TRANSFORMATION-INVARIANT FUNCTIONS

FOR APPLICATIONS IN PATTERN RECOGNITION

Hiromitsu Hama

1983

Faculty of Engineering Science

Osaka University

TRANSFORAMTION-INVARIANT FUNCTIONS

FOR APPLICATIONS IN PATTERN RECOGNITION

by

HIROMITSU HAMA

TRANSFORMATION-INVARIANT FUNCTIONS FOR APPLICATIONS IN PATTERN RECOGNITION

by

HIROMITSU HAMA

Submitted in partial fulfillment of
the requirement for the degree of
DOCTOR OF ENGINEERING
(Information and Computer Sciences)
at
OSAKA UNIVERSITY
TOYONAKA, OSAKA, JAPAN
1983

ACKNOWLEDGEMENT

The author would like to express his deepest gratitude to Prof. Kokichi Tanaka who has given great encouragement and provided invaluable guidance since the author's undergraduate course.

He is heartily grateful to Prof. Tadao Kasami, Prof. Nobuki Tokura, Prof. Toshio Fujisawa, Prof. Kensuke Takashima of Department of Information and Computer Sciences, and Prof. Saburo Tsuji of Department of Control Engineering for their invaluable and pertinent suggestions.

He wishes to express his warmest thanks to Prof. Kazumi Yamashita of Osaka City University who gave generous support and constant guidance during the entire course of this work.

His cordial thanks are due to Prof. Masamichi Shimura of Tokyo Institutes of Technology for the motivations of this work and a good deal of helpful advice.

To Prof. Shigeru Okuda, Prof. Keiichi Ando, Prof. Takaaki Okumoto of Osaka City University he tenders his sincere thanks for their good instruction and useful suggestions.

Especially must he express his gratitude to Prof. Tadahiro Kitahashi of Toyohashi University of Technology for his generous support.

He is deeply grateful to Associate Prof. Shinichi Tamura, Dr. Minoru Tanaka of Osaka University, Prof. Masaharu Mizumoto, Prof. Masakazu Ozawa, Associate Prof. Tetsuo Asano of Osaka Electro-Communication University, Associate Prof. Akihiro Takeuchi of Wakayama University for their association with seminars and discussion.

He expresses a special gratitude to Mr. Mutsumi Sato of Kinki University and Dr. Yoneharu Fujita of Osaka Prefecture University for their fruitful discussions and hearty helps.

It is very happy for the author to thank Associate Prof. Syoichi Hosokawa, Dr. Jiro Okamoto, Dr. Chikato Fujiwara of Osaka City University for their constant encouragement and several relevant suggestions.

He wishes to express his thanks to Mr. Hideo Kawai for his useful helps. He is very grateful to Miss Taneyo Ukawa for her hearty kindness and generous assistance.

Finally, the author's personal notes of gratitude go to his parents and his wife for their moral support and encouragement.

PREFACE

This thesis intends to investigate transformation-invariant functions for applications in pattern recognition. Man does easily recognize very complicated patterns, even if they are distorted, tilted, rotated, and enlarged. To machines, it is quite difficult to identify patterns, which have wide variations in size, style, orientation, and so on, with one another. The functions which are invariant to admissible transformations, if found, can ease postprocessing for pattern recognition. The author has attacked this field and spent several years to arrive at the results described in this thesis.

Some transformation-invariant functions are constructed on a multi-layer series-coupled machine. 'Perceptrons' were first proposed as a model for neural networks with learning capacity. Random connections between the first layer and the second may be interesting from the view-point of modeling neural networks. They, however, suffer from the lower capabilities because of inefficient use of the coupling. From an engineering standpoint, it is tried to build up the systematic connections. The processes of composing transformation-invariant functions are applied to Boolean functions, Walsh-Hadamard power spectrums, and Fourier spectrums. Subset methods, which can be considered one kind of template matching methods, are introduced to achieve economy of computation time and required memory. The effectiveness of the theory is confirmed through computer simulations.

CONTENTS

LIST OF SYMBOLS

CHAPTER 2

X	$:=[x(0),x(1),---,x(n-1)]$, input pattern
V	$:=\{x(0),x(1),---,x(n-1)\}$, set of variables
F	$:=\{f(0),f(1),---,f(m-1)\}$, set of partial functions
$F(X)$	:F-image of X
$S(f)$	:support of f
$\|S\|$	:degree of S
B	$:=\{0,1\}$
B^n	:n-cube
$\beta(n)$	:set of all mappings from B^n to B
$1[P]$	:=1, if P is true. =0, if P is false.
$\#[P]$	:number of elements satisfying P
$OL(\psi)$	:L-order of ψ
$OT(\psi)$	:T-order of ψ
MF	:F-matrix
$M'F$	:G-invariant F-matrix
$f(AND,S)$	:AND-mask on S
$f(OR,S)$	:OR-mask on S
$f(PARITY,S)$	:$PARITY$-mask on S
$F(AND)$	:set of all AND-masks
$F(OR)$	:set of all OR-masks
$F(PARITY)$	:set of all $PARITY$-masks
G	:some transformation group
g	:element of G

CHAPTER 3

$w(k,X)$	: k-th coefficient of (WHT)
$H(n)$	: $2^n \times 2^n$ Hadamard matrix or $k^n \times k^n$ complex-valued Hadamard matrix
$H(n,k)$	: k-th row vector of Hadamard matrix
$W(X)$	:= $(1/N)X \cdot H(n)$, Walsh-Hadamard transform (WHT)
$\otimes$	: Kronecker product
$P(n)$	:= $[p(0),p(1),---,p(n)]$, 1-dimensional translation-invariant (WHT) power spectrum
$\{p(i,j)\}$	: 2-dimensional translation-invariant (WHT) power spectrum
$\mathbf{1}$	:= $[\underbrace{1,1,---,1}_{2^i}]$
$\mathbf{0}$	:= $[\underbrace{0,0,---,0}_{2^n-2^{n-i}}]$
$Q(n)$	:= $[q(0),q(1),---,q(n)]$, enlargement and reduction-invariant (WHT) power spectrum
$\{q(i,j)\}$	: 2-dimensional enlargement and reduction-invariant (WHT) power spectrum
Gv	:= $\{gv, gv^2\}$, vertical symmetry transformations
Gh	:= $\{gh, gh^2\}$, horizontal symmetry transformations
Gd	:= $\{gd, gd^2\}$, diagonal symmetry transformations
Gr	:= $\{gr, gr^2, gr^3, gr^4\}$, rotations by multiples of 90°
$\{s(i,j)\}$	: (WHT) power spectrum invariant to translations, enlargements, reductions, rotations by multiples of 90°, and so on

z	$:=\cos(2\pi/k)+j\cdot\sin(2\pi/k)$
R	$:=[1, z, ---, z^{k-1}]$
$g(l)$	:translation by l elements
$A(n)$	$:=(1/N)\overline{H(n)}\cdot T(n)\cdot H(n)$

CHAPTER 4

$f(x)\leftrightarrow F(k)$	:Fourier transform pair
$f(x)\overset{r}{\leftrightarrow}F(k)$	:modified Fourier transform pair
$Fa(k)$	:Fourier transform of $f(x-a)$
$Fs(k)$	:translation-invariant Fourier spectrum
$Fb(k)$	:Fourier transform of $f(x/b)$
$Fe(k)$	:enlargement and reduction-invariant Fourier spectrum
$Fd(kr, k\theta)$	:modified Fourier transform of $f(r, \theta-d)$
$Fr(kr, k\theta)$	:rotation-invariant Fourier spectrum
$Fn(k)$	:negative-positive reversion-invariant Fourier spectrum
$Fa(k)$	:axis-symmetry transformation-invariant Fourier spectrum
c	:central frequency
$c0$	:c of the reference pattern
Δs	$:=[arg(F(1))]_{2\pi}/2\pi$
Δe	$:=c0/c$
Δr	$:=arg(F(0,1))$
$\Delta r0$	:Δr of the reference pattern

CHAPTER 5

Aj	:subpartition
$\mathcal{A}$	:set of subpartitions
Bi	:partition
$\mathcal{B}$	:set of partitions
$\theta(N)$	:threshold for noise
$\theta(C)$	:threshold for covering features
$\theta(M)$	:threshold for covered bits by subpatterns
$\theta(P)$	:threshold for $m(P)$
$\theta(F)$	:threshold for $m(F)$
$\theta(U)$	:threshold for $n(P)$
$\theta(A)$	:threshold for shifted versions of a feature
¶	:neiborhood
$l(P)$	$:=m(p)+n(P)$
$m(P)$	:number of specified covered bits in a pattern
$n(P)$	:number of specified uncovered bits in a pattern
$l(F)$	$:=m(F)+n(F)$
$m(F)$	:number of specified covered bits in a feature
$n(F)$	:number of specified uncovered bits in a feature
$g(l)$	:translation by l bits
F-table	:matrix with row Fi and column mj

CHAPTER 1
INTRODUCTION

Character recognition, speech recognition, medical diagnosis, and remote sensing appear to be one of the most interesting applications in the field of pattern recognition. Though they have an abundant literature, there is still a wide gap between human and machine in recognition of patterns without restriction. We do easily recognize complicated patterns, even if they are distorted. We can make out a character, even if it is tilted and written on any place of a sheet of paper and it is large or small. To machines, a large character and a small one are quite different. In optical character readers (OCR) and speech recognition machines on the market, some restraints are put on an input pattern. Input patterns had better be allowed to have wide variations in size, style, orientation, and so on within admissible transformations.

According to Felix Klein's mathematical viewpoint, every interesting geometrical property is invariant to any element of some transformation group. An input pattern is often normarized through appropriate preprocessing. Group-invariant functions, if found, could ease postprocessing. This paper aims at a basic research to achieve some transformation group-invariant system for applications in pattern recognition. Pattern recognition machines are generally designed to tolerate certain transformations and noise. According to the interesting indication by Ullmann [1], [2], a transformation changes the spacial arrangement of points of a pattern without changing their associated values, as typified by a distortion on a rubber sheet. Noise, on the other hand, changes the values, but does not change the spacial arrangement of the points. There are a number of techniques to deal with noisy and transformed patterns [1]-[9], one of which is a subset method investigated later. Transformations are divided into two classes, local ones such as distortions and global ones such as translations, enlargements, reductions, rotations and so on.

The theory of computational geometry was first developed by Minsky and Papert [10], [11], and next extended to 'Analog Perceptrons' with real-valued inputs and output by Uesaka [12]. They have made theoretical investigations of what is called perceptrons. Perceptrons were first suggested by Rosenblat [13] as a model for neural networks with learning capacity. Since then, various models of perceptrons have been proposed. We have been carried formal and mathematical investigations in a three-layer series-coupled machine without feedback loop and lateral coupling [14]-[25]. One of the most important problems in this type of machine is determination of connections between the first layer and the second. The connections are decisive on the capacity of the machine. Random connections may be interesting from the viewpoint of modeling neural networks in a biological system. They, however, suffer from the lower capabilities relative to the number of connections because of inefficient use of the coupling. From an engineering standpoint, it will be necessary to build up systematic connections with consideration of applications.

With this in mind, we make theoretical researches on the efficient construction of this type of machine in Chapter 2. The functions on the second layer are called *partial functions*, which are mainly Boolean functions in this chapter. First, the necessary and sufficient condition for a *complete system*, of which any function can be expressed by a linear threshold function, is derived. Among complete systems, the system which has the minimum number of connections between the first layer and the second is considered, which is called a *minimum complete system*. For parallel computation machines such as perceptrons, it is important what range each partial function should deal with. We define the locality of computation by T-order and L-order. A minimum complete system proposed here is useful for decision of these orders.

Next, we introduce a transformation group G on an input space, and show how to construct functions which are invariant under G. As a method of realizing such *G-invariant functions*, Minsky and Papert [11] have already presented a method of equating the coefficients of partial functions in the same G-equivalence class. This is summarized in 'Group-Invariance Theorem', which is one of the most useful mathematical tools. In contrast to their method, this paper describes a method by which every partial function itself has G-invariant property. Then any function on the last layer is of course G-invariant, since the function is expressed by a linear threshold function of partial functions. If a family F of partial functions is linearly independent and closed under G, the coefficients of partial functions depend only on the G-equivalence class in a linear expression of a G-invariant function. In other words, two partial functions have the same coefficient, if they belong to the same G-equivalence class. This is the necessary and sufficient condition here. This was the sufficient condition in the system of Minsky and Papert, since there was tolerance about expression of functions. The family of partial functions, which is G-invariant and has a linear expression of any G-invariant function, is called a *G-invariant complete system*. Such a system can make distinctions between G-nonequivalent patterns. It is investigated how to realize a G-invariant complete system.

In Chapter 3 Walsh functions [26]-[31] are used as partial functions. Walsh-Hadamard transform (WHT) has the advantage of computational simplicity when compared with Fourier transform. The Fourier power spectrum is invariant to translations, but the (WHT) power spectrum is not. The (WHT) for image processing has been discussed by Andrews and Hunt [32]. The axis-symmetry-histograms developed by Alexandridis and Klinger [33] are invariant to translations and rotations by multiples of 90°, but they require normalization of an input pattern through the Fourier transform. The (WHT) power spectrum developed by Ahmed, Rao and Abdussatar [31], [34] is invariant to translations, which is obtained directly through the (WHT). But it is not invariant to enlargements, reductions, and so on.

In this chapter it is described how to develop the (WHT) power spectrum to be invariant under all of translations, enlargements, reductions, rotations by multiples of 90°, and so on [35]. A composing process is introduced, which is available for obtaining the modified (WHT) power spectrum having these transformations-invariant properties. It is based on the Group-Invariance Theorem developed in Chapter 2. The main idea is to make a linear combination of G-equivalent functions. First, a $G1$-invariant (WHT) power spectrum is constructed. Next, a permutation group on the $G1$-invariant (WHT) power spectrum caused by $G2$ operating on an input pattern is found. Thus we arrive at a (WHT) power spectrum being constant under both $G1$ and $G2$, where $G1$ and $G2$ are some transformation groups. Repeating this process, a power spectrum can be derived from the (WHT) power spectrum that is invariant to a more general group of transformations. The real-valued (WHT) is mainly dealt with, which is convenient in treating transformations in the form of $2^{\pm m}$. When the complex-valued (WHT) is adopted for convenience of treating transformations in the form of $k^{\pm m}$, the G-invariant (WHT) power spectrum is obtained in the same composing process [36], [37]. Many interesting results are got through the composing process [35]-[47].

The G-invariant (WHT) power spectrum developed here may be regarded as a feature. It is also possible to adopt any other functions besides power spectrums. There are many other translation-invariant functions such as R-transform [48], [49] which is computed only through addition and subtraction, M-transform [50] which is done only through logical product and sum, BIFORE power spectrum [34], and so on. As pointed out by Arazi [51], these transformation-invariant functions are constant under other transformations besides objective transformations. In other words, they are not G-invariant complete systems. The Fourier transform is convenient in treating more general transformations.

In Chapter 4 Fourier sinusoids are used as partial functions. Changes in Fourier spectrum of an input pattern are investigated under several transformations such as translations, enlargements, reductions, rotations, and so on. Then the Fourier spectrum is developed to be invariant under these transformations [52]. It is well known that Fourier power spectrum and auto-correlation functions are translation-invariant [53] etc.. They preserve only amplitude information and are not a translation-invariant complete system. On one hand the importance of phase information in image processing is pointed out by Oppenheim and others [54]-[56]. The spectrum developed here is a transformation-invariant complete system and any essential information is not lost. We can regenerate (a representative of the class of) an input pattern through the inverse Fourier transform. Parameters introduced in this chapter represent the degree of translation, enlargement, reduction, and rotation, so they can be used for normalization of an input pattern. The normalization is less affected by local distortion and noise, since the Fourier transform is a global transform depending on the whole pattern. The efficiency of our theory is confirmed through computer simulation.

A simple template-matching recognition technique by using an associative memory was presented by Yau and Yang [57]. Classical template matching is of limited usefulness in advanced picture analysis systems. However, a hierachical template matching can be used even in cases where patterns are subject to distortions [5]. The process of finding best matching takes a long time. It may save us computation time and required memory to determine the correspondence between small parts of an input pattern instead of the whole pattern. The idea of using subsets has been developed by several people [1], [2], [4], [5], [6], [58]. Chapter 5 aims at a basic research to achieve automatic design of recognition and representation systems of highly variable patterns by a *subset method*. The design of features usually takes a great deal of human effort. It is desirable to automate the determination of features. In an example, our algorithm automatically extracts a few features from several patterns of the portion where the upper stroke joins the bottom loop in closed-loop 6's written without constraint. *Admissible distortions* are given *a priori*, but it is also a problem how to determine them. The investigation as to learning algorithm for them was conducted by Ullmann [1], [59].

In general, the set of automatically determined features may include redundant features. Features could be linearly ordered by a criterion of the estimated error or entropy [60]. Elashoff and others [61] showed that for optimum selection of a subset of features, the features generally may not be evaluated independently. A counter example was given to a possible claim that the best subset of features must contain the best single feature [62]. After learning of features, some of them are selected to obtain a *min-max-cover* which covers as many bits of patterns in the training set as possible. The selection problem is represented by a table (*F*-table). The table is an extension of a prime implicant table which has been used in a classical problem of minimizing a Boolean function, that is, selection of the minimum cost subset of prime implicants of a Boolean function. There are a great number of techniques for minimum or nonminimum irredundant solutions [63]-[73]. Three well-known reduction rules, which are based on *row dominance*, *column dominance* and *row essentiality* [65],[66], allow, in general, large simplification in the determination of min-max-covers. Differing from prime implicants, redundant features in the *F*-table are not always useless but often useful for increasing the reliability in recognition.

After obtaining a mim-max-cover, a classifier function is constructed in the form of a *product-of-sums*, whose geometrical interpretation will be given later, of features in the min-max-cover. Then, we introduce a product which works as a design tool for the sums in the product-of-sums. All patterns in the training set are correctly classified into the corresponding classes by the classifier function. Patterns in the *on-set* of it can be generated by possible combinations of features. A fundamental weakness of product-of-sums is pointed out by Ullmann [1], [2], [58]. On the contrary to the weakness, it has a good point of flexibility and economy. The reliability of a classifier function can be increased by using overlapping covers. Therefore product-of-sums expression is adopted here. The aspects of constructing a classifier function are shown in a simple example.

CHAPTER 2

MULTI-LAYER SERIES-COUPLED MACHINES

2.1 Introduction

In this chapter, theoretical investigations on the construction of multi-layer (mainly three-layer) series-coupled machines will be conducted. The first layer is an input space (an input pattern) and the last is an output space (a linear threshold function or a linear combination of partial functions). Functions on the middle layers are called *partial functions*. Our main purpose is to find a family of partial functions which is invariant under some transformation group G and is able to make distinctions between G-nonequivalent patterns. Such a family is called a *G-invariant complete system*. The system is generally required to have the additional property that any G-invariant function has a linear expression by it.

First, mathematical terms and definitions will be given in preparation for the following discussions. Next, the necessary and sufficient condition for a *complete system*, by which any function has a linear expression, will be given. The complete system which has the minimum number of connections between the first layer and the second is called a *minimum complete system*. It will be considered how to construct such a system. Lastly, 'Group-Invariance Theorem' and other theorems, which are useful mathematical tools to construct group-invariant functions, will be introduced. A composition method of a group-invariant complete system will be investigated.

2.2 Multi-Layer Series-Coupled Machines

The general scheme of a three-layer series-coupled machine is illustrated in Fig. 2.1. The first layer V is an input space consisting n points. The second F is a family of functions which are called *partial functions*. The values of partial functions are computed independently of one another. This is highly important for parallel computation. The last layer ψ is represented by a linear threshold function of partial functions or a linear combination of them.

Geometric patterns are often drawn by using black points and white ones. So we assume that an input pattern $X=[x(0),x(1),---,x(n-1)]$ is expressed by an n-vector of 1's and 0's which correspond to black points and white ones. We use $x(i)$ instead of x_i to avoid complicated subscripts. Let B be the set $\{0,1\}$, an n-cube B^n be the direct product of N B's, and $\beta(n)$ be the set of all mappings from B^n to B. X is an element of B^n and any function is an element of $\beta(n)$. The set of variables is expressed by $V=\{x(0),x(1),---,x(n-1)\}$ and the set of partial functions by $F=\{f(0), f(1), ---, f(m-1)\}$ or simply by $F=\{f\}$. Every partial function is a mapping from B^n to B. Let $f(i,X)$ denote the value of the i-th partial function for X, then an m-vector $F(X)=[f(0,X),f(1,X),---, f(m-1,X)]$ is called the F-image of X.

Let $|S|$ denote the number of members in a set S and be called the *degree* of S. If a subset S of V satisfies the following conditions, we call it the *support* of f and express it by $S(f)$. $|S(f)|$ is called the *degree* of f.

1. $f(X)=f(Y)$ for any X, and it is satisfied that

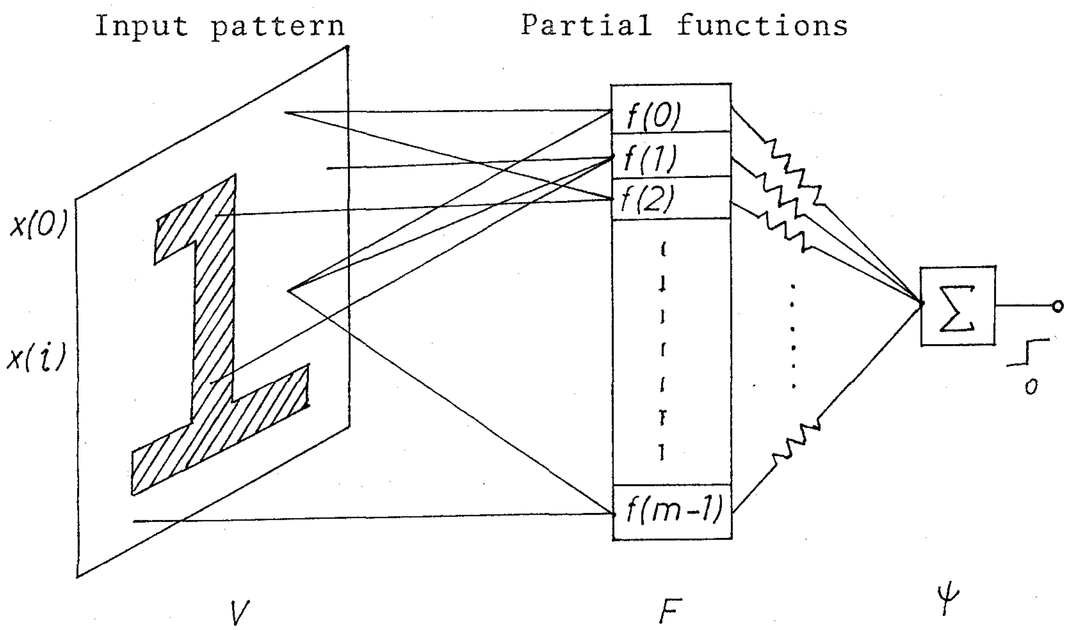


Fig. 2.1. A three-layer series-coupled machine.

$$y(i)=\begin{cases} x(i), & \text{if } y(i) \text{ is in } S, \\ \text{constant}, & \text{if } y(i) \text{ is in } V-S. \end{cases} \tag{2.1}$$

2. We can't choose S' which is a proper subset of S such that $f(X)=f(Z)$ and

$$z(i)=\begin{cases} x(i), & \text{if } z(i) \text{ is in } S', \\ \text{constant}, & \text{if } z(i) \text{ is in } V-S' \end{cases} \tag{2.2}$$

where $i=0,1,---,n-1$. For example, $S(x(0)\vee x(1))=\{x(0),x(1)\}$, $|S(x(0)\vee x(1))|=2$, $S((x(0)\vee x(1))\wedge x(1))=\{x(1)\}$, $|S((x(0)\vee x(1))\wedge x(1))|=1$, where $\vee$ and $\wedge$ denote ordinary Boolean notations.

A function $\psi(X)$ on the third layer has an output of 1 or 0 according to whether an input pattern X satisfies a proposition P or not. ψ is a member of $\beta(n)$ and usually expressed by a linear threshold function. Let $1[P]$ be defined as

$$1[P]=\begin{cases} 1, & \text{if } P \text{ is true} \\ 0, & \text{if } P \text{ is false}. \end{cases} \tag{2.3}$$

For instance, $1[0\leqq 1]=1$, $1[0>1]=0$. If there exist real numbers $a(i)$'s such that

$$\psi(X)=1[\textstyle\sum_{i=0}^{m-1} a(i)f(i,X)>0] \tag{2.4}$$

for any X in B^n then (2.4) is called a *T-expression* of ψ by F. We sometimes rewrite this more simply as

$$\psi=1[\textstyle\sum_{f\in F} a(f)f>0]. \tag{2.5}$$

The family of subsets of $\beta(n)$ which give a T-expression of ψ is deoted by $T(\psi)$. The *T-order* of ψ is defined as

$$OT(\psi)= \min_{F\in T(\psi)} \max_{f\in F} |S(f)|. \tag{2.6}$$

Similarly, if there exist real numbers $a(i)$'s such that

$$\psi(X)=\sum_{i=0}^{m-1} a(i)f(i,X). \tag{2.7}$$

for any X in B^n, then (2.7) is called a *L-expression* of ψ by F and simply rewritten as

$$\psi=\sum_{f\in F} a(f)f. \tag{2.8}$$

The family of subsets of $\beta(n)$ which give a L-expression of ψ is denoted by $L(\psi)$. The *L-order* of ψ is defined as

$$OL(\psi)=\min_{F\in L(\psi)} \max_{f\in F} |S(f)|. \tag{2.9}$$

Both $OT(\psi)$ and $OL(\psi)$ depend only on ψ and not on the choice of F. As seen from the definition, if a family gives a L-expression of ψ, it also gives a T-expression of ψ. The converse to this, however, is not always true. Consequently, for any ψ

$$OT(\psi)\leqq OL(\psi)\leqq |S(\psi)|. \tag{2.10}$$

When F is an infinite set, (2.5) and (2.8) are defined by quadratic mean convergence, that is,

$$\psi=1[\lim_{m\to\infty}\sum_{i=0}^{m-1} a(i)f(i)>0] \tag{2.5'}$$

$$\psi=\lim_{m\to\infty}\sum_{i=0}^{m-1} a(i)f(i). \tag{2.8'}$$

The orders $OT(\psi)$ and $OL(\psi)$ express an important concept of locality in parallel computation. If ψ has small orders, the partial functions needed are easy to compute in parallel, because each depends only on a small part of the whole input pattern. To illustrate one simple concept of locality, we state a fact about convexity. An input pattern X fails to be *convex* if and only if there exist three points such that $x(i2)$ is in the line segment joining $x(i1)$ and $x(i3)$, and $x(i1)=x(i3)=1$, $x(i2)=0$. Thus we can test for convexity by examining all triplets of points. If all the triplets pass the test then X is convex. If at least one triplet meets all the conditions above, then X is not convex, that is, *concave*. All the tests can be done independently. Fig 2.2 shows a convex pattern and a concave one. Then the order of 1[X is convex] is less than or equal to three. A function ψ is said to be *global*, if the order of Ψ is n. There exist two and only two global functions $x(0)\oplus x(1)\oplus{-}{-}{-}\oplus x(n-1)$ and $\overline{x(0)\oplus x(1)\oplus{-}{-}{-}\oplus x(n-1)}$ [14], where "$\oplus$" denotes 'exclusive or'.

The following three transformations are called *isomorphisms*:(1) exchanging the i-th element of X for the j-th, (2) negation of the i-th element, (3) negation of the function. Let

$$X=[x(0),x(1),{-}{-},x(i),{-}{-},x(j),{-}{-},x(n-1)]$$

$$X1=[x(0),x(1),{-}{-},x(j),{-}{-},x(i),{-}{-},x(n-1)]$$

$$X2=[x(0),x(1),{-}{-},\overline{x(i)},{-}{-},{-}{-}{-}{-},{-}{-},x(n-1)]$$

$$\psi1(X)=\psi(X1),\quad \psi2(X)=\psi(X2),\quad \psi3(X)=\overline{\psi(X)} \tag{2.11}$$

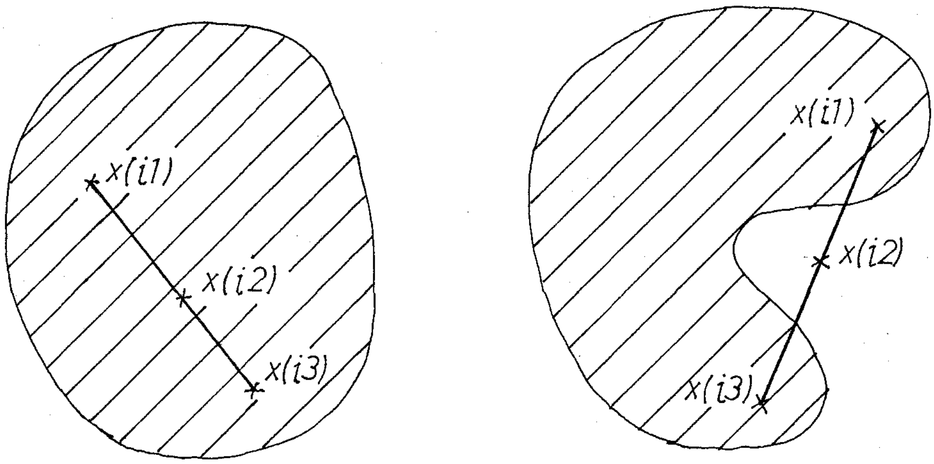
then ψ, $\psi1$, $\psi2$, and $\psi3$ are also called *isomorphism functions* and

$$OT(\psi)=OT(\psi1)=OT(\psi2)=OT(\psi3)$$

$$OL(\psi)=OL(\psi1)=OL(\psi2)=OL(\psi3)\,. \tag{2.12}$$

In other words, the order is unchanged by isomorphisms. This is easy to understand without proof [20], [21]. The set $\{F(X)\}$ of all F-images is also unchanged by isomorphisms [15].

Before we prove the theorem which gives the upper limit of the orders, we state a few propositions.



A convex pattern A concave pattern

Fig. 2.2. A convex pattern and a concave pattern.

[Proposition 2.1] Assume that ψ has two T-expressions as follows:

$$\psi=1[a(0)f(0)+\sum_{i=0}^{m-1} a(i)f(i)>0]$$
$$=1[b(0)f(0)+\sum_{i=0}^{m-1} b(i)f(i)>0] \qquad (2.13)$$

where $a(0)b(0)<0$. Then there exists a T-expression of ψ by $F-\{f(0)\}$:

$$\psi=1[\sum_{i=1}^{m-1}((a(i)/|a(0)|)+(b(i)/|b(0)|))f(i)>0]$$
$$F=\{f(0),f(1),---,f(m-1)\}. \qquad (2.14)$$

(Proof) When $\psi(X)=1$, we have

$$(a(0)/|a(0)|)f(0)+\sum_{i=1}^{m-1}(a(i)/|a(0)|)f(i)>0$$
$$(b(0)/|b(0)|)f(0)+\sum_{i=1}^{m-1}(b(i)/|b(0)|)\ (i)>0. \qquad (2.15)$$

Because $a(0)b(0)<0$, we obtain $(a(0)/|a(0)|)+(b(0)/|b(0)|)=0$. Therefore

$$\sum_{i=1}^{m-1}((a(i)/|a(0)|)+(b(i)/|b(0)|))f(i)>0. \qquad (2.16)$$

When $\psi(X)=0$, we conclude in the same way that

$$\sum_{i-1}^{m-1}((a(i)/|a(0)|)+(b(i)/|b(0)|))f(i)\leqq 0. \qquad (2.17)$$

QED.

[Proposition 2. 2] Let $\psi(AND,l)=f1\wedge f2\wedge---\wedge fl=\bigwedge_{i=1}^{l} fi$, $\psi(OR,l)=f1\vee f2\vee ---\vee fl=\bigvee_{i=1}^{l} fi$ and $\psi(PARITY,l)=f1\oplus f2\oplus---\oplus fl=\bigoplus_{i=1}^{l} fi$, then we have

$$OT(\psi(AND,l))\leqq \min_i(\max_{j\neq i}(OT(fi),OL(fj)),\quad OL(\psi(AND,l))\leqq\sum_{i=1}^{l} OL(fi)$$

$$OT(\psi(OR,l))\leqq \min_i(\max_{j\neq i}(OT(fi),OL(fj)),\quad OL(\psi(OR,l))\leqq\sum_{i=1}^{l} OL(fi)$$

$$OT(\psi(PARITY,l))\leqq\sum_{i=1}^{l} OT(fi),\quad OL(\psi(PARITY,l))\leqq\sum_{i=1}^{l} OL(fi). \qquad (2.18)$$

When $S(fi)$ and $S(fj)$ are disjoint for any i and j $(i\neq j)$, we obtain

$$OT(\psi(AND,l))\geqq\max_i OT(fi),\quad OL(\psi(AND,l))=\sum_{i=1}^{l} OL(fi)$$

$$OT(\psi(OR,l)) \geqq \max_i OT(fi),\quad OL(\psi(OR,l))=\sum_{i=1}^{l} OL(fi)$$

$$OT(\psi(PARITY,l))=\sum_{i=1}^{l} OT(fi),\quad OL(\psi(PARITY,l))=\sum_{i=1}^{l} OL(fi). \tag{2.19}$$

It is same thing with any isomorphism function.

(Proof) We briefly prove the assertion for $f1 \wedge f2 (l=2)$ in the simplest case. Suppose that

$$\begin{aligned} f1 &= \sum_{i=0}^{m-1} a(i)f(i) = 1[\sum_{i=0}^{m-1} b(i)f(i)>0] \\ f2 &= \sum_{j=0}^{m-1} c(j)f(j) = 1[\sum_{j=0}^{m-1} d(j)f(j)>0] \end{aligned} \tag{2.20}$$

then

$$\begin{aligned} \psi(AND,2) = f1 \wedge f2 &= 1[\sum_{i=0}^{m-1} b(i)f(i) + N2 \sum_{j=0}^{m-1} c(j)f(j) > N2] \\ &= 1[\sum_{j=0}^{m-1} d(j)f(j) + N1 \sum_{i=0}^{m-1} a(i)f(i) > N1] \\ &= (\sum_{i=0}^{m-1} a(i)f(i))(\sum_{j=0}^{m-1} c(j)f(j)) \\ &= \sum_{i,j=0}^{m-1} a(i)c(j)(f(i)\cdot f(j)) \end{aligned} \tag{2.21}$$

where $N1 \geqq \max_X(\sum_{j=0}^{m-1} d(j)f(j,X))$, $N2 \geqq \max_X(\sum_{i=0}^{m-1} b(i)f(i,X))$. Therefore we conclude that

$$\begin{aligned} &OT(\psi(AND,2)) = OT(f1 \wedge f2) \leqq \min(\max(OL(f1),OT(f2)),\max(OL(f2),OT(f1))) \\ &OL(\psi(AND,2)) = OL(f1 \wedge f2) \leqq OL(f1)+OL(f2). \end{aligned} \tag{2.22}$$

Suppose that $S(f1)$ and $S(f2)$ are disjoint. Since $\psi(X)=f1(X)$ for any X such that $f2(X)=1$, we obtain

$$OT(\psi(AND,2)) \geqq OT(f1). \tag{2.23}$$

Similarly we have

$$OT(\psi(AND,2)) \geqq OT(f2). \tag{2.24}$$

Hence

$$OT(\psi(AND,2)) \geqq \max(OT(f1),OT(f2)). \tag{2.25}$$

The equality is not satisfied in general. On the other hand we have

$$OL(f1\wedge f2)=\max_{i,j(a(i)c(j)\neq 0)}|S(f(i))\cup S(f(j))|$$

$$=\max_{i(a(i)\neq 0)}|S(f(i)|+\max_{j(c(j)\neq 0)}|S(f(j)|$$

$$=OL(f1)+OL(f2) \qquad (2.26)$$

since $S(fi)\cap S(fj)$ is an empty set from the supposition. The proof about the orders of $f1\vee f2$ and $f1\oplus f2$ is given in the similar way [21]. QED.

There exist partial functions $f1$ and $f2$ of order 1 such that $f1\wedge f2$ and $f1\vee f2$ become of rather large order, that is, the orders are not of bounded order as $|V|$ becomes large [11]. If there is not an input pattern X such that $e(X1)<e(X)<e(X2)$ for a function e which is not generally a Boolean function, the pair of two patterns $X1$ and $X2$ is called a *boundary*. Where $X1$ and $X2$ belong to different classes, that is, $[\psi(X1), \psi(x2)]=[1,0]$ or $[0,1]$. Furthermore it is supposed that there is no pair such that $e(X1)=e(X2)$.

[Theorem 2.1](Upper Limit of T-order) If there are k boundaries, then

$$OT(\psi)\leqq k\cdot\max_i|S(f(i))| \qquad (2.27)$$

where $e(X)=\sum_i a(i)f(i,X)$.

(Proof) Let k boundaries be expressed by $(e(X(2j-1)),e(X(2j))$ and $e(X(2j-1))<b(j)<e(X(2j))$ $(j=1,2,---,k)$, then we obtain

$$\psi(X)=1[\prod_{j=1}^{k}(e(X)-b(j))>0] \text{ or } 1[\prod_{j=1}^{k}(e(X)-b(j))<0]. \qquad (2.28)$$

Therefore we conclude that

$$OT(\psi)\leqq k\max_i|S(f(i))|. \qquad (2.29)$$

QED.

[Corollary 2. 1] If there are k boundaries, then $OT(\psi)\leqq k$, where $e(X)=\sum_i a(i)x(i)$. This is the special case of Theorem 2.1 when $f(i,X)=x(i)$.

[Corollary 2.2] If $|\{X:\psi(X)=0\}|=k$ or $|\{X:\psi(X)=1\}|=k$, then $OT(\psi)\leqq 2k-1$.

[Example 2.1] In Corollary 2.1 the number k of boundaries depends on $\{a(i)\}$. When $|V|=n\leqq 3$, it is seen from testing representatives of isomorphism functions that we can choose $\{a(i)\}$ such that $OT(\psi)=k$ except for one [20]. Fig 2. 3 shows representative functions when $n=3$ and a black point means an input pattern point X such that $\psi(X)=1$. Table 2.1 shows one example of $a(i)$'s which satisfy that $OT(\psi)=k$. $\psi 11$ is omitted because there do not exist suitable $a(i)$'s.

2.3 Complete Systems

If one can choose $a(i)$'s such that they satisfy $\sum_{i=0}^{m-1} a(i)f(i)=0$ and are not simultaneously zero, $F=\{f(i)\}(i=0,1,---,m-1)$ is called *linearly dependent*, otherwise *linearly independent*. If for any ψ in $\beta(n)$ there exists at least one T-expression of ψ by F, F is called a *T-complete system*. A *L-complete system* is defined in the same way. Later we know a T-complete system and a L-complete system are the same thing, so we will simply call them *complete systems*. In this section we describe first the necessary and sufficient condition for a complete system. Next we do how to construct a *minimum complete system* which has the minimum number of connections between the first layer and the second. The system is useful to decide orders of a function. The fact that F is a complete system implies that F-images of all elements in B^n are arbitrarily separable and F-images of two or more elements in B^n do not coincide with

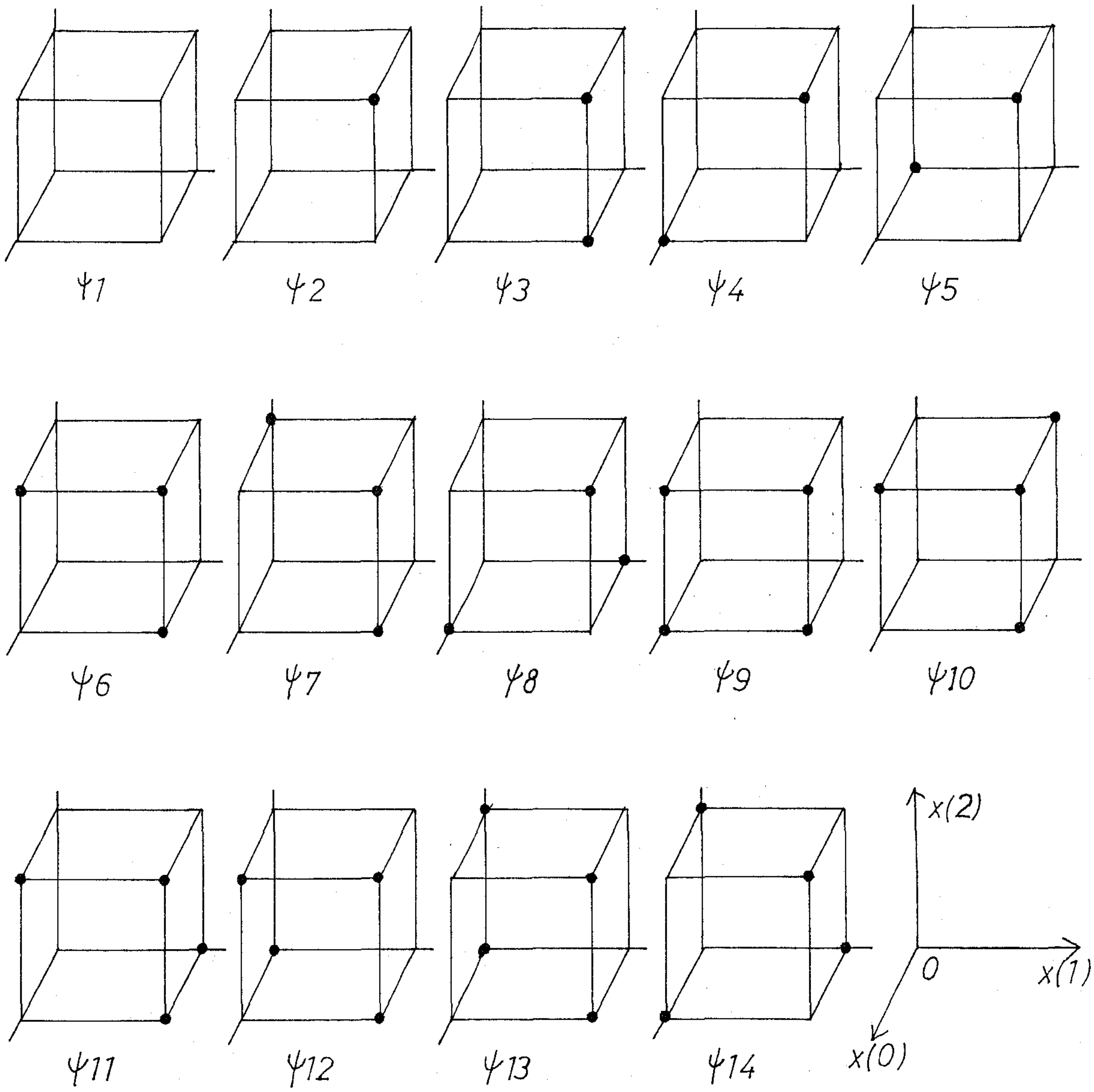


Fig. 2.3. Representatives of Boolean functions with three variables. (• :on-set).

Table 2.1. T-order and $a(i)$'s.

T-order	functions	$a(0)$	$a(1)$	$a(2)$	k
0	$\psi 1$	0	0	0	0
1	$\psi 2$	1	1	1	1
1	$\psi 3$	1	1	0	1
1	$\psi 6$	2	1	1	1
1	$\psi 9$	1	0	0	1
1	$\psi 10$	1	1	1	1
2	$\psi 4$	2	-1	1	2
2	$\psi 5$	1	1	1	2
2	$\psi 7$	2	2	-1	2
2	$\psi 8$	1	1	-1	2
2	$\psi 12$	2	1	1	2
2	$\psi 13$	1	1	0	2
3	$\psi 14$	1	1	1	3

each other.

[Lemma 2.1] The necessary and sufficient condition for N points to be arbitrarily linearly separable is that those N points span an $(N-1)$-dimensional space, that is, they do not stay on an $(N-2)$-dimensional space.

(Proof) Cover and others have determined the number of linear dichotomies of N points [74]. Using their results, this is easy to prove. QED.

Let $F=\{f(0),\ f(1),\ ---, f(m-1)\}$ and $B^n=\{X(0),X(1),---,X(N-1)\}(N=2^n)$, then a F-matix MF is defined as

$$MF=\begin{bmatrix} f(0,X(0)) & ,---, & f(m-1,X(0)) \\ --- & ,---, & --- \\ f(0,X(N-1)) & ,---, & f(m-1,X(N-1)) \end{bmatrix}. \tag{2.30}$$

When $N=m$, the determinant of the above matrix is denoted by $|MF|$. In the above definition the i-th row vector $[f(0,X(i)),---,f(m-1,X(i))]$ represents the F-image of $X(i)$ and the j-th column vector $[f(j,X(0)),---,f(j,X(N-1))]^t$ represents the partial function vector $f(i,B^n)$. Where the superscript "t" means a transposed vector.

[Theorem 2.2] (Necessary and Sufficient Condition for a Complete System) The necessary and sufficient condition for F to be a complete system is that $rank(MF)=N$. When $N=m$, this condition can be rewritten as $|MF|=0$.

(Proof) From Lemma 2.1 the necessary and sufficient condition for F to be a complete system is that the N row vectors of MF are linearly independent. Consequently, the condition is $rank(MF)=N$. QED.

When the number of partial functions in a complete system exceeds N, it is possible to choose N linearly independent partial functions so that they constitute a complete system.

[Corollary 2.3] A complete system F gives a L-expression to any ψ in $\beta(n)$. When the number of partial functions is minimum, that is, $|F|=N$, the expression is unique.

(Proof) Let $f(i,B^n)=[f(i,X(0)),f(i,X(1)),---,f(i,X(N-1))]^t$. From Theorem 2. 2 $rank(MF)=N$. Consequently, for any N-dimensional vector $\psi(B^n)$ there are certain real numbers $a(i)$'s such that $\psi(B^n)=\sum_i a(i)f(i,B^n)$. In other word, F gives a L-expression to ψ. When $|F|=N$, partial functions in F are linearly independent and $a(i)$'s are uniquely determined. QED.

From Corollary 2.3 it is seen that a T-complete system and a L-complete system are equivalent. Let for a subset S of V $J(S)$ be the set of all partial functions any of whose supports is a proper subset of S, and $I(S)=\{F|F\subset J(S)\}$ be the family of all subsets of $J(S)$. If $I(S)$ and $L(f)$ are disjoint for a partial function f whose support is S, that is, $S(f)=S$, then the partial function f is called a *mask* on S. For example, $f=x(0)\vee x(1)$ is a mask on $\{x(0),x(1)\}$, but $f=x(0)+x(1)$ is not. There are many masks on S. A mask is defined here in wider meaning than that of Minsky and others [11]. They used a mask only in the meaning of a AND-mask defined here. The functions $f(AND,S)$, $f(OR,S)$, $f(PARITY,S)$, and $f(MASK,S)$ are defined as

$$f(AND,S)=x(0)\wedge x(1)\wedge---\wedge x(k-1)=\bigwedge_{i=0}^{k-1} x(i)=1[\sum_{i=0}^{k-1} x(i)>k]$$

$$f(OR,S)=x(0)\vee x(1)\vee---\vee x(k-1)=\bigvee_{i=0}^{k-1} x(i)=1[\sum_{i=0}^{k-1} x(i)>1]$$

$$f(PARITY,S)=x(0)\oplus x(1)\oplus---\oplus x(k-1)=\bigoplus_{i=0}^{k-1} x(i)$$

$$=1[\sum_{i=0}^{k-1} x(i)=\text{ an odd number}]$$

$f(MASK,S)=$a mask on S

$$S=\{x(0),x(1),---,x(k-1)\}. \tag{2.31}$$

and they are called AND-, OR-, PARITY-mask, and simply a mask, respectively. If S is an empty set, then they are defined as $f(AND,S)=f(OR,S)=f(PARITY,S)=f(MASK,S)=1$. $F(AND)$, $F(OR)$, $F(PARITY)$, and $F(MASK)$ are defined by

$$F(AND)=\{f(AND,S)\,|\,S \text{ is a subset of } V\}$$
$$F(OR)=\{f(OR,S)\,|\,S \text{ is a subset of } V\}$$
$$F(PARITY)=\{f(PARITY,S)\,|\,S \text{ is a subset of } V\}$$
$$F(MASK)=\{f(MASK,S)\,|\,S \text{ is a subset of } V\} \tag{2.32}$$

and we have

$$|F(AND)|=|F(OR)|=|F(PARITY)|=|F(MASK)|=2^n. \tag{2.33}$$

$F(AND)$, $F(OR)$, and $F(PARITY)$ are special cases of $F(MASK)$.

[Proposition 2.3] Let $S=\{x(0),x(1),---,x(k-1)\}$, and $F(AND,l,S)$, $F(OR,l,S)$, and $F(PARITY,l,S)$ be defined as

$$F(AND,l,S)=\{f(AND,S')\,|\,S'\subset S,\ |S'|=l\}$$
$$F(OR,l,S)=\{f(OR,S')\,|\,S'\subset S,\ |S'|=l\}$$
$$F(PARITY,l,S)=\{f(PARITY,S')\,|\,S'\subset S,\ |S'|=l\} \tag{2.34}$$

then we have the following relations among AND-, OR-, and PARITY-masks:

$$f(AND,S)=\sum_{l=1}^{k}\{(-1)^{l-1}\sum_{f\in F(OR,l,S)} f\}$$
$$=(1/2)^{k-1}\sum_{l=1}^{k}\{(-1)^{l-1}\sum_{f\in F(PARITY,l,S)} f\}$$
$$f(OR,S)=\sum_{l=1}^{k}\{(-1)^{l-1}\sum_{f\in F(AND,l,S)} f\}$$
$$=(1/2)^{k-1}\sum_{l=1}^{k}\sum_{f\in F(PARITY,l,S)} f\}$$
$$f(PARITY,S)=\sum_{l=1}^{k}\{(-2)^{l-1}\sum_{f\in F(AND,l,S)} f\}$$
$$=(-1)^{k-1}\sum_{l=1}^{k}\{(-2)^{l-1}\sum_{f\in F(OR,l,S)} f\} \tag{2.35}$$

(Proof) Let $x(i)$'s $(i=0,1,2)$ be in B, then we have

$$x(0)\wedge x(1)=x(0)+x(1)-(x(0)\vee x(1))=(1/2)(x(0)+x(1)-(x(0)\oplus x(1)))$$

$$x(0)\vee x(1)=x(0)+x(1)-(x(0)\wedge x(1))=(1/2)(x(0)+x(1)+(x(0)\oplus x(1)))$$
$$x(0)\oplus x(1)=x(0)+x(1)-2(x(0)\wedge x(1))=(-1)(x(0)+x(1)-2(x(0)\vee x(1)))$$
$$(x(0)\pm x(1))\wedge x(2)=(x(0)\wedge x(2))\pm(x(1)\wedge x(2))$$
$$(x(0)\pm x(1))\vee x(2)=(x(0)\vee x(2))\pm(x(1)\vee x(2))\mp x(2)$$
$$(x(0)\pm x(1))\oplus x(2)=(x(0)\oplus x(2))\pm(x(1)\oplus x(2))\mp x(2) \qquad (2.36)$$

where the double signs($\pm$, $\mp$) are taken in the same order. Using these relations, (2.35) can be easily proved through mathematical induction [15], [24]. QED.

Let $F(AND,l,F(i))$, $F(OR,l,F(i))$, $\psi(OR\text{-}AND,l,F(i))$, and $\psi(AND\text{-}OR,l,F(i))$ be defined as

$$F(AND,l,F(i))=\{f\,|\,f=f(i1)\wedge f(i2)\wedge\text{---}\wedge f(il)\}$$
$$F(OR,l,F(i))=\{f\,|\,f=f(i1)\vee f(i2)\vee\text{---}\vee f(il)\}$$
$$\psi(OR\text{-}AND,l,F(i))=\bigvee_{f\in F(AND,l,F(i))} f$$
$$\psi(AND\text{-}OR,l,F(i))=\bigwedge_{f\in F(OR,l,F(i))} f \qquad (2.37)$$

where $\{f(i1),f(i2),\text{---},f(il)\}\subset F(i)\subset F$, $1\leqq l\leqq|F(i)|$, $F(AND,1,F(i))=F(OR,1,F(i))=F(i)$. From Proposition 2.3 we have

$$\bigvee_{f\in F(i)} f=\sum_{l=1}^{|F(i)|}(-1)^{l-1}\sum_{f\in F(AND,l,F(l))} f$$
$$\bigwedge_{f\in F(i)} f=\sum_{l=1}^{|F(i)|}(-1)^{l-1}\sum_{f\in F(OR,l,F(l))} f. \qquad (2.38)$$

[Proposition 2.4] For any $F(i)\subset F$ we have

$$\sum_{f\in F(i)} f=\sum_{l=1}^{|F(i)|}\psi(OR\text{-}AND,l,F(i))=\sum_{l=1}^{|F(i)|}\psi(AND\text{-}OR,l,F(i)). \qquad (2.39)$$

(Proof) Let $\#[P]$ be the number of elements which satisfy the condition P. For example, $\#[x\text{=an odd number}, 0<x<10]=5$. We assume without loss of generality that $\#[f(X)=1, f\in F(i)]=k$ for an input pattern X.

(1) If $k=0$, then each side in (2.39) is equal to zero.

(2) If $1\leqq k\leqq|F(i)|$, then the left side in (2. 39) is equal to k. If

$1 \leq k1 \leq k$ and $|F(i)|-k+1 \leq k2 \leq |F(i)|$, then we have

$$\psi(OR,k1,F(i),X)=1,\ \ \psi(AND,k2,F(i),X)=1. \tag{2.40}$$

If $k+1 \leq k1 \leq |F(i)|$ *and* $1 \leq k2 \leq |F(i)|-k$, then we obtain

$$\psi(OR,k1,F(i),X)=0,\ \ \psi(AND,k2,F(i),X)=0. \tag{2.41}$$

These aspects are shown in Fig 2.4. Hence

$$\#[\psi(OR,k1,F(i),X)=1]=\#[\psi(AND,k2,F(i),X)=1]=k$$

$$\#[\psi(AND,k1,F(i),X)=0]=\#[\psi(OR,k2,F(i),X)=0]=|F(i)|-k. \tag{2.42}$$

Therefore we can conclude that each side in (2.39) is equal to k. QED.

[Theorem 2.3] $f(AND,S)$, $f(OR,S)$, and $f(PARITY,S)$ are masks on S. For any function ψ in $\beta(n)$, the L-expressions of ψ by $F(AND)$, $F(OR)$, and $F(PARITY)$ exist and are unique. Generally speaking, the L-expression of ψ by $F(MASK)$ always uniquely exists.

(Proof) It is well known that any function ψ in $\beta(n)$ can be written uniquely in the disjunctive normal form. We can rewrite ψ using the arithmetic sum(+) instead of the logical sum($\vee$). Furthermore any nagation $\overline{x(i)}$ can be replaced by $1-x(i)$. Applying this repeatedly, all negations can be removed. When the same terms have been collected together, we have the L-expression of ψ by $F(AND)$. From Theorem 2.2 it must be satisfied that $rank(MF(AND))=2^n$. Since $|F(AND)|=2^n$, $F(AND)$ is linearly independent. This means that $F(AND)$ is a linearly independent complete system. It is easily known that $F(OR)$ and $F(PARITY)$ are also linearly independent complete systems, using Proposition 2.3. It is the same thing with $F(MASK)$.

Suppose that $f(AND,S)$ is not a mask, then we could have the L-expression of $f(AND, S)$ by F' such that $S(f)$ is a proper subset of S for

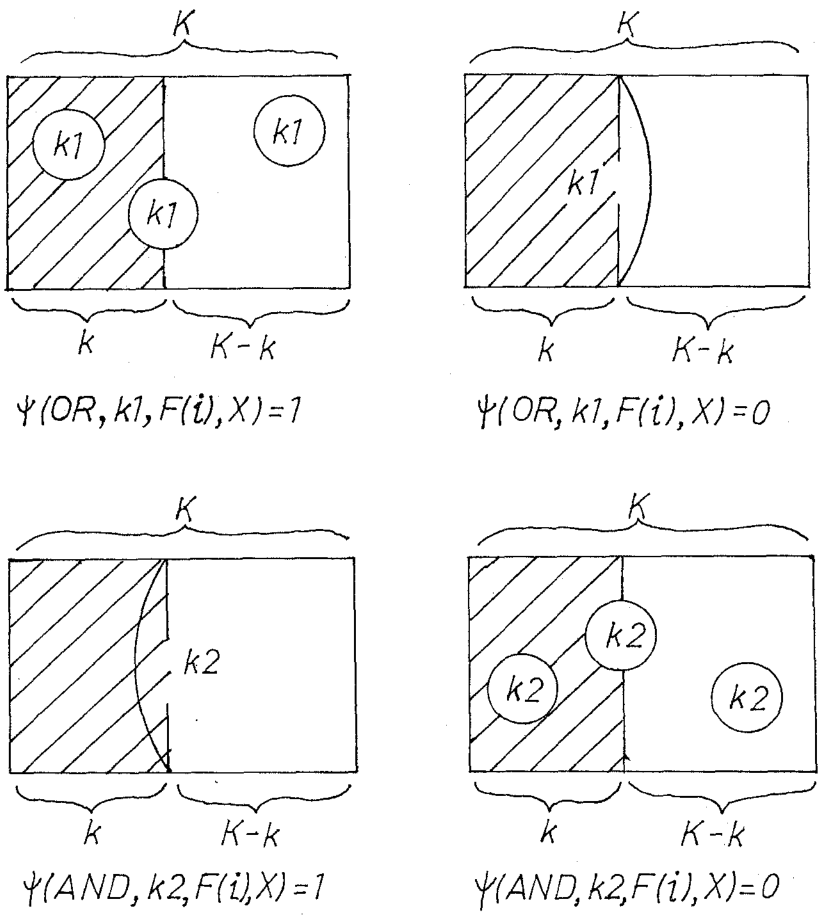


Fig. 2.4. Values of $\psi(OR,k1,F(i),X)$ and $\psi(AND,k2,F(i),X)$.

$(K=|F(i)|)$.

any f in F'. From the above proved fact, any f in F' has a L-expression by $\{f(AND, S') | S' \subset S\}$. Therefore $f(AND,S)$ has a L-expression by $\{f(AND, S') | S' \subset S\}$ as follows:

$$f(AND,S)= \sum_{S' \subset S} a(S')f(AND,S'). \tag{2.43}$$

This contradicts that $F(AND)$ is linearly independent, hence proves that $f(AND,S)$ is a mask. It is proved in the same way that $f(OR,S)$ and $f(PARITY,S)$ are masks. QED.

[Theorem 2.4] Assume that $S(f1)=S(f2)=S$, $f2$ is a mask and there is a L-expression of $f1$ such that

$$f1= \sum_{f \in F} a(f)f+a2 \cdot f2 \tag{2.44}$$

where $F \in I(S)$. Then the necessary and sufficient condition for $f1$ to be a mask is that $a2 \neq 0$ for any L-expresion of $f1$.

(Proof) Necessity:If $a2=0$, then it is obvious that $f1$ can not be a mask.

Sufficiency:If $a2 \neq 0$, then we obtain

$$f2=(1/a2)f1+ \sum_{f \in F} (-a(f)/a2)f \tag{2.45}$$

Suppose that $f1$ is not a mask. Then it is derived from (2.5) that $f2$ is not a mask. This contradicts the conditions, and therefore $f1$ is a mask.

QED.

[Corollary 2.4] Let f be a mask, then any function given by isomorphism is a mask.

[Corollary 2.5] Masks with different supports are linearly independent.

2.4 Minimum Complete Systems

A *minimum complete system* is defined as a complete system F such that the number $\sum_{f \in F} |S(f)|$ of connections between the first layer and the second is minimum.

[Theorem 2. 5](A Minimum Complete System) *F(MASK)* is a minimum complete system. Conversely, any element of a minimum complete system should be a mask.

(Proof) It has been proved that *F(MASK)* is a complete system. Since there is a mask f such that $S(f)=V$, there is at least one partial function should be in a complete system whose support is V. As the definition of a mask is taken into consideration, the partial function must be a mask. Furthermore there is a mask f such that $S(f)=V-\{x(i)\}$. Since masks with differnt supports are linearly independent, a complete system needs at least $({}_nC_n+{}_nC_{n-1})$ partial functions to have a L-expression of a mask f such that $|S(f)|\geqq n-1$. Where ${}_nC_n$ and ${}_nC_{n-1}$ mean combinations. Then the number of connections requested is at least ${}_nC_n\times n+{}_nC_{n-1}\times(n-1)$. If we continue this computation, we can conclude that the number of connections is more than or equal to $\sum_{i=0}^{n}{}_nC_i\times i=n\times 2^{n-1}$. On the other hand, the number of interconnections in *F(MASK)* is minimum, since $\sum_{f\in F(MASK)}|S(f)|=n\times 2^{n-1}$. Hence *F(MASK)* is a minimum complete system. It is obvious from linear independence of masks with different supports that any element of a minimum complete system must be a mask. QED.

[Corollary 2.6] For any ψ there are $F1$ and $F2$ such that

$$F1\subset F(MASK),\ F1\in T(\psi),\ OT(\psi)=\max_{f\in F1}|S(f)|$$

$$F2\subset F(MASK),\ F2\in L(\psi),\ OL(\psi)=\max_{f\in F2}|S(f)| \tag{2.46}$$

From Corollary 2.6 it is enough to test only subsets of *F(MASK)* instead of F in (2.6) and (2.9) in order to get $OT(\psi)$ and $OL(\psi)$. A minimum complete system defined above has the minimum number of interconnections. When a multi-stage coupled partial functions are used, the number becomes much less [16]. This is shown in brief here.

[Proposition 2.5] We can construct a complete system by using partial functions f such that $|S(f)|\leq k(k\geq 2)$ and f's are at most h-stage coupled. Where h is the smallest integer that is not under $log_k n$.

(Proof) We can construct any mask f by using one stage-coupled partial functions with order k, where $|S(f)|\leq k$. Assume that we have already got any mask whose order is less than or equal to l. Adding one more stage and using only partial functions f's such that $|S(f)|\leq k$, we can have any mask $f(MASK)$ such that $l+1\leq|S(f(MASK))|\leq min(k\times l,n)$. These facts lead the proposition. QED.

[Theorem 2.6](Multi-Stage Coupled Minimum Complete S-stems) When $k=2$ in Proposition 2.5, we have a complete system which has the minimum number of interconnections among multi-stage coupled complete systems. Then the number of interconnections is $2(2^n-1)-n$ and the number of stages is the smallest integer h that is not under $log_2 n$.

(Proof) With partial functions whose degree is 1, we can construct masks whose degree is 1. But we can not do masks whose degree is greater than 1. Even if we add one more stage in the proof of Proposition 2.5 using partial functions whose degree is 1, we can not obtain any masks f such that $|S(f)|\geq l+1$. If we use (2^n-1-n) partial functions whose degree is 2, any mask can be constructed by multi-stages. Then total number of connections is

$$1\times n+2\times(2^n-1-n)=2(2^n-1)-n. \tag{2.47}$$

QED.

Compared with a minimum complete system defined before, the numbers of interconnections are in the ratio of

$$(2(2^n-1)-n)/n2^{n-1}\cong 4/n(n>>1). \tag{2.48}$$

(2.48) decreases in inverse ratio to n. The system proposed in Theorem 2.6 is very convenient, since it can be constructed by using uniform components and moreover the number of components requested is minimum.

[Example 2.2] When $n=3$, a two-stage coupled minimum complete system $F1$ is as follows;

$$F1=\{f(i)\}(i=0,1,---,7)$$

$$f(0)=1,\ f(1)=x(0),\ f(2)=x(1),\ f(3)=x(2),\ f(4)=x(0)\wedge x(1)$$

$$f(5)=x(1)\wedge x(2),\ f(6)=x(2)\wedge x(0),\ f(7)=x(1)\wedge f(6)(=x(0)\wedge x(1)\wedge x(2))$$

$$\sum_{f\in F}|S(f)|=0+1+1+1+2+2+2+2=11,\ 2(2^3-1)-3=11 \tag{2.49}$$

For comparison, a minimum complete system $F2$ is shown.

$$F2=\{f(0),f(1),---,f(7)\},\ f(7)=x(0)\wedge x(1)\wedge x(2)$$

$$\sum_{f\in F}|S(f)|=0+1+1+1+2+2+2+3=12,\ 3\times 2^{3-1}=12. \tag{2.50}$$

$F1$ is shown in Fig. 2.5.

Let Hamming distance between F-images of $X1$ and $X2$ be expressed by $HD(F(X1)-F(X2))$, then we have the next theorem.

[Theorem 2.7] For $X1$ and $X2$ in $B^n(X1\neq X2)$ we have

$$\min_{X1,X2} HD(F(AND,X1)-F(AND,X2))=\min_{X1,X2} HD(F(OR,X1)-F(OR,X2))=1$$

$$\max_{X1,X2} HD(F(AND,X1)-F(AND,X2))=\max_{X1,X2} HD(F(OR,X1)-F(OR,X2))=2^n-1$$

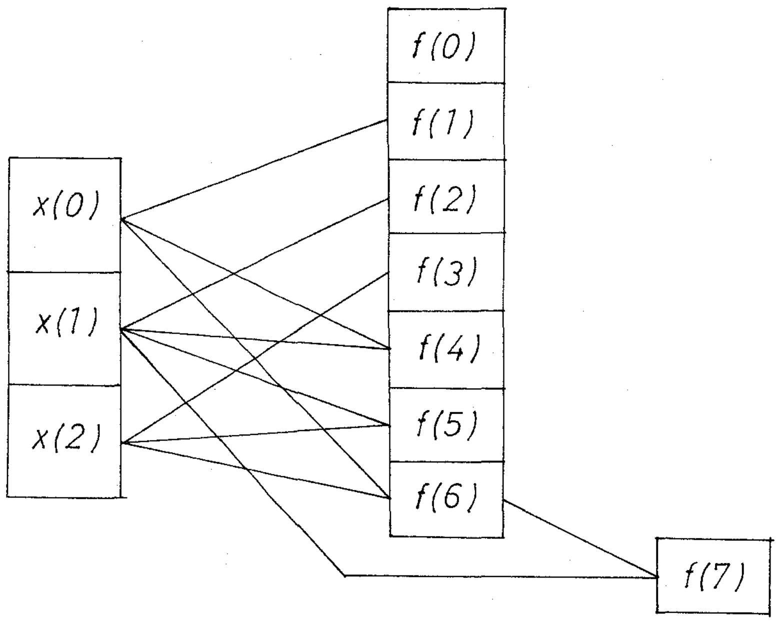


Fig. 2.5. A two-stage coupled minimum complete system when n=3.

$$\min_{X1,X2} HD(F(PARITY,X1)-F(PARITY,X2))$$

$$= \max_{X1,X2} HD(F(PARITY,X1)-F(PARITY,X2))=2^{n-1}\text{(constant)}. \qquad (2.51)$$

(Proof) First, we consider the case of $F(AND)$-images. There is no pair of $X1$ and $X2$ such that $F(AND,X1)=F(AND,X2)$ and $X1\neq X2$, since $F(AND)$ is a complete system. On the other hand, $F(AND,X)$ is a 2^n-dimensional vector. Hence we have

$$1\leqq HD(F(AND,X1)-F(AND,X2))\leqq 2^n-1. \qquad (2.52)$$

Let $X1=[0,0,---,0]$, $X2=[1,0,0,---,0]$ and $X3=[1,1,---,1]$, then we have

$$HD(F(AND,X1)-F(AND,X2))=1$$

$$HD(F(AND,X1)-F(AND,X3))=2^n-1. \qquad (2.53)$$

It is the same wih $F(OR)$-images, since $F(AND)$ and $F(OR)$ are isomorphisms.

Next, we consider the case of $F(PARITY)$-images. We can assume without loss of generallity that $X1=[x(0),x(1),--,x(k-1),x(k),--,x(n-1)]$, $X2=[\overline{x(0)},\overline{x(1)},--,\overline{x(k-1)},x(k),--,x(n-1)]$ $(0\leqq k-1\leqq n-1)$ for $X1$ and $X2$.

$$HD(F(PARITY,X1)-F(PARITY,X2))$$

$$= \sum_{i=1}^{k1} \#(f(PARITY,S)\,||S\cap\{x(0),x(1),---,x(k-1)\}|=i,$$

i is an odd number)

$$=2^{n-k}{}_kC_1+2^{n-k}{}_kC_3+---+2^{n-k}{}_kC_{k1}=2^{n-k}(2^k/2)=2^{n-1} \qquad (2.54)$$

where $k1=k$, if k is an odd number, otherwise $k1=k-1$ [23]. QED.

[Corollary 2.7]

$$\min_{X1,X2} HD(F(PARITY,X1)-F(PARITY,X2))> \min_{X1,X2} HD(F(*,X1)-F(*,X2))$$

$$*=AND \text{ or } OR, \text{ and } n\geqq 2. \qquad (2.55)$$

In linear separation reliability is generally estimated by the minimum distance from points of on-set and off-set to the separation plane. When reliability of a multi-layer series-coupled machine is estimated by the minimum Hamming distance between F-images, from Corollary 2.7 $F(PARITY)$ is the most reliable system compared with $F(AND)$ and $F(OR)$. But the reliability must be appraised in the total system. The minimum Hamming distance may be one criterion of it.

The above technique is applicable to construct a complete system on Hilbert space $L^2(0,1)$ [25]. The notations are used in the similar meaning to the foregoing. If any function $\psi(x)$ in $L^2(0,1)$ can be expressed by a series expansion as follows, F is called a complete system:

$$\psi(X)=\sum_{f\in F} a(f)f(x),\quad 0\leqq x<1 \tag{2.56}$$

where the equality signifies quadratic mean convergence. The family of Walsh functions [26] is a well-known complete system, which is represented as $\{\omega(i)\}(i=0,1,---)$. Where the numbering is based on that of Palay [27]. Rademacher functions are defined as

$$r(0,x)=\begin{cases}1, & \text{if } 0\leqq x<1/2\\ 0, & \text{if } 1/2\leqq x<1\end{cases}$$

$$r(0,x+1)=r(0,x)$$

$$r(n,x)=r(0,2^n x),\quad n=1,2,--- \tag{2.57}$$

where the range is changed from the set $\{-1,1\}$ to the interval $(-\infty<r(n,x)<\infty)$. Let $F(AND)$, $F(OR)$, and $F(PARITY)$ be defined as

$$F(AND)=\{f(AND,i)\},\quad F(OR)=\{f(OR,i)\},\quad F(PARITY,i)=\{f(PARITY,i)\},$$

$$f(AND,0)=f(OR,0)=f(PARITY,0)=1$$

$$f(AND,i)=r(i(1))\wedge r(i(2))\wedge---\wedge r(i(k))$$

$$f(OR,i)=r(i(1))\vee r(i(2))\vee\cdots\vee r(i(k))$$

$$f(PARITY,i)=r(i(1))\oplus r(i(2))\oplus\cdots\oplus r(i(k)) \tag{2.58}$$

where $i=1,2,3,\cdots$, $i=2^{i(1)}+2^{i(2)}+\cdots+2^{i(k)}$ and $i(l)<i(l+1)$ $(l=1,2,\cdots,k-1)$. Moreover let $Si=\{i(1),i(2),\cdots,i(k)\}$, $l1=\sum_{i(l)\in Si\cup Sj} 2^{i(l)}$, $l2=\sum_{i(l)\in(Si\cup Sj-Si\cap Sj)} 2^{i(l)}$, then we have

$$f(AND,i)\wedge f(AND,j)=f(AND,l1)$$

$$f(OR,i)\vee f(OR,j)=\begin{cases} f(OR,l1), & \text{if } i\cdot j\neq 0 \\ 1, & \text{if } i\cdot j=0 \end{cases}$$

$$f(PARITY,i)\oplus f(PARITY,j)=f(PARITY,l2)$$

$$\int_0^1 f(AND,i,x)\,dx=1/2^k$$

$$\int_0^1 f(OR,i,x)\,dx=1-1/2^k$$

$$\int_0^1 f(PARITY,i,x)\,dx=1/2 \tag{2.59}$$

The third equation in (2.59) corresponds to the following equation about Walsh functions:

$$\omega(i)\times\omega(j)=\omega(i\oplus j) \tag{2.60}$$

where $i=2^{i1}+2^{i2}+\cdots+2^{ik}$, $j=2^{j1}+2^{j2}+\cdots+2^{jk}$, $i\oplus j=2^{i1\oplus j1}+2^{i2\oplus j2}+\cdots+2^{ik\oplus jk}$.

Let $F(AND,l)$, $F(OR,l)$, and $F(PARITY,l)$ be defined as

$$F(AND,l)=\{\bigwedge_{f\in S} f\},\ F(OR,l)=\{\bigvee_{f\in S} f\},\ F(PARITY,l)=\{\bigoplus_{f\in S} f\} \tag{2.61}$$

where $S\subset\{r(i(1)),r(i(2)),\cdots,r(i(k))\}$ and $|S|=l$. Using (2.35) it is known that there is the ensuing relation between the above defined functions and Walsh functions.

$$\omega(0)=f(AND,0)=f(OR,0)=f(PARITY,0)=1$$

$$\omega(i)=2\sum_{l=1}^{k}\{(-2)^{l-1}\sum_{f\in F(AND,l)} f\}-1$$
$$=2(-1)^{k-1}\sum_{l=1}^{k}\{(-2)^{l-1}\sum_{f\in F(OR,l)} f\}-1$$

$$=2f(PARITY,i)-1 \tag{2.62}$$

Therefore $F(AND)$, $F(OR)$, and $F(PARITY)$ are complete systems, since any Walsh function can be expressed into finite series by any of them. Let k denote the number of Rademacher functions which is requested to compute $f(AND,i)$, $f(OR,i)$, and $F(PARITY,i)$, $i=0,1,---,2^k-1$. $F(AND)$, $F(OR)$, and $F(PARITY)$ have the similar property to a minimum complete system. Let us take an example by $F(AND)$. Any Walsh function $\omega(l)$ for $l \leqq 2^k-1$ has series expansion by $F(AND,0)$ $F(AND,1)$ --- $F(AND,k)$. Furthermore the number of interconnections is

$$0+1\times{}_kC_1+2\times{}_kC_2+---+k\times{}_kC_k=k\cdot 2^{k-1} \tag{2.63}$$

This is minimum among complete systems on $L^2(0,1)$.

2.5 Group-Invariant Complete Systems

The set of all one-to-one mappings from B^n to B^n forms a group, where composition of mappings is taken as operation in the group. Let G be a subgroup of such a group. When

$$f(X)=f(g\circ X) \tag{2.64}$$

holds for any g in G and any X in B^n, f is called a *G-invariant function*. If there is an element g in G such that

$$f'(X)=f(g\circ G) \tag{2.65}$$

for any X in B^n, it is said that f' and f are *G-equivalent*. This is rewritten as follows: $f' \overset{G}{=} f$. (2.65) is abbreviated to "$f'=g\circ f$". "$X' \overset{G}{=} X$" is defined in the same way. The relation $\overset{G}{=}$ forms an equivalence relation and decomposes F and B^n into G-equivalence classes:

$F=F(1)\cup F(2)\cup---\cup F(l1)$, $F(i)\cap F(j)=$an empty set, $(i\neq j)$

$B^n = B^n(1) \cup B^n(2) \cup \cdots \cup B^n(l2)$, $B^n(i) \cap B^n(j)$=an empty set, $(i \neq j)$ (2.66)

Where F is closed under G and any $|B^n(i)|$ is an divisor of G. When $g \circ f$ is in F for any g in G and any f in F, it is said that F is *closed under* G. If $f' \overset{G}{=} f$, then there is $F(i)$ such that f' and f are in $F(i)$. Otherwise, $i \neq j$ for $F(i)$ and $F(j)$ such that f' is in $F(i)$ and f is in $F(j)$. If $X' \overset{G}{=} X$, then there is $B^n(i)$ such that X' and X are in $B^n(i)$. Otherwise, $i \neq j$ for $B^n(i)$ and $B^n(j)$ such that X' is in $B^n(i)$ and X is in $B^n(j)$. If the following two conditions are satisfied, a family FG is called a *G-invariant complete system.*

(1) FG is in $T(\psi)$ for any G-inariant function ψ, and

(2) any f in FG is G-invariant.

In this section we consider FG which is constructed from a linearly independent complete system F, where F is closed under G.

Let F be a family of G-invariant partial functions and $\{X1, X2, \cdots, Xl2\}$ be representatives of G-equivalence classes $\{B^n(1), B^n(2), \cdots, B^n(l2)\}$, then the G-invariant F-matrix $M'F$ is defined as

$$M'F = \begin{pmatrix} f(0,X1), \cdots, f(m-1,X1) \\ \cdots \quad , \cdots, \quad \cdots \\ f(0,Xl2), \cdots, f(m-1,Xl2) \end{pmatrix} . \tag{2.67}$$

When $l2=m$, the determinant of the above matrix is denoted by $|M'F|$.

[Theorem 2.8](Necessary and Sufficient Condition for a G-invariant Complete Systems) The necessary and sufficient condition for a G-invariant complete system is that $rank(M'F)=l2$. When $|F|=l2$, this condition can be rewritten as $|M'F| \neq 0$.

(Proof) From Lemma 2.1 the necessary and sufficient condition for a G-

invariant complete system is that the $l2$ row vectors of $M'F$ are linearly independent. This condition can be written as $rank(M'F)=l2$. When $|F|=l2$, it is written as $|M'F|\neq 0$. QED.

[Corollary 2.8] A G-invariant complete system F gives a L-expression to any G-invariant function. When the number of G-invariant partial functions is minimum, that is, $|F|=l2$, the L-expression is unique.

(Proof) This is proved in the same way as that of Corollary 2.3. QED.

From Corollary 2. 8 it is seen that "F is in $T(\psi)$ for any G-invariant function ψ" and "F is in $L(\psi)$ for any G-invariant function ψ" are equivalent.

The composing process of G-invariant functions is shown by the diagram in Fig. 2.6. The diagram shows that one can compose G-invariant functions in two stages. At the first stage partial functions are computed independently of one another. Next they are combined through 'G-Invariance Theorem' to obtain G-invariant functions. Now we state G-invariance theorem and its corollaries which are imporant and powerful mathematical tools. The theorem was first proposed by Minsky and others [11]. Since they used T-expression, equivalence of coefficients was a sufficient condition and there was some tolerance. Since we use a L-expression to compose G-invariant systems, it is a necessary and sufficient condition. The same may be said of Boolean operations instead of arithmetic sum in a L-expression.

[Theorem 2. 9](G-Invariance Theorem) Let $F=\{f(i)\}(i=0,1,---,m-1)$ be a linearly independent complete systm which is closed under G and finite. Suppose that ψ is a G-invariant function and expressed as

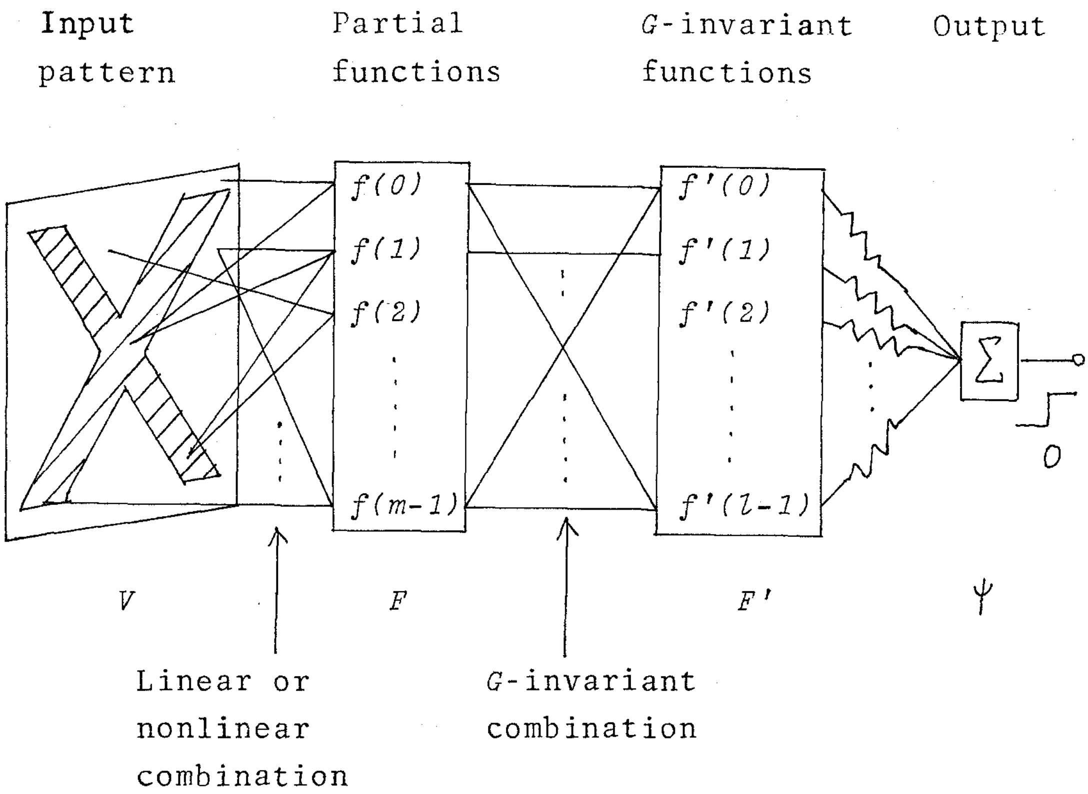


Fig. 2.6. Composing Process of G-invariant functions.

$$\psi = \sum_{i=0}^{m-1} a(i)f(i) = \bigvee_{i=0}^{m-1} (b(i)f(AND,i))$$
$$= \bigwedge_{i=0}^{m-1} (c(i)f(OR,i)),\ b(i) \text{ and } c(i) \text{ in } B \tag{2.68}$$

Then the coefficients depend only on G-equivalence classes, that is,

$$\text{if } f(i)\overset{G}{=}f(j), \text{ then } a(i)=a(j),\ b(i)=b(j),\ c(i)=c(j). \tag{2.69}$$

The reverse is also right, that is, if (2.69) is satisfied, then ψ is G-invariant.

(Proof) The proof about a L-expression is made in the same way with that of Minsky and others. We will prove only the case when $\psi = \bigvee_{i=0}^{m-1} (b(i)$ $f(AND, i))$, since we can do similarly the case when $\psi = \bigwedge_{i=0}^{m-1} (c(1)f(OR,$ $i))$ [24], [36].

Necessity: We use $b(f)$ instead of $b(i)$ to avoid complicated subscripts. Any element g in G defines a one-to-one correspondence $f \leftrightarrow g \circ f$ and $F(AND)$ is closed under G. Therefore, we have

$$\psi(X) = \bigvee_{f \in F(AND)} b(f)f(X) = \bigvee_{f \in F(AND)} b(g \circ f) g \circ f(X) \tag{2.70}$$

Since for any g in G the inverse g^{-1} is in G and ψ is G-invariant, we obtain

$$\psi(g^{-1} \circ X) = \bigvee_{f \in F(AND)} b(g \circ f)f(X) = \psi(X) \tag{2.71}$$

If it is supposed that $b(g \circ f') = b(f')$, then for X in $\{X \mid f'(X)=1\} \cap \{X \mid \bigvee_{f \neq f'} f(X)=0\}$ we have

$$\psi(g^{-1} \circ X) = b(g \circ f')f'(X) \vee (\bigvee_{f \neq f'} b(g \circ f)f(X)) = b(g \circ f')$$
$$\psi(X) = b(f')f'(X) \vee (\bigvee_{f \neq f'} b(f)f(X)) = b(f') \tag{2.72}$$

that is, $\psi(g^{-1} \circ X) = \psi(X)$, and this contradicts that ψ is G-invariant. Hence we can conclude that $b(g \circ f') = b(f')$ for any g in G and any f' in $F(AND)$.

Sufficiency: Let us assume that $b(g \circ f') = b(f')$ for any g in G and any f'

in $F(AND)$. From (2.70) we have

$$\psi(g\circ X)= \bigvee_{f\in F(AND)} b(g^{-1}\circ f)g\, f(g^{-1}\circ X)$$
$$= \bigvee_{f\in F(AND)} b(g^{-1}\circ f)f(X) \tag{2.73}$$

From the assumption we obtain

$$\psi(g\circ X)= \bigvee_{f\in F(AND)} b(f)f(X)=\psi(X). \tag{2.74}$$

Hence ψ is G-invariant. QED.

If F is an orthogonal complete system, even if F is an infinite set, then it is easily understood from the process of the proof that the theorem is satisfied.

[Corollary 2.9] For any f in $\beta(n)$, $f1$, $f2$, $f3$, and $f4$ written as follows are G-invariant functions:

$$f1= \sum_{f\in G} g\circ f,\ f2=\bigvee_{g\in G} g\circ f$$
$$f3= \bigwedge_{g\in G} g\circ f,\ f4=\bigoplus_{g\in G} g\circ f. \tag{2.75}$$

When $f1,f2,---,fk$ are G-invariant, ψ defined as follows is G-invariant:

$$\psi=f(f1,f2,---,fk) \tag{2.76}$$

where f is an arbitrary function.

If G is a permutation group on V, then V is decomposed into G-equivalence classes.

$$V=V(1)\cup V(2)\cup ---\cup V(l3)$$
$$V(i)\cap V(j)=\text{an empty set},\ i\neq j \tag{2.77}$$

If $x(i1)\overset{G}{=}x(i2)$, then there is $V(i)$ such that $x(i1)$ and $x(i2)$ are in $V(i)$ and $|V(k)|$ is a divisor of $|G|$. Otherwise, $i\neq j$ for $V(i)$ and $V(j)$ such that $x(i1)$ is in $V(i)$ and $x(i2)$ is in $V(j)$. $F(AND)$, $F(OR)$, and $F(PARITY)$ defined above are closed under any permutation group. If all members of V are G-equivalent, that is, $l3=1$ in (2.77), then G is called a global group. Translation and rotation are one kind of permutation groups on V. The group of all translations is a global group, but that of all rotations is not. $l2$ in (2.66) is not always equal to one, even if $l3=1$. On the contrary, $l3$ is always equal to one, if $l2=1$. In the case when $l2=1$, discrimination has no meaning, since all input patterns are G-equivalent.

[Example 2.3](G-Equivalence Decomposition)

(1) Decomposition induced by a cyclic group:Let $V=\{x(0),x(1),x(2)\}$, $g1=\begin{pmatrix}0,1,2\\1,2,0\end{pmatrix}$ and $G1=\{g1,g1^2,g1^3\}$, then $G1$ is a global group and V and B^3 are decomposed as follows:

$$V=\{x(0),x(1),x(2)\}=V(1)$$

$$B^3=\bigcup_{x(0)\leqq x(1)\leqq x(2)}\{[x(0),x(1),x(2)],[x(1),x(2),x(0)],\\ [x(2),x(0),x(1)]\}. \tag{2.78}$$

(2) Decomposition induced by another group:Let $V=\{x(0),x(1),x(2)\}$, $g2=\begin{pmatrix}0,1,2\\2,1,0\end{pmatrix}$ and $G2=\{g2,g2^2\}$, then V and B^3 are decomposed as follows:

$$V=\{x(0),x(3)\}\cup\{x(1)\}=V(1)\cup V(2)$$

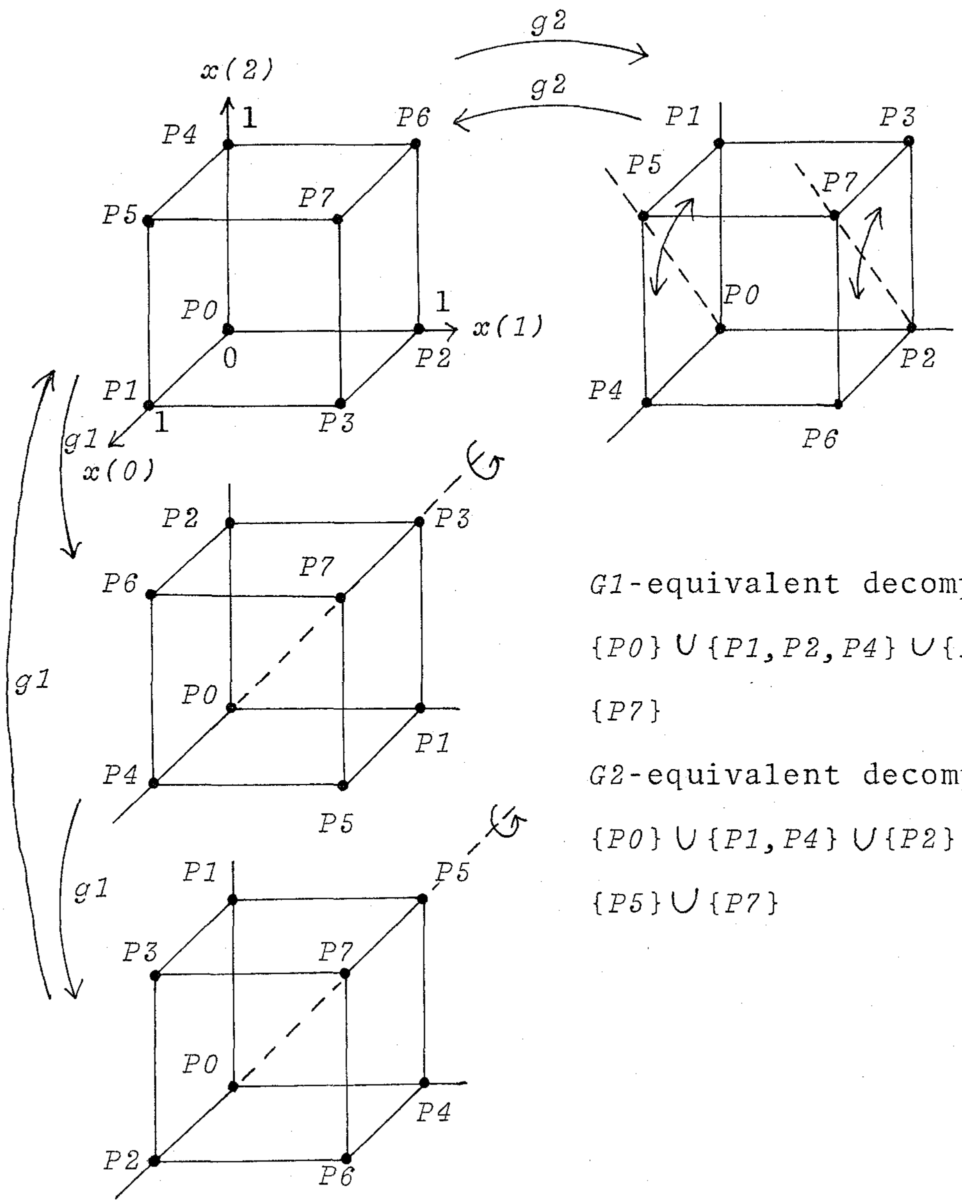
$$B^3=\bigcup_{x(0)\leqq x(2)}\{[x(0),x(1),x(2)],[x(2),x(1),x(0)]\}. \tag{2.79}$$

Where $g1^3$ and $g2^2$ are identity operators, and it is shown in Fig 2.7 that $g1$ and $g2$ operating on a pattern X in B^3 yield another pattern $g1\circ X$ and $g2\circ X$, respectively, and a subset of B^3 is decomposed into $G1$-equivalence classes and $G2$-equivalence ones by $G1$ and $G2$, respectively.

[Theorem 2.10](Support and Degree of G-invariant Functions) We can assume without loss of generality that the following three conditions are satisfied:

(1) ψ is a G-invariant function, and

(2) there is $x(i)$ such that $x(i)$ is in $S(\psi)\cap V(j)$, $j=1,2,---,l$, and



G1-equivalent decomposition:

$\{P0\} \cup \{P1, P2, P4\} \cup \{P3, P5, P6\} \cup \{P7\}$

G2-equivalent decomposition:

$\{P0\} \cup \{P1, P4\} \cup \{P2\} \cup \{P3, P6\} \cup \{P5\} \cup \{P7\}$

Fig. 2.7. *G1*-equivalent decomposition and *G2*-equivalent decomposition.

(3) there can not exist $x(i)$ such that $x(i)$ is in $S(\psi)\cap\{V(l+1)\cup V(l+2)\cup$ $---\cup V(l3)\}$, where G is a permutation group and $l\leqq l3$.

Then $S(\psi)$ is decomposed into G-equivalence classes as follows:

$$S(\psi)=V(1)\cup V(2)\cup---\cup V(l). \tag{2.80}$$

Since $v(1),V(2),---,V(l)$ are G-equivalence classes, we have

$$|S(\psi)|=\sum_{i=1}^{l}|V(i)|=|V(1)|+|V(2)|+---+|V(l)|. \tag{2.81}$$

(Proof) There exists F which is a linearly independent minimum complete system and closed under G. Then we have a L-expression of ψ by F.

$$\psi=\sum_{f\in F}a(f)f,\ S(\psi)=\bigcup_{a(f)\neq 0}S(f). \tag{2.82}$$

From Theorem 2.9 and (2.66) this is rewritten as follows:

$$\psi=\sum_{i=1}^{l1}a(i)(\sum_{f\in F(i)}f),\ S(\psi)=\bigcup_{a(i)\neq 0}(\bigcup_{f\in F(i)}S(f)) \tag{2.83}$$

where $F(i)$ is a G-equivalence class. If it is satisfied that $x(i1)$ is in $S(\psi)\cap V(j)$ $(j\leqq l)$, then there are g in G and $F(i)$ such that $x(i2)=g\circ x(i1)$ for any $x(i2)$ in $V(j)$, $x(i1)$ in $S(f)$, and f in $F(i)$. Hence

$$V(i)\subset\bigcup_{g\in G}S(g\circ f)=\bigcup_{f\in F(i)}S(f)=S(\sum_{f\in F(i)}f) \tag{2.84}$$

hence

$$\bigcup_{i=1}^{l}V(i)\subset\bigcup_{a(i)\neq 0}(\bigcup_{f\in F(i)}S(f))=S(\psi). \tag{2.85}$$

From the condition (3) we have

$$S(\psi)\cap\{V(l+1)\cup V(l+2)\cup---\cup V(l3)\}=\text{an empty set}$$

$$S(\psi)\subset V-\bigcup_{i=l+1}^{l3}V(i)=\bigcup_{i=1}^{l}V(i). \tag{2.86}$$

From (2.84) and (2.86) we can conclude that

$$S(\psi)=\bigcup_{i=1}^{l}V(i). \tag{2.87}$$

Since $V(i)$ and $V(j)$ $(i\neq j)$ are disjoint, then we obtain

$$|S(\psi)|=\sum_{i=1}^{l}|V(i)|. \tag{2.88}$$

QED.

If G is a global group, then the support of any G-invariant function is V except for a constant function. For exmaple, supports of translation invariant functions are V, but those of rotation invariant functions are not always V. $F(AND)$, $F(OR)$, and $F(PARITY)$ are all linearly independent minimum complete systems and closed under any permutation group. Then they are decomposed into G-equivalence classes:

$$F(AND)=\bigcup_{i=1}^{l1} F(AND,i),\ F(OR)=\bigcup_{i=1}^{l1} F(OR,i)$$
$$F(PARITY)=\bigcup_{i=1}^{l1} F(PARITY,i). \tag{2.89}$$

$F1(AND)$, $F2(AND)$, $F1(OR)$, and $F2(OR)$ are defined as:

$$F1(AND)=\{fi\,|\,fi=\sum_{g\in G} g\circ f'=\sum_{f\in F(AND,i)} f,\ f' \text{ in } F(AND,i)\}$$
$$F2(AND)=\{fi\,|\,fi=\bigvee_{g\in G} g\circ f'=\bigvee_{f\in F(AND,i)} f,\ f' \text{ in } F(AND,i)\}$$
$$F1(OR)=\{fi\,|\,fi=\sum_{g\in G} g\circ f'=\sum_{f\in F(OR,i)} f,\ f' \text{ in } F(OR,i)\}$$
$$F2(OR)=\{fi\,|\,fi=\bigwedge_{g\in G} g\circ f'=\bigwedge_{f\in F(OR,i)} f,\ f' \text{ in } F(OR,i)\}$$
$$F1(PARITY)=\{fi\,|\,fi=\sum_{g\in G} g\circ f'=\sum_{f\in F(PARITY,i)} f,\ f' \text{ in } F(PARITY,i)\} \tag{2.90}$$

where f' is a representative of G-equivalence classes $F(AND,i)$, $F(OR,i)$, and $F(PARITY,i)$. It will be shown that all the families in (2.90) are G-invariant linearly independent complete systems.

[Theorem 2.11](G-Invariant Linearly Independent Complete S-stems)

(1) Let F be a linearly independent complete system which is closed under G, then FG defined by the following equation is a G-invariant linearly independent complete system:

$$FG=\{f\,|\,f=\sum_{g\in G} g\circ f',\ f' \text{ in } F(i)\} \tag{2.91}$$

where f' is a representative of a G-equivalence class $F(i)$. Therefore $F1(AND)$, $F1(OR)$, and $F1(PARITY)$ are G-invariant linearly independent complete systems.

(2) $F2(AND)$ and $F2(OR)$ are also G-invariant linearly independent complete systems.

(Proof) (1) It is easy to see that any f in FG is G-invariant. Now we prove that FG is in $L(\psi)$ for any G-invariant function ψ. Since F is a complete system, ψ has a L-expression by F:

$$\psi=\sum_{f\in F} a(f)f. \tag{2.92}$$

Then we have

$$\begin{aligned} &g\circ\psi=\sum_{f\in F} a(f)g\circ f \\ &\sum_{g\in G} g\circ\psi=\sum_{g\in G}\sum_{f\in F} a(f)g\circ f \end{aligned} \tag{2.93}$$

Since ψ is G-invariant, we obtain

$$|G|\psi=\sum_{f\in F}\sum_{g\in G} a(g^{-1}\circ f)f=\sum_{f\in F}\sum_{g\in G} a(g\circ f)f \tag{2.94}$$

hence

$$\begin{aligned} &\psi=\sum_{f\in F}(1/|G|)\sum_{g\in G} a(g\circ f)f=\sum_{f\in F} b(f)f \\ &b(f)=(1/|G|)\sum_{g\in G} a(g\circ f) \end{aligned} \tag{2.95}$$

On the other hand, $\{g\circ f|f \text{ is in } F\}=F$ and from Theorem 2.9 we have

$$\psi=\sum_{i=1}^{l1} b(f')(\sum_{g\in G} g\circ f') \tag{2.96}$$

where f' is a representative of a G-equivalence class $F(i)$. We can rewrite (2.96) as follows:

$$\psi=\sum_{f\in FG} c(f)f,\ c(f)=b(f'),\ f=\sum_{g\in G} g\circ f' \tag{2.97}$$

(2.97) means that FG is in $L(\psi)$. Since F is linearly independent, so is FG, too.

(2) It is also easy to see that any f in $F2(AND)$ is G-invariant. Now it is shown that $rank(M'F2(AND))=l2$, that is, $F2(AND)$ is linearly independent. Where $l2$ is the number of G-equivalence classes of input patterns.

Assume that $F2(AND)$ is not linearly independent, then there exist non-zero coefficients such that

$$\sum_{f \in F'} a(f)f=0, \ F' \subset F2(AND) \tag{2.98}$$

Let f' be one of members in F' such that $|S(f')|$ is minimum, where $f'=\bigvee_{f \in F(AND,i)} f$. From (2.38) and (2.90) f' can be written as

$$\begin{aligned} f' &= \bigvee_{f \in F(AND,i)} f \\ &= \sum_{f \in F(AND,i)} f + \sum_{j=2}^{|F(AND,i)|} (-1)^{j-1} \sum_{f \in F(AND,i,j)} f \end{aligned} \tag{2.99}$$

From (2.98) and (2.99) we have

$$\begin{aligned} \sum_{f \in F(AND,i)} f = & \sum_{f \in F', f \neq f'} (a(f)/s(f'))f \\ & - \sum_{j=2}^{|F(AND,i)|} (-1)^{j-1} \sum_{f \in F(AND,i,j)} f \end{aligned} \tag{2.100}$$

where $F(AND, i, j)=\{f|f=f(i1) \wedge f(i2) \wedge --- \wedge f(ij)\}$, $S(\psi)=\{f(i1), f(i2), ---, f(ij)\} \subset F(AND,i)$. Let $|S(f)|=k$ for f in $F(AND,i)$, then (2.100) means that a linear combination of masks with degree k has a L-expression by masks whose degrees are k or more and whose supports are different from $S(f)$'s. This contradicts Corollary 2. 5, that is, the linear independence of masks with different supports. Cosequently, $F2$ is linearly independent, and $rank(M'F2(AND))=l2$. From Theorem 2.8 we can conclude that $F2(AND)$ is a G-invariant linearly independent complete system. It is the same with $F2(OR)$. Using Proposition 2.4, these may be proved in a different way [40]. QED.

[Corollary 2.10] Any G-invariant function has unique L-expressions by FG, $F1(AND)$, $F2(AND)$, $F1(OR)$, $F2(OR)$, and $F1(PARITY)$.

By applying Theorem 2.11, a G-invariant complete system can be easily constructed from the viewpoint of mathematics for any permutation group G, for instance, translations, enlargements, reductions, rotations, and so on.

[Example 2.4] Let $V=\{x(0),x(1),x(2)\}$ and $G=\{$any permutation on $V\}$, then $F1(AND)$, $F2(AND)$, $F1(OR)$, and $F2(OR)$ defined as follows are all G-invariant linearly independent complete systems:

$$F1(AND)=\{f(AND,11),f(AND,12),f(AND,f13),f(AND,14)\}$$

$$F2(AND)=\{f(AND,21),f(AND,22),f(AND,f23),f(AND,24)\}$$

$$F1(OR)=\{f(OR,11),f(OR,12),f(OR,f13),f(OR,14)\}$$

$$F2(OR)=\{f(OR,21),f(OR,22),f(OR,f23),f(OR,24)\} \qquad (2.101)$$

where

$$f(AND,11)=f(AND,21)=f(OR,11)=f(OR,21)=1$$

$$f(AND,12)=f(OR,12)=x(0)+x(1)+x(2)$$

$$f(AND,13)=x(0)\wedge x(1)+x(1)\wedge x(2)+x(2)\wedge x(0)$$

$$f(AND,14)=f(AND,24)=f(OR,22)=x(0)\wedge x(1)\wedge x(2)$$

$$f(AND,22)=f(OR,14)=f(OR,24)=x(0)\vee x(1)\vee x(2)$$

$$f(AND,23)=f(OR,23)=(x(0)\wedge x(1))\vee(x(1)\wedge x(2))\vee(x(2)\wedge x(0))$$

$$=(x(0)\vee x(1))\wedge(x(1)\vee x(2))\wedge(x(2)\vee x(0))$$

$$f(AND,13)=(x(0)\vee x(1))+(x(1)\vee x(2))+(x(2)\vee x(0)). \qquad (2.102)$$

The values of the functions in these systems are shown in Table 2.2, which shows that the functions are G-invariant. This is also shown in Fig. 2.8.

2.6 Conclusions

In parallel computing machines such as multi-layer series-coupled machines investigated here it is important how wide range each partial function should process. This is the problem of locality of a function.

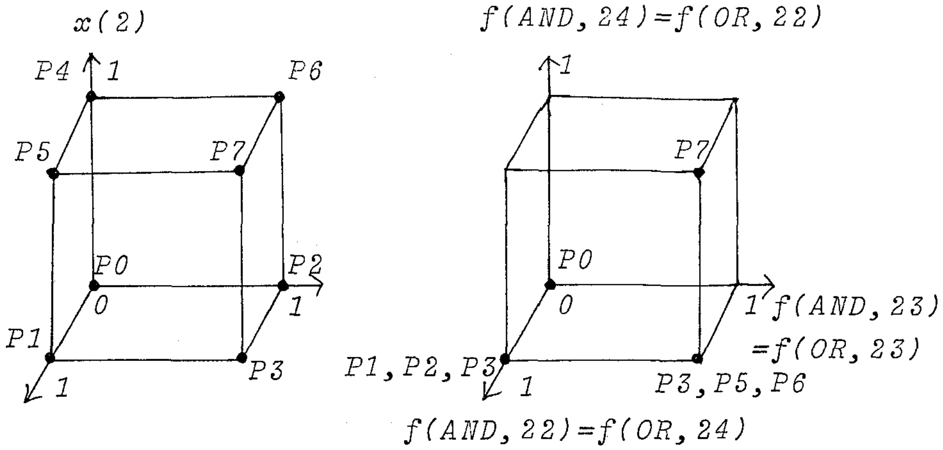


Fig. 2.8. *F2(AND)*-images and *F2(OR)*-images and *G*-equivalence decomposition: $B^3=\{P0\} \cup \{P1,P2,P4\} \cup \{P3,P5,P6\} \cup \{P7\}$.

Table 2.2. Values of *G*-invariant functions.

Patterns	f(AND,12) =f(OR,12)	f(AND,13)	f(AND,14) =f(AND,24) =f(OR,22)	f(AND,22) =f(OR,14) =f(OR,24)	f(AND,23) =f(OR,23)	f(OR,13)
P0(000)	0	0	0	0	0	0
P1(001)	1	0	0	1	0	2
P2(010)	1	0	0	1	0	2
P4(100)	1	0	0	1	0	2
P3(011)	2	1	0	1	1	3
P5(101)	2	1	0	1	1	3
P6(110)	2	1	0	1	1	3
P7(111)	3	3	1	1	1	3

The minimum complete system introduced in 2.4 is useful in the determination of such ranges. We represented the locality by *T*-order and *L*-order. In practical pattern recognition machines, however, the locality may be redefined in more suitable meaning for their structures. Moreover from the viewpoint of the number of partial functions it becomes difficult to realize a group-invariant complete system by electronic parts. It will be the next problem how to select the best subset of a group-invariant complete system.

CHAPTER 3

WALSH-HADAMARD POWER SPECTRUMS INVARIANT TO CERTAIN TRANSFORMATIONS

3.1 Introduction

In Chapter 2 it was considered how to construct G-invariant functions in multi-layer series-coupled machines which mainly dealt with Boolean functions. In this chapter Walsh functions are used as partial functions, and it is investigated how to construct G-invariant functions, especially G-invariant power spectrums. Walsh functions were introduced by Walsh in 1923 [26]. They can be generated recursively, are orthonormal, and form a closed set [26]-[31]. They have been used in several applications because of the simplicity of square waves. Walsh-Hadamard transform (WHT) has the advantage of computational simplicity when compared with Fourier transform. It is well known that the Fourier power spectrum is invariant to translations. Let $w(k,X)$ be the k-th (WHT) coefficient. Then the set $\{w^2(k,X)\}$ is not invariant to translations. The (WHT) power spectrum developed by Ahmed, Rao and Abdussattar [34] is invariant to translations, which is obtained through the (WHT). The fast algorithm for the power spectrum was also presented by them.

Here a composing process is introduced that produces functions which are invariant to some transformation groups. It will be described how to develop the (WHT) power spectrum to be invariant to enlargements, reductions, rotations by multiples of 90°, and some other transformations. Then it turns out that 'Group-Invariance Theorem' and other theorems developed in Chapter 2 are useful mathematical tools. Using 'Group-Invariance Theorem', it is known that there are the second degree G-invariant functions besides the power spectrum by Ahmed and others. According to the process presented here, a G-invariant complete system seems easy to make mathematically, but in general, difficult to make practically because of the huge number of functions required. Power spectrums which we chiefly adopt may be regarded as subsets of a G-invariant complete system.

Parallel computation is taken into consideration, but locality is not, because (WHT) itself is a global transform. We mainly treat transformations in the form of $2^{\pm n}$. When we treat ones in the form of $k^{\pm n}$, it is convenient to adopt complex-valued Walsh-Hadamard transform (CWHT). Then the (CWHT) power spectrum is developed to be invariant to the above described transformations, too.

3.2 Walsh-Hadamard Transform

There are three types of orderings for the Walsh functions [75] . One of them is used here, that is, natural or Hadamard ordering. Hadamard matrices can be recursively generated as follows:

$$H(0)=[1]$$

$$H(n)=\left[\begin{array}{c|c} H(n-1) & H(n-1) \\ \hline H(n-1) & -H(n-1) \end{array}\right] = H^n(1) \tag{3.1}$$

where $H^n(1)$ is the n succesive Kronecker product of $H(1)$. Let $H(n,k)$ denote the k-th row vector of $H(n)$, $k=0,1,---,2^n-1$. The Hadamard matrices satisfy the following:

$$H(n)=H(n-m)\otimes H(m)$$

$$(U\otimes V)(H(n-m)\otimes H(m))=(U\cdot H(n-m))\otimes(V\cdot H(m)) \quad (3.2)$$

where $0\leqq m\leqq n$; U and V are a 2^{n-m}-vector and a 2^m-vector, respectively; and the notation "$\otimes$" denotes the Kronecker product. These can be proved by taking the definitions of the Hadamard matrices and the Kronecker product into consideration.

Let $\{x(i)\}$ denote an N-periodic sequence $x(i)(i=0,1,---,N-1)$, of real numbers, and $\{x(i)\}$ be represented by means of an N-vector X:

$$X=[x(0),x(1),---,x(N-1)] \quad (3.3)$$

where $N=2^n$. Walsh-Hadamard transform (WHT) of an input pattern X is defined as

$$W(X)=(1/N)X\cdot H(n) \quad (3.4)$$

where $W(X)=[w(0,\ X),w(1,X),---,w(N-1,X)]$ and $w(k,X)$ is the k-th (WHT) coefficient. The inverse transform is defined as

$$X=W(X)\cdot H(n). \quad (3.5)$$

From (3.4) it follows that

$$W(X)\cdot W(X)^t=(1/N)X\cdot X^t \quad (3.6)$$

where "t" denotes the transposed vector. The right side of (3.6) represents the average power of the input pattern. Although Fourier power spectrum is invariant to translations, the set $\{w^2(i,X)\}$ is not. A composing process is proposed in preparation for the development of the (WHT) power spectrum.

3.3 Composing Process

The process to obtain a $G1 \otimes G2$-invariant power spectrum in several stages is shown by Fig. 3.1. First, an input pattern is transformed through the (WHT), and the (WHT) coefficients are squared. Second, the results are combined according to a certain function to obtain a $G1$-invariant power spectrum. Lastly, we find a permutation group on the $G1$-invariant power spectrum caused by $G2$ operating on the input pattern, and then we combine the $G1$-invarinant power spectrum to be $G2$-invariant in the same way and arrive at a $G1 \otimes G2$-invariant power spectrum. $G1$ and $G2$ are some transformation groups, and the product $G1 \otimes G2$ is defined by

$$G1 \otimes G2 = \{g \mid g = g1 \circ g2,\ g1 \text{ and } g2 \text{ are in } G1 \cup G2\}. \tag{3.7}$$

When functions in N-dimensional Hilbert space are used instead of Boolean functions with N variables, Theorem 2.9 and Corollary 2.9 are also satisfied. Then the corollary is rewritten as follows.

[Corollary 2.9'] For any f in N-dimensional Hilbert space, ψ written as follows is G-invariant:

$$\psi = \sum_{g \in G} g \circ f. \tag{3.8}$$

When $f(1)$, $f(2)$, ---, $f(k)$ are G-invariant, ψ defined as follows is also G-invariant:

$$\psi = \sum_{i=1}^{k} a(i) f(i) \tag{3.9}$$

where $a(i)$'s$(i=1,2,---,k)$ are arbitrary real numbers. We have a more general expression:

$$\psi = f(f(1), f(2), ---, f(k)) \tag{3.10}$$

where f is an arbitrary function.

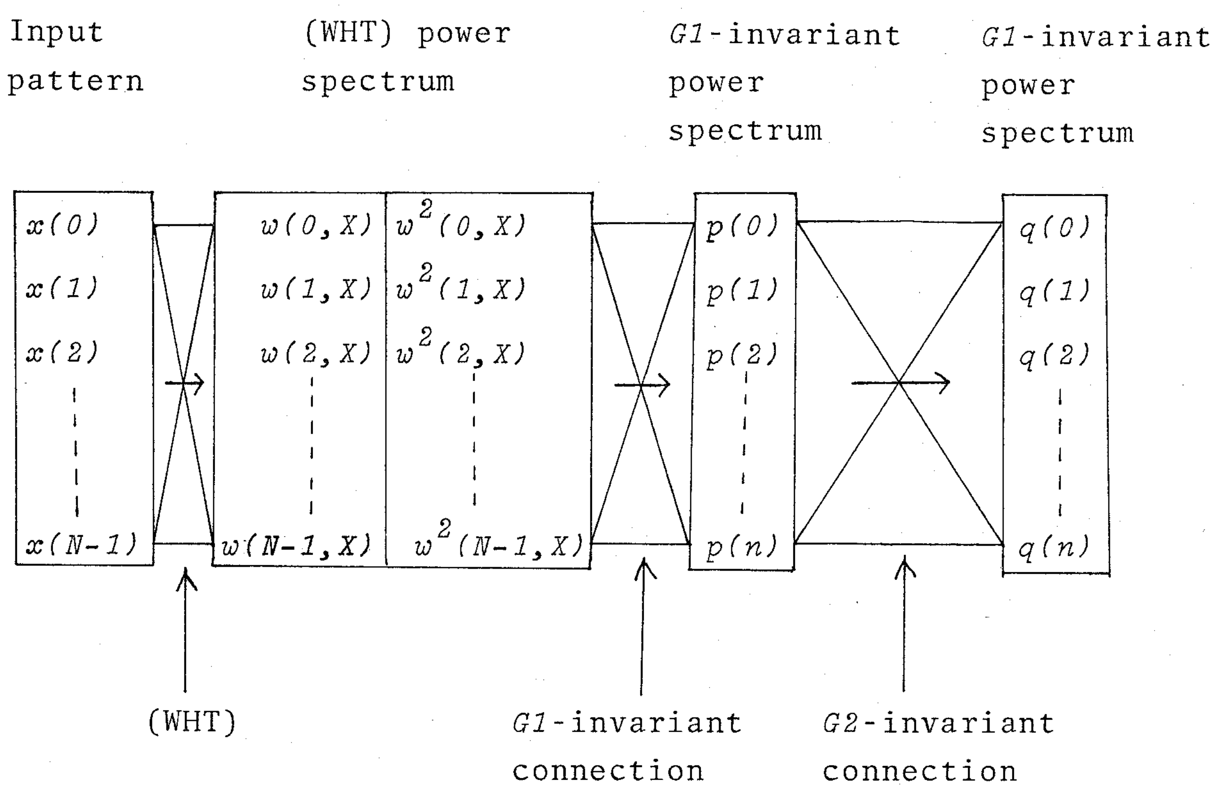


Fig. 3.1. Process to obtain $G1 \otimes G2$-invariant power spectrum.

Let functions $Fl(f(1),f(2),\text{---},f(L))(l=1,2,\text{---},L)$ be defined by

$$Fl(f(1),f(2),\text{---},f(L))=\begin{cases}\sum_{i=1}^{L} f^2(i)/A, & \text{if } A\neq 0,\ l=1\\ 2\sum_{i=1}^{L-l+1} f(i)f(i+l-1)/A, & \text{if } A\neq 0,\ l=2,3,\text{---},L\\ 0, & \text{if } A=0,\ l=1,2,\text{---},L\end{cases} \tag{3.11}$$

where $A=\sum_{i=1}^{L} f(i)$, then we have

$$\sum_{l=1}^{L} Fl(f(1),f(2),\text{---},f(L))=\sum_{i=1}^{L} f(i)$$
$$Fl(f(1),f(2),\text{---},f(L))\geqq 0\ (l=1,2,\text{---},L),\ \text{if } f(i)\geqq 0\ (i=1,2,\text{---},L). \tag{3.12}$$

Hence $\{Fl(f(1),f(2),\text{---},f(L))\}(l=1,2,\text{---},L)$ may be regarded as a developed power spectrum, if $\{f(i)\}(i=1,2,\text{---},L)$ is a power spectrum.

3.4 Translation($G1$)-Invariant (WHT) Power Spectrums

A translation-invariant power spectrum is illustrated by considering the case when $n=3$. Using Corollary 2.9' the second degree terms are combined to make translation-invariant functions:

$$f(l)=\sum_{k=0}^{7} x(k)x(k\oplus l) \tag{3.13}$$

where $l=0,1,\text{---},4$, and $\oplus$ is used for modulo eight addition. The $f(l)(l=0,1,\text{---},4)$ can be expressed by the set $\{w(k,X)\}(k=0,1,\text{---},N-1)$:

$$f(0)=8\sum_{k=0}^{7} w^2(k,X)$$
$$f(1)=4\{2w^2(0,X)-2w^2(1,X)+w^2(4,X)-w^2(5,X)+w^2(6,X)-w^2(7,X)$$
$$-2w(4,X)w(7,X)+2w(5,X)w(6,X)\}$$
$$f(2)=8\{w^2(0,X)+w^2(1,X)-w^2(2,X)-w^2(3,X)\}$$
$$f(3)=4\{2w^2(0,X)-2w^2(1,X)-w^2(4,X)+w^2(5,X)-w^2(6,X)+w^2(7,X)$$
$$+2w(4,X)w(7,X)-2w(5,X)w(6,X)\}$$

$$f(4)=4\{w^2(0,X)+w^2(1,X)+w^2(2,X)+w^2(3,X)-w^2(4,X)-w^2(5,X)-w^2(6,X)$$
$$-w^2(7,X)\}. \qquad (3.14)$$

Suppose that a function f has a linear expression by the second degree terms of $x(i)$'s$(i=0,1,----,7)$. From Theorem 2.9 f has a linear expression by $\{f(i)\}(l=0,1,---,4)$, if it is translation-invariant. Inversely f cannot be translation-invarint, if it cannot have a linear expression by $\{f(l)\}(l=0,1,---,4)$.

The adequate linear combination of the $f(l)$'s$(l=0,1,--,4)$ leads to a translation-invariant power spectrum. Adding $f(1)$ to $f(3)$ removes the cross terms $w(4,X)w(7,X)$ and $w(5,X),w(6,X)$:

$$f(1)+f(3)=16\{w^2(0,X)-w^2(1,X)\}. \qquad (3.15)$$

Then $\{w^2(k,X)\}$ is grouped as $\{w^2(0,X)\}\cup\{w^2(1,X)\}\cup\{w^2(2,X),w^2(3,X)\}\cup\{w^2(4,X), w^2(5, X), w^2(6,X),w^2(7,X)\}$. The $w^2(k,X)$'s in the same group have the same coefficients in the expressions of $f(0)$, $f(2)$, $f(4)$, and $f(1)+f(3)$. Thus we obtain

$$p(0)=(1/8^2)(f(0)+2f(1)+2f(2)+2f(3)+2f(4))=w^2(0,X)$$
$$p(1)=(1/8^2)(f(0)-2f(1)+2f(2)-2f(3)+2f(4))=w^2(1,X)$$
$$p(2)=(2/8^2)(f(0)-2f(2)+2f(4))=w^2(2,X)+w^2(3,X)$$
$$p(3)=(4/8^2)(f(0)-2f(4))=w^2(4,X)+w^2(5,X)+w^2(6,X)+w^2(7,X) \qquad (3.16)$$

The remainder may be expressed by

$$p(4)=(8/8^2)(f(1)-f(3))=w^2(4,X)-w^2(5,X)+w^2(6,X)-w^2(7,X)$$
$$-2w(4,X)w(7,X)+2w(5,X)w(6,X). \qquad (3.17)$$

$\{p(i)\}(i=0,1,2,3,\neq 4)$ turns out to be the very same power spectrum as that developed by Ahmed and others [34].

From the first-degree terms we obtain only one translation-invariant function which is linearly independent:

$$p(5)=(1/8)\sum_{k=0}^{7} x(k)=w(0,X). \tag{3.18}$$

As pointed out in [51], $\{p(i)\}(i=0,1,2,3)$ is not a translation-invariant complete system; that is, there are other shiftings under which $\{p(i)\}(i=0,1,2,3)$ is invariant. This is the same even if $\{p(i)\}(i=0,1,---,5)$ is used istead of $\{p(i)\}(i=0,1,2,3)$. In order to obtain a complete system we must usually use the higher-degree terms besides the first and the second-degree terms.

The generalization of the power spectrum mentioned above is straightforward as seen in [31], [34]:

$$p(0)=w^2(0,X)$$
$$p(s)=\sum_{i=2^{s-1}}^{2^s-1} w^2(i,X),\ s=1,2,---,n. \tag{3.19}$$

The generalization in the two-dimensional case is as follows:

$$p(0,0)=w^2(0,0,X), \qquad p(0,t)=\sum_{j=2^{t-1}}^{2^t-1} w^2(0,j,X)$$
$$p(s,0)=\sum_{i=2^{s-1}}^{2^s-1} w^2(i,0,X),\ p(s,t)=\sum_{i=2^{s-1}}^{2^s-1}\sum_{j=2^{t-1}}^{2^t-1} w^2(i,j,x), \tag{3.20}$$

where $s=1,2,---,n$ and $t=1,2,---,m$. The developed (WHT) power spectrums are expressed as:

$$P(n)=[p(0),p(1),---,p(n)]$$
$$P(n,m)=\begin{pmatrix} p(0,0),p(0,1),---,p(0,m) \\ ---\ ,\ ---\ ,---,\ --- \\ p(n,0),p(n,1),---,p(n,m) \end{pmatrix}. \tag{3.21}$$

When we want to take notice of the input pattern, we write as follows:

$$P(n,X)=[p(0,X),p(1,X),---,p(n,X)]$$

$$P(n,m,X)=\begin{pmatrix} p(0,0,X),p(0,1,X),---,p(0,m,X) \\ --- \quad , \quad --- \quad ,---, \quad --- \\ p(n,0,X),p(n,1,X),---,p(n,m,X) \end{pmatrix}. \tag{3.22}$$

This can be extended to any number of dimensions [31], [34].

3.5 Enlargement and Reduction($G2$)-Invariant (WHT) Power Spectrums

The changing aspects of the (WHT) coefficients under enlargements and reductions of an input pattern are investigated. An input pattern X is called a 2^i*-time enlargeable pattern*, if

$$X=[\underbrace{U,U,---,U}_{2^i}] \tag{3.23}$$

and $U=[x(0),x(1),---,x(2^{n-i}-1)]$. Similarly an input pattern Y is called a $1/2^i$*-time reducible pattern,* if $Y=[y(0)\mathbf{1},y(2^i)\mathbf{1},---,y(2^n-2^i)\mathbf{1}]$ and

$$\mathbf{1}=[\underbrace{1,1,---,1}_{2^i}]. \tag{3.24}$$

The 2^i*-time enlarged pattern* of X and the $1/2^i$*-time reduced pattern of* Y are defined as $[x(0)\mathbf{1},x(1)\mathbf{1},---,x(2^{n-i}-1)\mathbf{1}]=U\otimes\mathbf{1}$ and

$$[\underbrace{U',U',---,U'}_{2^i}]=\mathbf{1}\otimes U', \tag{3.25}$$

respectively, where

$$\mathbf{1}=[\underbrace{1,1,---,1}_{2^i}] \tag{3.26}$$

and $U'=[y(0),y(2^i),---,y(2^n-2^i)]$. For example, the 2-time enlarged pattern of X=[01230123] is [00112233], the 1/2-time reduced pattern of Y=[0 0112233] is [01230123], and so on.

From (3.1) the k-th row vector $H(n,k)$ of $H(n)$ is expressed as

$$H(n,k)=\begin{cases}[H(n-1,k),\ H(n-1,k)], & k=0,1,---,2^{n-1}-1\\ [H(n-1,k),-H(n-1,k)], & k=2^{n-1},2^{n-1}+1,---,2^{n}-1.\end{cases} \tag{3.27}$$

The upper half row vectors of $H(n)$ are 2-time enlargeable patterns. The fact of (3.27) induces recursively the following expression:

$$H(n,k)=\begin{cases}[H(0,k),---,H(0,k)]=[1,1,---,1], & k=0,\ 2^{n}\text{-time enlargeable},\\ [H(1,k),---,H(1,k)]=[1,-1,---,1,-1], & k=1,\ 2^{n-1}\text{-time enlargeable},\\ [H(2,k),---,H(2,k)], & k=2,3,\ 2^{n-2}\text{-time enlargeable},\\ [H(3,k),---,H(3,k)], & k=,5,6,7,\ 2^{n-3}\text{-time enlargeable},\\ ---\quad---\quad--- & \\ [H(n-1,k),H(n-1,k)], & k=2^{n-2},2^{n-2}+1,---,2^{n-1}-1,\ 2\text{-time enlargeable},\\ [H(n-1,k),-H(n-1,k)], & k=2^{n-1},2^{n-1}+1,---,2^{n}-1.\end{cases} \tag{3.28}$$

On the other hand from (3.2) $H(n)$ is also expressed as

$$H(n)=H(n-i)\otimes H(i)=H(i)\otimes H(n-i). \tag{3.29}$$

Taking account of that

$$H(i,0)=[\underbrace{1,1,---,1}_{2^{i}}], \tag{3.30}$$

we obtain that for a multiple k of 2^{i} $H(n,k)$ is a $1/2^{i}$-time reducible pattern. For an even number k the k-th row vector $H(n,k)$ of $H(n)$ is a 1/2-time reducible pattern. From (3.29) the 2^{i}-time enlarged pattern of $H(n,l)$ is $H(n,2^{i}\cdot l)$, where $l=0,1,---,2^{n-i}-1$. The 2-time enlarged pattern of $H(n,l)$ is $H(n,2l)$. These facts are well illustrated by considering the case when $n=3$:

4-time enlargement

$$H(3)=\left(\begin{array}{rrrrrrrr}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1
\end{array}\right)
\begin{array}{l}
H(3,0) \\ H(3,1) \\ H(3,2) \\ H(3,3) \\ H(3,4) \\ H(3,5) \\ H(3,6) \\ H(3,7)
\end{array}
\quad
\begin{array}{l}
\\ \\ *1 \\ \\ \\ *2 \\ \\ \\
\end{array}
\tag{3.31}$$

2-time enlargement

(*1:2-time enlargeable patterns, *2:1/2-time reducible patterns)

A 2^i-time enlargeable pattern X and the 2^i-time enlarged pattern Y of it are expressed as

$$X=[\underbrace{U,U,---,U}_{2^i}]=[\underbrace{1,1,---,1}_{2^i}]\otimes U=\mathbf{1}\times U$$

$$Y=[x(0)\mathbf{1},x(1)\mathbf{1},---,x(2^{n-i}-1)\mathbf{1}]=[x(0),x(1),---,x(2^{n-i}-1)]\otimes\mathbf{1}=U\otimes\mathbf{1}$$

$$U=[x(0),x(1),---,x(2^{n-i}-1)]$$

$$\mathbf{1}=[\underbrace{1,1,---,1}_{2^i}]. \tag{3.32}$$

The (WHT)'s of X and Y are as follows:

$$W(X)=(1/N)X\cdot H(n)=(1/N)[U,U,---,U]\cdot(H(i)\otimes H(n-i))$$

$$=(1/N)[U,U,---,U]\cdot\left(\begin{array}{c|c}
\begin{array}{cc|c} H(n-i), & H(n-i) & \vdots \\ H(n-i), & -H(n-i) & \vdots \\ \hline \multicolumn{2}{c}{------} & \end{array} & H(n-1) \\
\hline
H(n-1) & -H(n-1)
\end{array}\right)$$

$$=(1/N)(\mathbf{1}\cdot H(i))\otimes(U\cdot H(n-i))=(1/2^{n-i})[\underbrace{1,0,0,---,0}_{2^i}]\otimes(U\cdot H(n-i))$$

$$W(Y)=(1/N)Y\cdot H(n)=(1/N)(U\otimes\mathbf{1})(H(n-i)\otimes H(i))$$

$$=(1/N)[x(0)\mathbf{1},x(1)\mathbf{1},---,x(2^{n-i}-1)\mathbf{1}]\cdot\left[\begin{array}{c|c}\begin{array}{cc}H(i), & H(i)\\ H(i), & -H(i)\\ \hline & \\ \end{array} & H(n-1)\\ \hline H(n-1) & -H(n-1)\end{array}\right]$$

$$=(1/N)(U\cdot H(n-i))\otimes(\mathbf{1}\cdot H(i))$$

$$=(1/2^{n-i})(U\cdot H(n-i))\otimes[\underbrace{1,0,0,---,0}_{2^i}]. \tag{3.33}$$

From (3.33) we conclude that

$$w(k,X)=\begin{cases}\text{value of the } k\text{-th element of } (1/2^{n-i})U\cdot H(n-i), \\ \quad 0\leqq k\leqq 2^{n-i}-1\\ 0,\ 2^{n-i}\leqq k\leqq 2^n-1\end{cases}$$

$$w(k,Y)=\begin{cases}\text{value of the } k\text{-th element of } (1/2^{n-i})U\cdot H(n-i), \\ \quad k=l\cdot 2^i (l=0,1,---,2^{n-i}-1)\\ 0,\ k \text{ is not a multiple of } 2^i\end{cases} \tag{3.34}$$

and furthermore

$$w(k,X)=w(2^i\cdot k,Y),\ 0\leqq k\leqq 2^{n-i}-1. \tag{3.35}$$

These aspects are well summarized on a $n\times 2^{n-1}$-matrix $C(X)=[\{c(i,j,X)\}]$ defined recursively by

$$c(1,j,X)=w(2^{n-1}+j-1,X)$$

$$c(i,j,X)=\begin{cases}w(k/2,X), \text{ if } c(i-1,j,X)=w(k,X),\ k \text{ is an even number}\\ 0, \text{ otherwise}\end{cases} \tag{3.36}$$

where $i=2,3,---,n$ and $j=1,2,---,2^{n-1}$. 2^i-time enlargement and $1/2^i$-time reduction of an input pattern cause upward shift by i rows and downward shift by i rows, respectively. $w(0,X)$ is invariant under any shuffling of the elements of an input pattern. In the case when $n=3$, the matrix is as follows:

$$C(X)=\begin{pmatrix} w(4,X), w(5,X), w(6,X), w(7,X) \\ w(2,X), \quad 0 \quad , w(3,X), \quad 0 \\ w(1,x), \quad 0 \quad , \quad 0 \quad , \quad 0 \end{pmatrix} . \quad \begin{array}{l} \text{enlargement} \uparrow \\ \qquad\qquad \downarrow \text{reduction} \end{array} \tag{3.37}$$

Let Y be a 2-time enlarged pattern of a 2-time enlargeable pattern X, and then

$$C(Y)=\begin{pmatrix} w(4,Y), 0, w(6,Y), 0 \\ w(2,Y), 0, \quad 0 \quad , 0 \\ \quad 0 \quad , 0, \quad 0 \quad , 0 \end{pmatrix} = \begin{pmatrix} w(2,X), 0, w(3,X), 0 \\ w(1,X), 0, \quad 0 \quad , 0 \\ \quad 0 \quad , 0, \quad 0 \quad , 0 \end{pmatrix} . \tag{3.38}$$

As seen in (3.38), 2-time enlargement of an input pattern causes upward shift of the elements of $C(X)$. Then an enlargement and reduction-invariant power spectrum $\{q(i)\}(i=0,1,---,7)$ can be obtained by the method introduced in 3.3:

$$\begin{aligned} &q(0)=w^2(0,X), && q(3)=F1(w^2(3,X),w^2(6,X)) \\ &q(1)=F1(w^2(1,X),w^2(2,X),w^2(4,X)), && q(6)=F2(w^2(3,X),w^2(6,X)) \\ &q(2)=F2(w^2(1,X),w^2(2,X),w^2(4,X)), && q(5)=w^2(5,X) \\ &q(4)=F3(w^2(1,X),w^2(2,X),w^2(4,X)), && q(7)=w^2(7,X). \end{aligned} \tag{3.39}$$

The average power of $\{x(k)\}$ is

$$Pav= \sum_{i=0}^{7} q(i)= \sum_{k=0}^{7} w^2(k,X)=(1/8) \sum_{k=0}^{7} x^2(k). \tag{3.40}$$

There are many other functions which are enlargement and reduction invariant and have linear expressions by the second-degree terms of the $w(k, X)(k=0,1,---,7)$. Taking the following into consideration in this case

$$p(0)=w^2(0,X),\ \ p(2)=w^2(2,X)+w^2(3,X)$$
$$p(1)=w^2(1,X),\ \ p(3)=w^2(,X)+w^2(5,X)+w^2(6,X)+w^2(7,X), \qquad (3.41)$$

we have

$$P(X)=[p(0,X),p(1,X),p(2,X),0]=[p(0,Y),p(2,Y),p(3,Y),0]$$
$$P(Y)=[p(0,Y),0,p(2,Y),p(3,Y)]=[p(0,X),0,p(1,X),p(2,X)]$$
$$P=[p(0),p(1),p(2),p(3)] \qquad (3.42)$$

$\dashrightarrow$ enlargement

$\longleftarrow$ reduction.

This means that 2-time enlargement of an input pattern X causes shift of $P(X)$ toward the right by one element except $p(0,X)$. $p(0,X)$ is always invariant to any permutation of an input pattern. This has been proved in another way [41]. Therefore we arrive at a power spectrum $\{q(i)\}(i=0,1,2,3)$ which is invariant to translations, enlargements, and reductions:

$$q(0)=p(0),\ \ q(1)=F1(p(1),p(2),p(3)),\ \ q(2)=F2(p(1),p(2),p(3)),$$
$$q(3)=F3(p(1),p(2),p(3))$$
$$Pav=\sum_{i=0}^{3} q(i)=\ \sum_{i=0}^{3} p(i)=\sum_{k=0}^{7} w^2(k,X)=(1/8)\sum_{k=0}^{7} x^2(k). \qquad (3.43)$$

Generally the statements following(3.42) are also valid for any natural number n. This is known from (3.19) and (3.35) in the same way as used in the case when $n=3$. Then a translation, enlargement, and reduction-invariant power spectrum can be obtained:

$$q(0)=p(0),\ \ q(1)=Fl(p(1),p(2),---,p(n)),\ \ l=1,2,--,n. \qquad (3.44)$$

In the case when an input pattern X is a $2^n\times2^m$-matrix, these results are developed as follows. Let $G2$ denote the group of horizontal enlargements and reductions and vertical ones, and let $G22$, the subgroup of $G2$, denote any member of which is enlargement or reduction formed on the same scale horizontally and vertically. Examples of $G2$-equivalent and $G22$-equivalent patterns are shown in Fig. 3.2. The values in Fig. 3.2 are the $p(i,j)$ multiplied by 64 for simplicity. 1's and -1's are blacked and blanked, respectively. In the same manner, the variation of P caused by $G22$ and $G2$ are illustrated in Fig. 3.3 for the case when $n=m=3$. We cite changes of $P(X)$ in Fig. 3.2 for more detailed explanation:

$$P(X2)=\left(\begin{array}{c|cccc} p(0,0,X1) & p(0,2,X1) & \leftarrow p(0,3,X1) & \leftarrow 0 \\ \hline p(2,0,X1) & p(2,2,X1) & p(2,3,X1) & 0 \\ \uparrow & & & \\ p(3,0,X1) & p(3,2,X1) & p(3,3,X1) & 0 \\ \uparrow & & & \\ 0 & 0 & 0 & 0 \end{array}\right). \tag{3.45}$$

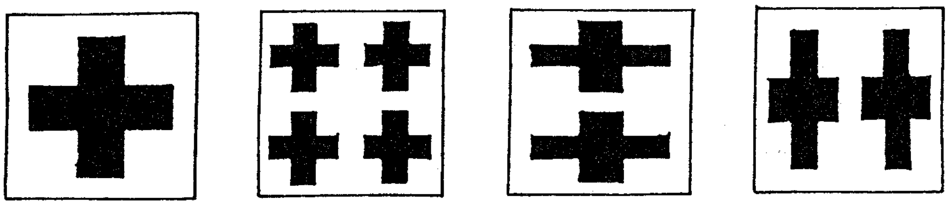
$\{p(i,j)\}$ is classified into $G22$-equivalence classes: $\{p(0,0)\}\cup\{p(0,1),p(0,2),p(0,3)\}\cup\{p(1,0),p(2,0),p(3,0)\}\cup\{p(1,1),p(2,2),p(3,3)\}\cup\{p(1,2),p(2,3)\}\cup\{p(2,1),p(3,2)\}\cup\{p(1,3)\}\cup\{p(3,1)\}$. The second-degree terms of $p(i,j)$ in the same $G22$-equivalence classes are also classified into $G22$-equivalence classes, for example, $\{p(0,1),p(0,2),p(0,3)\}\rightarrow\{p^2(0,1),p^2(0,2),p^2(0,3)\}\cup\{p(0,1)\cdot p(0,2),p(0,2)\cdot p(0,3)\}\cup\{p(0,1)\cdot p(0,3)\}$. From Corollary 2.9' the $q(0,i)(i=1,2,3)$ defined by the following expression are $G22$-invariant:

$$\begin{aligned} q(0,1)&=(p^2(0,1)+p^2(0,2)+p^2(0,3))/(p(0,1)+p(0,2)+p(0,3)) \\ q(0,2)&=2(p(0,1)\cdot p(0,2)+p(0,2)\cdot p(0,3))/(p(0,1)+p(0,2)+p(0,3)) \\ q(0,3)&=2p(0,1)\cdot p(0,3)/(p(0,1)+p(0,2)+p(0,3)). \end{aligned} \tag{3.46}$$

Let $F(1\otimes 22)=\{q(i,j)\}(i=0,1,2,3,\ j=0,1,2,3)$ be defined by

$$q(0,0)=p(0,0),\quad q(0,j)=Fj(p(0,1),p(0,2),p(0,3)),$$

$$q(i,0)=Fi(p(1,0),p(2,0),p(3,0))$$



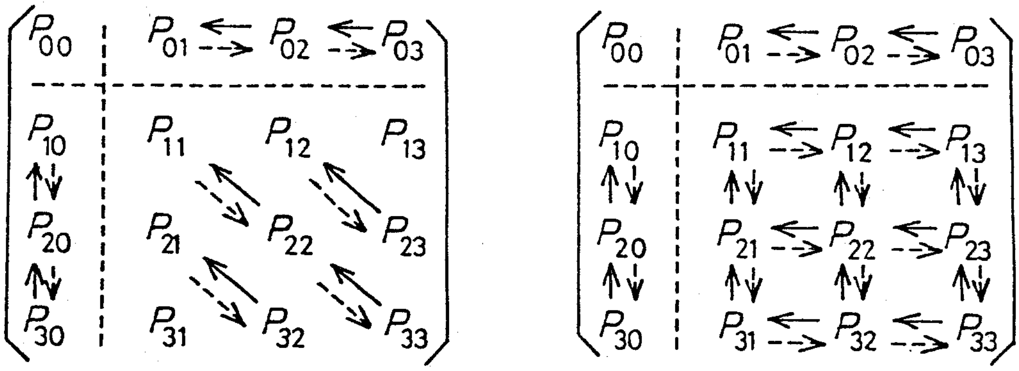
X_1 X_2 X_3 X_4

$$\begin{pmatrix} 9 & 0 & 1 & 18 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 9 & 2 \\ 18 & 0 & 2 & 4 \end{pmatrix} \quad \begin{pmatrix} 9 & 1 & 18 & 0 \\ 1 & 9 & 2 & 0 \\ 18 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 9 & 0 & 1 & 18 \\ 1 & 0 & 9 & 2 \\ 18 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 9 & 1 & 18 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 9 & 2 & 0 \\ 18 & 2 & 4 & 0 \end{pmatrix}$$

$$X_1 \overset{G_2}{=} X_2 \overset{G_2}{=} X_3 \overset{G_2}{=} X_4$$

$$X_1 \overset{G_{22}}{=} X_2 \overset{G_{22}}{\neq} X_3 \overset{G_{22}}{\neq} X_4$$

Fig. 3.2. *G2*-equivalent patterns and *G22*-equivalent patterns.



G_{22} : $\leftarrow$ ENLARGEMENT
$\rightarrow$ REDUCTION

G_2 : $\leftarrow$ HORIZONTAL ENLARGEMENT
$\rightarrow$ HORIZONTAL REDUCTION
$\downarrow$ VERTICAL ENLARGEMENT
$\uparrow$ VERTICAL REDUCTION

Fig. 3.3. Shifts of elements of *P* caused by *G22* and *G2*.

$$q(i,j)=F_{j1+1}(p(i-j1,j-j1),p(i-j1+1,j-j1+1),\text{--},p(i,j),\text{--},$$
$$p(i+j2,j+j2)) \tag{3.47}$$

where $1\leqq i-j1$, $1\leqq j-j1$, $i+j2\leqq 3$, $j+j2\leqq 3$, $i=1,2,3$, $j=1,2,3$. $F(1\otimes 22)$ is a $G1\otimes G22$-invariant power spectrum. For any numbers n and m we obtain

$$q(0,0)=p(0,0),\ q(0,j)=Fj(p(0,1),p(0,2),\text{---},p(0,m)),$$
$$q(i,0)=Fi(p(1,0),p(2,0),\text{---},p(n,0)),$$
$$q(i,\ j)=F_{i1+1}(p(i-j1,j-j1),p(i-j1+1,j-j1+1),\text{--},p(i,j),\text{--},p(i+j2,i+j2)) \tag{3.48}$$

where $1\leqq i-j1$, $1\leqq j-j1$, $i+j2\leqq n$, $j+j1\leqq m$, $i=1,2,\text{---},n$, $j=1,2,\text{---},m$. In the same way we obtain a $G1\otimes G2$-invariant power spectrum. Let $F(1\otimes 2)=\{q(i,j)\}$ $(i=0,1,\text{---}5,\ j=0,1,\text{---},5)$ be defined by

$$q(0,0)=p(0,0),\ q(0,j)=Fj(p(0,1),p(0,2),p(0,3)),$$
$$q(i,0)=Fi(p(1,0),p(2,0),p(3,0)),$$
$$q(1,1)=\sum_{k=1}^{3}\sum_{l=1}^{3}p^2(k.l)/A,\ q(1,2)=D(0,1)/A,$$
$$q(1,3)=D(0,2)/A,\ q(2,1)=D(1,0)/A,\ q(2,2)=D(1,1)/A,$$
$$q(2,3)=D(1,2)/A,\ q(2,4)=D(1,-2)/A,\ q(3,1)=D(2,0)/A,$$
$$q(3,2)=D(2,1)/A,\ q(3,3)=D(2,2)/A,\ q(4,2)=D(2,-1)/A,$$
$$q(4,4)=D(1,-1)/A,\ q(5,5)=D(2,-2)/A \tag{3.49}$$

where $i=1,\ 2,3$, $j=1,2,3$, $A=\sum_{k=1}^{3}\sum_{l=1}^{3}p(k,l)=0$, the $q(i,j)$ undefined in (3.49) are equal to zero, and $D(k,l)$'s are given by

$$D(k,l)=2\sum_{r}\sum_{s}p(r,s)p(r+k,s+l),\ 1\leqq r,s,r+k,s+l\leqq 3. \tag{3.50}$$

The $q(i,j)(i=1,2,\text{---},5,\ j=1,2,\text{---},5)$ are equal to zero, if $A=0$. Then $F(1\otimes 2)$ is a $G1\otimes G2$-invariant power spectrum. Let the $F'(i,j)(i=1,2,\text{---},L1,\ j=1,2,\text{---},L2)$ be the $G2$-equivalence classes of the second-degree terms of $p(i,j)(i=1,2,\text{---},n,\ j=1,2,\text{---},m)$, where some of them may be equal to the empty set and the $F'(i,j)$'s are symmetric with the $p(i,j)$ if $n=m$ and $L1=L2$. Exceptionally, we define that $F'(1,1)=\{(1/2)p^2(i,j)\}(i=1,2,---,n, j=1,2,---,m)$. For any numbers n and m we have

$$q(0,0)=p(0,0),\ q(0,j)=Fj(p(0,1),p(0,2),---,p(0,m))$$
$$q(i,0)=Fi(p(1,0),p(2,0),---,p(n,0)),\ q(i,j)=2\sum_{f\in F'(i,j)} f/A \tag{3.51}$$

where $i=1,2,---,L1$, $j=1,2,---,L2$, and

$$A=\sum_{k=1}^{n}\sum_{l=1}^{m} p(k,l)=0. \tag{3.52}$$

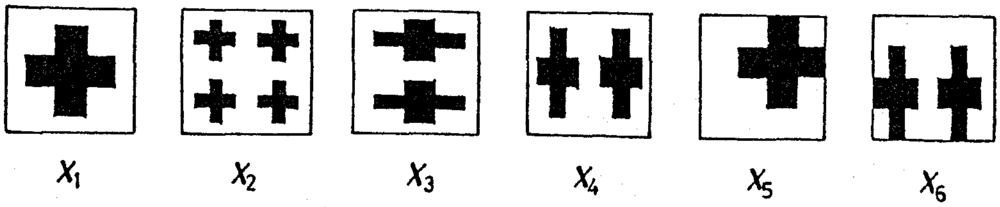
The $q(i,j)(i=1,2,---,L1,\ j=1,2,---,L2)$ are equal to zero, if $A=0$.

$$Pav=\sum_{q\in F(1\otimes 22)} q = \sum_{q\in F(1\otimes 2)} q=\sum_{i=0}^{n}\sum_{j=0}^{m} p(i,j). \tag{3.53}$$

These aspects are shown ih Fig. 3.4. The values in Fig. 3.4(b) and (c) are the $q(i,j)$ multiplied by 64. Each element $x(k,l)(k=0,1,---,7,\ l=0,1,---,7)$ of $Xi(i=1,2,---,6)$ is located at the intersection of row k and column l. 1's and -1's are blacked and blanked, respectively. The results in a 2-dimensional case can be easily extended to any number of dimensions.

3.6 Rotation by Multiples of 90°, Symmetry Transformation, and Exchanging 1 for -1($G3$)-Invariant (WHT) Power Spectrums

Let a horizontal symmetry transformation operator on an input pattern X be denoted by gh, $Gh=\{gh,gh^2\}$, a vertical one be denoted by gv, $Gv=\{gv,gv^2\}$, and a diagonal one be denoted by gd, $Gd=\{gd,gd^2\}$. Then gh^2, gv^2, and gd^2 are identity operators. Symmetry transformation of an input pattern yields only sign change of some (WHT) coefficients, and does not have any other changes because of the symmetry of row vectors



$$X_1 \overset{G_1 \otimes G_{22}}{=} X_2 \overset{G_1 \otimes G_{22}}{=} X_5 \overset{G_1 \otimes G_{22}}{\neq} X_4 \overset{G_1 \otimes G_{22}}{=} X_6$$

$$X_1 \overset{G_1 \otimes G_2}{=} X_2 \overset{G_1 \otimes G_2}{=} X_3 \overset{G_1 \otimes G_2}{=} X_4 \overset{G_1 \otimes G_2}{=} X_5 \overset{G_1 \otimes G_2}{=} X_6$$

(a)

q_{ij}	$i \backslash j$	0	1	2	3	$64\Sigma q_{ij}$
X_1	0	9.00	17.11	1.89	0	64
X_2	1	17.11	7.46	2.00	0	
X_5	2	1.89	2.00	5.54	0	
	3	0	0	0	0	
X_4	0	9.00	17.11	1.89	0	64
X_6	1	17.11	2.00	0	0	
	2	1.89	7.46	0	0	
	3	0	2.00	5.54	0	

(b)

q_{ij}	$i \backslash j$	0	1	2	3	4	5	$64\Sigma q_{ij}$
X_1	0	9.00	17.11	1.89	0	–	–	64
X_2	1	17.11	6.18	3.06	0	–	–	
X_3	2	1.89	3.06	4.23	0	0	–	
X_4	3	0	0	0	0	–	–	
X_5	4	–	–	0	–	0.47	–	
X_6	5	–	–	–	–	–	0	

(c)

Fig. 3.4. $G1 \otimes G22$-invariant power spectrum and $G1 \otimes G2$-invariant power spectrum: (a) Input patterns, (b) $G1 \otimes G22$-invariant power spectrum, (c) $G1 \otimes G2$-invariant power spectrum.

$H(n,k)$. Hence, squares of the coeffcients are invariant to symmetry transformation. Therefore the set $\{w^2(i,j,X)\}$ is $Gh\otimes Gv$-invariant. Let 90° rotation operator be denoted by gr and $Gr=\{gr,gr^2,gr^3,gr^4\}$, then gr^4 is an identity operator. We notice that gd and gr operate only on a $N\times N$ square matrix $X(N=2^n)$.

It is seen in Fig. 3.5 that gr is equivalent to $gv\circ gd$ and $gd\circ gh$, that is,

$$gr\circ X=gv\circ gd\circ X=gd\circ gh\circ X. \tag{3.54}$$

Therefore we have

$$\begin{aligned} w(i,j,X1)&=w(i,j,X2)=w(j,i,X3) \\ &=\begin{cases} w(j,i,X), & \text{if } \sum_{k=0}^{n-1} i(k) \text{ is an even number} \\ -w(j,i,X), & \text{otherwise} \end{cases} \end{aligned} \tag{3.55}$$

where $X1=gr\circ X$, $X2=gd\circ X3$, $X3=gh\circ X$, $i=\sum_{k=0}^{n-1} 2^k\, i(k)$, $i(k)=0$ or 1, $i=0,1,$ $---,2^n-1$, $j=0,1,---,2^n-1$. Let $G3=Gh\otimes Gv\otimes Gd$, then it is known from (3.54) that Gr is included in $G3$. From (3.48), (3.51), and (3.55) we obtain a $G1\otimes Gl\otimes G3$-invariant power spectrum $F(1\otimes l\otimes 3)=\{s(i,j)\}$ defined by

$$s(i,j)=\begin{cases} F1(q(i,j),q(j,i)), & \text{if } i>j \\ q(i,j), & \text{if } i=j \\ F2(q(i,j),q(j,i)), & \text{if } i<j \end{cases} \tag{3.56}$$

and

$$Pav=\sum_{s\in F(1\otimes l\otimes 3)} s=\sum_{q\in F(1\otimes l)} q=\sum_{i=0}^{n}\sum_{j=0}^{n} p(i,j) \tag{3.57}$$

where the $q(i,j)$ are in $F(1\otimes l)$ and $l=2,22$.

Geometric patterns are often drawn by using 1's and -1's which correspond to black and white points, respectively. Let ge be an exchanging operator 1 for -1 and $G4=\{ge,ge^2\}$, then

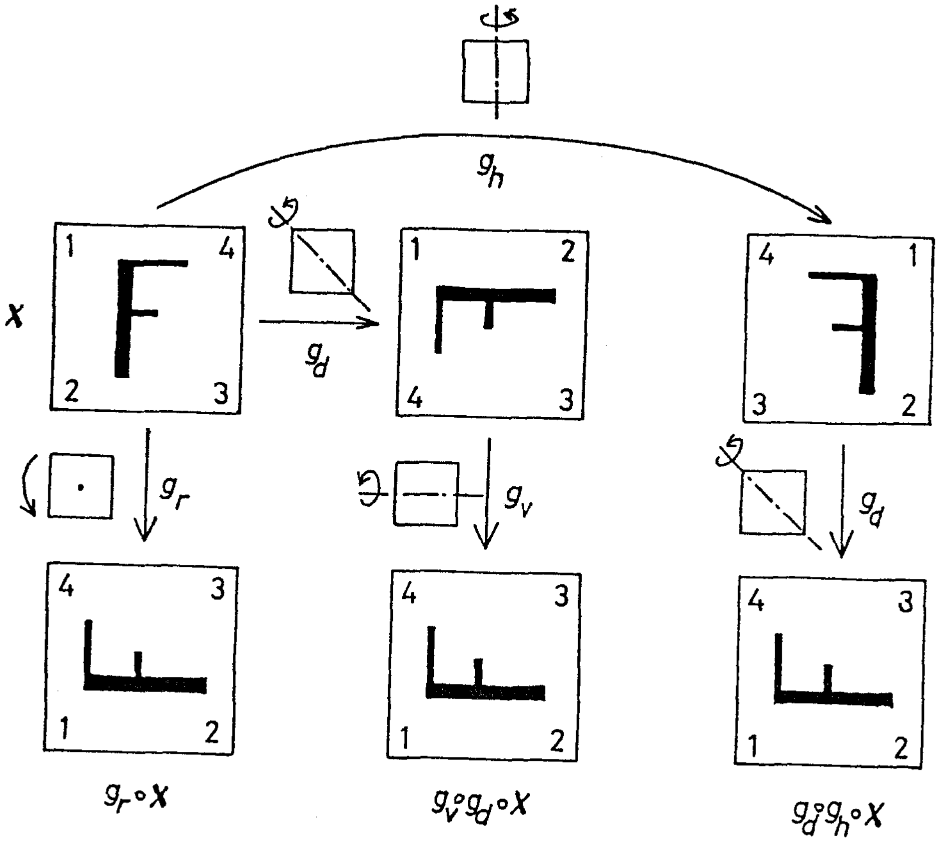


Fig. 3.5. Aspects of operations *gr*, *gv∘gd*, and *gd∘gh*.

$$ge \circ X = -X, \quad W(X1) = -W(X) \tag{3.58}$$

where $X1=ge\circ X$. From (3.58) it follows that the $w^2(i,j,X)$'s are invariant to exchanging the sign of an input pattern.

The aspects of group-invariant power spectrums described above are shown in Table 3.1. The marks o and ◎ are used to show whether or not the families of some group-invariant functions have equal values with respect to input patterns $X1,X2,---,X12$ in the same column. To take $G1$-invariant power spectrum $P(X)$, for example, $P(X1)=P(X2)=P(X4)=P(X5)=P(X6)$ and $P(X1)\neq P(X3)\neq P(X7)$ in general. It is observed from Table 3.1 that the $G1\otimes G2\otimes G3$-invariant power spectrum $F(1\otimes 2\otimes 3)$ is a developed power spectrum to be invariant to translations, enlargements, reductions, symmetry transformations, rotations by multiples of 90°, and exchanging the sign of an input pattern.

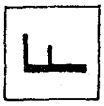
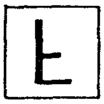
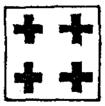
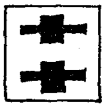
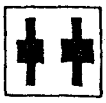
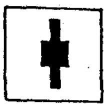
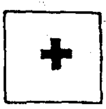
3.7 Enlargement and Reduction($G2'$)-Invariant (WHT) Power Spectrums

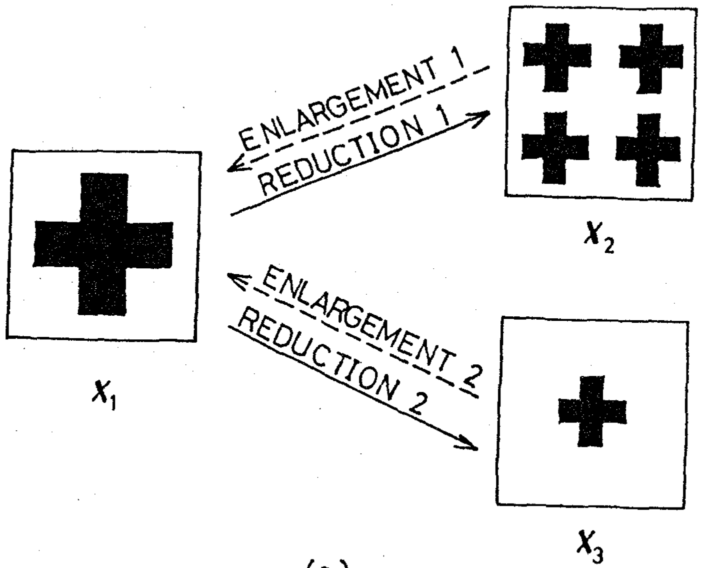
Enlargement and reduction have been already defined in the foregoing section. For instance, reduction of $X1$ yields $X2$ in Fig. 3.6. Such a definition makes mathematical analysis easy, but does not make natural sense for man. Therefore the definition is newly introduced in this section as shown in Fig. 3.6. An input pattern X is called a 2^i-time enlargeable pattern, if $X=[U,\mathbf{0}]$, $U=[x(0),x(1),---,x(2^{n-i}-1)]$ and

$$\mathbf{0}=[\underbrace{0,0,---,0}_{2^n-2^{n-i}}]. \tag{3.59}$$

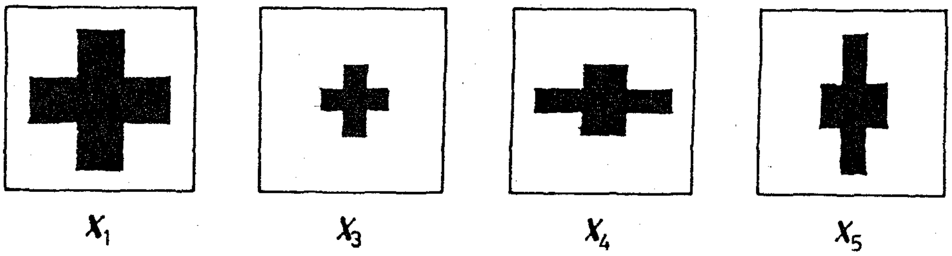
For example, $X=[12000000]$ is a 2^2-time enlargeable pattern, but $X=[01200000]$ is not. A $1/2^i$-time reducible pattern is defined in the same

Table 3.1
Aspects of group-invariant power spectrums and other group-invariant families.

Transformation Groups / Input Patterns		Group-Invariant Power Spectra					Group-Invariant Families	
		G_1	$G_1 \otimes G_{22}$	$G_1 \otimes G_2$	$G_1 \otimes G_{22} \otimes G_3$	$G_1 \otimes G_2 \otimes G_3$	$G_1 \otimes G_{22}' \otimes G_3$	$G_1 \otimes G_2' \otimes G_3$
X_1		○	○	○	○	○	○	○
X_2		○	○	○	○	○	○	○
X_3		×	×	×	○	○	○	○
X_4		○	○	○	○	○	○	○
X_5		○	○	○	○	○	○	○
X_6		○	○	○	○	○	○	○
X_7		×	◎	◎	◎	◎	×	×
X_8		×	×	◎	×	◎	×	×
X_9		×	×	◎	×	◎	×	×
X_{10}		×	◎	◎	◎	◎	◎	◎
X_{11}		×	×	×	×	×	×	◎
X_{12}		×	×	×	×	×	◎	◎



(a)



$$16\begin{pmatrix} 25 & 0 & 1 & 18 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 9 & 2 \\ 18 & 0 & 2 & 4 \end{pmatrix} \quad \begin{pmatrix} 25 & 1 & 18 & 44 \\ 1 & 9 & 2 & 12 \\ 18 & 2 & 4 & 24 \\ 44 & 12 & 24 & 80 \end{pmatrix} \quad 4\begin{pmatrix} 25 & 0 & 1 & 18 \\ 1 & 0 & 9 & 2 \\ 18 & 0 & 2 & 4 \\ 44 & 0 & 12 & 44 \end{pmatrix} \quad 4\begin{pmatrix} 25 & 1 & 18 & 44 \\ 0 & 0 & 0 & 0 \\ 1 & 9 & 2 & 12 \\ 18 & 2 & 4 & 24 \end{pmatrix}$$

$$X_1 \overset{G_{22}}{=} X_3 \overset{G_{22}}{\neq} X_4 \overset{G_{22}}{\neq} X_5$$
$$X_1 \overset{G_2}{=} X_3 \overset{G_2}{=} X_4 \overset{G_2}{=} X_5$$

(b)

Fig. 3.6. Explanation of $G2'$ and $G22'$: (a) Two different definitions of enlargements and reductions, (b) $G2'$-equivalent patterns and $G22'$-equivalent patterns.

manner as that of the foregoing section, that is, an input pattern Y is a $1/2^{i}$-time reducible pattern, if

$$Y=[y(0)\mathbf{1},y(2^{i})\mathbf{1},---,y(2^{n}-2^{i})\mathbf{1}],\ \mathbf{1}=[\underbrace{1,1,---,1}_{2^{i}}]. \tag{3.60}$$

A 2^{i}-time enlarged pattern of X and $1/2^{i}$-time reduced pattern of Y **are defined as** $[x(0)\mathbf{1},x(1)\mathbf{1},---,x(2^{n-i}-1)\mathbf{1}]$ and $[U',\mathbf{0}]$, respectively, where

$$U'=[y(0),y(2^{i}),---,y(2^{n}-2^{i})],\ \mathbf{1}=[\underbrace{1,1,---,1}_{2^{i}}]$$
$$\mathbf{0}=[\underbrace{0,0,---,0}_{2^{n}-2^{n-i}}]. \tag{3.61}$$

For instance, the 2-time exlarged pattern of X=[12000000] is [1122000], and the 2^{2}-time enlarged pattern of it is [11112222]. Reversely the 1/2-time reduced pattern of Y=[11223300] is [12300000] and so on. Under the definitions there is obviously no enlargement and reduction-invariant power spectrum, because the average power of an input pattern itself is changed by the transformations. Therefore we give up to obtain an enlargement and reduction-invariant power spectrum and try to get enlargement and reduction-invariant functions from the second-degree terms.

Let the 2^{i}-time enlarged pattern of 2^{i}**-time enlargeable pattern** X **be** Y, then Y is a $1/2^{i}$-time reducible pattern and $1/2^{i}$-time reduced pattern of Y is X. X and Y can be expressed as $X=[U,\mathbf{0}]$, $Y=U\otimes\mathbf{1}$, where

$$U=[x(0),x(1),---,x(2^{n-i}-1)],\ \mathbf{0}=[\underbrace{0,0,---,0}_{2^{n}-2^{n-i}}]$$
$$\mathbf{1}=[\underbrace{1,1,---,1}_{2^{i}}]. \tag{3.62}$$

The (WHT) of X is as follows:

$$W(X)=(1/N)X\cdot H(n)=(1/N)[U,\mathbf{0}]\cdot(H(i)\otimes H(n-i))$$

$$=(1/N)[U,0]\left(\begin{array}{cc|c} H(n-i), & H(n-i) & \\ H(n-i), & -H(n-i) & H(n-1) \\ \hline \multicolumn{2}{c|}{H(n-1)} & -H(n-1) \end{array}\right)$$

$$=(1/N)[U\cdot H(n-i),U\cdot H(n-i),---,U\cdot H(n-i)]$$

$$=(1/N)\mathbf{1}\otimes(U\cdot H(n-i)) \tag{3.63}$$

where

$$\mathbf{1}=[\underbrace{1,1,---,1}_{2^i}] \tag{3.64}$$

From (3.63) we have

$$W(k,X)=\begin{cases} \text{value of the } k\text{-th element of } (1/N)U\cdot H(n-i), & 0\leqq k\leqq 2^{n-i}-1 \\ w(l,X), & 2^{n-i}\leqq k\leqq 2^n-1 \end{cases} \tag{3.65}$$

where l is k **modulo** 2^{n-i}. Regarding Y, from (3.33) we have

$$W(Y)=(2^i/N)(U\cdot H(n-i))\otimes[\underbrace{1,0,---,0}_{2^i}]. \tag{3.33'}$$

(3.63) and (3.33') lead that

$$w(k,X)=\begin{cases} 2^i\cdot w(l,X), & \text{if } k=2^i\cdot l \\ 0, & \text{otherwise} \end{cases} \tag{3.66}$$

where $l=0,1,---,2^{n-i}-1$.

These aspects are well-illustrated by utilizing the matrix $C(X)$ defined already in the case when $n=3$. Let X be a 2-time enlargeable pattern, and let Y the 2-time enlarged pattern of it. Then we have

$$C(Y)=\begin{pmatrix} w(4,Y),0,w(6,Y),0 \\ w(2,Y),0,\ \ 0\ \ ,0 \\ 0\ \ ,0,\ \ 0\ \ ,0 \end{pmatrix}=\begin{pmatrix} 2w(2,X),0,2w(3,X),0 \\ 2w(1,X),0,\ \ 0\ \ ,0 \\ 0\ \ ,0,\ \ 0\ \ ,0 \end{pmatrix}$$

$$w(0,Y)=2w(0,X),\ w(0,X)=w(4,X),\ w(1,X)=w(5,X),$$
$$w(2,X)=w(6,X),\ w(3,X)=w(7,X),\ \sum_{k=0}^{7} w^2(k,Y)=2\sum_{k=0}^{7} w^2(k,X). \quad (3.67)$$

Let X be a 2^2-time enlargeable pattern, and let Z the 2^2-time enlarged pattern of it. Then we have

$$C(Z)=\begin{pmatrix} w(4,Z),0,0,0 \\ 0\quad ,0,0,0 \\ 0\quad ,0,0,0 \end{pmatrix} = \begin{pmatrix} 4w(1,X),0,0,0 \\ 0\quad ,0,0,0 \\ 0\quad ,0,0,0 \end{pmatrix}$$

$$w(0,Z)=4w(0,X),\ w(0,X)=w(2,X)=w(4,X)=w(6,X),$$
$$w(1,X)=w(3,X)=w(5,X)=w(7,X),\ \sum_{k=0}^{7} w^2(k,Z)=4\sum_{k=0}^{7} w^2(k,X). \quad (3.68)$$

(3.67) and (3.68) are obtained from (3.65) and (3.66).

Furthermore, it follows that

$$p(s,X)=\sum_{j=0}^{s-1} p(j,X) \quad (3.69)$$

where X is a 2^i-time enlargeable pattern, and $n-i+1\leqq s\leqq n$. The following formulas are asymptotically led from (3.69):

$$p(n-i+1,X)=\sum_{j=0}^{n-i} p(j,X)$$
$$p(n-i+2,X)=\sum_{j=0}^{n-i+1} p(j,X)=2p(n-i+1,X)$$
$$\text{---} \qquad \text{---}$$
$$p(n,X)=2^i\cdot p(n-i+1,X). \quad (3.70)$$

From (3.65) we have

$$p(0,Y)=2^{2i}\cdot p(0,X),\ p(s,Y)=\begin{cases} 2^{2i}\cdot p(s-i,X), & \text{if } i+1\leqq s\leqq n \\ 0, & \text{if } 1\leqq s\leqq i \end{cases}$$

$$\sum_{s=0}^{n} p(s,Y)=2^i \sum_{s=0}^{n} p(s,X). \quad (3.71)$$

when Y is the 2^i-time enlarged pattern of X. In the case when $n=3$ and $i=1$, we obtain

$$P(X)=[p(0,X),p(1,X),p(2,X),p(3,X)]$$

$$=[p(0,X),p(1,X),p(2,X),p(0,X)+p(1,X)+p(2,X)]$$

$$P(Y)=[p(0,Y),0,p(2,Y),p(3,Y)]$$

$$=[4p(0,X),0,4p(1,X),4p(2,X)]. \quad (3.72)$$

These aspects are illustrated in Fig. 3.6 in the 2-dimensional case when $n=m=3$. The values in Fig. 3.6 are the $p(i,j)$ multiplied by 64^2 for simplicity. 1's and 0's are blacked and blanked, respectively, and we should notice that 0's are used in Fig. 3.6 instead of -1's.

Let $P'=[p'(0),p'(1),---,p'(n)]$ be defined by

$$p'(0)=p(0)^2/A^2,\ p'(s)=p(s)(p(s)-\sum_{j=0}^{s-1}p(j))/A^2 \quad (3.73)$$

where $s=1,2,---,n$, and $A=\sum_{j=0}^{n} p(j)$. Then 2^i-time enlargement and $1/2^i$-time reduction of an input pattern cause shifts on P' by i elements toward the right and the left, respectively, except $P'(0)$. $P'(0)$ is invariant to the transformations. Therefore translation, enlargement, and reduction-invariant functions are obtained as

$$q'(0)=p'(0),\ q'(j)=Fj(p'(1),p'(2),---,p'(n)),\ 1\leqq j\leqq n. \quad (3.74)$$

We note that $\{q'(j)\}(j=0,1,---,n)$ is not a power spectrum. If we treat an input pattern whose elements have only the values of 1 and -1 and redefine enlargement and reduction by changing 0 for -1, then we can get a translation, enlargement, and reduction-invariant power spectrum. In the 2-dimensional case P' is defined by

$$P'(s,t)=p^2(s,t)(p(s,t)-\sum_{i<s} p(i,t))(p(s,t)-\sum_{j<t} p(s,j))/A^4 \quad (3.75)$$

where $s=0,1,---,n$; $t=0,1,---,m$; and

$$A=\sum_{i=0}^{n}\sum_{j=0}^{m} p(i,j). \quad (3.76)$$

Let $G'2$ and $G'22$ be defined in the similar method to those of the foregoing section. Then the $\{q'(i,j)\}$, defined by replacing $p(i,j)$ in(3.48) and (3.51) with the $p'(i,j)$, are $G1\otimes G'22$-invariant and $G1\otimes G'2$-invariant, respectively. $G1\otimes G'22\otimes G3$-invariant functions and $G1\otimes G'2\otimes G3$-invariant functions can be also obtained in the same way as those of the foregoing section. The aspects of these group-invariant functions are shown in Table 3.1. The above discussion has been restricted to real-valued input patterns and functions, but they can be easily extended to complex-valued ones [36], [37].

3.8 Complex-Valued Walsh-Hadamard Transform

(WHT) whose elements are ±1 has been discussed above, but in this section complex-valued (WHT) is introduced. In the former case one can easily define the transformations in the form of $2^{\pm n}$, but in the latter case the transformations in the form of $k^{\pm n}$. Now we can also get the (WHT) power spectrum invariant to translations, enlargements, and reductions in the form of $k^{\pm n}$. Then the above discussions can be considered as the special case when $k=2$. Let z and R be defined as:

$$z=\cos(2\pi/k)+j\cdot\sin(2\pi/k),\ j^2=-1$$

$$R=H(1,1)=(1,z,z^2,---,z^{k-1}). \tag{3.77}$$

Complex-valued Hadamard matrices can be recursively generated as follows:

$$H(0)=[1]$$

$$H(n)=\begin{pmatrix} H(n-1), & H(n-1),---, & H(n-1) \\ H(n-1), & z\cdot H(n-1),---,z^{k-1} & \cdot H(n-1) \\ ---, & ---\ \ ,---, & --- \\ H(n-1),z^{k-1} & \cdot H(n-1),---, & z\cdot H(n-1) \end{pmatrix}=H^n(1). \tag{3.78}$$

where $H^n(1)$ is the n succesive Kroneker product of $H(1)$. Let $H(n,i)$ denote the i-th row vector of $H(n)$. $H(n)$ is an orthonormal and symmetric matrix, that is, $H(n)\cdot\overline{H(n)}^t=N\cdot E(n)$, and $H(n)^t=H(n)$, where $N=k^n$, $E(n)$ is an $N\times N$-unit matrix, and "$\overline{}$" means a conjugate matrix or vector. $H(1)$ is expressed as follows:

$$H(1)=\begin{pmatrix} 1,1,------,1 \\ R \\ R*R \\ --- \\ R*R*---*R \end{pmatrix} = \begin{pmatrix} R^0 \\ R^1 \\ R^2 \\ -- \\ R^{k-1} \end{pmatrix} \tag{3.79}$$

where "*" means the product of corresponding elements and R^i means that each element is raised to the i-th power. Complex-valued Walsh-Hadamard transform (CWHT) of an input pattern X is defined as:

$$W(X)=(1/N)X\cdot H(n) \tag{3.80}$$

where $W(X)=[w(0,X),w(1,X),---,w(N-1,X)]$ and $N=k^n$. The inverse transform is defined as:

$$X=W(X)\cdot\overline{H(n)}. \tag{3.81}$$

From (3.80) it follows that

$$W(X)\cdot\overline{W(X)}^t=(1/N)X\cdot\overline{X}^t. \tag{3.82}$$

Fast algorithm for (CWHT) is nearly equal to that for (WHT) [44]. The some properties of complex-valued Walsh functions and Hadamard matrices are related in [36]. These aspects are shown in Fig. 3.7 in a simple example when $n=2$, $k=3$.

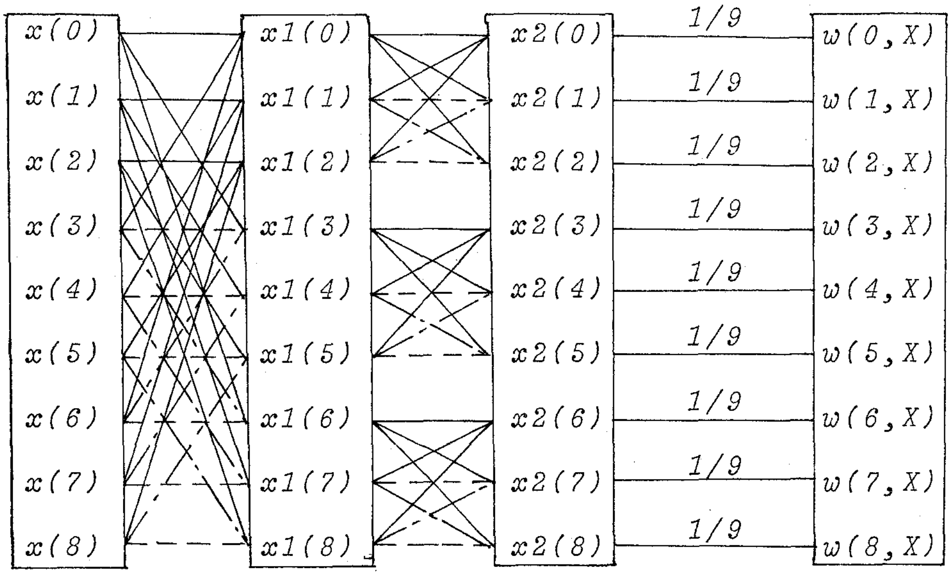


Fig. 3.7. Fast algorithm for (CWHT), when $k=3$, $n=2$. Where $\omega^3=1$, $\omega\neq 1$, $\xrightarrow{1}$ $\overset{\omega}{\dashrightarrow}$ $\overset{\omega^2}{\dashrightarrow}$.

3.9 Translation($G1$)-Invariant (CWHT) Power Spectrums

In the complex-valued case we can also obtain the corresponding power spectrum to that developed by Ahmed and others [34]. Let $g(l)$ be a translation by l elements, then $g(l)\circ X$ is the translated X by l elements. $N\times N$-**matrices** $T(n)$ and $M(n)$ are defined as:

$$T(n)=\begin{pmatrix} 010---0 \\ 001---0 \\ --- \\ 000---1 \\ 100---0 \end{pmatrix},\quad M(n)=\begin{pmatrix} 00---0 \\ 00---0 \\ --- \\ 00---0 \\ 11---1 \end{pmatrix} \tag{3.83}$$

where $N=k^n$. It is plan to see that $T(n)=\overline{T(n)}$ and $M(n)=\overline{M(n)}$.

$$g(l)\circ X=X\cdot T(n)^l$$

$$W(g(l)\circ X)=(1/N)X\cdot T(n)^l\cdot H(n)=(1/N^2)X\cdot H(n)\cdot\overline{H(n)}\cdot T(n)^l\cdot H(n)$$

$$=W(X)(1/N)\overline{H(n)}\cdot T(n)^l\cdot H(n). \tag{3.84}$$

Let $A(n)=(1/N)\overline{H(n)}\cdot T(n)\cdot H(n)$. Now let us consider the matrix $A(n)$. The product "*" of an N-vector and an $N\times N$-matrix is defined by the product "*" of the vector and the each row vector, which is defined already. Then we have

$$T(1)\cdot H(1)=R*H(1) \tag{3.85}$$

hence

$$T(n)\cdot H(n)=T(n)\cdot(H(1)\otimes H(n-1))$$

$$=\begin{pmatrix} T(n-1)\cdot H(n-1), & T(N-1)\cdot H(n-1), ---, & T(n-1)\cdot H(n-1) \\ T(n-1)\cdot H(n-1), & z\cdot T(N-1)\cdot H(n-1), ---, & z^{k-1}\cdot T(n-1)\cdot H(n-1) \\ --- , & --- , ---, & --- \\ T(n-1)\cdot H(n-1), & z^{k-1}\cdot T(N-1)\cdot H(n-1), ---, & z\cdot T(n-1)\cdot H(n-1) \end{pmatrix}$$

$$
-\begin{pmatrix} R^0\otimes M(n-1) \\ R^1\otimes M(n-1) \\ --- \\ R^{k-1}\otimes M(n-1) \end{pmatrix} + \begin{pmatrix} R^1\otimes M(n-1) \\ R^2\otimes M(n-1) \\ --- \\ R^0\otimes M(n-1) \end{pmatrix}
$$

$$
=H(1)\otimes(T(n-1)\cdot H(n-1))+(R-R^0)*[R^0,R^1---,R^{k-1}]^t\otimes M(n-1)
$$

$$
=H(1)\otimes(T(n-1)\cdot H(n-1))+((R-R^0)*H(1))\otimes M(n-1)
$$

$$
=H(1)\otimes(T(n-1)\cdot H(n-1))+(T(1)\cdot H(1)-H(1))\otimes M(n-1)
$$

$$
=H(1)\otimes(T(n-1)\cdot H(n-1)-M(n-1))+(T(1)\cdot H(1))\otimes M(n-1). \tag{3.86}
$$

Here one should notice that the last row vector of $T(n-1)\cdot H(n-1)$ is $H(n-1,0)=[1,1,---,1]$, since the first row vector of $H(n-1)$ is $H(n-1,0)$.

Let $A=[\{a(i,j)\}]$ and $B=[\{b(i,j)\}]$ are $N1\times N1$-matrices and C and D are $N2\times N2$-matrices, then we obtain

$$
(A\cdot B)\otimes(C\cdot D)= \begin{pmatrix} a(1,1)C,--, & a(1,N1)C \\ --- \quad ,--, & --- \\ a(N1,1)C,--, & a(N1,N1)C \end{pmatrix} \begin{pmatrix} b(1,1)D,---, & b(1,N1)D \\ --- \quad ,---, & --- \\ b(N1,1)D,---, & b(N1,N1)D \end{pmatrix}
$$

$$
=(A\otimes C)(B\otimes D). \tag{3.87}
$$

From (3.86) and (3.87) we have

$$
A(n)=(1/N)\overline{H(n)}\cdot T(n)\cdot H(n)
$$

$$
=(1/N)(\overline{H(1)}\otimes\overline{H(n-1)})(H(1)\otimes(T(n-1)\cdot H(n-1)-M(n-1))+(T(1)\cdot H(1))\otimes M(n-1))
$$

$$
=(1/N)(kE(1)\otimes(\overline{H(n-1)}\cdot T(n-1)\cdot H(n-1)-\overline{H(n-1)}\cdot M(n-1))
$$

$$
+\ k\begin{pmatrix} 1,0,0,--,0,0 \\ 0,z,0,--,0,0 \\ ---- \\ 0,0,0,--,0,z^{k-1} \end{pmatrix}\otimes\ \overline{(H(n-1)}\cdot M(n-1))). \tag{3.88}
$$

Let $D(n-1,i)$ be defined as

$$D(n-1,i)=(1/k^{n-1})(H(n-1)\cdot T(n-1)\cdot H(n-1)-H(n-1)\cdot M(n-1)$$
$$+z^{i}\cdot H(n-1)\cdot M(n-1)) \tag{3.89}$$

where $i=0,1,---,k-1$. Then we have

$$A(n)=\begin{pmatrix} D(n-1,0), & 0 & ,--,0 \\ 0 & ,D(n-1,1) & ,--,0 \\ --- & ,\ --- & ,--,-- \\ 0 & ,\ 0 & ,--,D(n-1,k-1) \end{pmatrix}, \tag{3.90}$$

that is, $A(n)$ is expressed by the direct sum of $D(n-1,0),D(n-1,1),---,$ $D(n-1,k-1)$. On one hand we obtain

$$D(n-1,i)\cdot\overline{D(n-1,i)}^{t}=(1/k^{n-1})^{2}\overline{H(n-1)}(T(n-1)\cdot H(n-1)+(z^{i}-1)M(n-1))\cdot$$
$$(\overline{H(n-1)}\cdot\overline{T(n-1)}^{t}+(\bar{z}^{-i}-1)M(n-1)^{t})\cdot H(n-1). \tag{3.91}$$

Since $T(n)\cdot T(n)^{t}=E(n)$ and the row vectors of $H(n-1)$ are orthogonal, we have

$$T(n-1)\cdot H(n-1)\cdot\overline{H(n-1)}\cdot\overline{T(n-1)}^{t}=k^{n-1}E(n-1)$$

$$(z^{i}-1)T(n-1)\cdot H(n-1)\cdot M(n-1)^{t}$$
$$=k^{n-1}(z^{-i}-1)T(n-1)\begin{pmatrix} 0,---,0,1 \\ 0,---,0,0 \\ --- \\ 0,---,0,0 \end{pmatrix} =k^{n-1}(z^{i}-1)\begin{pmatrix} 0,---,0,0 \\ 0,---,0,0 \\ --- \\ 0,---,0,1 \end{pmatrix}$$

$$(z^{i}-1)M(n-1)\cdot\overline{H(n-1)}\cdot\overline{T(n-1)}^{t}$$
$$=k^{n-1}(z^{i}-1)\begin{pmatrix} 0,---,0,0 \\ --- \\ 0,---,0,0 \\ 1,---,0,0 \end{pmatrix}T(n-1)^{t} =k^{n-1}(z^{i}-1)\begin{pmatrix} 0,---,0,0 \\ --- \\ 0,---,0,0 \\ 0,---,0,1 \end{pmatrix}$$

$$(z^{i}-1)(z^{-i}-1)M(n-1)\cdot M(n-1)^{t}=k^{n-1}(2-z^{-i}-z^{i})\begin{pmatrix} 0,---,0,0 \\ --- \\ 0,---,0,0 \\ 0,---,0,1 \end{pmatrix}. \tag{3.92}$$

Hence

$$D(n-1,i)\cdot\overline{D(n-1,i)}^{t}=(1/k^{n-1})^{2}\overline{H(n-1)}k^{n-1}E(n-1)\cdot H(n-1)=E(n-1). \tag{3.93}$$

(3.92) is satisfied for any $i(i=0,1,---,k-1)$, and in the special case when $i=0$ we have

$$D(n-1,0)=(1/k^{n-1})\overline{H(n-1)}\cdot T(n-1)\cdot H(n-1)=A(n-1). \tag{3.94}$$

In other words, $A(n)$ is decomposed into the direct sum of k pieces of $D(n-1,i)$, furthermore $D(n-1,0)=A(n-1)$ is also decomposed into direct sum of k pieces of $D(n-2,i)$. Thus $D(m,0)(m=n-1,n-2,---,2,1)$ can be decomposed into the direct sum of smaller matrices. From (3.84), (3.90), and (3.94), we have

$$W(g(1)\circ X)=W(X)\cdot A(n),\quad W(j,g(1)\circ X)=W(j,X)\cdot D(n-1,j)$$
$$W(i,j,g(1)\circ X)=W(i,j,X)\cdot D(i-1,j) \tag{3.95}$$

where

$$W(j,X)=[\{w(i1,X)\}]\ (i1=k^{n-1}\cdot j,\ k^{n-1}\cdot j+1,---,k^{n-1}\cdot(j+1)-1,$$
$$j=0,1,---,k-1)$$
$$W(i,j,X)=[\{w(i2,X)\}]\ (i2=k^{i-1}\cdot(j-1),k^{i-1}\cdot(j-1)+1,---,k^{i-1}\cdot j-1,$$
$$j=1,2,---,k-1). \tag{3.96}$$

Taking (3.93) into consideration, we have

$$p(i,j)=\sum_{i2}|w(i2,X)|^{2}$$
$$=W(i,j,X)\cdot\overline{W(i,j,X)}^{t}=W(i,j,g(1)\circ X)\cdot\overline{W(i,j,g(1)\circ X)}^{t}. \tag{3.97}$$

Since we have

$$p(0)+\sum_{i=1}^{n}\sum_{j=1}^{k-1}p(i,j)=\sum_{i=0}^{N-1}|w(i,X)|^2=\sum_{i=0}^{N-1}|x(i)|^2, \quad (3.98)$$

the set $\{p(0)\}\cup\{p(i,j)\}(i=1,2,---,n,\ j=1,2,----,k-1)$ is a translation-invariant power spectrum. Where $N=k^n$. From the properties of complex-valued Walsh functions we get the expressions corresponding to (3.34) and (3.35) for a k^i-time enlargeable pattern X and a $1/k^i$-time reducible pattern Y:

$$w(l,X)=0, \quad \text{if } l\geqq k^{n-i}$$

$$w(l,Y)=0, \quad \text{if } l \text{ is not a multiple of } k^i. \quad (3.99)$$

When Y is a k^i-time enlarged pattern of X, we have

$$w(l,X)=w(k^i\cdot l,Y) \quad (3.100)$$

where $l=0,\ 1,--,k^{n-i}-1$. Therefore k^i-time enlargements and $1/k^i$-time reductions on X cause parallel translations by i elements toward the upper parts and the lower on P defines as follows, respectively:

$$P=\left(\begin{array}{c|c} p(0) & p(n,1),p(n,2),---,p(n,k-1) \\ & ---\ ,\ ---\ ,---,\ --- \\ & p(1,1),p(1,2),---,p(1,k-1) \end{array}\right). \begin{array}{c} \text{enlargements} \uparrow \\ \downarrow \text{reductions} \end{array} \quad (3.101)$$

The case when $n=k=3$ is shown in the following:

$$P=\left(\begin{array}{c|c} p(0) & p(3,1),p(3,2) \\ & p(2,1),p(2,2) \\ & p(1,1),p(1,2) \end{array}\right)$$

$$=\left(\begin{array}{c|ccccccccc|ccccccccc} 0 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 & 25 & 26 \\ \hline & 3 & & & 4 & & & 5 & & & 6 & & & 7 & & & 8 & & \\ \hline & 1 & & & & & & & & & 2 & & & & & & & & \end{array}\right) \quad (3.102)$$

where $p(0)$ and $p(i,j)$'s$(i=1,2,3,\ j=1,2)$ are the square sums of the corresponding groups, for example, $p(0)=|w(0)|^2$, $p(2,1)=|w(3)|^2+|w(4)|^2+$

$|w(5)|^2$, and so on. Then we can get a $G1\otimes G2$-invariant power spectrum $\{q(0)\}\cup\{q(l,r)\}(l=1,2,---,n,\ r=1,2,---,k-1)$ as follows:

$$q(0)=p(0)$$

$$q(l,r)=Fl(p(1,r),p(2,r),---,p(n,r)) \tag{3.103}$$

where $q(0)+\sum_{l=1}^{n}\sum_{r=1}^{k-1} q(l,r)=\sum_{i=1}^{n}\sum_{j=1}^{k-1} p(i,j)=\sum_{i=1}^{N-1}|x(i)|^2$. The development of the power spectrum to 2-dimensional case is easily led in a similar manner to 3.4 and 3.5.

3.10 Conclusions

A composing process of some transformation group-invariant functions and the application to the (WHT) power spectrum have been presented. The main idea is to find a permutation group on a family of some functions caused by the transformation operating on an input pattern . Using the process, the (WHT) power spectrum are developed to be unchangeable by translations, enlargements, reductions, rotations by multiples of 90°, and symmetry transformations. Using polar coordinates (r,θ) instead of orthogonal ones, we can define any rotations besides rotations by multiples of 90°. Then every rotation may be regarded as translation toward θ-direction and enlargement and reduction as exponential shifting toward r-direction. With of this convenience the new problem arises, that is, how to define translations of the elements of an input pattern. We have the alternative of orthogonal or polarcoordinates, complying with needs.

Since the power spectrums may be regarded as a proper subset of a group-invariant complete system, it cannot perfectly make distinctions between the group-nonequivalent patterns. For example, P is translation-invariant and also sign exchanging-invariant at the same time. But generally it seems almost impossible to make up a group-invariant complete system of hardware when the number of functions in the system is taken into account. Therefore it becomes very important to select appropriately a subset of the system. We mainly adopted power spectrums, but it is also possible to adopt any other functions besides power spectrums. As seen in 3.7 there is no group-invariant power spectrum in some cases. This depends on the transformation group. Although this chapter is limited to the applications to the (WHT) power spectrums on a discrete input space, the discussions will enhance the further research of group-invariant functions.

CHAPTER 4

FOURIER SPECTRUMS INVARIANT TO CERTAIN TRANSFORMATIONS

4.1 Introduction

Fourier sinusoids are used as partial functions in this chapter. Fourier spectrums (not power spectrums) are developed to be unchangeable under several transformations such as translations, enlargements, reductions, and so on. Although the Fourier transform takes more computation time than the Walsh-Hadamard transform, it is more convenient in treating general transformations. Fourier power spectrum, auto-correlation function, and the (WHT) power spectrum developed in Chapter 3 are translation-invariant. But they are not translation-invariant complete systems. The spectrums proposed here are transformation-invariant complete systems, and preserve any essential information without any loss. Therefore (a representative of the class of) an input pattern can be regenerated through the inverse Fourier transform. Parameters introduced here represent the degree of transformations, so they can be used for normalization of an input pattern. The normalization is less affected by local distortion and low energy noise, since the Fourier transform is a global transform. Through computer simulation these aspects are shown and the efficiency of our theory is confirmed.

4.2 Input Patterns and the Fourier Transform

Let $f(x)(-\infty<x<\infty)$ be an input pattern, and an N-vector $X=[f(x(0)), f(x(1)), ---, f(x(N-1))]$ be the sampled input pattern, where $x(i)=i/N$.

We suppose the following two cases:

(1) $f(x)$ is a periodic function with a period 1,

(2) $f(x)$ is an aperiodic function, and $f(x)=0$ for any x $(x<0,\ x\geqq 1)$.

In both cases transformation-invariant functions are obtained in the similar way. Fourier spectrum $F(k)(k$:integer$)$ means the former case and $F(z)(z$:real$)$ does the latter. Fourier spectrum can be computed in a short time by the fast Fourier transform (FFT).

Here one problem rises when we treat such transformations as enlargements and reductions. It is illustrated in a simple example. In Fig. 4.1 $f2(x)$ is the 2-time enlarged pattern of $f1(x)$. $X1$ and $X2$ are the corresponding vectors, respectively. Almost all computer simulations on picture processing deal with sampled patterns on a mesh space. Then it is more natural to consider $X3$ as the 2-time enlarged pattern of $X1$. For another example, the 2-time enlarged pattern of $X4$ is $X5$. Therefore in this chapter we deal only with step functions whose values does not change within sampling intervals. The FFT can not be applied to this case as it is. But with a little contrivance it becomes applicable and saves us very much computation time. Let $f(x)\leftrightarrow F(k)$ be the Fourier transform pair:

$$f(x)=\sum_{k=-\infty}^{\infty} F(k)e^{j2\pi kx}$$

$$F(k)=\int_0^1 f(x)e^{-j2\pi kx}dx=\sum_{i=0}^{N-1} f(x(i))\int_{x(i)}^{x(i+1)} e^{-j2\pi kx}dx$$

$$=(1/j2\pi k)\sum_{i=0}^{N-1}\{f(x(i+1))-f(x(i))\}e^{-j2\pi kx(i)}$$

$$(k\neq 0,\ x(N)=x(0),\ j^2=-1)$$

$$F(0)=(1/N)\sum_{i=0}^{N-1} f(x(i)). \tag{4.1}$$

(4.1) means that $F(k)(k\neq 0)$ is the coefficients divided by $j2\pi k$ which is

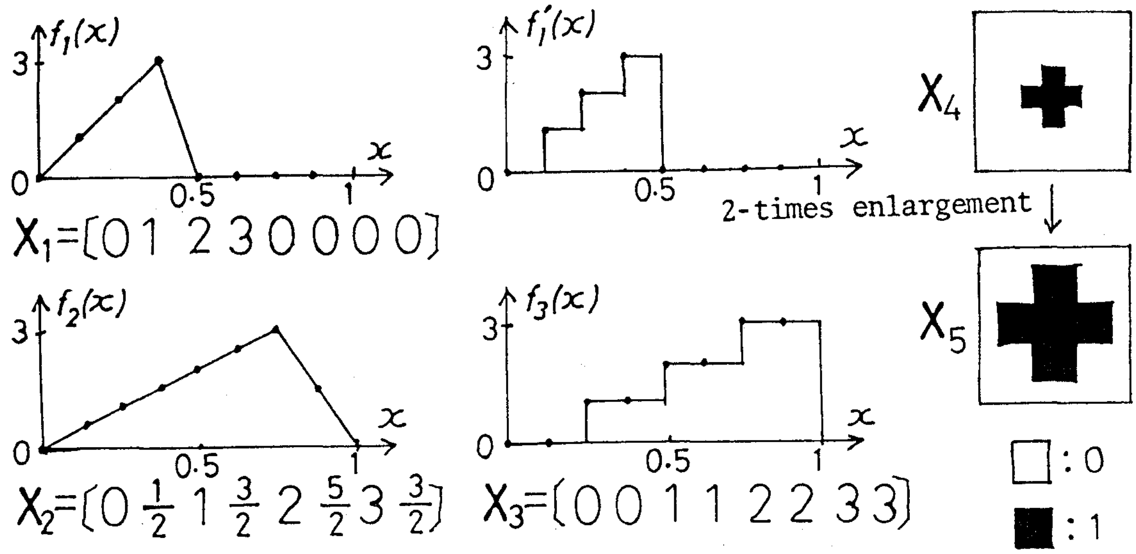


Fig. 4.1. Input patterns and their vector expressions.

obtained from the discrete Fourier transform of difference between adjacent sampled components. That is, the FFT becomes applicable when $N=2^n$.

4.3 Translation-Invariant Spectrums

In this section a translation-invariant spectrum is obtained. For simplicity, the discussion is restricted to a 1-dimensional case, but it is easily extensible into higher-dimensional cases. It is well known that auto-correlation functions $\phi_{ff}(\tau)$ and the Fourier power spectrum $|F(k)|^2$ are translation-invariant. When $f=f(x)f(x+\tau)$ and arithmetic summation $\sum$ is exchanged for integral $\int$ in Corollary 2.9, $f1$ and $\phi_{ff}(\tau)$ are the same. $\phi_{ff}(\tau)$ and $|F(k)|^2$ are a Fourier transform pair and both of them reserve only amplitude information and no phase one. In other words, they are translation-invariant, but not translation-invariant complete systems.

To construct a translation-invariant complete system, we examine changes of the phase angles caused translations of an input pattern. Let $Fa(k)$ be the Fourier transform of $f(x-a)$ which is a translated pattern of $f(x)$ by a.

$$Fa(k)=\int_0^1 f(x-a)e^{-j2\pi kx}dx=e^{-j2\pi ak}F(k) \tag{4.2}$$

(4.2) means that amplitude components are the same and phase components change by $-2\pi ak$. As a calculative result of a phase angle, $[arg(F(k))-2\pi ak]_{2\pi}$ is obtained. Where $arg(\ )$ means an argument and $[\ ]_{2\pi}$ does mod 2π. Since the interval $(0\leqq x<1)$ is an object of our discussion as the domain of an input pattern $f(x)$, the phase angle of $Fa(1)$ can be considered a parameter of location of an input pattern. Using this fact, a translation-invariant spectrum is constructed as follows:

$$|Fs(k)|=|F(k)|$$

$$arg(Fs(k))=[arg(F(k))-k(arg(F(1))-\pi)]_{2\pi} \quad (k\neq 0)$$

$$Fs(0)=F(0),\ \Delta s=[arg(F(1))]_{2\pi}/2\pi \tag{4.3}$$

$Fs(k)$ is the Fourier specturm which is translated so that $arg(Fs(1))=\pi$. This has close relations to translating the input pattern $f(x)$ so that the center of gravity may be 0.5. Δs can be regarded as a parameter of location of an input pattern and $0\leq\Delta s<1$. For example, for X=[11000000], $F(1)=0.00507e^{j7\pi}/4$, $arg(F(1))-\pi=3\pi/4$, therefore $\Delta s=1/8$(=the center of gravity) and the pattern is translated toward the right by 3 elements(= $3\pi/4\times 1/2\pi\times 8$). Through the inverse Fourier transform of $Fs(k)$ the input pattern [00011000] is obtained, and the center of gravity is 0.5.

It is proved as follows that $Fs(k)$ is a translation-invarinat complete system:

(Proof) Invariance:Let $f(k)\leftrightarrow Fs(k)$. For amplitude we have

$$|Fs(k)|=|Fa(k)|=|F(k)|. \tag{4.4}$$

For argument we have

$$arg(Fs(k))=[arg(Fa(k))-k(arg(Fa(1))-\pi)]_{2\pi}$$

$$=[arg(F(k))-k(arg(F(1))-\pi)]_{2\pi} \quad (k\neq 0)$$

$$Fs(0)=Fa(0)=F(0). \tag{4.5}$$

Completeness: Suppose that $Fs1(k)=Fs2(k)$, **and** $Fs1(k)$ and $Fs2(k)$ are $Fs(k)$'s obtained from G-nonequivalent functions $f1$ and $f2$ $(f1\overset{G}{\neq}f2)$, respectively. Let $f1(x)\leftrightarrow F1(k)$, $f2(x)\leftrightarrow F2(k)$.

$$|F1(k)|=|Fs1(k)|=|Fs2(k)|=|F2(k)|$$

$$[arg(F1(k))-k(F1(1)-\pi)]_{2\pi}=arg(Fs1(k))$$

$$=arg(Fs2(k))=[arg(F2(k))-k(F2(1)-k)]_{2\pi}. \tag{4.6}$$

Hence

$$f1(x-(F1(1)-\pi)/2\pi) \leftrightarrow Fs1(k)$$

$$f2(x-(F2(1)-\pi)/2\pi) \leftrightarrow Fs2(k). \tag{4.7}$$

Since $Fs1(k)=Fs2(k)$ from the supposition, we obtain

$$f1(x-(F2(1)-\pi)/2\pi)=f2(x-(F2(1)-\pi)/2\pi). \tag{4.8}$$

(4.8) means that $f1\overset{G}{=}f2$. This is inconsistent with the supposition. QED.

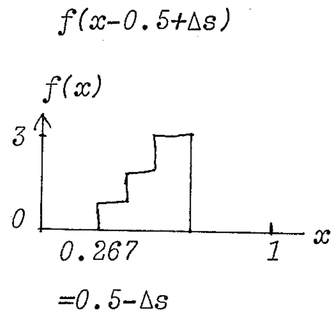
Deciding the phase angle of $F(1)$, the input pattern is reproduced through the inverse Fourier transform. Though the phase angle of the frequency 1 is picked out as the standard on the above, that of another frequency k can be adopted, too. For instance, we cannot but do so when $|F(1)|=0$. Note that the decision of the phase angle has the freedom of k. To utilize the phase angles for normalization of an input pattern, the freedom is got rid of by adopting several phase angles of the standard patterns at one time. The example of a translation-invariant spectrum is shown in Table 4.1. The parameter Δs expresses the location of an input pattern. The normalized patterns $f(x-0.5+\Delta s)$ by using Δs are the same for the three input patterns, and so are the translation-invariant spectrum $Fs(i)$'s$(i=0,1,2,3)$.

4.4 Enlargement and Reduction-Invariant Spectrums

In this section two kinds of enlargements and reductions are introduced, and we obtain the spectrums invariant to them. First, we consider periodic functions with a period 1. When $f(x+1/b)=f(x)$ for an natural number b and an arbitrary real number x, we say that $f(x)$ is b-time

Table 4.1. An example of translation-invariant spectrum.

Input patterns	$F(1)$	Δs	$Fs(0)$	$Fs(1)$	$Fs(2)$	$Fs(3)$
[12300000]	$0.613e^{j0.767\times 2\pi}$	0.233	0.75	$0.613e^{j\pi}$	$0.318e^{j6.07}$	$0.161e^{j1.90}$
[01230000]	$0.613e^{j0.642\times 2\pi}$	0.358 $=0.233+\frac{1}{8}$	0.75	$0.613e^{j\pi}$	$0.318e^{j6.07}$	$0.161e^{j1.90}$
[00123000]	$0.613e^{j0.517\times 2\pi}$	0.488 $\cong 0.233+\frac{2}{8}$	0.75	$0.613e^{j\pi}$	$0.318e^{j6.07}$	$0.161e^{j1.90}$



enlargeable and $f(x/b)$ is the b-time enlarged pattern of $f(x)$. It is conversely said that $f(x)$ is the $1/b$-time reduced pattern of $f(x/b)$. For instance, [01230123] is 2-time enlargeable, and the 2-time enlarged pattern is [00112233]. Let $f(x)\leftrightarrow F(k)$, $f(x/b)\leftrightarrow Fb(k)$, then we obtain

$$\begin{aligned} F(k) &= \int_0^1 f(x) e^{-j2\pi kx} dx = \sum_{i=0}^{b-1} \int_{i/b}^{(i+1)/b} f(x) e^{-j2\pi kx} dx \\ &= \int_0^{1/b} f(x) e^{-j2\pi kx} \sum_{i=0}^{b-1} e^{-j2\pi ki/b} dx \\ &= (1/b) \int_0^1 f(x/b) e^{-j2\pi kx/b} \sum_{i=0}^{b-1} e^{-j2\pi ki/b} dx. \end{aligned} \tag{4.9}$$

Since

$$\sum_{i=0}^{b-1} e^{-2\pi ki/b} = \begin{cases} b, & \text{if } k \text{ is a multiple of } b \\ 0, & \text{otherwise,} \end{cases} \tag{4.10}$$

we have

$$F(k) = \begin{cases} Fb(k/b), & \text{if } k \text{ is a multiple of } b \\ 0, & \text{otherwise.} \end{cases} \tag{4.11}$$

Saying in other words, a b-time enlargement of an input pattern causes a translation of frequency components $F(k)$'s into $F(k/b)$'s. This is shown in a simple example of Table 4.2.

Let the abscissa be $\log k$ and the ordinate be $F(k)$, then a b-time enlargement becomes a translation by $-\log b$. We can construct enlargement and reduction-invariant fucntions, when translation-invariant functions are composed on such an axis of co-ordinates. For example, let us construct enlargement and reduction-invariant functions by the method of Theorem 2.9:

$$F' = \{F(0),\ \sum_{k=1}^{\infty} F(k),\ F^2(0),\ \sum_{k=1}^{\infty} F^2(k),\ \sum_{k=1}^{\infty} F(l\cdot k)F(m\cdot k), ---\}$$
$$(l=1,2,---,\ m=l+1,l+2,---). \tag{4.12}$$

According to the above definition, the 1/2-time reduced pattern of

Table 4.2. 2-time enlargement and change of Fourier spectrum.

Input patterns	F(0)	F(1)	F(2)	F(3)	F(4)
[12301230]	1.5	0	-0.637	0	-0.318j
[11223300]	1.5	-0.637	-0.318j	0.212	0

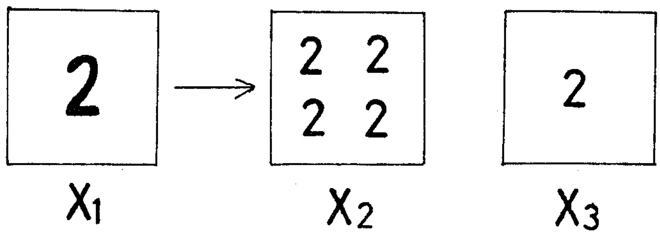


Fig. 4.2. Two kinds of reduced patterns.

$X1$ is $X2$ as shown in Fig. 4.2. Such a definition makes mathematical analysis easy, but it is unnatural for human feeling. It is more natural and moreover has wider practial application areas in pattern recognition and picture processing that we define the 1/2-time reduced pattern of $X1$ as $X3$. Let $f(x)=0(x<0,\ 1\leqq x)$, the Fourier transform of $f(x)$ be $F(z)$, and the central frequency c be defined as folllows:

$$c:Q(c)=P/2,\ \ Q(c-\varepsilon)<P/2\ \ (\varepsilon>0)$$
$$Q(t)=\int_{-t}^{t}|F(z)|^2dz\ \ (t>0)$$
$$P=Q(\infty)=\int_0^1|f(x)|^2dx. \tag{4.13}$$

An input pattern $f(x)$ is always $1/b$-time reducible $(b>1)$. $f(bx)$ is called the $1/b$-time reduced pattern of $f(x)$. If there exists a real number a such that

$$f((x-a)/b)=0\ \ (x<0,\ 1\leqq x,\ |a|<1) \tag{4.14}$$

we say that $f(x)$ is b-time enlargeable and $f((x-a)/b)$ is the b-time enlarged pattern of $f(x)$. For instance, [01230000] is 2-time enlargeable, and the 2-time enlarged pattern of it is [00112233]. For simplicity, suppose that $a=0$, and let us investigate the change brought about b-time enlargement. Let $f(x)\leftrightarrow F(z)$, $f(x/b)\leftrightarrow Fb(z)$, then we have

$$Fb(z)=\int_0^{\infty} f(x/b)e^{-j2\pi zx}dx=b\cdot F(bz). \tag{4.15}$$

Let the central frequencies of $f(x)$ and $f(x/b)$ be $c1$ and $c2$, respectively.

$$c1:\int_{-c1}^{c1}|F(z)|^2dz=P/2$$
$$c2:\int_{-c2}^{c2}|Fb(z)|^2dz=b\int_{-bc2}^{bc2}|F(z)|^2dz=Pb/2=b\cdot P/2$$
$$Pb=\int_{-\infty}^{\infty}|Fb(z)|^2dz=b\int_{-\infty}^{\infty}|F(z)|^2dz=b\cdot P. \tag{4.16}$$

Hence, it follows that $c1=b\cdot c2$. In the same way as the parameter Δs of the pattern location is obtained from the phase angle of frequency 1, the parameter Δe of the pattern size is done from the central frequency. Let the central frequency of the standard pattern be $c0$, then the enlargement and reduction-invariant spectrum is got in the following equation:

$$\Delta e = c0/c$$

$$Fe(z) = F(z \cdot c/c0)/P$$

$$P = \int_0^1 |f(z)|^2 dz = \int_{-\infty}^{\infty} |F(z)|^2 dz. \tag{4.17}$$

Where $f(x)$ and $K \cdot f(x)$ are included in the same equivalence class. For example, [12300000]=[24600000]. This means that the pictures of photograph and TV are in the same equivalence class, even if the brightness is changed.

It is proved as follows that $Fe(z)$ is an enlargement and reduction-invariant complete system:

(proof) Invariance: Let $f(x) \leftrightarrow F(z)$, $f(x/b) \leftrightarrow Fb(z)$, and Pb, P, $c1$, and $c2$ be defined in (4.16) and (4.17), then

$$Fe(z) = Fb(z \cdot c2/c0)/Pb = F(z \cdot c1/c0)/P. \tag{4.18}$$

Completeness: The proof is done by a reduction to absurdity. Let $Fe1(z)$ and $Fe2(z)$ be $Fe(z)$'s obtained from G-nonequivalent pair $f1$, $f2$ ($f1 \overset{G}{\neq} f2$), and $c1$ and $c2$ be the central frequency of $f1(x)$ and $f2(x)$, respectively. Suppose that $Fe1(z) = Fe2(z)$, $f1(x) \leftrightarrow F1(z)$, $f2(x) \leftrightarrow F2(z)$, $f3(x) \leftrightarrow Fe1(z) = Fe2(z)$.

$$f3(x) \leftrightarrow Fe1(z) = F1(z \cdot c1/c0)/P1 \leftrightarrow f1(x/b1)/(b1 \cdot P1) \quad (b1 = c1/c0)$$

$$f3(x) \leftrightarrow Fe2(z) = F2(z \cdot c2/c0)/P2 \leftrightarrow f2(x/b2)/(b2 \cdot P2) \quad (b2 = c2/c0). \tag{4.19}$$

This means that $f3(x)$ is the $b1$-time enlarged pattern of $f1(x)/(b1 \cdot P1)$, that is, $f3 \overset{G}{=} f1$. In the same way it is obtained that $f3 \overset{G}{=} f2$. Since we have $f1 \overset{G}{\neq} f2$, this contradicts to the supposition. QED.

An example of an enlargement and reduction-invariant spectrum is shown in Table 4.3. The changes of the Fourier spectrum and the central frequency caused by an enlargement are seen in the table. $Fe(z)$'s$(z=0,1,$ ---, $4)$ are the same for the two input patterns. $f(x)$ becomes $f(x/b-a)$, after b-time enlargement and translating by a. Let $f(x) \leftrightarrow F(z)$, then $f(x/b-a) \leftrightarrow b \cdot e^{-j2\pi ab} F(bz)$. First, the enlargement and reduction-invariant spectrum $Fe(z)=F(z \cdot c/c0)/P$ is obtained. Next, the translation-invariant spectrums are got by using the phase angle of $Fe(1)$. Thus we arrive at a translation, enlargement, and reduction-invarinant spectrum. The above discussion is easily extended to a 2-dimensional case.

4.5 Rotation and Other Transformation-Invariant Spectrums

Rotation can be considered translation of the angle(θ), when a 2-dimensional pattern is expressed by polar coordinates (r,θ). For an input pattern $f(r,\theta)$, $F(kr,k\theta)$ is defined as follows, and such a pair is written $f(r, \theta) \overset{r}{\leftrightarrow} F(kr,k\theta)$:

$$F(kr,k\theta)= \int_0^1 \int_0^1 f(r,\theta) e^{-j2\pi k_r r} e^{-j2\pi k_\theta \theta} r dr d\theta . \tag{4.20}$$

Let us note that (4.20) is different from the Fourier spectrum expressed by polar coordinates. In this a phase angle is treated as if it is one axis of orthogonal coordinates. Therefore the same technique as used to obtain a translation-invariant spectrum is applicable for obtaining a rotation-invariant spectrum. Let $f(r,\theta-d) \overset{r}{\leftrightarrow} Fd(kr,k\theta)$, then we have

$$Fd(kr,k\theta)= \int_0^1 \int_0^1 f(r,\theta-d) e^{-j2\pi k_r r} e^{-j2\pi k_\theta \theta} r dr d\theta = e^{-j2\pi d k_\theta} F(kr,k\theta) . \tag{4.21}$$

Table 4.3. An example of enlargement and reduction-invariant spectrum.

Input patterns	F(0)	F(1)	F(2)	F(3)	F(4)	c	Fe(0)	Fe(1)	Fe(2)	Fe(3)	Fe(4)
[12300000]	0.75	0.066 -0.61j	-0.318	0.128 +0.1j	-0.159j	0.854=c0	0.429	0.038 +0.348j	-0.182	0.073 +0.855j	-0.091j
	×2		×2	×2		1/2					
[11223300]	1.50	-0.637	-0.318j	0.212	0	0.428	0.429	0.038 +0.348j	-0.182	0.073 +0.855j	-0.091j

In the similar way to 4.3, the phase angle of $F(0,1)(kr=0,k\theta=1)$ can be regarded as the parameter of inclination. It is expressed as $\Delta r=arg(F(0,1))$. Let $\Delta r0$ be the parameter of a reference pattern and Δr be one of an input pattern, then the input pattern can be normalized by rotating by $\theta 0(=\Delta r0-\Delta r)$. Rotation of an input pattern by $\theta 0$ causes rotation of the Fourier spectrum by $\theta 0$. Hence $Fr(kr,k\theta)$ defined in the following is a rotation-invariant spectrum:

$$Fr(kr,k\theta)=F(kx\cdot cos(\theta 0)+ky\cdot sin(\theta 0),-kx\cdot sin(\theta 0)+ky\cdot cos(\theta 0)) \quad (4.22)$$

where, $f(x,y)\leftrightarrow F(kx,ky)$. It is easily proved that $Fr(kx,ky)$ is a rotation-invariant complete system in the same manner as that of a translation-invariant spectrum. For an input pattern $f(x)(0\leqq f(x)\leqq 1)$, the negative-positive reversed pattern is defined as $f'(x)=1-f(x)$. Then $Fn(k)$ defined in the following is a negative-positive reversion-invariant spectrum:

$$Fn(k)=\begin{cases} F(k), & \text{if } Re(F(1))>0 \\ -F(k), & \text{if } Re(F(1))<0,\ k\neq 0 \\ 1-F(0), & \text{if } Re(F(1))<0,\ k=0 \end{cases} \quad (4.23)$$

where $Re(\ )$ means the real part.

Axis-symmetry transformation is defined as $f''(x)=f(1-x)=f(-x)$, where $f(x)$ is a periodic function with period 1. The following $Fa(k)$ is an axis-symmetry transformation-invariant spectrum:

$$Fa(k)=Re(F(k))+j\cdot Im((F(k))sgn(Im(F(1))) \quad (4.24)$$

where $Im(\ )$ and $sgn(\ )$ mean the imaginary part and the sign, respectively. When $Re(F(1))=0$ and $Im(F(1))=0$ in (4.23) and (4.24), respectively, the transformation-invariant spectrums must be redefined by using higher order nonzero components. It is also easily proved that *Fn(k)* and *Fa(k)* are transformation-invariant complete systems.

4.6 Computer Simulation

In this section we have a try of computer simulation of the modified Fourier spectrum invariant to translations, enlargements, reductions, and rotations. Using the spectrum, some cartoons are classified into two classes. Since the Fourier transform is a global transform, it is less affected by local distortion and noise. We take an example of segmentation in Fig. 4.3. From a maximum flame method *Segmentation 1* is obtained. On the other hand, if the energy of noise is low, *Segmentation 2* becomes possible by the method proposed here.

The aspects of obtaining the invariant spectrum to various transformations are shown through computer simulation. Ten patterns in Fig. 4.4 are used as original data. *X2,X3,---,X8* are got through adding noise, translating, and tilting after making equal or reduced size copies from *X1*. *X9* and *X10* belonging to another class are used for comparison with them. *X1,X2,---,X10* are expressed by binary patterns sampled on a 512×512 mesh. First, we obtain the parameter Δr of tilting. For restriction of memory size and computation time, Δr is got by using 128×128 patterns whose components are arithmetic summation of 4×4 points, where the origin is the gravity center. Let the parameter of *X1* be $\Delta r0$, then $\Delta r0-\Delta r$ is shown in Table 4.4. Because of summing 4×4 points, the precision of them is bad, but the error angles are within the scope of $\pm 10°$. The rotation of an input pattern can be roughly estimated. Rotating the

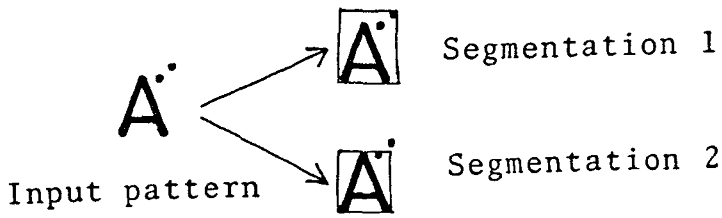


Fig. 4.3. Segmentations of an input pattern.

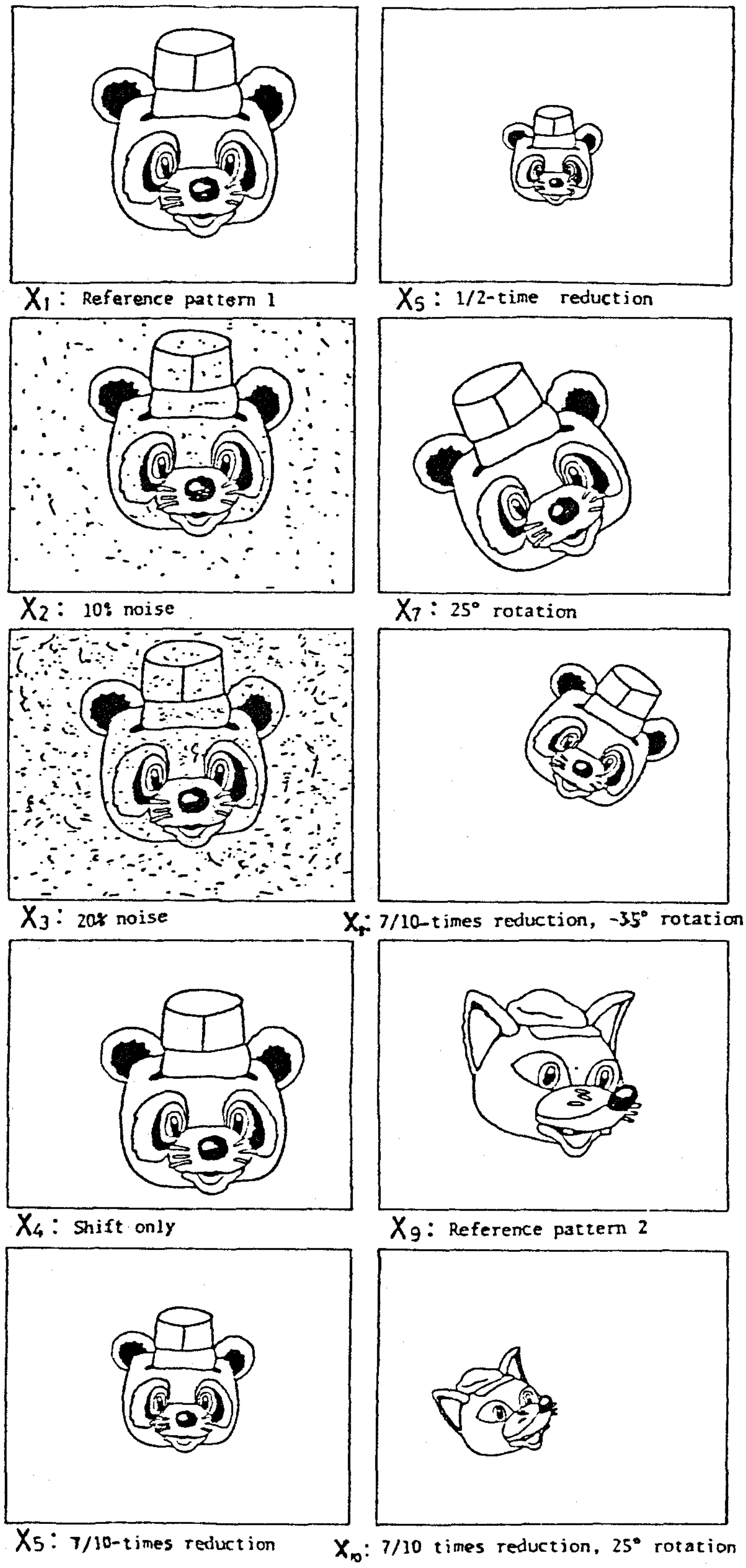


Fig. 4.4. Input patterns.

Table 4.4 Parameters of transformations and Euclid distance between transformation-invariant Fourier spectrums.

Input patterns	$\Delta r0-\Delta r$	real angle	Δex, Δey	real ratio	Δsx, Δsy	Distance from $X1$	Distance from $X9$	Classification
$X1$	$0°$	$0°$	1.00, 1.00	1.0	0.45, 0.35	0	0.230	$X1$
$X2$	$-6.4°$	$0°$	1.08, 1.10	1.0	0.45, 0.30	0.089	0.220	$X1$
$X3$	$3.5°$	$0°$	1.22, 1.22	1.0	0.41, 0.31	0.164	0.202	$X1$
$X4$	$-2.8°$	$0°$	1.00, 0.94	1.0	0.45, 0.39	0.026	0.226	$X1$
$X5$	$-9.8°$	$0°$	0.70, 0.71	0.7	0.48, 0.50	0.085	0.182	$X1$
$X6$	$-3.9°$	$0°$	0.50, 0.50	0.5	0.47, 0.51	0.164	0.211	$X1$
$X7$	$19.5°$	$25°$	0.90, 1.14	1.0	0.26, 0.48	0.048	0.214	$X1$
$X8$	$-38.0°$	$-30°$	0.72, 0.78	0.7	0.68, 0.34	0.085	0.189	$X1$
$X9$	$29.6°$	$25°$	0.72, 0.78	0.7	0.50, 0.48	0.230	0	$X9$
$X10$	52.3		0.54, 0.55		0.27, 0.59	0.366	0.164	$X9$

Fourier spectrum of a 2-dimensional pattern by $\Delta r0$-Δr, the rotation-invariant spectrum can be got. For simplicity, we normalize beforehand an input pattern by rotating. To save memory size and computation time, we compute only the Fourier spectrums of two 1-dimensional patterns which are obtained as follows:

$$Xk=[\{x(i,j)\}](i,j=1,2,---,512)$$
$$X=[\{x(i)\}](x(i)=\sum_{j=1}^{512} x(i,j),\ i=1,2,---,512)$$
$$Y=[\{y(i)\}](y(j)=\sum_{i=1}^{512} x(i,j),\ j=1,2,---,512). \tag{4.25}$$

Through the method mensioned in 3.2, Δsx, Δsy, Δex, and Δey are computed from the Fourier spectrums of X and Y. Where the origin (0,0) is the upper left corner and the subscripts x and y mean parameters of x-direction and y-direction, respectively. The Fourier transform is done after a (4096-512)-vector [0,0,---,0] is added to X and Y to be 4096-vectors. After that, the Fourier spectrums are obtained by trapezoidal approximation from the FFT, so Δex and Δey can be computed. The results are shown in Table 4.4. Δe's of $X2$ and $X3$ are 10%-20% larger than the real pattern size. Since we can consider that input patterns are expanded by adding noise, this tendency is acceptable to some extent. Δe's of the patterns which are transformed only through rotations and reductions are fairly accurate, that is, nearly equal to the ratio of the real pattern size. Some of patterns which are rotated are influenced by the bad precision of normarization. Δs's are nearly equal to pattern locations. After the enlargement and reduction-invariant spectrum is obtained by using Δe, the translation-invariant spectrum is obtained by using the phase angle of frequency 1. We cannot reverse the order, because translations do not change the size, but enlargements and reductions change the location. Euclidean distances between above obtained transformation-invariant spectrums are shwon in Table 4.4. In this table we see that ten patterns in Fig. 4.4 are classified correctly into two classes by the nearest neighborhood method.

4.7 Conclusions

Invariant spectrums to translations, enlargements, reductions, and rotations were developed through the Fourier spectrum. Changes of the Fourier spectrum caused by the transformations of an input pattern were investigated, and a transformation-invariant complete system was constructed. Transformations of an input pattern bring about changes of the parameters introduced here. Since these parameters are computed through the Fourier transform of a global operation, they are steady under local distortions and noise. As seen in computer simulation, the parameters of location, size, and inclination were extracted approximatly correctly. Using these parameters, it becomes possible to do normalization of input patterns which is stable under the above transformations and noise. When a step function is transformed through Fourier transform, the fast Fourier transform (FFT) becomes applicable by using the difference $f(x(i+1))-f(x(i))$ of values of neiboring sampling points. This accomplishes a considerable economy of computation time. In computer simulation all the 2-dimensional patterns were distinguished. Although we used only $Fg(zx, 0)$ and $Fg(0, zy)$ instead of all members $Fg(zx, zy)$'s of the transformation-invariant complete system, ten patterns under several transformations are all classified correctly.

CHAPTER 5

SUBSET METHODS

5.1 Introdunction

In the foregoing chapters global transformations are mainly treated. In this chapter local transformations, which are usually called distortions, are treated. Distortions may include global transformations. Subset methods, which can be regarded as one kind of modified template matching methods, are introduced to construct systems unaffected by certain distortions. Any kind of distortions should not be tolerated, but only a particular class of distortions, which we call *admissible distortions*, should be.

According to a template matching technique, an unknown pattern is recognized by deciding whether the unknown pattern matches one of templates for differently shaped specimens of each pattern class within an admissible distortion. The process of optimizing the match may take a long time. The pattern should be subjected in turn to each of a large number of distortions. It may save computation time and required memory to determine which points can correspond to which by a method that uses a set of small parts of the pattern and the template. Besides that any part of a pattern will be less affected by a distortion than the whole. For this reason, it is easier for machines to recognize features which are local properties of patterns than to recognize whole distorted patterns. The design of features usually takes a great deal of human efforts. It is desirable to automate the determination of features.

Therefore this chapter starts with discussion of subset methods and learning algorithm for automatic feature extraction. Feature extraction may be the optimum selection of common subpatterns to many patterns. The results of computer simulation in a simple example will be shown to illustrate how the algorithm works.

In general, the set of automatically determined features may include redundant features. The features generally may not be evaluated independently. After learning of features, some of them are selected to obtain a *min-max-cover* which covers as many bits of patterns in the trainig set as possible. The selection problem is represented by a *F*-table. After obtaining a min-max-cover, a classifier function is constructed in the form of a *product-of-sums* of features in the min-max-cover.

5.2 Subset Methods

In this section some definitions are given, and a subset method is introduced. For simplicity, let us consider a problem where an input pattern is expressed in a binary form of a n-vector $X=[x(0),x(1),---,x(n-1)]$ of 0's and 1's. It is desired to determine automatically a set of features by a subset method. Here admissible distortions are given *a priori*. In practice the number of admissible distortions is so large that we can not store them explicitly. One way to overcome this difficulty is work with subset methods. Every admissible distortion is expressed by a *partition* which was introduced by Ullmann [1]. A partition Bi on T is a set of non-overlapping subsets of T whose union is T, where T is the set of variables of two patterns. A partition Bi is said to be *applicable* to a pair of patterns, if the labelled bit locations with the same letter have the same value. For example, $B1$ is a partition on T which has 20 subsets as shown in Fig. 5.1(b). In the figure bit locations which are labelled with the same letter belong to the same subset of T, and ones which are not labelled with any letter have no restriction from $B1$. $B1$ is applicable to 2^{20} pairs of patterns. A pair of patterns $X1$ and $Y1$ of patterns is one of such pairs of patterns in Fig. 5.1(c). Another example is shown in Fig. 5.1(d), (e). A partition can be regarded as a many-to-many mapping of individual bits of patterns. If Bi is applicable to a pair of patterns X and Y, then we can think of Bi as an operator which changes X into Y and we write $Y=Bi \circ X$ and $X=Bi^{-1} \circ Y$, where Bi^{-1} is the inverse of Bi. For instance, $Y1=B1 \circ X1$, $X1=B1^{-1} \circ Y1$, and $Y2=B2 \circ X2$ in Fig. 5.1.

To work with a subset method, let us suppose that a partition can be constructed by a *possible combination* of *subpartitions* which are smaller sets than partitions. We say that a partition Bi is one of possible combinations of subpartitions Aj's$(j=1,2,---,k)$, if every labelled subset of $Aj(j=1,2,---,k)$ is included in one of labelled subsets of Bi and every labelled bit location of Bi is included in at least one Aj's. For instance, a partition $B2$ in Fig. 5.1(d) is constructed by a possible combination of subpartitions $A1$, $A2$, and $A3$ as shown in Fig. 5.2. It often occurs that there is not only one but also many possible combinations which construct the same partition. Note that a subpartition is nonpositional and a partition is positional. A partition takes the place

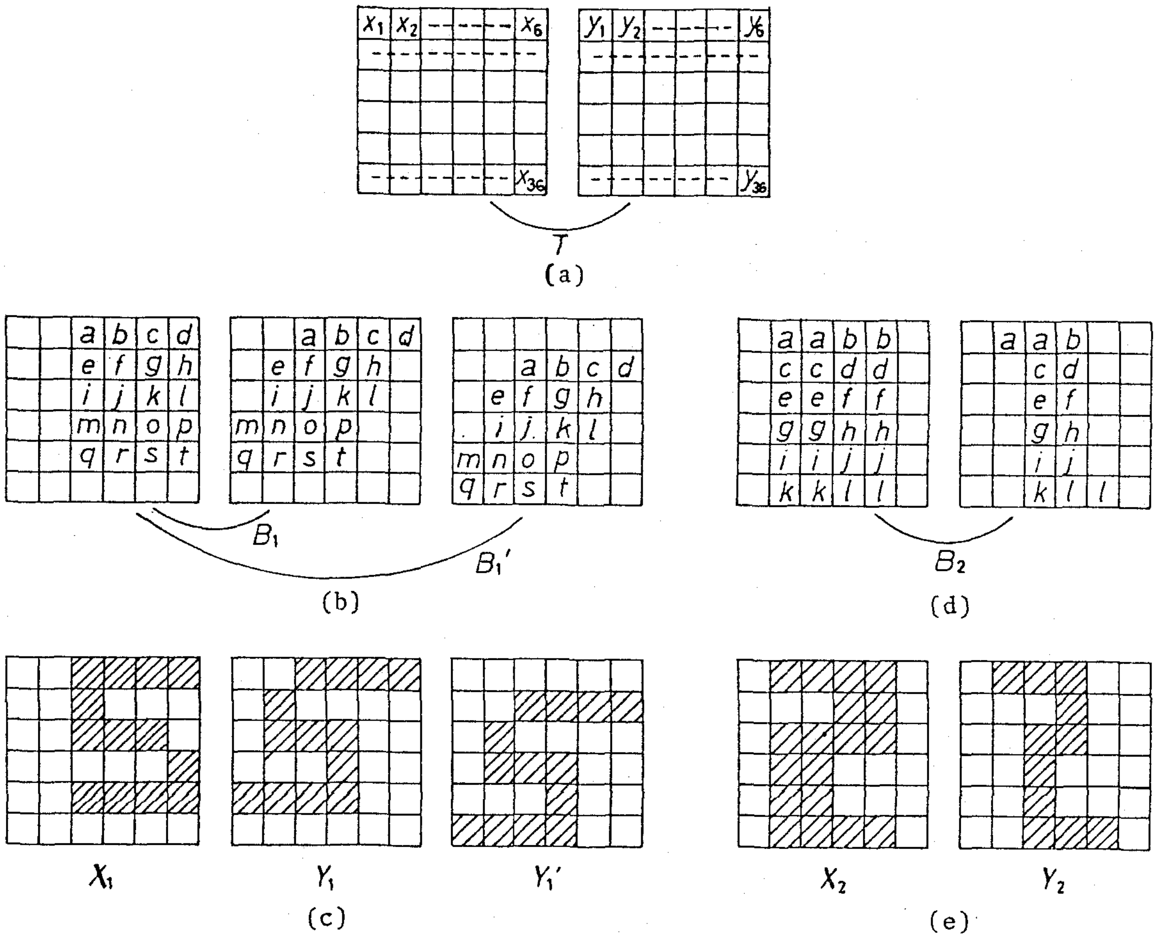
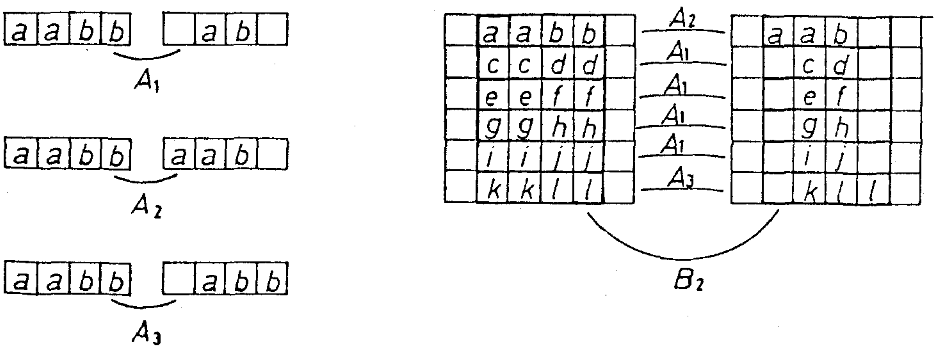


Fig. 5.1 Examples of partitions.



(a) Subpartitions. (b) A partition.

Fig. 5.2. A possible combination of subpartitions.

of an admissible distortion. Economy is gained by using possible combinations of subpartitions instead of partitions. Unreliability is caused by some kind of misrecognition as investigated by Ullmann [2], [58], in exchange for economy, as long as rather large subpartitions are not used. Possible combination of small subpartitions may not hold topological equivalence.

Let $\boldsymbol{A}$ be the set of admissible subpartitions which are *a priori* given and $\boldsymbol{B}$ be the set of partitions which are constructed by possible combinations of subpartitions in $\boldsymbol{A}$. It is natural to assume that $\boldsymbol{A}$ includes an identity operator $A0$ and for any subpartition Ai in $\boldsymbol{A}$ the inverse Ai^{-1} is also in $\boldsymbol{A}$. This is the same with $\boldsymbol{B}$. Let us naturally suppose that a shifted version Bi' of a partition Bi in $\boldsymbol{B}$ is also in $\boldsymbol{B}$. For example, $B1'$ is in $\boldsymbol{B}$, if $B1$ is in $\boldsymbol{B}$ in Fig. 5.1(b). It is not usually satisfied that there exists a partition $B3$ in $\boldsymbol{B}$ such that $X3=B3 \circ X1$, where $X2=B1 \circ X1$, $X3=B2 \circ X2$, and $B1$ and $B2$ are in $\boldsymbol{B}$. But it is always satisfied in the case where at least one of $B1$ and $B2$ is an identity operator or a translation. Accordingly $\boldsymbol{B}$ does not make a *group*. The problem deciding whether two patterns $X1$ and $X2$ are within an admissible distortion becomes equivalent to the problem deciding whether there is a partition Bi in $\boldsymbol{B}$ such that $X2=Bi \circ X1$. In other words, it is whether there is a possible combination constructing a partition which is the set of subsets of $X1$ and $X2$.

It is convenient to say that the set of 1's and 0's which are the same in two patterns is the *common subpattern* to the two patterns without any distortions. For instance, the common subpattern to [10110001] and [10011001] is [10-1-001], where "-" means an unspecified bit. The common subpattern to [10-1-001] and [10110011] is [10-1-0-1]. We say that for two patterns $X1$ and $X2$ and for two partitions $B1$ and $B2$ the set of 1's and 0's which are the same in two distorted patterns $B1 \circ X1$ and $B2 \circ X2$ is a common subpattern to those two patterns $X1$ and $X2$ within an admissible distortion. There are usually many common subpatterns to two patterns within admissible distortions. For instance, the common subpattern to [10110001] and [10011001] are [10-1-001], [-01100--], [--1-0---], and so on, where admissible distortions are only translations. Since the shifted versions of every partition in B are included in B, it is sufficient that the shifted versions of a common subpattern are represented by an appropriate subpattern. This subpattern may be a candidate for a feature on some conditions.

$\#[x|P]$ is defined already in Chapter 2. For example, $\#[x|2<x<5$, x is an integer$]=2$. To allow for noise we use a threshold $\theta(N)$. A subpattern F is *contained* in a pattern X, if there is partitions $B1$ and $B2$ in $\boldsymbol{B}$ such that $n(F) \leqq \theta(N)$. Where $n(F)=\#[\boldsymbol{x}|x$ is a bit in the subpattern F which is not "-", x corresponds to the bit in the distorted subpattern $B1 \circ F$ which is not the same as the corresponding bit in the distorted pattern $B2 \circ X]$. For example, [01100---] is contained in [10110001] and not contained in [10001101], if admissible distortions are only translations and $\theta(N)=0$. [01100---], however, is contained in [10001101], if admissible distortions are only transaltions and $\theta(N)=1$. Adding the subpartitions indicated in Fig. 5.3(a), [01100---] is contained in [1001100], even if $\theta(N)=0$. These aspects are shown in Fig. 5.3(b).

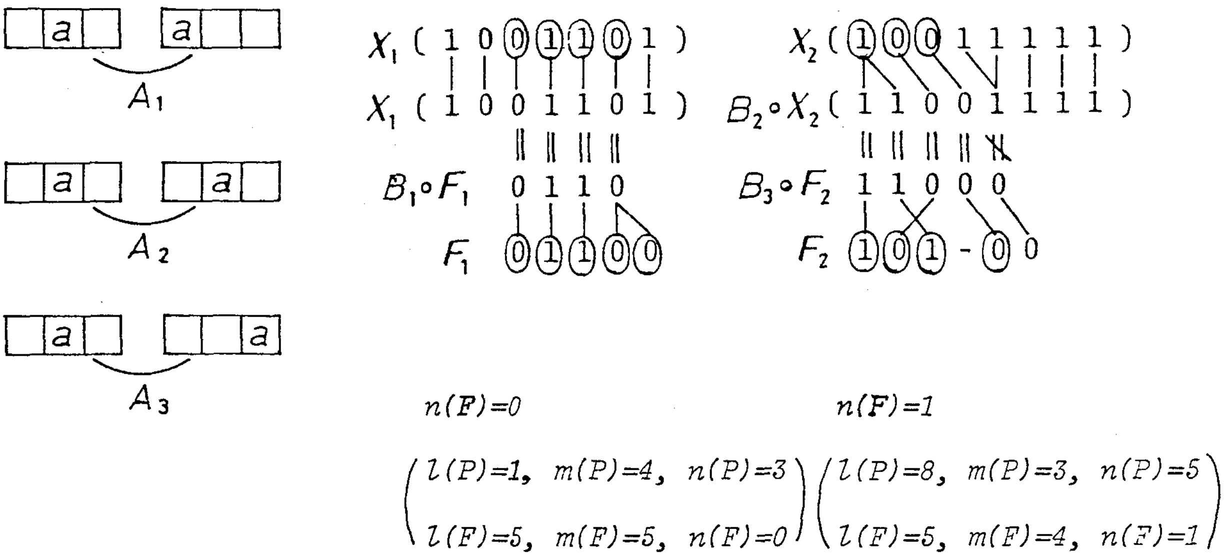


Fig. 5.3. Subpatterns contained in patterns.

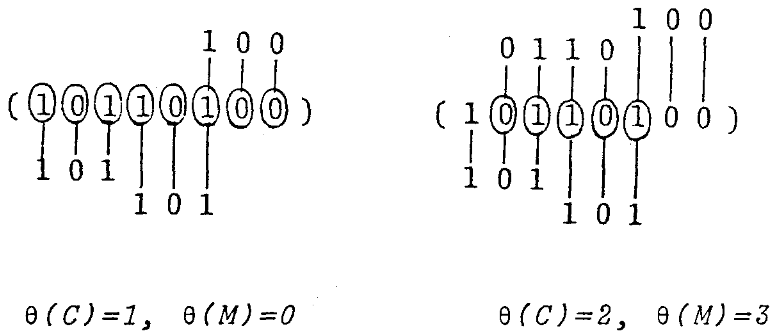


Fig. 5.4. Patterns covered by subpatterns.

A bit $x(i)$ in a pattern X is *covered* by a subpattern F, if it is satisfied that the corresponding bit in the distorted pattern $B1 \circ X$ to $x(i)$ is the same as the corresponding bit in the distorted subpattern $B2 \circ F$. Covering by a pattern is similarly defined. For instance, encircled bits in Fig. 5.3(b) are covered by a subpattern or a pattern. We say that a pattern is *covered* by the set of features, if almost every bit in the pattern is covered by more than or equal to $\theta(C)$ of features in the set and less than or equal to $\theta(M)$ of bits in the pattern are not. For example, [10110100] is covered by {[101-----], [100-----]}, if $\theta(C)=1$, $\theta(M)=0$. [10110100] is covered by {[101-----],[100-----],[0110----]}, if $\theta(C)=2$, $\theta(M)=3$. These aspects are shown in Fig. 5.4, where admissible distortions are translations and encircled bits indicate covered bits. Let $l(P)=\#[x|x$ is a bit in a pattern which is not "-"], $m(P)=\#[x|x$ is a covered bit in a pattern which is not "-"], $n(P)=\#[x|x$ is an uncovered bit in a pattern which is not "-"], then $l(P)=m(P)+n(P)$. $l(F)$, $m(F)$, and $n(F)$ are defined similarly to $l(P)$, $m(P)$, and $n(P)$, respectively, by exchanging a pattern for a subpattern, then $l(F)=m(F)+n(F)$. $l(P)$ and $l(F)$ depend only on a pattern and on a subpattern, respectively, but $m(P)$, $n(P)$, $m(F)$, and $n(F)$ depend on partitions, too. The examples of these notations are shown in Fig. 5.3(b).

5.3 Automatic Feature Extraction

Learning algorithm for automatic feature extraction is investigated in this section. Feature extraction may be to select some of common subpatterns to many patterns. It is not necessary that a feature is contained in all patterns in the class. Let us attach the following hypothetical conditions to features:

(F.1) the feature size $l(F)$ should be greater than or equal to $\theta(F)$, that is, $l(F) \geqq \theta(F)$;

(F.2) if a feature is contained in a pattern, then $m(P) \geqq \theta(P)$, $n(P) \leqq \theta(U)$;

(F.3) more than or equal to $\theta(A)$ of shifted versions of the feature should be contained in the pattern within the neiborhood ¶ of the position where the feature is located;

(F.4) a feature should be contained in many patterns,

where $\theta(F)$, $\theta(P)$, $\theta(U)$, and $\theta(A)$ are predetermined thresholds. If we want all bits in a pattern to be covered by a feature, then we set them as follows: $\theta(P)=l(P)=n$, $\theta(U)=0$.

A common subpattern F to two patterns $X1$ and $X2$ is a *candidate common subpattern* for a feature, if it satisfies the following conditions:

(C.1) the number $l(F)$ of the specified bits in F is greater than or equal to $\theta(F)$, that is, $l(F) \geqq \theta(F)$;

(C.2) the number $m(P1)$ of the covered bits in $X1$ and the number $m(P2)$ of the covered bits in $X2$ are greater then or equal to $\theta(P)$, and the number $n(P1)$ of the uncovered bits in $X1$ and the number $n(P2)$ of the uncovered bits in $X2$ are less than or equal to $\theta(U)$, that is, $m(P1)$, $m(P2) \geqq \theta(P)$, and $n(P1)$, $n(P2) \leqq \theta(U)$;

(C.3) there are more than $\theta(A)$ of common subpatterns to $X1$ and $X2$ which satisfy the above two conditions (C.1) and (C.2) within the neiborhood ¶ of the position where F is located;

(C.4) F is contained in at least two patterns.

These four conditions (C.1)-(C.4) correspond to the conditions (F.1)-(F.4), respectively. Generally, however, (C.3) and (C.4) are looser conditions than (F.3) and (F.4), respectively, and (C.4) is always satisfied apparently. (C.3) and (F.3) are the same, if only the shifted versions of the subpattern F are checked within the neiborhood ¶. It might be better to impose this restriction on a candidate common subpattern, but we do not do so for saving computation time.

There are usually some candidate common subpattern within the neiborhood. From the condition (C.3) they may be regarded as the distorted versions of the same features, and it is sufficient to select the optimum one among them. Here occurs a problem about the decision of the neiborhood ¶. Different features in ¶ may be gathered to make one feature, if it is too large. On the contrary, some distorted versions of the same feature may be redundantly selected as different features, if it is too small. So we should decide the range of neiborhood adequate to select one candidate common subpattern for one feature. We employ five criteria for selection one of the candidate common subpatterns within the neiborhood. We choose one such that

(P.1) the minimum of $m(P1)$ and $m(P2)$ is greatest;

(P.2) $(m(P1)+m(P2))$ is the greatest;

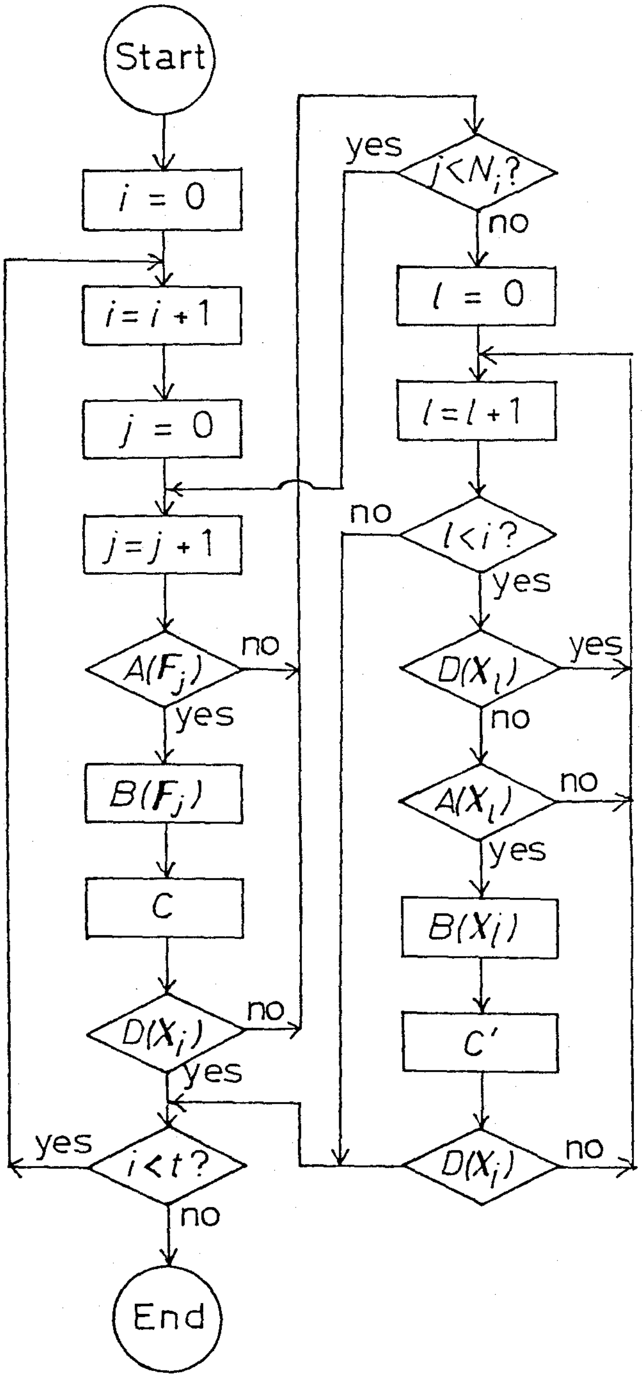
(P.3) the maximum of $n(P1)$ and $n(P2)$ is the least;

(P.4) the partition are constructed by the greatest number of subpartitions which are identity operators;

(P.5) the amount of the translations is smallest, where (P.1) has the first priority, and (P.1) has priority over (P.2), and (P.2) over (P.3), and so on. According to the priority ordering, the extracted features may differ to some extent, but not so large.

Learning algorithm for feature extraction is introduced by using the above defined notations and conditions. Let $\{X1, X2, ---, Xt\}$ be the training set and $\{F1, F2, ---, Fs\}$ be the set of features which have been extracted by the i-th step. Fig. 5.5 is a flow diagram for the learning algorithm. This algorithm differs from that of Sterns [1] in the point that it allows for distortions and noise within certain limits. To extract effectively features by this algorithm, the size and orientation of patterns must be normalized in advance, and thickness in character recognition must be too. The product "*" which is used in our learning algorithm is defined by Table 5.1. The table shows the value of the i-th bit of $X*Y$, where $X=[x(0), x(1), ---, x(n-1)]$, $Y=[y(0), y(1), ---, y(n-1)]$, and $1d$ and $0d$ mean that the bit has been distorted by at least one partition in B except for an identity operator and translations. In computation of common subpatterns, for simplicity, we do not distort further the bit with subscript d by any partition except for an identity operator and translations. We may distort by a partition in B the bits which have no subscript. Therefore a common subpattern to a subpattern F and a pattern $X1$ is also common to patterns $X2$ and $X3$, where F is a common subpattern to $X2$ and $X3$.

Let us take a very simple example in Fig. 5.6 to see how our learning algorithm works. In the example admissible distortions are only horizontal translations and the trainig set is $\{X1, X2, X3\}$, where the



Block A(Y): Is there any candidate common subapttern to $\boldsymbol{X}_i$ and Y? Check the conditions (C.1)-(C.4).

Block B(Y): Select one candidate common subpattern $\boldsymbol{F}$ to $\boldsymbol{X}_i$ and Y which is the best according to the priority (P.1)-(P.5).

Block C : If the candidate common subpattern $\boldsymbol{F}$ is contained in all the patterns which contain $\boldsymbol{F}_j$, then $\boldsymbol{F}_j$ is eliminated from the feature set and $\boldsymbol{F}$ is added into it. Otherwise, $\boldsymbol{F}$ is eliminated.

Block C' : Add $\boldsymbol{F}$ into the feature set

Block D($\boldsymbol{X}_i$): Is almost every bit in $\boldsymbol{X}_i$ covered by the feature set?

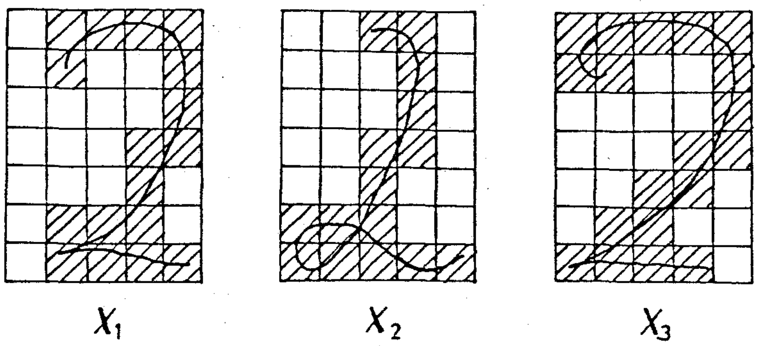
Fig. 5.5. Flow diagram for feature extraction.

Table 5.1. Definition of $x(i)*y(i)$.

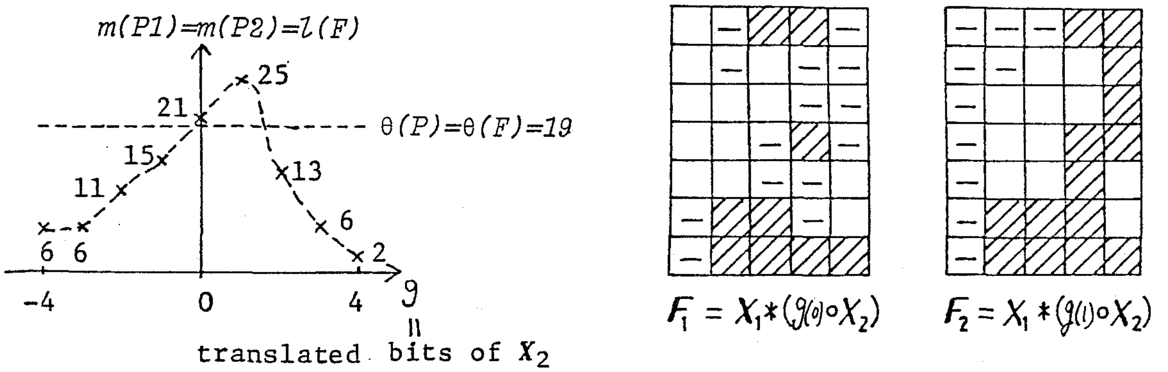
$y(i)$ \ $x(i)$	0	0_d	1	1_d	-
0	0	0_d	-	-	-
0_d	0_d	0_d	-	-	-
1	-	-	1	1_d	-
1_d	-	-	1_d	1_d	-
-	-	-	-	-	-

Table 5.2. Definition of $f(i) \circledast f'(i)$.

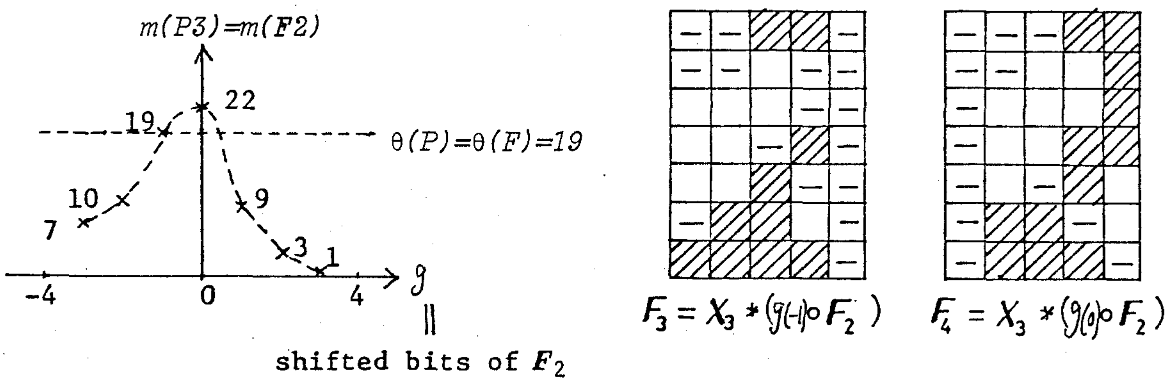
$f(i)$ \ $f'(i)$	0	0_d	1	1_d	-
0	0	0_d	×	×	0
0_d	0_d	0_d	×	×	0_d
1	×	×	1	1_d	1
1_d	×	×	1_d	1_d	1_d
-	0	0_d	1	1_d	-



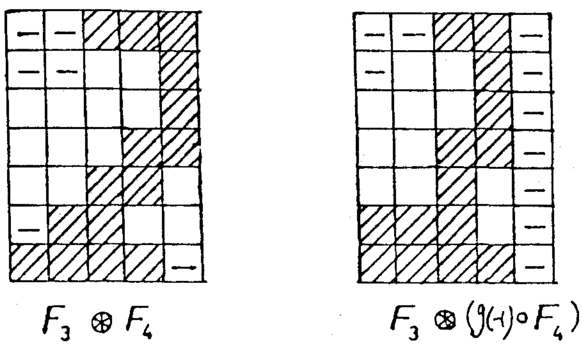
(a) Patterns in the training set.



(b) Candidate common subpatterns to X_1 and X_2.



(c) Candidate common subpatterns to X_3 and F_2.

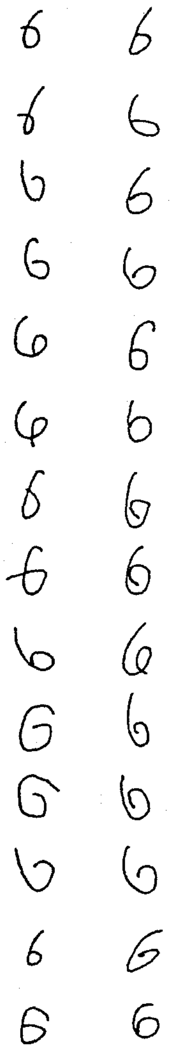


(d) Possible combinations of F_3 and F_4.

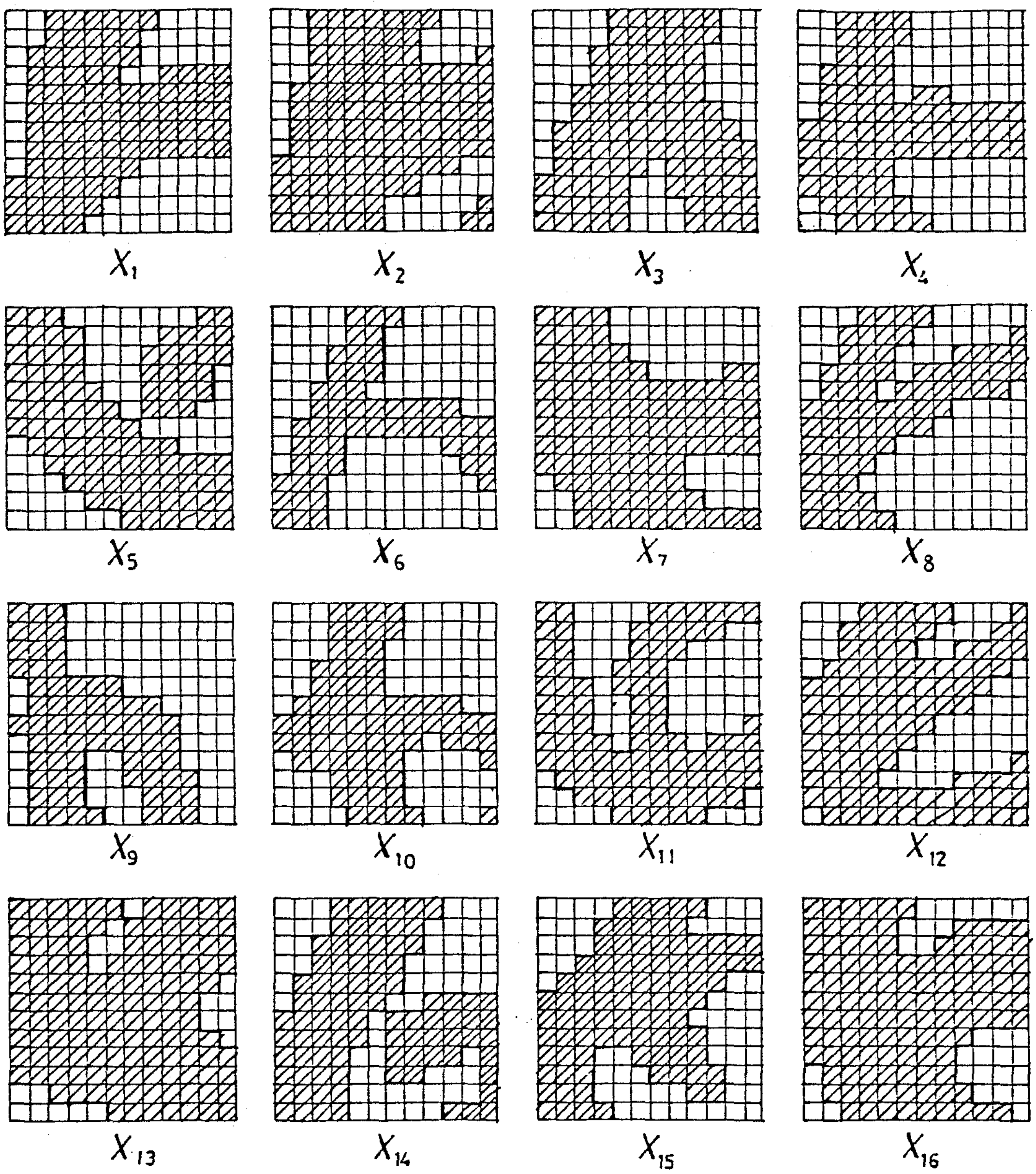
Fig. 5.6. Examples of feature extraction.

hatched bits are 1's and the white bite are 0's. The thresholds are as follows: $\theta(N)=0$, $\theta(C)=1$, $\theta(M)=16$, $\theta(P)=\theta(F)=19$, $\theta(U)=16$, $\theta(A)=1$ and ¶= {±0 **bit**}. Fig. 5.6(b) shows the values of $m(P1)=m(P2)$ and the candidate common subpatterns to $X1$ and $X2$, where $g(i)$ means tanslating operator by i bits. There is no candidate common subpatterns to $X3$ and $F1$. Fig. 5.6(c) shows the values of $m(P3)(=m(F2))$ and the candidate common subpattern to $X3$ and $F1$. Thus we obtain two features $F3$ and $F4$ which are common to and contained in $X1$, $X2$, $X3$. Then all patterns in the training set can be generated by possible combinations of $F3$ and $F4$. The generation is carried out by the product "$\circledast$" defined by Table 5.2. In the table "×" means prohibited combination, and the value is that of the i-th bit of $F \circledast F'$, where $F=[f(0),f(1),---,f(n-1)]$ and $F'=[f'(0),f'(1),---, f'(n-1)]$. The patterns which contain $F3$ and $F4$ are obtained by filling each bit location labelled by "-" with a black bit or a white bit in Fig. 5.6(d).

Next a more complicated example is considered. Fig. 5.7(a) shows copies of the numeral '6' written without constraint by different people [1]. Fig. 5.7 shows a few example of portion where the upper stroke joins the bottom loop in closed-loop 6's. $\{X1,X2,---,X16\}$ is used as the training set after normalization of thickness. Every subpartition is obtained by 90° rotations and combinations of basic subpartitions $A0$ and $A1$ as shown in Fig. 5.8(a), (b). In this example we direct our attention only to black bits and $l(P)$, $m(P)$, $n(P)$, $l(F)$, $m(F)$, and $n(F)$ are newly defined as the number of covered or uncovered black bits. The thresholds are set as follows: $\theta(N)=0$, $\theta(C)=1$, $\theta(M)=\theta(U)=30$, $\theta(P)=40$, $\theta(F)=25$,

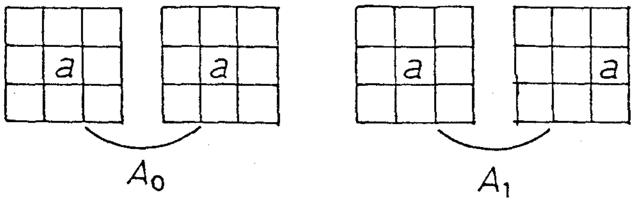


(a)

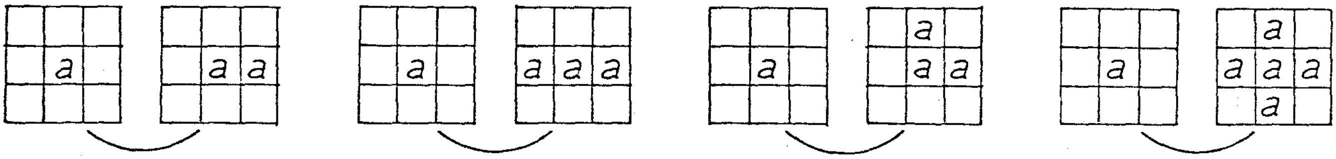


(b)

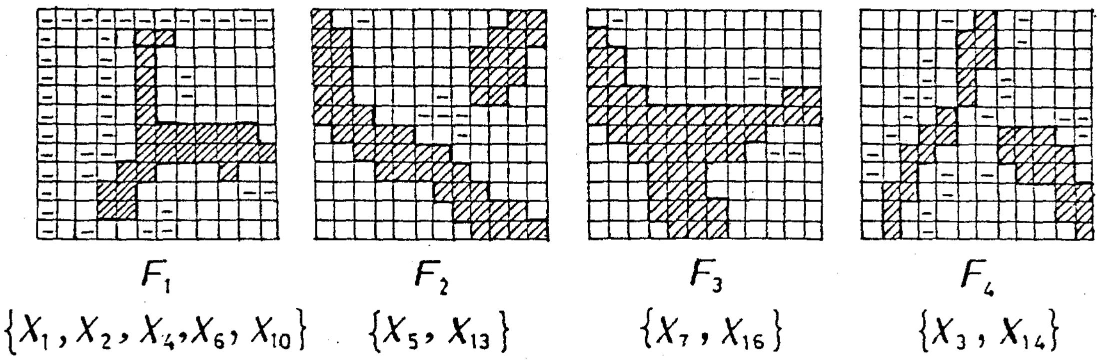
Fig. 5.7. Patterns in the training set.



(a) Basic subpartitions.



(b) Examples of subpartitions.



(c) Extracted features.

Fig. 5.8. Subpatterns and extracted features.

$\theta(A)=5$, and ¶={±3 bits}. The extracted features are shown in Fig. 5.8(c). The patterns which are covered by these features are indicated under the features.

The process of thickness normalization is explained in more detail. First, patterns in Fig. 5.7(b) are thinned to be patterns in Fig. 5.9(b). The thinning process as follows:

(Step 1) If there is a portion which matches one of windows in Fig. 5.9(a) and the 90° rotated versions of them, then the central 1 is changed to 2.

(Step 2) After (Step 1) is applied to as many portions as possible, all 2's are changed to 0. Repeat (Step 1) and (Step 2) until these can not be applicable to any portion.

Thus we obtain the thinned patterns which are shown in Fig. 5.9(b). Next, those patterns are thickened to be patterns in Fig. 5.10(b). The thickening process is as follows:

(Step 1) If there is a portion which matches the window in Fig. 5.10(a), then the central 0 is changed to 2.

(Step 2) After (Step 1) is applied to as many portions as possible, all 2's are changed to 1. (Step 2) is applied only once.

The thickened patterns of patterns in Fig. 5.9(b) may be considered as the thickness normalized patterns of patterns in Fig. 5.7(b). The set of those patterns is used as the training set. According to the feature extraction algorithm, we obtain the features *F1*, *F2*, *F3*, and *F4*. *F1* is contained in *X1*, *X2*, *X4*, *X6*, and *X10*, *F2* in *X5* and *X13*, *F3* in *X7* and *X16*, and *F4* in *X3* and *X14*. Covered and uncovered black bits of patterns

1	1	1
	1	
	0	

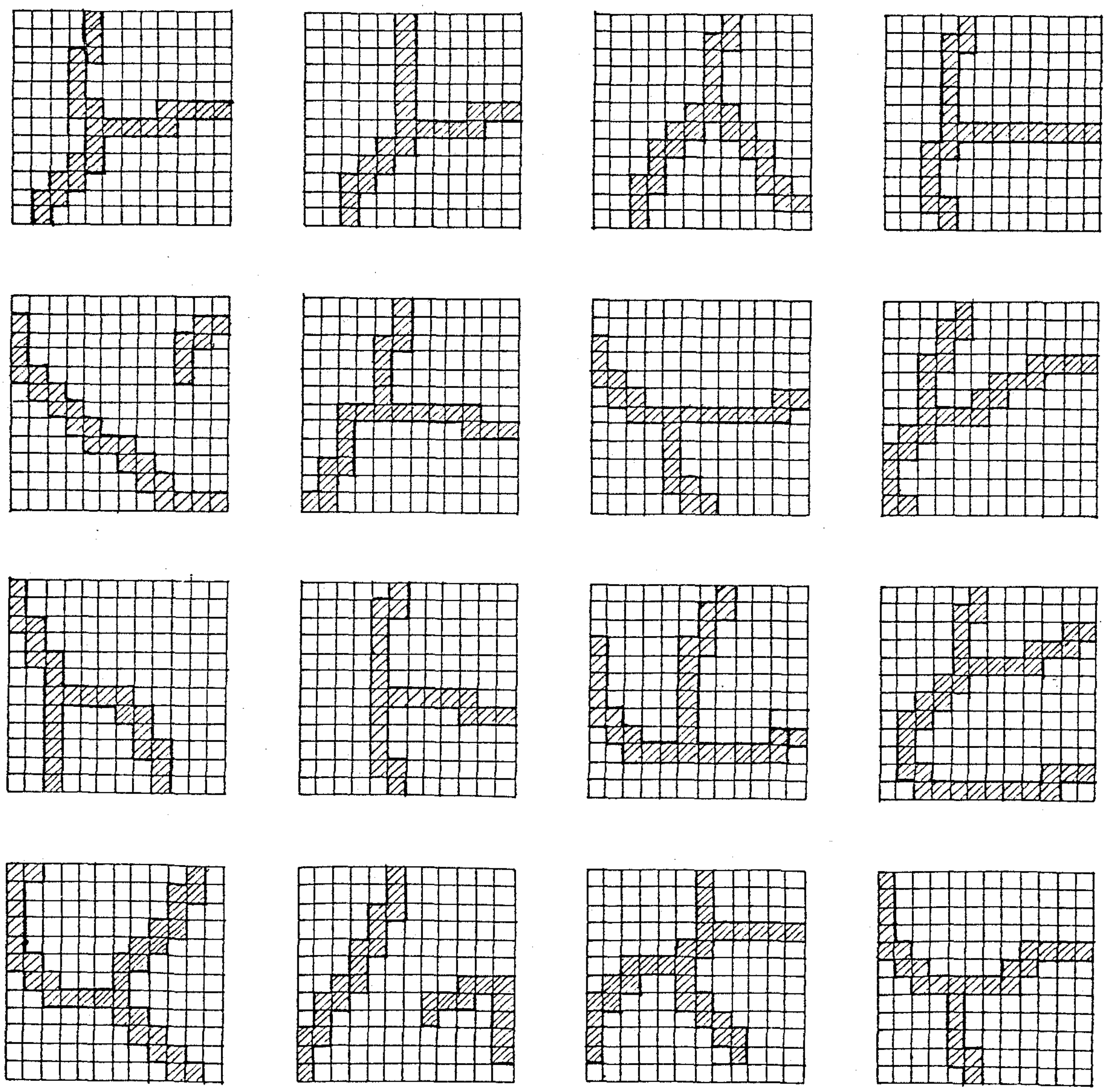
2	1	2
q	1	q
	0	

2	1	1
q	1	
	0	

1	1	2
	1	q
	0	

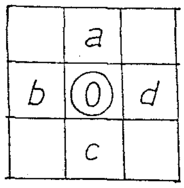
	p	1
q	1	p
	q	

(a) Windows for thinning. p = 1 or 2, q = 0 or 2.



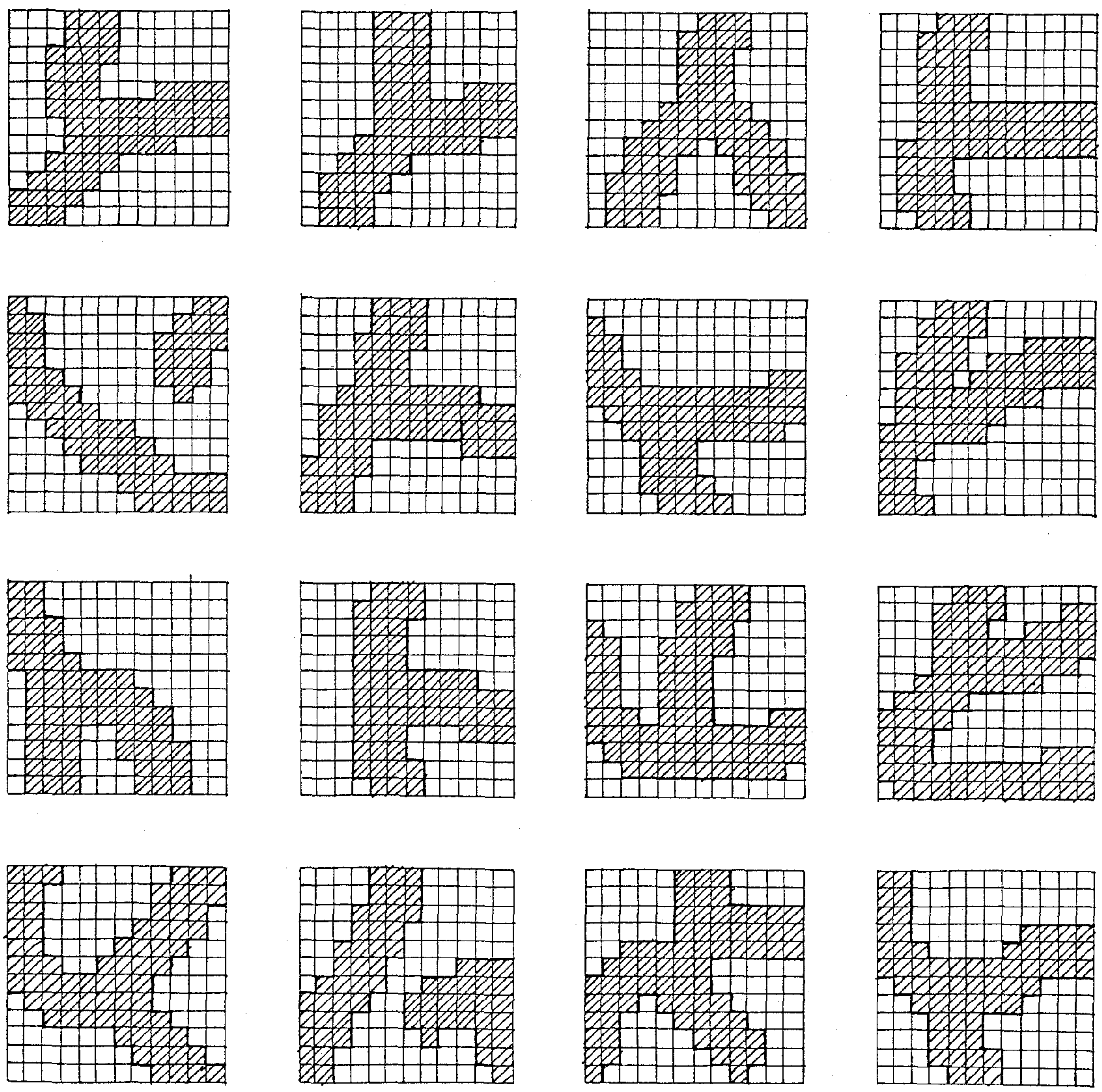
(b) Thinned patterns.

Fig. 5.9. Thinning of patterns.



(a) A window for thickening.

a or *b* or *c* or *d* is equal to 1.



(b) Thickened patterns.

Fig. 5.10. Normalization of thickness of patterns.

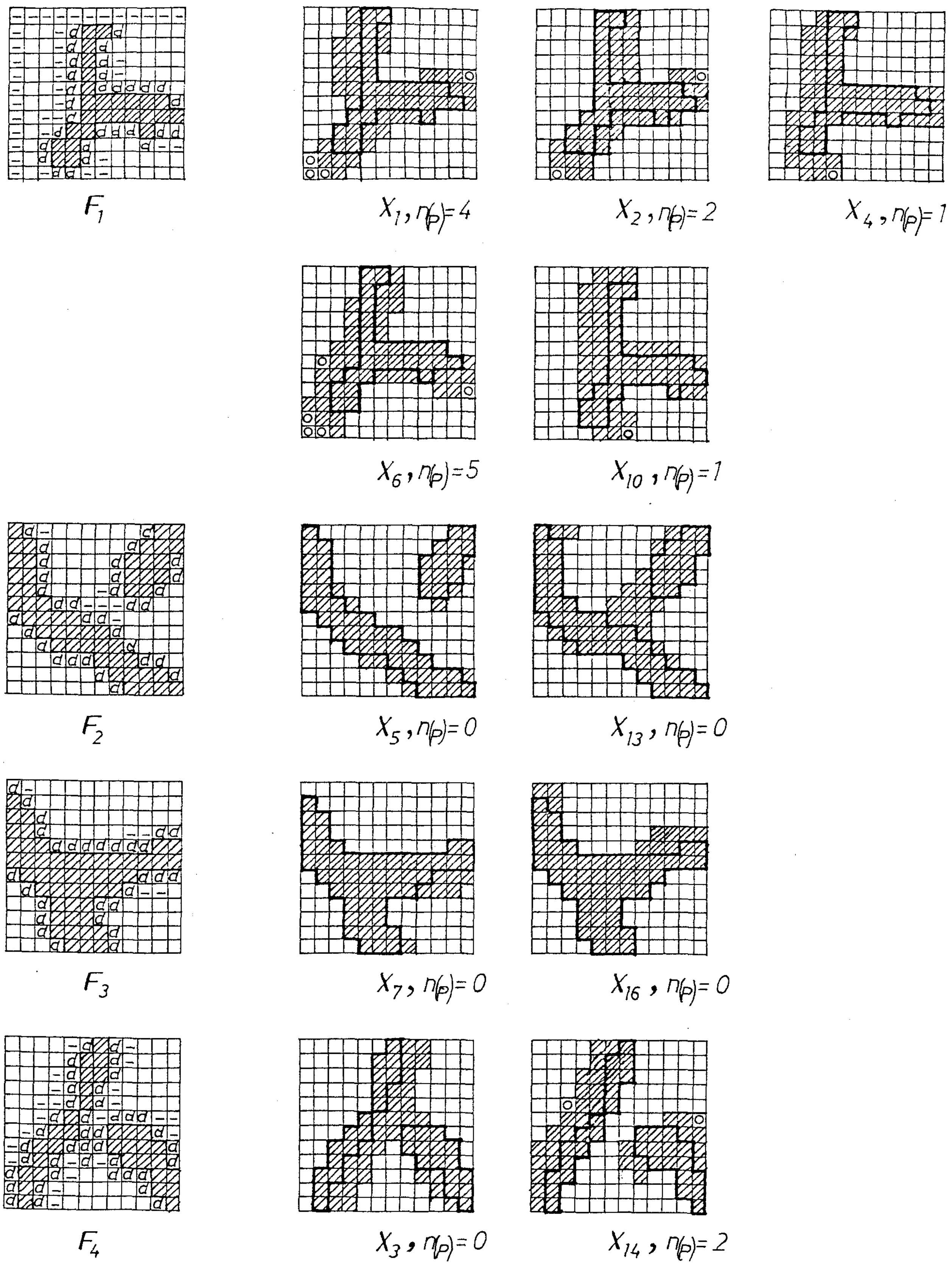


Fig. 5.11. Extracted features and covered bits of patterns.

Features
- □ : white bits(0)
- ▨ : black bits(1)
- ⊟ : unspecified bits
- d : distorted black bits(1_d)

Patterns
- □ : white bits(0)
- ▨ : covered black bits
- ○ : uncovered black bits

by those features are shown in Fig. 5.11.

It depends on the values of the thresholds as well as the range of the neighborhood what features are extracted and which patterns are covered by the features. It is one problem how to decide appropriately them in advance. One solution is to iterate the learning process as we change them increasingly or decreasingly toward the looser condition for feature extraction. In spite of the desirable results this method takes a long time. After determination of features the next problem rises, that is, how to construct a classifier function by using them.

5.4 Selection of Features and Construction of Classifier Functions

In this section we consider a method to select some features out of the extracted features and construct a classifier function. The selection and the construction have a close relation to each other and they should not be done separately. The same training set as that employed in extracting features is used in the process for them. As seen in the paper [62], it is not always right that the best subset of features must contain the i-th best features $(i=1,2,---,k)$, where k is the number of features in the subset. The selection problem is approached by means of a table called a F-table, which is an extension of a prime implicant table used in minimizing a Boolean function. A technique for F-table reduction is very similar to that for selection of a minimum cost subset of prime implicants of a Boolean function. A *F-table* has a row Fi for every selected feature Fi, and a column mj for every bit of patterns in the training set: 1(0) is placed at the intersection $a(i,j)$ of Fi and *mj*, if *Fi* can cover (can not cover) *mj*. *Fi* and *mj* in a *F*-table work as a prime implicant and as a minterm in a classical prime implicant table, respectively. A column *mj* is said to be *covered* by a subset of features, if it has 1's in more than $\theta(C)$ of *Fi*'s$(i=1,2,---k)$. A *max-cover* is the subset of features *Fi*'s which covers as many columns as possible. A *min-max-cover* is the max-cover the number of features in which is minimum. Finding a min-max-cover must be a kind of quasi-optimum feature selection.

Reduction of a *F*-table can be obtaioned by a technique, which is based on generalized rules of row dominance, column dominance, and row essentiality [65], [67]. These rules also allow, in general, large simplification in reducing a *F*-table without any modification if $\theta(C)=1$, and with some modification if $\theta(C)\geq 2$. For example, the same reducing technique as that in [67] can be applicable to a *F*-table with the modification such that the constant terms $s(i)$ in a *P*-table are set to be $\theta(C)$ instead of 1.

(Rule 1) Eliminate from the *F*-table each column which has no 1 in every row. A row *Fi* is an *essential* row, if there is at least one column *mj* for which

$$\sum_i a(i,j)\leqq\theta(C). \qquad (5.1)$$

Every essential row must be a member of any min-max-cover. For simplicity we mainly deal with the case when $\theta(C)=1$.

(Rule 2) Delete every essential row *Fi* from the table and take it as a member of a min-max-cover. Eliminate from the table each column *mj* such that *Fi* covers *mj*, that is, $a(i,j)=1$. A row *Fi* *dominates* a row *Fh*, if in

every column the entry of Fi is greater than or equal to the one of Fh, that is, $a(i,j)\geqq a(h,j)$.

(Rule 3) Eliminate from the table each row Fh which is dominated by at least one row Fi. A set of rows $\{Fi\}(i=i1,i2,---,ik)$ *dominates* a set of rows $\{Fh\}(h=h1,h2,---,hl)$, if in every column it is satisfied that

$$\sum_{i=i1}^{ik} a(i,j)\geqq \sum_{h=h1}^{hl} a(h,j) \quad (k\leqq l). \tag{5.2}$$

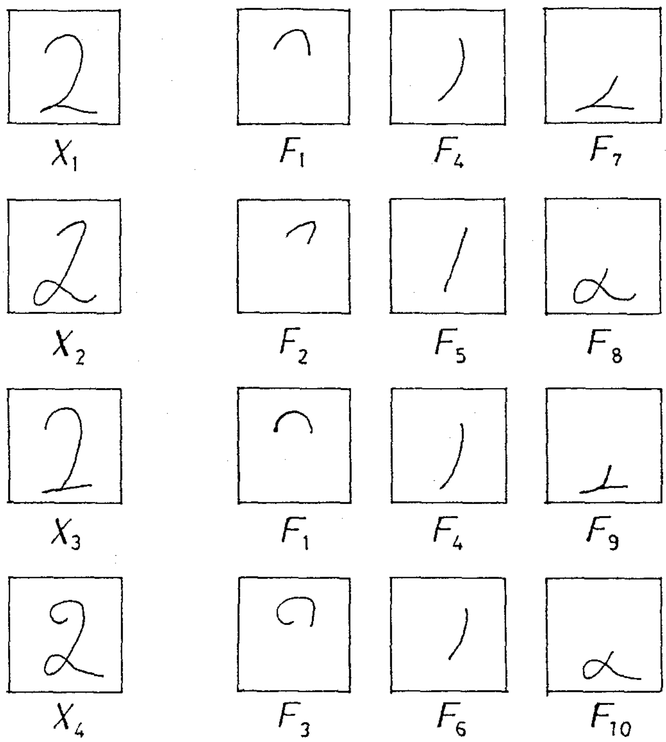
(Rule 4) Eliminate from the table each column which is in Fh's$(h=h1,h2,---,hl)$ and not in Fi's$(i=i1,i2,---,ik)$, that is, $\{Fh\}-\{Fi\}$. A column mj *dominates* a column mh, if in every row the entry of mj is greater than or equal to the one of mh, that is, $a(i,j)\geqq a(i,h)$.

(Rule 5) Eliminate from the table each column mj which dominates at least one column.

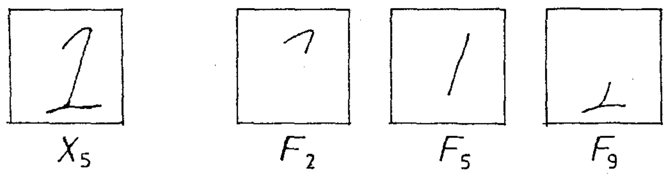
After these rules have been applied as many times as possible, a reduced F-table is obtained. Several methods [63]-[73], which have been used to make a minimum cover for prime implicant tables can be employed to determine a min-max-cover for any F-table. We express a classifier function in the form of a product-of-sums whose geometrical meaning and weakness are expressed in the similar example to that shown by Ullmann [1], [2]. Let us suppose that four 2's in a training set are divided into features as shown in Fig. 5.12. Every member in the class "2" should contain one of $F1$, $F2$, and $F3$, one of $F4$, $F5$, and $F6$, and one of $F7$, $F8$, $F9$, and $F10$. Then a classifier functions $\phi 1$ for the class "2" might be expressed by a product-of-sums as follows:

$$\phi 1=(F1\vee F2\vee F3)\wedge(F4\vee F5\vee F6)\wedge(F7\vee F8\vee F9\vee F10) \tag{5.3}$$

where



(a) Four 2's divided into three features.



(b) One '1' divided into three features that are in (a)

Fig. 5.12. Explanatory examples of a product-of-sums.

$$Fi(X)=\begin{cases}1, & \text{if a pattern } X \text{ contains a feature } Fi\\ 0, & \text{otherwise}\end{cases} \quad (5.4)$$

for $i=1,2,---,10$. As shown in [1], [2], [58], a fundamental weakness of product-of-sums expression arises from the fact that a pattern which satisfies the product-of-sums is not necessarily a member of the class. The weakness is illustrated in Fig. 5.12(b). The character '1' in Fig. 5 12(b) satisfies (5. 3) and it is misrecognized as "2". The use of the absence of features and the use of N-tuples of features may lessen this weakness. For instance, $\phi 2$ expressed as follows does not make such a misrecognition:

$$\phi 2=\phi 1\wedge(\overline{F2\wedge F9}) \quad (5.5)$$

where "—" is the negation. $\phi 3$ using 2-tuples of features does not make such a misrecognition, too:

$$\phi 3=((F1\wedge F4)\vee(F2\wedge F5)\vee(F3\wedge F6))\vee((F4\wedge F7)\vee(F5\wedge F8)\vee(F4\wedge F9)\vee(f6\wedge F10)). \quad (5.6)$$

The closest approximation to the training set is obtained by $\phi 4$ defined as

$$\phi 4=(F1\wedge F4\wedge F7)\vee(F2\wedge F5\wedge F8)\vee(F1\wedge F4\wedge F9)\vee(F3\wedge F6\wedge F10). \quad (5.7)$$

$\phi 4$ also avoids such a misrecognition. We could use the logical sum of all features like

$$\phi 5=F1\vee F2\vee---\vee F10. \quad (5.8)$$

However, the approximation by $\phi 5$ seems too rough. Let $O(\phi i)$ be called the *on-set* of a function ϕi and defined by $O(\phi i)=\{X|\phi i(X)=1\}$, then the aspect of the relations among $O(\phi 1),O(\phi 2),---,O(\phi 5)$ is shown in Fig. 5. 13. $O(\phi i)$'s$(i=1, 2,---,5)$ contain every pattern in the training set. On

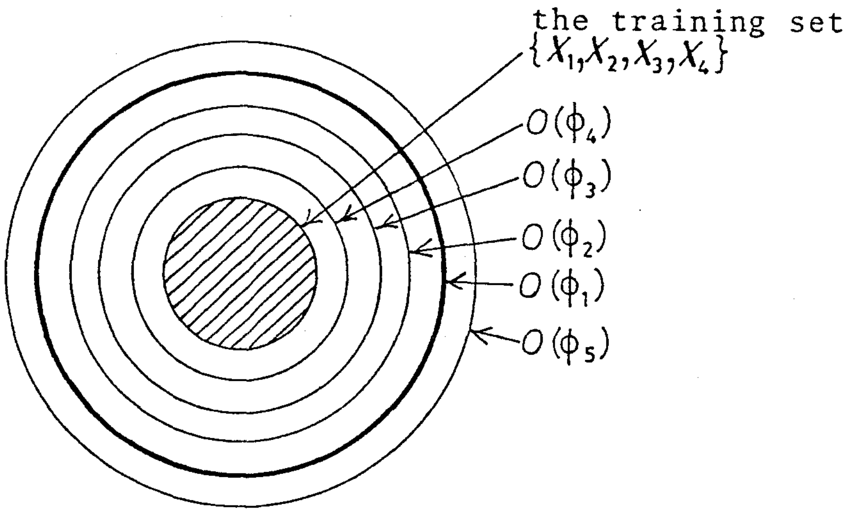


Fig. 5.13. Relation among the training set and $O(\phi_i)$'s.

the contrary to the weakness, product-of-sums expression has a good point of flexibility and economy. The reliability of a classifier function can be increased by using overlapping covers. Therefore we adopt product-of-sums expression.

Let us introduce the product "$\otimes$" which works as a design tool for sums in the product-of-sums. The product $W1 \otimes W2$ of two families $W1$ and $W2$ is defined as

$$W1 \otimes W2 = \{\{ti, sj\}\} \ (i=1,2,--,l,\ j=1,2,---,m) \tag{5.9}$$

where $W1$ and $W2$ are two families of sets of features and are expressed as $W1=\{t1,t2,---,tl\}$ and $W2=\{s1,s2,---,sm\}$. Let W' be the maximum subset of W such that every member of W' does not include any other member of W, and "$W1 \Rightarrow W2$" means that "$W2=W2'$". For example,

$$\begin{aligned}&\{\{F1\},\{F2\},\{F3,F4\}\} \otimes \{\{F1\},\{F3\}\}\\ &=\{\{F1\},\{F1,F2\},\{F1,F3\},\{F2,F3\},\{F3,F4\},\{F1,F3,F4\}\}\\ &\Rightarrow \{\{F1\},\{F2,F3\},\{F3,F4\}\}.\end{aligned} \tag{5.10}$$

Now we can introduce the procedure to determine a classifier function. The sum of features in every member of W' obtained as follows is a candidate for a sum in a product-of-sums:

$$W=W(X1) \otimes W(X2) \otimes --- \otimes W(Xt) \Rightarrow W' \tag{5.11}$$

where $\{X1,X2,---,Xt\}$ is the training set and $W(Xi)(i=1,2,---,t)$ is the set of features in a min-max-cover which are contained in Xi. The sum of features in every member of W' is always one for $Xi(i=1,2,---,t)$. Then we can obtain a classifier function by a product-of-sums such that as many features in the min-max-cover as possible appear in at least one sum. The on-set of the product-of-sums constructed in this way always includes the training set $\{X1,X2,---,Xt\}$.

Let us take an example and see how a product-of-sums is constructed by the procedure introduced here. Table 5.3 shows a F-table. $\{X1,X2,---, X5\}$ is the training set and Fi's$(i=1,2,---,14)$ are extracted features which are contained in N patterns, where admissible distortions are only translations and $\theta(N)=0$, $\theta(C)=1$, $\theta(M)=\theta(U)=4$, $\theta(P)=\theta(F)=4$, $\theta(A)=1$ and $\P=\{\pm 0$ bit$\}$. After *Rule 5* is applied to the F-table in Table 5.3(a), there is no reduction rule applicable to the reduced table. We select $F1$ which is contained in all patterns in the training set, and eliminate from the table the row $F1$ and all the column mj's covered by $F1$. Then the reduction rules are applied again and we obtain the further reduced table which is shown in Table 5.3(b). Repeating these process we obtain two min-max-covers $\{F1,F3,F4,F13\}$ and $\{F1,F3,F5,F13\}$. Every bit of patterns in the training set is covered by these min-max-covers. In Table 5.3(c) 1's are put to show that the corresponding pattern contains the corresponding feature in a min-max-cover. Let us construct a classifier function by using the former min-max-cover:

$$W(X1)=W(X2)=W(X3)=\{\{F1\},\{F3\},\{F4\}\}$$

$$W(X4)=W(X5)=\{\{F1\},\{F13\}\}$$

$$W=W(X1)\otimes W(X2)\otimes---\otimes W(X5)\Rightarrow\{\{F1\},\{F3,F13\},\{F4,F13\}\}. \tag{5.12}$$

Then we can obtain a classifier function as follows:

$$\phi=F1\wedge(F3\vee F13)\wedge(F4\vee F13). \tag{5.13}$$

Patterns in the on-set $O(\phi)$ are generated by possible combinations of features which satisfy (5.13). For example, we have

Table 5.3. A F-table and min-max-covers.

	Patterns		X_1								X_2								X_3								X_4								X_5							
			(1 0 1 1 0 0 0 1)								(1 0 0 1 1 0 0 1)								(1 0 0 0 1 1 0 1)								(1 0 1 0 1 0 0 1)								(1 0 0 1 0 1 0 1)							
N	Features	F_i \ m_j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
5	10------01	1	1	1					1	1	1	1					1	1	1	1					1	1	1	1					1	1	1	1					1	1
4	10-01	2				1	1		1	1									1	1		1	1				1	1		1	1							1	1		1	1
3	0110	3		1	1	1	1						1	1	1	1						1	1	1	1																	
	100-1	4				1	1	1		1	1	1	1		1				1	1	1		1																			
	1-001	5				1		1	1	1				1		1	1	1	1		1	1	1																			
	1001	6									1	1	1	1	1	1	1	1													1	1	1	1	1	1	1	1				
2	10001	7				1	1	1	1	1									1	1	1	1	1																			
	101-0	8	1	1	1		1																						1	1	1		1									
	0-101	9																				1		1	1	1										1		1	1	1		
	1-0-01	10			1		1		1	1																									1		1		1	1		
	10-0-1	11																	1	1		1		1					1	1		1		1								
	10101	12																									1	1	1	1	1							1	1	1	1	1
	01010	13																										1	1	1	1	1					1	1	1	1	1	
	10--01	14																											1	1			1	1	1	1			1	1		
dominated column				1		6	3			6		9			11	12		15	18		6	6	6		24			25		27			1			24			38		25	25

(a) A F-table.

F_i \ m_j	3	6	11	12	22	27	29	30	35	36	38
3	1		1	1	1						
4		1	1								
5		1		1							
6			1	1			1	1	1	1	
8	1					1	1				
9					1					1	1
10	1		1						1		1
11					1	1		1			
13						1	1	1	1	1	1

(b) A reduced F-table after selecting F_1.

	X_1	X_2	X_3	X_4	X_5
F_1	1	1	1	1	1
F_3	1	1	1		
F_4	1	1	1		
F_{13}				1	1

	X_1	X_2	X_3	X_4	X_5
F_1	1	1	1	1	1
F_3	1	1	1		
F_5	1	1	1		
F_{13}				1	1

(c) Features in min-max-covers and patterns in training set.

```
F1:[10----01] → [10110-01] → [10110001]=X1
     ||||          ||| |
  F3:0110       F4:100-1

F1:[10----01] → [10011001] → [10011001]=X2
      ||||        ||| |
   F3:0110     F4:100-1

F1:[10----01] → [10-01101] → [10001101]=X3
       ||||       ||| |
    F3:0110    F4:100-1

F1:[10----01] → [10101001]=X4
     |||||
 F13:01010

F1:[10----01] → [10010101]=X5
      |||||
  F13:01010                                    (5.14)
```

In this case the trainig set is equal to the on-set $O(\phi)$. Generally the former set is included in the latter set. For simplicity we set $\theta(C)=1$, but the higher reliability can be got by the use of the greater $\theta(C)$. It is not always necessary to find a min-max-cover first. After obtaining the sums of features we could select the sums so that as many bits of patterns in the training set as possible can be covered by the minimum number of the sums.

5.5 Conclusions

We have introduced subset methods for distorted pattern recognition, a technique for automatic determination of features, and for automatic construction of a classifier function. Patterns in the on-set of the classifier function can be generated by using possible combinations of features. The reliability of a classifier function may be rendered higher by the use of an overlapping cover, the use of the absence of features, and the use of N-tuples of features. Furthermore we can use positional features and accordingly some more information, if we store the locations of features in learning and allow features to move within the neighborhood of the locations in covering. Generally the higher reliability can be achieved at the cost of greater storage requirements. To attain a big reduction in storage requirements, the idea of *random superimposed coding* (known as *zatocoding*) has been adapted by Ullmann [76], [77]. This is an excellent technique worth reseaching.

As shown in the computer simulation in 5.3, the feature extraction algorithm may be applicable to unsupervised learning of a classifier function. In other words, patterns are classified into the same class, if they include the same candidate common subpatterns. The technique using subset methods is particularly efficient in handwritten character recognition. So binary patterns have been mainly dealt with, but patterns with gray level must be dealt with in speech recognition and in medical diagnosis such as vectorcardiograms, X-ray pictures, blood cells, blood-pressure wave recognition, and so on. It is possible to adapt the technique to patterns with gray level after some modification.

CHAPTER 6

CONCLUSIONS

Some transformation-invariant functions are presented in this thesis. It is said that every interesting geometrical property is invariant to any element of some transformation group. Since transformation-invariant functions can be considered one kind of features of an input pattern, the methods developed above are applicable to feature extraction. This paper is at the stage of preprocessing or normarization of patterns in pattern recognition. It is the next problem how to realize pattern recognition systems by using information obtained here.

Although immense progress has been made in regard to computer's ability of pattern recognition, there is still a wide gap between human and machine. It appears almost impossible to bridge over the gap without basic research. In this paper the problem of composing transformation-invariant functions for recognizing noisy distorted patterns is discussed. Partial functions are limited to Boolean functions, Walsh-Hadamard power spectrums, and Fourier spectrums on a discrete input space. The author, however, hopes the discussions will enhance the further research of transformation-invariant functions. Evidently people use more information, such as topology and context, rather than spectrums and so on, in recognizing an object. Under what transformations do they regard that the objective pattern class is invariant? This is not well-known yet. The transformations could help to design efficient pattern recognition systems. This is left for future work.

Von Neumann type of general purpose computers on the market seems inherently weak in picture processing from a viewpoint of computation time. Researches on special purpose computers for picture processing, such as the massively parallel processor being designed by NASA, are producing fruit. The development of software and hardware will facilitate the great progress of advanced pattern recognition systems. The author will be happy if the present paper will make some contributions to the study of pattern recognition.

LIST OF PUBLICATIONS AND TECHNICAL REPORTS

[1] Hama, H. and Yamashita, K.:"On Universal Families of Partial Functions for Perceptrons", *Papers of Technical Groups of IECE*, A71-43, Jul. 1971.

[2] Hama, H. and Yamashita, K.:"Perceptrons-The Families of Group-Invariant Partial Functions", *Papers of Technical Groups of IECE*, A71-81, 1971.

[3] Hama, H. and Yamashita, K.:"On Universal Families of Partial Functions for Perceptrons", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.12, p.69-78, Dec. 1971.

[4] Hama, H. and Yamashita, K.:"The Method to Compute the Coefficients of the Positive Normal Form and Other Problems", *Papers of Technical Groups of IECE*, A71-102, IT71-87, Jan. 1972.

[5] Hama, H. and Yamashita, K.:"Perceptrons-The Order of the Function of Two Functions-", *Papers of Technical Groups of IECE*, AL72-52, PRL72-50, Jul. 1972.

[6] Hama, H. and Yamashita, K.:"Families of Partial Functions for Perceptrons", *IECE Trans.*, vol.57-A, 7, p.481-488, Jul. 1974.

[7] Hama, H. and Yamashita, K.:"A Method for Composition of Two-Valued Universal Families", *IECE Trans.*, vol.58-D, 4, p.220-221, Apr. 1975.

[8] Yamashita, K. and Hama, H.:"Walsh Functions in Feature Extraction for Geometric Patterns (Part I)", *Papers of Technical Groups of IECE*, PRL75-17, p.73-84, Jun. 1975.

[9] Hama, H. and Yamashita, K.:"Walsh Functions in Feature Extraction for Geometric Patterns (Part II)", *Papers of Technical Groups of*

IECE, PRL75-18, p.85-96, Jun. 1975.

[10] Yamashita, K. and Hama, H.:"Group-Invariance Theorem for Extracting Geometrical Features of Patterns", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.16, p.83-96, Dec. 1975.

[11] Hama, H. and Yamashita, K.:"The Transformations-Invariant Modified Walsh Power Spectra for Extracting Geometrical Features of Patterns", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.17, p.87-99, Dec. 1976.

[12] Takahashi, T., Hama, H., and Yamashita, K.:"Linear Associative Machines", *Papers of Technical Groups of IEE*, IP-77-51, Oct. 1977.

[13] Yamashita, K and Hama, H.:"On the Theory in Feature Extraction for Geometrical Patterns (Part I)", *Progress in Cybernetics and Systems Research*, vol.IV, p.462-472, 1977.

[14] Hama, H. and Yamashita, K.:"On the Theory in Feature Extraction for Geometrical Patterns (Part II)", *Progress in Cybernetics and Systems Research*, vol.IV, p.473-483, 1977.

[15] Hama, H. and Yamashita, K.:"Walsh-Hadamard Power Spectra Invariant to Certain Transform Groups", *IEEE Trans.*, SMC-9, 4, p.227-237, Apr. 1979.

[16] Hama, H. and Yamashita, K.:"On a Subset Method for Representation of Distorted Pattern Classes", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.22, p.69-87, Dec. 1981.

[17] Hama, H. and Yamashita, K.:"Transformation Invariant Fourier Spectra and Normarization of Patterns", *IECE Trans.*, vol.65-D, 8, p.981-988, Aug. 1982.

REFERENCES

[1] Ullmann, J.R.:*Pattern Recognition Techniques*, London:Butterworths and New York:Crane Russak, 1973.

[2] Ullmann, J.R.:"Subset Methods for Recognizing Distorted Patterns", *IEEE Trans.*, SMC-7, 3, p.180-191, Mar. 1977.

[3] Bledsoe, W.W. and Browning, I.:"Pattern Recognition and Reading by Machine", *Proc. Eastern Joint Computer Conf.*, p.225-232, 1959.

[4] Riseman, E.M. and Hanson, A.R.:"A Contextual Postprocessing System for Error Correction Using Binary *n*-grams", *IEEE Trans.*, C-23, 5, p.480-493, May 1974.

[5] Vanderburg, G.J. and Rosenfeld, A.:"Two-Stage Template Matching", *IEEE Trans.*, C-26, 4, p.384-393, Apr. 1977.

[6] Egawa, K., Hama, H., and Yamashita, K.:"Recognition of Numarals by Subfeatures", *Records of Kansai Chapter Joint Convention of IECE*, G8-8, 1978.

[7] Hama, H. and Yamashita, K.:"On a Subset Method for Representation of Distorted Patttern Classes", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.22, p.69-87, Dec. 1981.

[8] Widrow, B.:"The 'Rubber-Mask' Technique---II. Pattern Storage and Recognition", *Pattern Recognition*, vol.5, 3, p.199-211, Sep. 1973.

[9] Ozawa, K. and Tanaka, K.:"Character Recognition Using Correlation Technique on Spacial Complex-Frequency Plane", *Second International Joint Conf. on Pattern Recognition*, Copenhagen, Aug. 1974.

[10] Minsky, M. and Papert, S. :"Linearly Unrecognizable pattern",

Proc. of Symposia in Applied Mathematics, Mathematical Aspects of Computer Science, XIX, p.176, 1967.

[11] Minsky, M. and Papert, S.:*Perceptrons-An Introduction to Computational Geometry*, MIT Press, 1969.

[12] Uesaka, Y.:"Analogue Perceptrons-On Additive Representation of Functions", *IECE Trans.*, vol.54-C, 7, p.586-593, July 1971.

[13] Rosenblatt, F.:*Principles of Neurodynamics*, Spartan, 1961.

[14] Hama, H., Shimura, M., and Tanaka, K.:"Existense Theorem of Global Functions for Perceptrons", *Records of National Convention of IECE*, 120, 1971.

[15] Hama, H. and Yamashita, K.:"On Universal Families of Partial Functions for Perceptrons", *Papers of Technical Groups of IECE*, A71-43, Jul. 1971.

[16] Hama, H. and Yamashita, K.:"Perceptrons(2)-A Family of Multi-Stage Coupled Partial Functions with Order 2-", *Records of Kansai Chapter Joint Convention of IECE*,G8-16, 1971.

[17] Hama, H. and Yamashita, K.:"On Universal Families of Partial Functions for Perceptrons", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.12, p.69-78, Dec. 1971.

[18] Hama, H. and Yamashita, K.:"Perceptrons-The Families of Group-Invariant Partial Functions", *Papers of Technical Groups of IECE*, A71-81, 1971.

[19] Hama, H. and Yamashita, K.:"The Method to Compute the Coefficients of the Positive Normal Form and Other Problems", *Papers of Technical Groups of IECE*, A71-102, IT71-87, Jan. 1972.

[20] Hama, H. and Yamashita, K.:"Structure of Boolean Functions with 3 Variables", *Records of Kansai Chapter Joint Convention of IECE*, G9-11, 1972.

[21] Hama, H. and Yamashita, K.:"Perceptrons-The Order of the Function of Two Functions-", *Papers of Technical Groups of IECE*, AL72-52, PRL72-50, Jul. 1972.

[22] Hama, H. and Yamashita, K.:"Perceptrons(4)- More-Than-One-in-a-Box Problem -", *Records of National Convention of IECE*, 1321, 1973.

[23] Hama, H. and Yamashita, K.:"Perceptrons(5)-Humming Distance Between Φ-Images(*R*)", *Records of Kansai Chapter Joint Convention of IECE*, G8-23, 1973.

[24] Hama, H. and Yamashita, K.:"Families of Partial Functions for Perceptrons", *IECE Trans.*, vol.57-A, 7, p.481-488, Jul. 1974.

[25] Hama, H. and Yamashita, K.:"A Method for Composition of Two-Valued Universal Families", *IECE Trans.*, vol.58-D, 4, p.220-221, Apr. 1975.

[26] Walsh, J.L.:"A Closed Set of Normal Orthogonal Functions", *Amer. J. Math.*, vol.55, p5-24, 1923.

[27] Paley, R.E.A.C.:"On Orthogonal Matrices", *J. Math. Phys.*, vol.12, p.311-320, 1933.

[28] Fine, N.J.:"On the Walsh Functions", *Trans. Amer. Math. Soc.*, vol.65, p.372-414, 1949.

[29] Harmuth, H.F.:*Transmission of Information by Orthogonal Functions*, New York:Springer-Verlag, 1972.

[30] Kimura, M.:"Walsh Functions and Their Applications", *JIEEJ Proc.*, vol.91, 4-6, Apr.-Jun. 1971.

[31] Ahmed, N. and Rao, K.R.:*Orthogonal Transforms for Digital Signal Processing*, New York:Springer-Verlag, 1975.

[32] Andrews, H.C. and Hunt, B.R.:*Digital Image Restoration*, Engleywood Cliffs, NJ:Prentice-Hall, 1977.

[33] Alexandridis, N.A. and Klinger, A.:"Walsh Orthogonal Functions in Geometrical Feature Extraction", *Proc. 1971 Sym. on Applications of Walsh Functions*, Naval Research Lab., Washington, DC, p.18-25, Apr. 1971.

[34] Ahmed, N., Rao, K.R. and Abdussatar,A.L.:"BIFORE or Hadamard Transform", *IEEE Trans.*, AU-19, 3, p.225-234, Sep. 1971.

[35] Hama, H. and Yamashita, K.:"Walsh-Hadamard Power Spectra Invariant to Certain Transform Groups", *IEEE Trans.*, SMC-9, 4, p.227-237, Apr. 1979.

[36] Yamashita, K. and Hama, H.:"Walsh Functions in Feature Extraction for Geometric Patterns (Part I)", *Papers of Technical Groups of IECE*, PRL75-17, p.73-84, Jun. 1975.

[37] Hama, H. and Yamashita, K.:"Walsh Functions in Feature Extraction for Geometric Patterns (Part II)", *Papers of Technical Groups of IECE*, PRL75-18, p.85-96, Jun. 1975.

[38] Yamashita, K. and Hama, H.:"Walsh Functions in Geometrical Feature Extraction of Patterns (Part I)- Group-Invariance Theorem -", *Submitted to IEEE International Symposium on Information Theory*, Jun. 1976.

[39] Hama, H. and Yamashita, K.:"Walsh Functions in Geometrical Feature Extraction of Patterns (Part II)- Group-Invariant Features -", *Sub-*

mitted to IEEE International Symposium on Information Theory, Jun. 1976.

[40] Yamashita, K. and Hama, H.:"On the Theory in Feature Extraction for Geometrical Patterns (Part I)", *Progress in Cybernetics and Systems Research*, vol.IV, p.462-472, 1977.

[41] Hama, H. and Yamashita, K.:"On the Theory in Feature Extraction for Geometrical Patterns (Part II)", *Progress in Cybernetics and Systems Research*, vol.IV, p.473-483, 1977.

[42] Hama, H. and Yamashita, K.:"Feature Extraction Invariant to Translation, Enlargements and Reductions", *Records of National Convention of IECE*, 1207, 1975.

[43] Tsujimura, H., Hama, H., and Yamashita, K.:"On Geometrical Feature Extraction(1)", *Records of Kansai Chapter Joint Convention of IECE*, G8-17, 1975.

[44] Hama, H., Tsujimura, H., and Yamashita, K.:"On Geometrical Feature Extraction(2)", *Records of Kansai Chapter Joint Convention of IECE*, G8-18, 1975.

[45] Yamashita, K. and Hama, H.:"Group-Invariance Theorem for Extracting Geometrical Features of Patterns", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.16, p.83-96, Dec. 1975.

[46] Tsujimura, H., Hama, H., and Yamashita, K.:"Transformation-Group Invariant Functions Through Hadamard Transform", *Records of Kansai Chapter Joint Convention of IECE*, G7-13, 1976.

[47] Hama, H. and Yamashita, K.:"The Transformations-Invariant Modified Walsh Power Spectra for Extracting Geometrical Features of Pat-

terns", *Memoirs of the Faculty of Engineering of Osaka City University*, vol.17, p.87-99, Dec. 1976.

[48] Reitboeck, H. and Brody, T.P.:"A Transform with Invariance under Cyclic Permutation for Applications in Pattern Recognition", *Inf. & Control*, vol.15, p.130-154, 1969.

[49] Kunt, M.:"On Computation of the Hadamard Transform and the *R* Transform in Ordered Form", *IEEE Trans.*, C-24, 11, p.1120-1121, Nov. 1975.

[50] Wagh, W.D. and Kanethar, S.V.:"A Class of Translation Invariant Transforms", *IEEE Trans.*, Assp-25, 2, p.203-205, Apr. 1977.

[51] Arazi, B.:"Hadamard Transforms of Some Specially Shuffled Signals", *IEEE Trans.*, Assp-23, 6, p.580-583, Dec. 1975.

[52] Hama, H. and Yamashiata, K.:"Transformation Invariant Fourier Spectra and Normarization of Patterns", *IECE Trans.*, vol.65-D, 8, p.981-988, Aug. 1982.

[53] Ikehara, T.:*Lecture of Applied Mathematics*, Gakujyutsu Tosho, 1966.

[54] Huang, T.S., Burnett, J.W. and Deczky, A.G.:"The Importance of Phase in Image Processing Filters", *IEEE Trans*, Assp-23, 6, p529-542, 1975.

[55] Oppenheim, A.V., Lim, J.S., Kopec, G.E., and Pohlig, S.C.:"Phase in Speech and Pictures", *Proc. IEEE Int. Conf. Assp.*, p.632-637, Apr. 1979.

[56] Oppenheim, A.V. and Lim, J.S.:"The Importance of Phase in Signals", *Proc. IEEE*, vol.69, 5, p.529-541, May 1981.

[57] Yau, S.S. and Yang, C.C.:"Pattern Recognition by Using an Associative Memory", *IEEE Trans.*, EC-15, 12, p.944-947, Dec. 1966.

[58] Ullmann, J.R.:"Correspondence in Character Recognition", in *Machine Perception of Patterns and Pictures*, Conf. Ser., 13, Institute of Physics, London, p.34-44, 1972.

[59] Ullmann, J.R.:"Transference of Learning between Recognition Classes", *IEEE Trans.*, C-21, 2, p.219-220, Feb. 1972.

[60] Hanakata, K.:"Feature Generation and Ordering for Pattern Description and Matching"

[61] Elashoff, J.D., Elashof, R.M., and Goldman, G.E.:"On the Choice of Variables in Classification Problems with Dichotomous Variables", *Biometrika*, vol.54, p.668-670, 1967.

[62] Toussaint, G.T.:"Note on Optimul Selection of Independent Binary-Valued Features for Pattern Recognition", *IEEE Trans.*, IT-17, 9, p.618, Sep. 1971.

[63] Mc Cluskey, E.J.:"Minimization of Boolean Functions", *Bell Syst Tech. J.*, vol.35, p.1417-1444, 1956.

[64] Quine, W.V.:"A way to simplify Truth Functions", *Amer. Math. Mon.* vol.62, p.627-631, 1956.

[65] Pyne, I.B. and Mc Clusky, E.J.:"An Essay on Prime Implicant Table", *J. Soc. Ind. Appl. Math.*, vol.9, p.604-631, Dec. 1961.

[66] Gimpel, J.F.:"A Reduction Technique for Prime Implicant Table", *1964 Proc. Fifth Annual Symp. on Switching Theory and Logical Design*, p.183-191, 1964.

[67] Luccio, F.:"A Method for the Selection of Prime Implicants", *IEEE Trans.*, EC-15, 2, p.205-212, Apr. 1966.

[68] Tison, P.:"Generation of Consensus Theory and Application to the

Minimization of Boolean Functions", *IEEE Trans.*, EC-16, 8, p.446-456, Aug. 1967.

[69] Das, S.R.:"An Approach for Simplifying Switching Functions by Utilizing the Cover Table Representation", *IEEE Trans.*, C-20, 3 p.355-359, Mar. 1971.

[70] Sureshchander:"Minimization of Switching Functions - A Fast Technique", *IEEE Trans.*, C-24, 7, p.753-756, Jul. 1975.

[71] Reusch, B.:"Generation of Prime Implicant Subfunctions and a Unifying Approach to the Covering Problem", *IEEE Trans.*, C-24, 9, p.924-930, Sep. 1975.

[72] Rhyne, V.T., Nol, P.S., Mckinney, M.H., and Pooch, V.W.:"A New Technique for the Fast Minimization of Switching Functions", *IEEE Trans.*, C-26, 8, p.757-762, Aug. 1977.

[73] Curtis, H.A.:"An Adjacency Table Methods of Deriving Minimal Sums", *IEEE Trans.*, C-26, 11, p.1136-1141, Nov. 1977.

[74] Nilson, N.J.:*Learning Machines*, McGraw Hill, 1965.

[75] Ahmed, N., Shreiber, H.H., and Lopresti, P.V.:"On Notation and Definition of Terms Related to a Class of Complete Orthogonal Functions", *IEEE Trans.*, EMC-15, 2, p.75-80, May 1973.

[76] Ullmann, J.R.:"Reduction of Storage Requirements of Bledsoe and Browning's *n*-tuple Method of Pattern Recognition", *Pattern Recognition*, Pergamon Press, vol.3, p.297-306, 1971.

[77] Ullmann, J.R.:"Pattern Recognition Using Degenerate Reference Data", *Proc. Joint Workshop on Pattern Recognition and Artificial Intelligence*, Cape Cod. Massachusetts, Jun. 1976.