligned}&\text{minimize } 2^{-1} w^T w,\\ &\text{subject to } y_i \cdot (\langle w,\ x\rangle + b) \geq 1,\ \forall i \in l.\end{aligned} \tag{6}$$

Suppose the pair (w^*, b^*) is the solution of primal optimization problem, the maximal margin hyperplane can be written as

$$H_{w^*, b^*} = \{x \in \Re^n: f_{w^*,b^*}(x) = \langle w^*, x\rangle + b^* = 0\},\ w^* \in R^n,\ b^* \in R \text{ and } \gamma_s = \| w^* \|^{-1}.$$

To solve the optimization problem, practically, one uses Lagrangian theorem to convert Eq. (6) to the dual optimization problem [12].

$$\begin{aligned}&\text{maximize } \sum_{i=1}^{l} z_i - 2^{-1}\sum_{i=1}^{l}\sum_{j=1}^{l} z_i z_j y_i y_j <x_i, x_j>,\\ &\text{subject to } \sum_{i=1}^{l} z_i y_i = 0 \text{ and } z_i \geq 0, \forall i \in l.\end{aligned} \tag{7}$$

Suppose z^* is the solution of dual optimization problem, one define the index set I_{SV} for support vectors, $I_{SV} = \{i \in l: z_i^* > 0\}$. The optimal normal vector w^* can be written as

$$w^* = \sum_{i=1}^{l} z_i^* y_i x = \sum_{i \in I_{SV}} z_i^* y_i x.$$

By means of the Karush-Kuhn-Tuker (KKT) conditions [13]

$$z_i^*(y_i <w^*, x> + y_i b^* - 1) = 0,\ y_i(<w^*, x> + b^*) - 1 \geq 0,\ z_i^* \geq 0. \tag{8}$$

The optimal bias b^* is represented as

$$b^* = y_k - <w^*, x_k> = y_k - <\sum_{i \in I_{SV}} z_i^* y_i x, x_k> = y_k - \sum_{i \in I_{SV}} z_i^* y_i <x, x_k>,$$

where k is an arbitrary element in I_{SV}. Hence the discriminant function is defined as

$$f_{w^*,b^*}(x) = <w^*, x> + b^* = \sum_{i \in I_{SV}} z_i^* y_i <x, x_i> + b^*. \tag{9}$$

For any index $i \in I_{SV}$, the vector x_i is called support vector, in which the corresponding Lagrange multipliers $z_i^* > 0$. The support vectors are the nearest data points away from hyperplane and they are the most difficult points to be classified.

Based on the KKT conditions, one has the following equation

$$y_i \cdot (<w^*, x_i> + b^*) = 1. \tag{10}$$

This equation reveals that when category $y_i = +1$, the quantity $<w^*, x_i> + b^* = +1$ or when $y_i = -1$, the quantity $<w^*, x_i> + b^* = -1$. Because the optimal hyperplane H_{w^*, b^*} can correctly separate the training set, hence one gets

$$y_i = \begin{cases} +1, & \text{if } f_{w^*,b^*} > 0 \\ -1, & \text{else} \end{cases}. \tag{11}$$

By the Eqs. (9), (10) and (11), the sign of discriminant function $\text{sgn}(f_{w^*,b^*}(x_i))$ is defined as follows:

$$\text{sgn}(f_{w^*,b^*}(x_i)) = \begin{cases} +1, & \text{if } <w^*, x_i> + b^* > 0 \\ -1, & \text{else} \end{cases}. \tag{12}$$

Since the sign of discriminant function can decide the data classified to "+ 1" category or "– 1" category, compared to Eq. (3), this discriminant function can also represent a linearly separable Boolean function. Hence, it can be used to represent WOSF. In other words, we take the inputs and outputs to form a training set $S = \{(x_i, y_i)\}_{i=1}^{l}$ from a truth table generated by a linearly separable Boolean function. By SVM training, we generate a maximal margin hyperplane H_{w^*,b^*}. This hyperplane has one optimal normal vector w^* and optimal bias b^*, they can form a discriminant function and be used to represent a WOSF. Because the truth table also represents lots of WOSF with the same outputs, hence the discriminant function can be used to represent all WOSF with the same outputs but different weight vectors and threshold values.

4. BF EQUIVALENT CLASS FOR WOSF

For a given WOSF (Ω_0, r_0), $[(\Omega_0, r_0)]_{BF}$ denote the equivalent class such that all the WOSF in this set generates the same linearly separable Boolean function. For this linearly separable Boolean function, we can find a unique maximal margin hyperplane (w^*, b^*) through SVM. Therefore we can regard the term (w^*, b^*) as the unique representative of the BF class $[(\Omega_0, r_0)]_{BF}$.

Let $(\Omega^*, r^*) \in [(\Omega_0, r_0)]_{BF}$ be a WOSF. In order to compromise the notation, we define function $F_{\Omega^*,r^*}(x)$: $\{-1, 1\}^n \to \{-1, 1\}$. From Eqs. (4)-(5), $F_{\Omega^*,r^*}(x)$ is rewritten to

$$F_{\Omega^*,r^*}(x) = \begin{cases} +1 & \text{if } <\Omega^*, x> + r^* \geq 0 \\ -1 & \text{else} \end{cases}.$$

In terms of discriminant function, this function can be formed as

$$\text{sgn}(F_{\Omega^*,r^*}(x)) = \text{sgn}(<\Omega^*, x> + r^*). \tag{13}$$

On the other hand, according to Eq. (12), the sign of discriminant function from SVM

$$\text{sgn}(f_{w^*,b^*}(x)) = \text{sgn}(<w^*, x> + b^*), \tag{14}$$

also generates the same Boolean function. As a consequence, we can regard the pair (w^*, b^*) as the unique representative of the equivalent class $[(\Omega_0, r_0)]_{BF}$ and denoted as $[(w^*, b^*)]_{BF}$. In other words, each element of the BF equivalent class $[(w^*, b^*)]_{BF}$ is a WOSF, and all of these elements have the same outputs. Under the concept of equivalent class and the property of unique representative through SVM, we will characterize the classes of all WOSF.

5. GLOBAL EQUIVALENT CLASS FOR WOSF

From above, we know that the WOSF (Ω, r) defined a linearly separable Boolean function and this Boolean function generates a training set $S = \{(x_i, y_i)\}_{i=1}^{l}$. The training set can generate the corresponding maximal margin hyperplane through SVM. This hyperplane, for simplicity denoted by (w, b), is the unique representative of the equivalent class and is denoted by $[(w, b)]_{BF}$. To explore other distinct equivalent classes without performing the same procedure of SVM, we operate this representative by simple sign change and permutation on the quantities $w^{\#}$ and $b^{\#}$, respectively. Without the SVM training, we can therefore save the cost of complicated SVM training process.

The margin of the hyperplane (w, b) is $\gamma_s = \| w \|^{-1}$ in which $w = (w_1, \ldots, w_n)$ is referred to as the normal vector. For fixed b, if 2 components of w are swapped and denoted by $\hat{w}$, we will obtain a new equivalent class of WOSF. For which $\hat{w}$ is the unique representative of the class in the sense of SVM classification and has the margin $\| \hat{w} \|^{-1} = \| w \|^{-1}$. Such operation is called the permutation on the components of w. Also, if the sign of one component of w is changed, we still obtain a new class that the new vector $\hat{w}$ is the unique representative, too. Besides, if w fixed and sign of b is altered, denoted as $\hat{b}$, we will obtain another new class. For these operations, the new hyperplanes still have margin $\| w \|^{-1}$. This suggests another equivalent relation.

Given a maximal margin hyperplane (w, b), we define the global equivalent class $[(w, b)]_G$ as the set of BF equivalent classes $[(w, b)]_{BF}$. If $\hat{w}$ is a new vector from permutation or sign change of w and $\hat{b}$ is a new bias from sign change of b, then $[(\hat{w}, \hat{b})]_{BF} \in [(w, b)]_G$. We choose the symbol $(w^{\#}, b^{\#})$ in which $w^{\#} = (w_1^{\#}, \ldots, w_n^{\#})$, $w_1^{\#} \geq w_2^{\#} \geq \ldots \geq w_n^{\#} \geq 0$, and $b^{\#} \geq 0$ as the representative of the global equivalent class $[(w, b)]_G$, and denoted by $[(w^{\#}, b^{\#})]_G$. It is easy seen that the representative is unique. It should be noted that each element in the global equivalent class is itself a BF equivalent class. A Boolean function can be specified by its outputs when the inputs are ordered in the common way. Therefore, each element in $[(w^{\#}, b^{\#})]_G$ can be regarded as a set of outputs of the corresponding linearly separable Boolean function. Next, we will calculate the corresponding outputs of all elements in $[(w^{\#}, b^{\#})]_G$.

Let $[(w, b)]_{BF}$ be an element in the global equivalent class $[(w^{\#}, b^{\#})]_G$ and the corresponding outputs of the linearly separable Boolean function be denoted by $y = (y_1, y_2, \ldots, y_l)$, where $l = 2^n$. When the components of w are permutated and sign changed, or the sign of b is changed, we will obtain a new hyperplane $(\hat{w}, \hat{b})$ and the corresponding outputs $\hat{y}$ of Boolean function that can be obtained directly from y.

Below is an example illustrating the above idea with $n = 3$. When normal vector $w^{\#} = (1\ 1\ 0)$ and bias $b^{\#} = 1$, the pair $(1\ 1\ 0), 1)$ is the unique representative of global equivalent class $[(1\ 1\ 0), 1]_G$. By permutation and sign change on $w^{\#}$, it yields 11 new vectors, $(1\ 1\ 0)$, $(0\ 1\ 1)$, …, and $(1\ -1\ 0)$. Similarly, the sign change on $b^{\#} = 1$, it yields the new bias $\hat{b} = -1$. Collecting all vectors, $w^{\#}$ and $\hat{w}$, and bias $b^{\#}$ and $\hat{b}$, we construct 24 combinations. Each of the combination is a hyperplane. Table 3 lists all of these hyperplanes. And as mentioned above, all of the 24 hyperplanes have the same maximum margins $1/\sqrt{2}$.

6. CORRESPONDING OUTPUTS

Every global class contains many BF classes. Each corresponding output of BF class in the global class can be explicitly derived. In the following propositions, we propose 3 translated formulas to directly generate the corresponding outputs of distinct BF class and the proofs are listed in the Appendix.

Let (w_1, b_1) and (w_2, b_2) denote 2 distinct WOSF derived from 2 distinct maximal margin hyperplanes, which are the representatives of BF classes $[(w_1, b_1)]_{BF}$ and $[(w_2, b_2)]_{BF}$, respectively. The normal vectors w_1 and w_2 are the n-dimensional vectors and the biases b_1 and b_2 are real numbers. Let y_1 and y_2 are respectively the outputs of $[(w_1, b_1)]_{BF}$ and $[(w_2, b_2)]_{BF}$, denoted as the form of vector.

Proposition 1 Let $b_2 = b_1$. If w_2 is a 1-component sign change of w_1, *i.e.*, $w_1 = [w_{11}, \ldots, w_{1j}, \ldots, w_{1n}]^T$ and $w_2 = [w_{11}, \ldots, -w_{1j}, \ldots, w_{1n}]^T$, then the relation between y_1 and y_2 can be modeled as

$$y_2 = P^{2^{(n-j)}} y_u + P^{-2^{(n-j)}} y_v, \tag{15}$$

where $y_1 = y_u + y_v$, $j = 1, \ldots, n$, and P is an $n \times n$ matrix of the form

$$P = \begin{bmatrix} 0 & 1 & . & . & . & 0 \\ . & . & 1 & \ddots & . & . \\ . & . & . & \ddots & \ddots & . \\ . & . & . & . & \ddots & . \\ 0 & . & . & . & . & 1 \\ 1 & 0 & . & . & . & 0 \end{bmatrix}_{n\times n}, \quad P^{2(n-j)} = P^0, P^2, P^4, \ldots$$

The terms y_u and y_v are defined as $y_u = \sum_{m=1}^{2^{j-1}} \sum_{u=0}^{2^{n-j}-1} (\tilde{y}_{1(2^{(n-j+1)}m-u)})$ and $y_v = \sum_{m=1}^{2^{j-1}} \sum_{v=2^{n-j}}^{2^{n-j+1}-1} (\tilde{y}_{1(2^{(n-j+1)}m-v)})$, where $\tilde{y}_{1j} = [0\ldots y_{1j}\ldots 0]^T$.

Proposition 2 Let $b_2 = b_1$. If w_2 is a 2-components $w_{1(j-1)}$, w_{1j} permutation of w_1, *i.e.*, $w_1 = [w_{11}, \ldots, w_{1(j-1)}w_{1j}, \ldots, w_{1n}]^T$ and $w_2 = [w_{11}, \ldots, w_{1j}w_{1(j-1)}, \ldots, w_{1n}]^T$, then the relation between y_1 and y_2 can be modeled as

$$y_2 = P^{2^{(n-j)}} y_u + P^{-2^{(n-j)}} y_v + \sum_{j \notin u,v} \tilde{y}_{1j}, \tag{16}$$

where $y_1 = y_u + y_v$, $j = 1, \ldots, n$, the matrix P is defined as proposition 1. The terms y_u and y_v are defined as $y_u = \sum_{m=odd}^{2^{j-1}-1} \sum_{u=-2^{n-j}}^{-1} (\tilde{y}_{1(2^{(n-j+1)}m-u)})$ and $y_v = \sum_{m=odd}^{2^{j-1}-1} \sum_{v=1-2^{n-j}}^{0} (\tilde{y}_{1(2^{(n-j+1)}m+v)})$, where $\tilde{y}_{1j} = [0\ldots y_{1j}\ldots 0]^T$.

Proposition 3 Let $w_2 = w_1$. If b_2 is the sign change of b_1, *i.e.*, $b_2 = -b_1$, then the relation between y_1 and y_2 can be modeled as

$$y_2 = NOT(M\, y_1), \tag{17}$$

where M is an $n \times n$ matrix of the form

$$M = \begin{bmatrix} 0 & 0 & . & . & 0 & 1 \\ 0 & . & . & . & 1 & 0 \\ . & . & \ddots & \cdot^{\cdot^{\cdot}} & . & . \\ . & . & \cdot^{\cdot^{\cdot}} & . & . & . \\ 0 & 1 & . & . & . & . \\ 1 & 0 & . & . & . & 0 \end{bmatrix}_{n \times n} .$$

7. EXPERIMENT AND ILLUSTRATION

We take arbitrary WOSF $F_{(131),4}(x)$ that has corresponding output (0 0 0 1 0 0 1 1) to illustrate. From WOSF $F_{(131),4}(x)$, composing all possible inputs x_i and outputs y_i to construct a training set $S = \{(x_i, y_i)\}^8_{i=1}$. By SVM training, we obtain a hyperplane with normal vector $w^* = (1\ 2\ 1)$ and bias $b^* = -1$. Therefore, the discriminant function can be written as $f_{(121),-1}(x) = <[1\ 2\ 1]^T, x> + (-1)$. Because the WOSF is characterized by outputs (00010011), the signed discriminant function $\text{sgn}(f_{(121),-1}(x))$ can represent all WOSF with outputs y = (00010011). Example, the WOSF $F_{(121),3}(x)$, $F_{(131),4}(x)$, $F_{(243),6}(x)$, $F_{(252),7}(x)$ and $F_{(456),10}(x)$ generate the same output, so they can are represented by $\text{sgn}(f_{(121),-1}(x))$. The detailed WOSF, hyperplane and BF class are listed in Table 1.

Table 1. WOSF representation and BF equivalent classes corresponding same outputs.

Output	(0 1 1 1 1 1 1 1)	(0 0 1 1 0 1 1 1)	(0 0 1 1 0 0 1 1)	(0 0 0 1 0 0 1 1)	(0 0 0 0 0 0 0 1)
WOSF	$\text{sgn}(f_{(111),2})$	$\text{sgn}(f_{(121),1})$	$\text{sgn}(f_{(010),0})$	$\text{sgn}(f_{(121),-1})$	$\text{sgn}(f_{(111),-2})$
BF	$[(111), 2]_{BF}$	$[(121), 1]_{BF}$	$[(101), 0]_{BF}$	$[(121), -1]_{BF}$	$[(111), -2]_{BF}$
(Ω, r)	((131), 1)	((131), 2)	((131), 3)	((131), 4)	((131), 5)
(Ω, r)	((324), 2)	((231), 3)	((120), 2)	((121), 3)	((111), 3)
(Ω, r)	((543), 3)	((243), 4)	((241), 4)	((243), 6)	((213), 6)
(Ω, r)	((474), 4)	((154), 5)	((252), 5)	((252), 7)	((121), 4)
(Ω, r)	((589), 5)	((364), 6)	((263), 6)	((465), 10)	((225), 9)

From Table 1, the first column shows 5 WOSF with the same outputs (01111111). From section 4, we construct the BF equivalent class $[(1\ 1\ 1), 2]_{BF}$ in which the pair ((1 1 1), 2) is unique representative. Note there are infinite elements in the class $[(1\ 1\ 1), 2]_{BF}$ and only 5 instances are listed in the column. The other columns in Table 1 show additional examples of equivalent classes listed in the third row.

For $n = 3$, there are $2^{(2^3)} = 256$ different Boolean functions, and these functions construct 256 different training sets. By SVM training, only 102 sets are linearly separable. The 102 sets generate 102 hyperplanes, each hyperplane is a representative of BF class. These BF classes constitute 5 distinct global classes, *i.e.*, $[(1\ 1\ 1), 2]_G$, $[(1\ 1\ 0), 1]_G$, $[(2\ 1\ 1), 1]_G$, $[(1\ 0\ 0), 0]_G$ and $[(1\ 1\ 1), 0]_G$. They contain 16, 24, 48, 6, and 8 *BF* classes, respectively. In Table 2, the 5 global classes and their element number are listed. For global class $[(1\ 1\ 0), 1]_G$, all elements are listed in Table 3. For global class $[(1\ 1\ 1), 0]_G$, all ele- ments are listed in Table 4.

Table 2. 5 global classes generated by 102 BF classes of order 3.

Num.	Global classes	BF classes number
1	$[(1\ 1\ 1), 2]_G$	16
2	$[(1\ 1\ 0), 1]_G$	24
3	$[(2\ 1\ 1), 1]_G$	48
4	$[(1\ 0\ 0), 0]_G$	6
5	$[(1\ 1\ 1), 0]_G$	8
Total		102

Table 3. The 24 elements of global class $[(1\ 1\ 0), 1]_G$ and corresponding outputs.

$[w, b]_{BF}$		outputs	$[w, b]_{BF}$		outputs
w	b		w	b	
(1 1 0)	1	(0 0 1 1 1 1 1 1)	(1 1 0)	– 1	(0 0 0 0 0 0 1 1)
(1 0 1)	1	(0 1 0 1 1 1 1 1)	(1 0 1)	– 1	(0 0 0 0 0 1 0 1)
(0 1 1)	1	(0 1 1 1 0 1 1 1)	(1 0 – 1)	– 1	(0 0 0 0 1 0 1 0)
(1 0 – 1)	1	(1 0 1 0 1 1 1 1)	(1 – 1 0)	– 1	(0 0 0 0 1 1 0 0)
(0 1 – 1)	1	(1 0 1 1 1 0 1 1)	(0 1 1)	– 1	(0 0 0 1 0 0 0 1)
(1 – 1 0)	1	(1 1 0 0 1 1 1 1)	(0 1 – 1)	– 1	(0 0 1 0 0 0 1 0)
(0 – 1 1)	1	(1 1 0 1 1 1 0 1)	(– 1 1 0)	– 1	(0 0 1 1 0 0 0 0)
(0 – 1 – 1)	1	(1 1 1 0 1 1 1 0)	(0 – 1 1)	– 1	(0 1 0 0 0 1 0 0)
(– 1 1 0)	1	(1 1 1 1 0 0 1 1)	(– 1 0 1)	– 1	(0 1 0 1 0 0 0 0)
(– 1 0 1)	1	(1 1 1 1 0 1 0 1)	(0 – 1 – 1)	– 1	(1 0 0 0 1 0 0 0)
(– 1 0 –1)	1	(1 1 1 1 1 0 1 0)	(– 1 0 – 1)	– 1	(1 0 1 0 0 0 0 0)
(–1 –1 0)	1	(1 1 1 1 1 1 0 0)	(– 1 – 1 0)	– 1	(1 1 0 0 0 0 0 0)

Table 4. The only 8 elements and corresponding outputs of global class $[(1\ 1\ 1), 0]_G$.

$[w, b]_{BF}$		outputs
w	b	
(1 1 1)	0	(0 0 0 1 0 1 1 1)
(1 1 – 1)	0	(0 0 1 0 1 0 1 1)
(1 – 1 1)	0	(0 1 0 0 1 1 0 1)
(– 1 1 1)	0	(0 1 1 1 0 0 0 1)
(1– 1 – 1)	0	(1 0 0 0 1 1 1 0)
(– 1 1 – 1)	0	(1 0 1 1 0 0 1 0)
(– 1 – 1 1)	0	(1 1 0 1 0 1 0 0)
(– 1 – 1 – 1)	0	(1 1 1 0 1 0 0 0)

When bias b is fixed, Table 5 shows the generated vectors and corresponding outputs by sign change and permutation on vector (2 1 1). For examples, the pair ((1 1 2), 1) is obtained by interchanging on 1st and 3rd components of vector (1 1 2), and the pair ((2 1 –1), 1) is obtained by sign change on the 3rd component of vector (2 1 1).

Next, we give an experiment to generate the corresponding output. In Table 6, we show these results. From the 4th row on left side, the BF class $[(2\ 1\ 1), 1]_{BF}$ has corresponding outputs y = (0 0 0 1 1 1 1 1). When we perform sign change on vector (2 1 1) to get a new vector (2 – 1 1), the new outputs can be calculated by Eq. (15) in proposition 1

Table 5. The generated vectors and corresponding outputs by sign change and permutation of normal vector (2 1 1), when *b* = 1.

w permutation		*w* sign change	
w	outputs	*w*	outputs
(2 1 1)	(0 0 0 1 1 1 1 1)	(2 1 1)	(0 0 0 1 1 1 1 1)
(1 2 1)	(0 0 1 1 0 1 1 1)	(2 1 – 1)	(0 0 1 0 1 1 1 1)
(1 1 2)	(0 1 0 1 0 1 1 1)	(2 – 1 1)	(0 1 0 0 1 1 1 1)
		(– 2 1 1)	(1 1 1 1 0 0 0 1)
		(2 – 1 – 1)	(1 0 0 0 1 1 1 1)
		(– 2 1 – 1)	(1 1 1 1 0 0 1 0)
		(– 2 – 1 1)	(1 1 1 1 0 1 0 0)
		(– 2 – 1 – 1)	(1 1 1 1 1 0 0 0)

Table 6. The generated vectors and corresponding outputs by sign change of bias *b* = 1 from global class [(2 1 1), 1]$_G$.

b = 1			*b* = - 1		
[*w*, *b*]$_{BF}$		outputs	[*w*, *b*]$_{BF}$		outputs
w	*b*		*w*	*b*	
(2 1 1)	1	(0 0 0 1 1 1 1 1)	(2 1 1)	- 1	(0 0 0 0 0 1 1 1)
(2 1 - 1)	1	(0 0 1 0 1 1 1 1)	(2 1 - 1)	- 1	(0 0 0 0 1 0 1 1)
(2 - 1 1)	1	(0 1 0 0 1 1 1 1)	(2 - 1 1)	- 1	(0 0 0 0 1 1 0 1)
(- 2 1 1)	1	(1 1 1 1 0 0 0 1)	(- 2 1 1)	- 1	(0 1 1 1 0 0 0 0)
(2 - 1- 1)	1	(1 0 0 0 1 1 1 1)	(2 – 1 - 1)	- 1	(0 0 0 0 1 1 1 0)
(- 2 1- 1)	1	(1 1 1 1 0 0 1 0)	(- 2 1 - 1)	- 1	(1 0 1 1 0 0 0 0)
(- 2- 1 1)	1	(1 1 1 1 0 1 0 0)	(- 2 - 1 1)	- 1	(1 1 0 1 0 0 0 0)
(- 2 - 1 - 1)	1	(1 1 1 1 1 0 0 0)	(- 2 – 1 - 1)	- 1	(1 1 1 0 0 0 0 0)

and get $\hat{y} = (0\ 0\ 1\ 0\ 1\ 1\ 1\ 1)$. In other words, we obtain a new BF class [(2 1 −1), 1]$_{BF}$ with corresponding outputs $\hat{y} = (0\ 0\ 1\ 0\ 1\ 1\ 1\ 1)$. Similarly, if we perform a permutation on (2 1 1) to obtain a new vector (1 2 1), the new outputs can be calculated from Eq. (16) in proposition 2 and get $\hat{y} = (0\ 0\ 1\ 1\ 0\ 1\ 1\ 1)$. This means that we get a new BF class [(1 2 1), 1]$_{BF}$ with corresponding outputs $\hat{y} = (0\ 0\ 1\ 1\ 0\ 1\ 1\ 1)$. When vector (2 1 1) fixed, we perform a sign change on bias $b^{\#} = 1$, by Eq. (17) on proposition 3, we get a new outputs $\hat{y} = (0\ 0\ 0\ 0\ 0\ 1\ 1\ 1)$. Therefore, we obtain a new BF class [(2 1 1), −1]$_{BF}$ with corresponding outputs $\hat{y} = (0\ 0\ 0\ 0\ 0\ 1\ 1\ 1)$.

8. CONCLUSION

For WOSF, one weight vector and threshold value can determine its property. Different weight vectors and threshold values can represent different WOSF, but the outputs may be the same. To characterize the WOSF, we utilize SVM technique to uniquely represent the WOSF. This method is based on the property of maximal margin classification of SVM. From SVM training, we can generate a maximal margin hyperplane which is formulated by an optimal normal vector and an optimal bias. The hyperplane defines a discriminant function and this function has the same outputs as those of the WOSF. In other words, we can only use the normal vector and bias to represent a lot of WOSF having the same outputs. Compared to the other representation of WOSF, the proposed method is obviously efficient and inexpensive.

We also propose an alternative method to characterize the WOSF. The characterization consists of two stages of equivalent classes. The first class, referred to as BF class, is constructed by all of the WOSF with same outputs. The second class, referred to as global class, is composed of the BF classes. In the global class, every representative of the BF class has the same weight vectors subject to component-wise permutation and sign changes. Also, every bias is the same subject sign changes. Based on the two classes, we can characterize all the WOSF with same order.

We have proved that the permutation and sign change operations can generate new BF classes in which each member produce the same outputs. For these BF classes, we also formulate the relation between the outputs and these operations. Three translated formulas are introduced to directly and fast generate new corresponding outputs and therefore, there is no need to build the truth tables or perform other machine learning process to calculate the outputs.

APPENDIX

Proposition 1

***Proof*:** Let w_2 is a vector of sign change of arbitrary element w_{1j} in w_1, *i.e.*, $w_2 = [w_{11}, \ldots, -w_{1j}, \ldots, w_{1n}]^T$. One formulate y_2 in terms of y_1. By the discriminant function $f^*(x_i) = \langle w^*, x_i\rangle + b^*$, the element y_{2i} of y_2 is written as

$$\begin{aligned} y_{2i} &= f_2(x_i) = \langle w_2, x_i\rangle + b_2 \\ &= [w_{11}, \ldots, -w_{1j}, \ldots, w_{1n}][x_{i1}, \ldots, x_{ij}, \ldots, x_{in}]^T + b_1 \\ &= (w_{11}x_{i1} + \ldots + (-w_{1j})x_{ij} + \ldots + w_{1n}x_{in}) + b_1 \\ &= [w_{11}, \ldots, w_{1j}, \ldots, w_{1n}][x_{i1}, \ldots, -x_{ij}, \ldots, x_{in}]^T + b_1. \end{aligned}$$

The input $x_i = [x_{i1}, \ldots, x_{ij}, \ldots, x_{in}]^T$, it has two conditions as follow

$$x_i = [x_{i1}, \ldots, -x_{ij}, \ldots, x_{in}]^T = \begin{cases} x_{(i+2^{n-j})}, & i = 2^{(n-j+1)}m - k_0 \\ x_{(i-2^{n-j})}, & i = 2^{(n-j+1)}m - k_1 \end{cases},$$

where $k_0 = 2^{n-j}, \ldots, (2^{n-j+1} - 1)$, $k_1 = 0, \ldots, (2^{n-j} - 1)$, and $m = 1, \ldots, 2^{j-1}$.

According to the condition, when $x_i = x_{(i+2^{n-j})}$, the component y_{2i} is represent as

$$y_{2i} = \langle w_1, x_{(i+2^{n-j})}\rangle + b_1 = y_{1(i+2^{n-j})} = y_{1(2^{(n-j+1)}m-k_0+2^{n-j})} = y_{1(2^{(n-j+1)}m-u)},$$

where $u = k_0 - 2^{n-j}$. Each existed independent component y_{1i} of y_1, one represent it as a vector $\tilde{y}_{1i} = [0, \ldots, y_{1i}, \ldots, 0]^T$. Collecting these vectors $\tilde{y}_{1i}$, one builds a new vector y_u

$$y_u = \sum_{m=1}^{2^{j-1}} \sum_{u=0}^{2^{n-j}-1} (\tilde{y}_{1(2^{(n-j+1)}m-u)}).$$

Also, by the operations of matrix P, the relationship of vectors $\tilde{y}_{1i}$ and $\tilde{y}_{2i}$ has the variation property, $y_{2i} = y_{1(i+j)}$ if $\tilde{y}_{2i} = P^j \tilde{y}_{1(i+j)}$. Based on this relation, one gets

$$\tilde{y}_{2(i)} = P^{2^{(n-j)}} \tilde{y}_{1(i+2^{(n-j)})} = P^{2^{(n-j)}} \tilde{y}_{1(2^{(n-j+1)}m-u)}.$$

Collecting all vectors $\tilde{y}_{2(i)}$, one gets a new result

$$\sum_{1}^{2^{n-1}} \tilde{y}_{2(i)} = P^{2^{(n-j)}} \left(\sum_{m=1}^{2^{j-1}} \sum_{u=0}^{2^{n-j}-1} (\tilde{y}_{1(2^{(n-j+1)}m-u)}) \right) = P^{2^{(n-j)}} y_u .$$

Similarly, when $x_i = x_{(i-2^{n-j})}$, the component y_{2i} is represented as

$$y_{2i} = <w_1, x_{(i-2^{n-j})}> + b_1 = y_{1(i-2^{n-j})} = y_{1(2^{(n-j+1)}m-k_1-2^{n-j})} = y_{1(2^{(n-j+1)}m-v)},$$

where $v = k_1 + 2^{n-j}$. To integrate all vectors $\tilde{y}_{1(2^{(n-j+1)}m-v)}$, one defines another vector y_v

$$y_v = \sum_{m=1}^{2^{j-1}} \sum_{v=2^{n-j}}^{2^{n-j+1}-1} (\tilde{y}_{1(2^{(n-j+1)}m-v)}),$$

based on the same property above, the vector $\tilde{y}_{2(i)}$ be written as

$$\tilde{y}_{2(i)} = P^{-2^{(n-j)}} \tilde{y}_{1(i-2^{(n-j)})} = P^{-2^{(n-j)}} \tilde{y}_{1(2^{(n-j+1)}m-v)}, \text{ hence}$$

$$\sum_{2^{n-1}+1}^{2^n} \tilde{y}_{2(i)} = P^{-2^{(n-j)}} \left(\sum_{m=1}^{2^{j-1}} \sum_{v=2^{n-j}}^{2^{n-j+1}-1} (\tilde{y}_{1(2^{(n-j+1)}m-v)}) \right) = P^{-2^{(n-j)}} y_v ,$$

where $y_1 = y_u + y_v$. By above deduction, one gets the relationship of y_1 and y_2

$$y_2 = \sum_{1}^{2^{n-1}} \tilde{y}_{2(i)} + \sum_{2^{n-1}+1}^{2^n} \tilde{y}_{2(i)} = P^{2^{(n-j)}} y_u + P^{-2^{(n-j)}} y_v .$$ ❑

Proposition 2

***Proof*:** Let w_2 is a 2-components $w_{1(j-1)}$, w_{1j} permutation of w_1, *i.e.*, $w_2 = [w_{11} \ldots w_{1j} w_{1(j-1)} \ldots w_{1n}]^T$. One formulate y_2 in terms of y_1. By the discriminant function $f^*(x_i) = <w^*, x_i> + b^*$, the element y_{2i} of y_2 is written as

$$\begin{aligned} y_{2i} &= f_2(x_i) = <w_2, x_i> + b_2 = [w_{11} \ldots w_{1j} w_{1(j-1)} \ldots w_{1n}][x_{i1} \ldots x_{i(j-1)} x_{ij} \ldots x_{in}]^T + b_1 \\ &= (w_{11}x_{i1} + \ldots + w_{1j}x_{i(j-1)} + w_{1(j-1)}x_{ij} + \ldots + w_{1n}x_{in}) + b_1 \\ &= [w_{11} \ldots w_{1j} \ldots w_{1n}][x_{i1} \ldots x_{ij} x_{i(j-1)} \ldots x_{in}]^T + b_1. \end{aligned}$$

The input x_i has following two conditions

$$x_i = [x_{i1}, \ldots, x_{ij} x_{i(j-1)}, \ldots, x_{in}]^T = \begin{cases} x_{(i+2^{n-j})},\ i = 2^{(n-j+1)} m - k^+ \text{ if } k^+ = 0, \ldots, 2^{n-j} - 1,\ m = 1, 3, \ldots, 2^{j-1} - 1) \\ x_{(i-2^{n-j})},\ i = 2^{(n-j+1)} m + k^- \text{ if } k^- = 1, \ldots, 2^{n-j},\ m = 1, 3, \ldots, 2^{j-1} - 1) \end{cases}.$$

Based on the condition, when $x_i = x_{(i+2^{n-j})}$, the component y_{2i} is represented as

$$y_{2i} = <w_1, x_{(i+2^{n-j})}> + b_1 = y_{1(i+2^{n-j})} = y_{1(2^{(n-j+1)} m - k^+ + 2^{n-j})} = y_{1(2^{(n-j+1)} m - l^+)},$$

where $l^+ = k^+ - 2^{n-j}$, $k^+ = 0, \ldots, (2^{n-j} - 1)$.

Each existed independent component y_{1i} of y_1, one might represent it as a vector $\tilde{y}_{1i} = [0 \ldots y_{1i} \ldots 0]^T$. To integrate all vectors of $\tilde{y}_{1i}$, one builds a new vector y_u

$$y_u = \sum_{\substack{m=1 \\ odd}}^{2^{j-1}-1} \sum_{l^+=-2^{n-j}}^{-1} (\tilde{y}_{1(2^{(n-j+1)} m - l^+)}).$$

Besides, by the operations of matrix P, the relationship of vectors $\tilde{y}_{1i}$ and $\tilde{y}_{2i}$ that has the variation property, $y_{2i} = y_{1(i+j)}$ if $\tilde{y}_{2i} = P^j \tilde{y}_{1(i+j)}$, hence one gets

$$\tilde{y}_{2(i)} = P^{2^{(n-j)}} \tilde{y}_{1(i+2^{(n-j)})} = P^{2^{(n-j)}} \tilde{y}_{1(2^{(n-j+1)} m - l^+)}.$$

Sum all vectors $\tilde{y}_{2(i)}$, one gets a new result

$$\sum_{i=1}^{2^n-1} \tilde{y}_{2(i)} = P^{2^{(n-j)}} \left(\sum_{m=1}^{2^{j-1}} \sum_{l^+=-2^{n-j}}^{-1} (\tilde{y}_{1(2^{(n-j+1)} m - l^+)}) \right) = P^{2^{(n-j)}} y_u.$$

Similarly, when $x_i = [x_{i1}, \ldots, x_{ij} x_{i(j-1)}, \ldots, x_{in}]^T = x_{(i-2^{n-j})}$, the component y_{2i} is varied to represent as

$$y_{2i} = <w_1, x_{(i-2^{n-j})}> + b_1 = y_{1(i-2^{n-j})} = y_{1(2^{(n-j+1)} m + k^- - 2^{n-j})} = y_{1(2^{(n-j+1)} m + l^-)},$$

where $l^- = k^- - 2^{n-j}$, $k^- = 1, \ldots, 2^{n-j}$.

Collecting all vectors $\tilde{y}_{1(2^{(n-j+1)} m + l^-)}$, one defines another vector y_v

$$y_v = \sum_{\substack{m=1 \\ odd}}^{2^{j-1}} \sum_{l^-=1-2^{n-j}}^{0} (\tilde{y}_{1(2^{(n-j+1)} m + l^-)}).$$

Based on the above property, the vector $\tilde{y}_{2(i)}$ be written as

$$\tilde{y}_{2(i)} = P^{-2^{(n-j)}} \tilde{y}_{1(i-2^{(n-j)})} = P^{-2^{(n-j)}} \tilde{y}_{1(2^{(n-j+1)} m + l^-)},\ \text{hence}$$

$$\sum_{2^{n-1}+1}^{2^n} \tilde{y}_{2(i)} = P^{-2^{(n-j)}} \left(\sum_{m=1}^{2^{j-1}} \sum_{l^-=2^{n-j}}^{2^{n-j+1}-1} (\tilde{y}_{1(2^{(n-j+1)}m+l^-)}) \right) = P^{-2^{(n-j)}} y_v.$$

By the deduction, one gets the relationship of vectors y_1 and y_2

$$y_2 = P^{2^{(n-j)}} y_u + P^{-2^{(n-j)}} y_v + \sum_{j \notin u,v} \tilde{y}_{1j}.$$ ❑

Proposition 3

***Proof*:** Let two weight vectors $w_2 = w_1$, *i.e.*, $[w_{21} \ldots w_{2j} \ldots w_{2n}]^T = [w_{11} \ldots w_{1j} \ldots w_{1n}]^T$. The input $x_i = [x_{i1}, \ldots, x_{ij}, \ldots, x_{in}]^T$, based on the variation property of input x_i, it can be represented as the form, $x_i = -x_{(2^n+1-i)}$. Because the bias $b_2 = -b_1$, by discriminant function $f^*(x_i) = <w^*, x_i> + b^*$, the element y_{2i} of y_2 is written as

$$\begin{aligned} y_{2i} &= f_2(x_i) = <w_2, x_i> + b_2 = <w_1, -x_{(2^n+1-i)}> - b_1 = -(<w_1, x_{(2^n+1-i)}> + b_1) \\ &= -f_1(x_{(2^n+1-i)}) = -y_{1(2^n+1-i)}. \end{aligned}$$

By the operations of matrix M, one gets the relationship of vectors y_1 and y_2

$$y_2 = NOT(M\, y_1).$$ ❑

REFERENCES

1. P. D. Wentd, "Nonrecursive and recursive stack filters and their filtering behavior," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 38, 2001, pp. 2099-2107.
2. L. Yin, J. Astola, and Y. Neuvo, "Adaptive multistage weighted order statistic filters for image restoration," in *Proceedings of SPIE*, Vol. 1451, 1991, pp. 216-227.
3. O. Yli-Harja, J. T. Astola, and Y. Neuvo, "Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation," *IEEE Transactions on Signal Processing*, Vol. 39, 1991, pp. 395-410.
4. P. Koivisto and H. Huttunen, "Design of weighted order statistic filters by training-based optimization," in *Proceedings of International Conference Symposium on Signal Processing and its Application*, 2001, pp. 13-16.
5. M. Ropert, F. M. de Saint-Martin, and D. Pele, "A new representation of weighted order statistic filters," *Signal Processing*, Vol. 54, 1996, pp. 201-206.
6. C. E. Savin, M. O. Ahmad, and M. N. S. Swamy, "Design of weighted order statistic filters using linearly separable stack-like architecture," in *Proceedings of the 37th Midwest Symposium on Circuits and Systems*, 1994, pp. 753-756.
7. N. S. T. Hirata and R. Hirata Jr., "Design of order statistic filters from examples," in *Proceedings of the 16th Brazilian Symposium on Computer Graphics and Image Processing*, 2003, pp. 222-229.
8. S. Marshall, "New direct design method for weighted order statistic filters," in *Proceedings of IEE International Conference on Vision, Image, and Signal Processing*, Vol. 151, 2001, pp. 1-8.

9. R. Lukac, K. N. Plataniotis, B. Smolka, and A. N. Venetsanopoulos, "Generalized selection weighted vector filters," *EURASIP Journal on Applied Signal Processing*, Vol. 12, 2004, pp. 1870-1885.
10. P. T. Yu and W. H. Liao, "Weighted order statistics filters-their classification, some properties, and conversion algorithm," *IEEE Transactions on Image Processing*, Vol. 42, 1994, pp. 2673-2691.
11. G. R. Arce, T. A. Hall, and K. E. Barner, "Permutation weighted order statistic filter lattices," *IEEE Transactions on Image Processing*, Vol. 4, 1995, pp. 1070-1083.
12. B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, 1992, pp. 144-152.
13. V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed., Springer-Verlag, New York, 1999.
14. Y. L. Lin, W. C. Teng, J. H. Jeng, and J. G. Hsieh, "Characterization of canonical robust template values for a class of uncoupled CNNs implementing linearly separable boolean functions," *WSEAS Transactions on Information Science and Applications*, Vol. 7, 2005, pp. 940-944.
15. C. C. Yao and P. T. Yu, "The optimal design of weighted order statistics filters by using support vector machines," *EURASIP Journal on Applied Signal Processing*, Vol. 1, 2006, pp. 1-13.
16. W. C. Chen and J. H. Jeng, "A representation of WOS filters through support vector machine," in *Proceedings of the 1st Pacific-Rim Symposium on Image and Video Technology*, 2006, pp. 1168-1176.

Wei-Chih Chen (陳為志) received the B.S. degree in Applied Mathematics from Cheng Kung University, Tainan, Taiwan, in 1993, received the M.S. degree in Electronic Engineering from the I-Shou University, Kaohsiung, Taiwan, in 2000. Currently, he is working toward the Ph.D. degree in Information Engineering at I-Shou University, Kaohsiung, Taiwan. His research interests include soft computing and machine learning, statistic and mathematical application.

Jyh-Horng Jeng (鄭志宏) received the B.S. and M.S. degrees in Mathematics from Tamkang University, Taiwan in 1982 and 1984, respectively, and the Ph.D. degree in Mathematics from State University of New York at Buffalo in 1996. He was as Senior Research Engineer with Chung Shan Institute of Science and Technology (CSIST) Taiwan, from 1984 to 1992. Currently, he is a Professor of Department of Information Engineering, I-Shou University, Taiwan. His research interests include multimedia applications, soft computing and machine learning, error correcting code, and VLSI architecture design.