SUPPORT VECTOR MACHINES ON THE SPACE OF WALSH FUNCTIONS AND THEIR PROPERTIES

A. Fazekas,[*] C. Kotropoulos, I. Buciu and I. Pitas

Department of Informatics, Aristotle University of Thessaloniki

GR-54006 Thessaloniki Box 451, Greece

`fattila@math.klte.hu,{costas,nelu,pitas}@zeus.csd.auth.gr`

Abstract

Support vector machine is a special kind of learning machines, proposed by Vapnik. The learning capability of support vector machines depends on the Vapnik-Chervonenkis dimension of the kernel function used. In this paper we construct a new kernel function for support vector machine, which is based on Walsh functions. We prove some theoretical results related to the VC-dimension of the support vector machines which are built in the space of the Walsh functions. First experimental results for face detection are reported.

1. Introduction

The Bayes likelihood ratio test yields the optimal classifier in the sense that it minimizes the probability of error [7]. However in order to construct the likelihood ratio, the conditional probability density function (pdf) for each class must be known. Although, there are several procedures for estimating a pdf from a finite number of observations [7], the problem of density estimation is ill-posed [1]. An alternative method to solve a two-class pattern recognition problem is to resort to example-based techniques, such as the support vector machines [1].

Support vector machines implement the following idea: Let us map the input vectors, which are the elements of the training set, onto a high-dimensional feature space through a mapping chosen a priori. In this space, we construct an optimal separating hyperplane expecting that after projection the pattern will be linearly separable. One can interpret the result of the binary decision as the input vector belongs to the given class or not. For example, in the face detection the input vector is a region of the digital image, the result of the binary decision is the answer to the following question:

Is this region a face or not?

In the general case the determination of the separating hyperplane is not easy, because the dimensionality of the feature space is high. However, in Hilbert spaces one can estimate the inner product of two vectors in the feature space, as a function of two vectors in the input space that it enables us to find the solution.

These expressions for inner products are referred as kernel functions. Some kernel functions are well-known, for example the polynomial, the radial, the sigmoid, etc. [1],[5]. In this paper we will construct a new kernel function that is based on the Walsh system. In this case, the Walsh transformation reveals the special symmetries of the input vectors. The Walsh transformation divides up the elements of the input vectors and calculates the sum of the elements, which are multiplied by either $+1$ or -1, of the parts. This transformation is useful in the pattern recognition [3].

The VC-dimension was defined by Vapnik [1] as the capacity factor of the support vector machine. Its knowledge is very important to control the behavior of the support vector machine. In this paper, we are going to prove some propositions related to the VC-dimension of the class of the Walsh functions.

The outline of the paper is as follows. Section 2 is a brief overview of the Walsh system. Section 3 explains the construction of the Walsh kernel. Section 4 gives the theoretical results on the VC-dimension of the class of the Walsh functions. Section 5 is devoted to higher dimensional Walsh kernels and their properties. Section 6 reports first experimental results on face detection.

2. The Walsh system

In the literature the term "Walsh functions" refers to one of three orthonormal systems: the Walsh-Paley system, the original Walsh system, or the Walsh-Kaczmarz system [2]. These systems contain the same functions and differ only in the enumeration. We will investigate the Walsh-Paley sys-

---

[*] A. Fazekas is on leave from Institute of Mathematics and Informatics, University of Debrecen, H-4010 Debrecen P.O.Box 12, Hungary.

tem, which will be referred as the Walsh system henceforth. For more details the interested reader may consult [2].

Notation 1. *We will denote the set of non-negative integers by $\mathbb{N}$, the set of positive integers by $\mathbb{N}^+$, the set of integers by $\mathbb{Z}$, and the set of real numbers by $\mathbb{R}$.*

Definition 2. *Let r be the function defined on $[0, 1)$ by*

$$r(x) = \begin{cases} 1, & x \in [0, \frac{1}{2}), \\ -1, & x \in [\frac{1}{2}, 1) \end{cases}$$

extended to $\mathbb{R}$ periodically with period 1. The Rademacher system $R = \{r_n, n \in \mathbb{N}\}$ is defined by

$$r_n(x) = r(2^n x), \qquad x \in \mathbb{R}, n \in \mathbb{N}.$$

Definition 3. *Given $n \in \mathbb{N}$ it is possible to write n uniquely as*

$$n = \sum_{k=0}^{\infty} n_k 2^k,$$

where either $n_k = 0$ or 1 for $k \in \mathbb{N}$. This expression will be called the binary expansion of n and the numbers n_k will be called the binary coefficients of n.

Definition 4. *The Walsh system $W = \{w_n, n \in \mathbb{N}\}$ is product of Rademacher functions in the following way. If $n \in \mathbb{N}$ has binary coefficients $\{n_k, k \in \mathbb{N}\}$ then*

$$w_n(x) = \prod_{k=0}^{\infty} r_k^{n_k}(x).$$

It is easy to see that this product is always finite, $w_0 = 1$ and $w_{2^n} = r_n$ for $n \in \mathbb{N}$. It is worth noticing that each Walsh function is piecewise constant with finitely many jump discontinuities on $[0, 1)$, and takes only the values of either $+1$ or -1.

3. The construction of the Walsh kernel

In this section we will define a new kernel function for support vector machines. The construction is based on Vapnik's idea in [1].

It is well-known, that the Walsh system is a complete orthonormal system on $[0, 1)$ and the Walsh system is a Schauder basis in L^p for $1 < p < \infty$ [2]. This linear space is a Hilbert space, where the inner product, which is denoted by $\langle \cdot, \cdot \rangle$, is the integral of the product of two functions over $[0, 1)$ [2].

It is well-known if f is an $\mathbb{R}$-valued, integrable function on the interval $[0, 1)$ then

$$f(x) = \sum_{k=0}^{\infty} a_k(f) w_k(x), \tag{1}$$

where $a_k = \langle f, w_k \rangle$.

Let N be a fixed element of the set $\mathbb{N}$, and

$$\mathbf{\Phi}_N(f) = (a_0(f), \dots, a_{N-1}(f)).$$

We would like to note it is convenient to imply the relations $N = 2^n$ and $n \geq 2$, where $n \in \mathbb{N}$. The reason for these relations is based on the theoretical characteristics of the Walsh functions.

Let f and g be $\mathbb{R}$-valued, integrable functions on the interval $[0, 1)$. Then the kernel function is

$$K_N(f, g) = \langle \mathbf{\Phi}_N(f), \mathbf{\Phi}_N(g) \rangle. \tag{2}$$

It is easy to see that $K_N(f, g)$ has the characteristics of the inner product in L^p, so $K_N(f, g)$ is an ordinary inner product in N-dimensional feature space.

It is easy to see, that $\mathbf{\Phi}(f)$ is a Walsh-transformation of the one-dimensional function f. This transformation is very informative, because it is the sum of the values of the function, which are multiplied with either $+1$ or -1 dependly on the co-ordinates of the points.

In digital signal processing the functions f and g are finite functions, whose domains are sets of connected finite subsets of $\mathbb{Z}$. In this case, the meaning of the function $\mathbf{\Phi}$ is the discrete Walsh transformation of the function f.

Let us consider the first elements of the vector $\mathbf{\Phi}(f)$. The first element is the sum of the values of the function in the points of the finite subset, the second element is the difference between the sum of the values of the functions on the first half interval, and the sum of the values of the functions on the second half interval. If the second element of the vector is equal to 0, then there is a balance between the avarage values of the function on these intervals. So the elements of this vector are special measures of the "symmetries" of the function, which can be useful for describing the signal. In digital image processing f is a 2-dimensional function. In this case, $\mathbf{\Phi}$ describes the "symmetries" of the image, which is represented by function f [3].

To determine the value of the kernel function $K_N(f, g)$ is not a difficult task, because some fast Walsh transformation are well-known in the literature [4].

To construct the support vector machine for the m-dimensional vector space it is enough to use the product of one-dimensional kernels, because a tensor of the kernel functions is a kernel function [1]. We will investigate the case of higher dimensional functions in Section 5.

4. The VC-dimension of the class of the Walsh functions

The Vapnik-Chervonenkis dimension has a very important role in the statistical learning. The VC-dimension of support vector machine characterizes the learning capacity of

the machine. With control of the VC-dimension one can avoid the overfitting of the support vector machine and one can minimize the expected value of the error [1]. So the knowledge of the VC-dimension of the class of the functions implemented by the machine is very important. In this section we investigate the VC-dimension of the class of Walsh functions. At first we quote some definitions from [1] which are important in order to understand the theoretical results given this section.

Definition 5. *An arbitrary $\{+1, -1\}$-valued function with domain $\mathbb{R}$ is called an indicator function.*

Definition 6. *Let f be an arbitrary indicator function. The sets $\{x \mid f(x) = +1, x \in \mathbb{R}\}$ and $\{x \mid f(x) = -1, x \in \mathbb{R}\}$ are the separated classes of the domain by using f.*

Definition 7. *The VC-dimension of a set of indicator functions is equal to the largest number h of points of the domain of the functions that can be separated into two different classes in all the 2^h possible ways using the functions of this set of functions. If for any n there exists a set of n points that can be shattered by the functions of the set, then the VC-dimension is equal to infinity.*

In the rest of the paper we assume that the domain of the Rademacher functions, the Walsh functions, and the indicator functions is the set $[0, 1)$.

Definition 8. *Let x be an arbitrary element of the interval $[0, 1)$. If x has the form $\frac{p}{2^n}$ for some $p, n \in \mathbb{N}$ ($0 \le p < 2^n$), we will call x dyadic rational in the interval $[0, 1)$.*

Definition 9. *Any $x \in [0, 1)$ can be written in the form*

$$x = \sum_{k=0}^{\infty} x_k 2^{-(k+1)},$$

where each x_k is equal to either 0 or 1. We will call it the dyadic expansion of x. When x is a dyadic rational there are two expressions of this form, one which terminates in 0's and one which terminates in 1's. In this case, the dyadic expansion of x will be the one which terminates in 0's.

Lemma 10. *Let x be arbitrary element of the interval $[0, 1)$ and $x_0, x_1, \ldots$ be the dyadic expansion of the x. Let n be arbitrary element of $\mathbb{N}$ and $n_0, n_1, \ldots$ binary coefficients of the n. Then the following equation is true*

$$w_n(x) = (-1)^{\sum_{k=0}^{\infty} n_k x_k}.$$

Proof. It is easy to prove using Definition 4. $\square$

Theorem 11. *The VC-dimension of the class of all the Walsh functions is equal to ∞.*

Proof. The proof will be constructive. Let h be an arbitrary element of the set $\mathbb{N}^+$. Let $\mathbf{x}_0, \ldots, \mathbf{x}_{2^h-1}$ be all the h-dimensional vectors with elements 0 and 1. Let us construct a matrix in the following form:

$$X = \left(\mathbf{x}_0, \ldots, \mathbf{x}_{2^h-1} \right).$$

Let us consider the rows of the matrix as a dyadic expansion of h numbers in the following form: The dyadic expansion of the ith number is $x_{i0}, \ldots, x_{i2^h-1}, 0, \ldots$. They are considered as points in the interval $[0, 1)$. Let us call them as $p_1, \ldots, p_h$.

Let $v_1, \ldots, v_h$ be an arbitrary sequence of the values $+1$ and -1. We will show that one of all the Walsh functions admits the values $v_1, \ldots, v_h$ at the points $p_1, \ldots, p_h$ of the interval $[0, 1)$. Let us select an index t which fulfils the following condition

$$x_{it} = \begin{cases} 0, & \text{if } v_i = 1, \\ 1, & \text{if } v_i = -1. \end{cases}$$

Let n be equal to 2^t. In this case $w_n(p_i) = (-1)^{\sum_{k=0}^{\infty} n_k x_{ik}}$ by Lemma 10. Because $n = 2^t$, only $n_t = 1$, and the other binary coefficients of the n are equal to 0. From this follows that $w_n(p_i) = (-1)^{x_{it}} = v_i$. So, the elements of the set W can shatter those points. Because the number h is an arbitrary element of the set $\mathbb{N}^+$, so the VC-dimension of the set of all the Walsh functions is equal to ∞. $\square$

The relations $N = 2^n$ and $n \ge 2$ are assumed in the following.

Lemma 12. *The elements of $W_N = \{w_k \mid k = 0, \ldots, N-1\}$ have $N-1$ jumps at most.*

Proof. The elements of W_N are the finite products of the Rademacher functions. By Definition 4 only the first n functions $r_k(x)$ play role in the product. Let J_k be the set of those points in $[0, 1)$ where $r_k(x)$ exhibits jump. It is easy to see $J_{k+1} \supset J_k$, for $k \in \mathbb{N}$. Because the functions of the Rademacher system are piecewise constant with finitely many jump discontinuities on $[0, 1)$, the elements of the W_N admit $|J_k|$ jumps at most. It is easy to see, that $r_n(x) \in W_N$, and the number of the jumps of $r_n(x)$ is equal to $N-1$. With this remark the lemma is proved. $\square$

Theorem 13. *The VC-dimension of the set $W_N = \{w_k \mid k = 0, \ldots, N-1\}$ equals n.*

Proof. From the proof of Theorem 11 we recall that the VC-dimension of the set W_N is not less than n, because we proved n points in the interval $[0, 1)$ can be shattered by one of the first 2^n Walsh functions.

On the other hand, by Lemma 12 a function in the set W_N admits $N-1$ jumps at most. Let us consider $n+1$

points $x_1, \ldots, x_n, x_{n+1}$ in the interval $[0, 1)$, which are in increasing order. Let f be a function of the following form:

$$f(x) = \begin{cases} 1, & \text{if } x = x_i \text{ and } (i \bmod 2) = 1, \\ -1, & \text{if } x = x_i \text{ and } (i \bmod 2) = 0, \end{cases}$$

where $1 \leq i \leq n+1$. Let us suppose we can find a function w_k in the set W_N for which the equation $f(x_i) = w_k(x_i)$ holds. In this case the function $w_k(x_i)$ exhibits at least N jumps. By Lemma 12, this is a contradiction, so the VC-dimension of the set W_N is not greater than n. $\square$

Definition 14. *Let $G = \{x_n \mid n \in \mathbb{N}\}$, where either $x_n = 0$ or $x_n = 1$. Set $I_0(x) = G$ for all $x \in G$. For each $x \in G$ and $n \in \mathbb{N}^+$ define*

$$I_n(x) = \{y \in G \mid y_i = x_i, 0 \leq i < n\}.$$

We call the sets of $I_n(x)$ the dyadic intervals of order n in $[0, 1)$.

Definition 15. *By a dyadic step function of order n we will mean a finite linear combination of characteristic functions of dyadic intervals of order n in $[0, 1)$.*

Notation 16. *We will use the following notation:*

$$\theta(x) = \begin{cases} 1, & \text{if } x >= 0, \\ -1, & \text{if } x < 0, \end{cases}$$

where $x \in \mathbb{R}$.

Definition 17. *Let $f_k(x)$ be $\mathbb{R}$-valued functions. We call the set of indicator functions $\theta(f_k(x) - t)$, where $t \in (\inf_x f_k(x), \sup_x f_k(x))$ the set of indicators for functions $f_k(x)$.*

Theorem 18. *The VC-dimension of the set $lin(W_N) = \{w(x) \mid w(x) = \alpha_0 w_0(x) + \ldots + \alpha_{N-1} w_{N-1}(x), \alpha_i \in \mathbb{R}\}$ is equal to N.*

Proof. Let f be an arbitrary finite linear combination of the elements in W_N. By Definition 4, the function f gets the following form:

$$f(x) = \sum_{i=0}^{N-1} \alpha_i w_i(x) = \sum_{i=0}^{N-1} \alpha_i \prod_{k=0}^{n-1} r_k^{i_k}(x),$$

where $\alpha_i \in \mathbb{R}$. So the function $f(x)$ is a finite linear combination of the finite products of the first n functions in the Rademacher system. From the proof of the Lemma 12 and because the elements of Rademacher system are piecewise constant with finitely many jump discontinuities on $[0, 1)$, the function $f(x)$ exhibits $N-1$ jumps at most. (By the continuity of the finite linear combination of continuous functions.) Let us suppose the function $f(x)$ exhibits $N - 1$ jumps.

From [2] it is well-known that any dyadic step function of order n is a finite linear combination of the elements of the set W_N. By Definition 17 it is enough to investigate the dyadic step functions of order n, which are indicator functions. Let $j_1, j_2, \ldots, j_{N-1}$ be points in increasing order in $[0, 1)$, where the function $f(x)$ exhibits a jump. Let x_i be a point in the interval $[j_i, j_{i+1})$ $(0 \leq i < N - 1)$, where $j_0 = 0$ and $j_N = 1$. These points can be shattered by the elements in $lin(W_N)$ by the previously established connection between the dyadic step functions and the finite linear combination of the elements in W_N. So the VC-dimension is not less than N.

Let us suppose the VC-dimension is greater than N. Let $x_0, x_1, \ldots, x_N$ be the points in the interval $[0, 1)$, which can be shattered by the finite linear combination of the elements in W_N. But in this case an element of $lin(W_N)$ exhibits N jumps, which is a contradiction, by Lemma 12. $\square$

Corollary 19. *The VC-dimension of the kernel function K_N is equal to N.*

In the following section we are going to investigate functions of higher dimensional spaces. The structure of the theorems will be similar to the one-dimensional case, because the results are based on Theorem 11 and Theorem 13.

5. The VC-dimension of the class of the higher dimensional Walsh functions

At first we are going to generalize the definition of the Walsh system (Definition 4) and the Walsh kernel (2) in the case m-dimensional space.

Definition 20. *The m-dimensional Walsh system $W^{(m)} = \{w_{(n_1, \ldots, n_m)}^{(m)} \mid n_i \in \mathbb{N}, m \in \mathbb{N}^+, m \geq 2, 1 \leq i \leq m\}$ is defined in the following way:*

$$w_{(n_1, \ldots, n_m)}^{(m)}(x_1, \ldots, x_m) = \prod_{k=1}^{m} w_{n_k}(x_k).$$

In the case of the m-dimensional Walsh system the form (1) gets the following form:

$$f(x_1, \ldots, x_m)$$
$$=$$
$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} a_{k_1, \ldots, k_m}(f) w_{(k_1, \ldots, k_m)}^{(m)}(x_1, \ldots, x_m).$$

Based on Definition 20 we can construct the function $\Phi_{(N_1, \ldots, N_m)}^{(m)}(f)$, which will give the values $a_{k_1, \ldots, k_m}(f)$, for $0 \leq k_i < N_i$ and $1 \leq i \leq m$. This function is a hypermatrix-valued function (or operator). We can define an inner product in the space of those hypermatrices as the

sum of the products of the corresponding elements of the hypermatrices. From this, it is easy to see the structure of the m-dimensional Walsh kernel.

Definition 21. *The m-dimensional Walsh kernel function gets the following form:*

$$K_{N_1,\ldots,N_m}^{(m)}(f,g) = \left\langle \boldsymbol{\Phi}_{(N_1,\ldots,N_m)}^{(m)}(f), \boldsymbol{\Phi}_{(N_1,\ldots,N_m)}^{(m)}(g) \right\rangle.$$

Theorem 22. *The VC-dimension of the class of all the m-dimensional Walsh functions is equal to ∞.*

Proof. Let $x_1, x_2, \ldots, x_n, \ldots$ be arbitrary points of the set $[0,1)$, which can be shattered by an element of the W. This system of the points exists by Theorem 11.

Let $x_1', x_2', \ldots, x_n', \ldots$ be the points, whose co-ordinates are equal to 0, except ith, which is equal to the co-ordinates of the points $x_1, x_2, \ldots, x_n, \ldots$, respectively. By Definition 20, the elements of $W^{(m)}$ can shatter these points. $\qquad\square$

The relations $N_i = 2^{n_i}$ and $n_i \geq 2$ $(1 \leq i \leq m)$ are assumed henceforth.

Theorem 23. *The VC-dimension of the set $W_{(N_1,\ldots,N_m)}^{(m)} = \{w_{(k_1,\ldots,k_m)}^{(m)} \mid k_i = 0,\ldots,N_i-1, 1 \leq i \leq m\}$ is equal to $\log_2(\prod_{i=1}^{m} N_i)$.*

Proof. At first, we will prove the VC-dimension of the set $W_{(N_1,\ldots,N_m)}^{(m)}$ is not less than $\log_2(\prod_{i=1}^{m} N_i)$. By Definition of VC-dimension it is enough to find $\log_2(\prod_{i=1}^{m} N_i)$ points in the set $[0,1)^m$, which can be shattered by the elements from $W_{(N_1,\ldots,N_m)}^{(m)}$.

Let us consider $\log_2(\prod_{i=1}^{m} N_i) = \sum_{i=1}^{m} n_i$ points $p_{k=\sum_{i=1}^{x-1} n_i + y}$ with the co-ordinates $(\mathrm{Pr}_1(p_k),\ldots,\mathrm{Pr}_m(p_k))$ in the set $[0,1)^m$ in the following form:

$$p_{(x,y)} = \begin{cases} (\frac{1}{2^{y+1}}, 0, \ldots, 0), & \text{if } x=1, 0 \leq y \leq n_1 - 1, \\ (0, \frac{1}{2^{y+1}}, 0, \ldots, 0), & \text{if } x=2, 0 \leq y \leq n_2 - 1, \\ \vdots & \\ (0, \ldots, 0, \frac{1}{2^{y+1}}), & \text{if } x=m, 0 \leq y \leq n_m - 1. \end{cases}$$

Let $v_{(1,0)}, \ldots, v_{(1,n_1-1)}, v_{(2,0)}, \ldots, v_{(m,n_m-1)}$ be an arbitrary sequence of the values $+1$ and -1. We will show that one from all the m-dimensional Walsh functions in $W_{(N_1,\ldots,N_m)}^{(m)}$ admits the values $v_{(1,0)}, \ldots, v_{(m,n_m-1)}$ at the points $p_{(1,0)}, \ldots, p_{(m,n_m-1)}$.

Let $u_{(x,0)}, \ldots, u_{(x,n_x-1)}, 0, \ldots$ be the binary expansion of u_x $(1 \leq x \leq m)$. Let us define the binary expansion of u_x in the following form:

$$u_{(x,y)} = \begin{cases} 0, & \text{if } v_{(x,y)} = 1, \\ 1, & \text{if } v_{(x,y)} = -1. \end{cases}$$

Then the function $w_{(u_1,\ldots,u_m)}^{(m)}(p_{(i,j)}) \in W_{(N_1,\ldots,N_m)}^{(m)}$ admits values $v_{(1,0)}, \ldots, v_{(m,n_m-1)}$ at the points $p_{(1,0)}, \ldots, p_{(m,n_m-1)}$ because by Lemma 10 and Definition 20.

$$w_{(u_1,\ldots,u_m)}^{(m)}(p_{(x,y)}) = \prod_{i=1}^{m}(-1)^{\sum_{k=1}^{\infty} u_{(i,k)} \mathrm{Pr}_i(p)_k}$$

From this follows

$$w_{(u_1,\ldots,u_m)}^{(m)}(p_{(x,y)}) = (-1)^{\sum_{k=0}^{\infty} u_{(1,k)} \mathrm{Pr}_1(p)_k} \cdot \ldots$$
$$\cdot (-1)^{\sum_{k=0}^{\infty} u_{(m,k)} \mathrm{Pr}_m(p)_k} =$$
$$(-1)^{\sum_{k=0}^{\infty} u_{(x,k)} \mathrm{Pr}_x(p)_k} =$$
$$(-1)^{u_{(x,y)} \mathrm{Pr}_x(p)_y} = (-1)^{u_{(x,y)}} = v_{(x,y)}.$$

On the other hand, let us consider $\log_2(\prod_{i=1}^{m} N_i) + 1$ points in the set $[0,1)^m$. Let us suppose we can find functions from the set $W_{(N_1,\ldots,N_m)}^{(m)}$, which can shatter these points.

It is very easy to see the number of the elements of $W_{(N_1,\ldots,N_m)}^{(m)}$ is equal to $\prod_{i=1}^{m} N_i$. The number of the all vectors with elements 0 and 1 of size $\log_2(\prod_{i=1}^{m} N_i) + 1$ is equal to $2^{\log_2(\prod_{i=1}^{m} N_i)+1} = 2 \cdot \prod_{i=1}^{m} N_i$, but there are only $\prod_{i=1}^{m} N_i$ functions, so the elements of the set $W_{(N_1,\ldots,N_m)}^{(m)}$ can not shatter these points. So the VC-dimension of the set $W_{(N_1,\ldots,N_m)}^{(m)}$ is not greater than $\log_2(\prod_{i=1}^{m} N_i)$.

Theorem 24. *The VC-dimension of the set $lin(W_{(N_1,\ldots,N_m)}^m)$ is equal to $\prod_{i=1}^{m} N_i$.*

Proof. The proof can be based on Theorem 18.

Corollary 25. *The VC-dimension of the kernel function $K_{(N_1,\ldots,N_m)}^{(m)}$ is equal to $\prod_{i=1}^{m} N_i$.*

6. Experimental results

For all experiments the `Mathlab` SVM toolbox developed by Steve Gunn was used [6]. For a complete test, several auxiliary routines have been added to the original toolbox.

A training data set of 46 images, 31 images containing a face and another 15 images with non-face patterns, is built. The images containing face patterns have been derived from the face database of IBERMATICA where several sources of degradation are modeled. All images in this database are recorded in 256 grey levels and they are of dimensions 320×240. The procedure for collecting face patterns is as follows. From each image a bounding rectangle of dimensions 128×128 pixels has been manually determined that includes the actual face. This area has been subsampled four times. At each subsampling, non-overlapping regions of 2×2 pixels are replaced by their average. Accordingly,

training patterns of dimensions 8×8 are built. The ground truth, that is, the class label $+1$ has been appended to each pattern. Similarly, 15 non-face patterns have been collected from images in the same way, and labeled by -1.

We have trained the three different SVMs indicated in Table 1. The trained SVMs have been applied to 414 test examples (249 face and 165 non-face) from the IBERMAT-ICA database that have not been included in the training set. The resolution of each test image has been reduced four times yielding a final image of dimensions 8×8. The test images are classified as non-face ones or face ones.

The following table gives the results on the test.

Table 1. Experimental Results

	Linear	Walsh	Polynomial
Time	2.3581	2.3432	2.5327
Number of Errors	9	8	7
Margin	0.66	4.58	2.17
Number of SV	15	12	8

7. Acknowledgement

References

[1] V.N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, 1998.

[2] F. Schipp, W.R. Wade, P. Simon and J. Pál. *Walsh Series: An Introduction to Dyadic Harmonic Analysis*. Adam Hilger, Budapest, 1990.

[3] A. Fazekas and A. Hajdu. Recognizing typeset documents using Walsh transformation, *Journal of Computing and Information Technology*, *accepted for publication*, 2000.

[4] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Addison-Wesley, New York, 1993.

[5] L. Devroye, L. Györfi and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.

[6] S. Gunn. Support vector machines for classification and regression, *ISIS Technical Report ISIS-1-98*, Image Speech Intelligent System Research Group, University of Southapton, 1998.

[7] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, 1990.