Group-theoretical model of feature extraction

Reiner Lenz

Linköping University, S-58183 Linköping, Sweden

Received August 11, 1988; accepted January 30, 1989

A group-theoretically motivated investigation of feature extraction is described. A feature extraction unit is defined as a complex-valued function on a signal space. It is assumed that the signal space possesses a group-theoretically defined regularity that we introduce. First the concept of a symmetrical signal space is derived. Feature mappings then are introduced on signal spaces and some properties of feature mappings on symmetrical signal spaces are investigated. Next the investigation is restricted to linear features, and an overview of all possible linear features is given. Also it is shown how a set of linear features can be used to construct a nonlinear feature that has the same value for all patterns in a class of similar patterns. These results are used to construct filter functions that can be used to detect patterns in two- and three-dimensional images independent of the orientation of the pattern in the image. Finally it is sketched briefly how the theory developed here can be applied to solve other, symmetrical problems in image processing.

1. INTRODUCTION

Feature extraction is one of the most important tasks in pattern recognition. In this paper a general model of a feature-extraction unit is investigated. Basically we shall assume that such a unit transforms an input signal \( s \) into a response signal \( f(s) \). We call \( f(s) \) the feature extracted from the input signal \( s \). For the purpose of this paper we can consider the input signal \( s \) as a part of an image detected by a camera or the retina. The feature-extraction unit is then a cell or a pattern-recognition algorithm that analyzes the incoming gray-value pattern. The computed result \( f(s) \) is a description of the incoming pattern \( s \). In the case in which we consider the feature-extraction unit to be a cell, we can say that the incoming signal \( s \) describes the output of the cells in the receptive field of the feature-extraction cell. The output \( f(s) \) is the reaction of the feature-extraction cell to the incoming signal. The purpose of this paper, however, is not a description of natural cells but the development of a mathematical model that is based on mathematically motivated criteria.

The feature-extraction unit obviously is specified completely if we know how the response \( f \) depends on the input \( s \). Designing such a unit is equivalent to constructing the mapping \( f \). Many approaches for designing such units use an ad hoc strategy depending on the problem at hand. Others study natural systems, such as the human vision system, to find out how well working systems have solved the problem. Many of these approaches have in common that they treat the set of input signals as just an unstructured set. In reality, however, we find that the space of input signals is often highly structured in the sense that the space can be partitioned into different subspaces, in which each subspace consists of essentially the same signals. In this paper we assume that the signal space can be partitioned into subspaces in such a way that all the signals in a subspace are connected to one another with the help of a group-theoretically defined transformation. We shall also say that the subspaces are invariant under the group operation or that the signal space is symmetrical.

As an example, consider a feature-extraction unit analyzing the gray-value distribution in the neighborhood of a point in an image. Assume further that the goal of this unit is to find out whether the incoming gray-value pattern is edgelike. We shall refer to this problem as the two-dimensional (2-D) edge-detection problem. In what follows, we shall denote by \( s_0, s_h \), and \( s_v \) the gray-value distributions of a vertical edge, a horizontal edge, and a noise pattern, respectively. In the case in which we assume that the signal space is unstructured, we would treat \( s_0, s_h \), and \( s_v \) as just three input patterns. We would completely ignore the additional information that \( s_0, s_h, \) and \( s_v \) have in common than \( s_0, s_h, \) and \( s_v \). In contrast to this unstructured approach, we shall in what follows describe a method that is based on the observation that an effective feature-extraction unit should use the additional information. In this theory we shall divide the space of input signals into subspaces of "essentially the same signals." (This expression is defined in Section 2.) We call such a subspace an invariant subspace. In the edge-detection example one such subspace would consist of all rotated versions of a fixed gray-value distribution. Now \( s_0, s_h \), and \( s_v \) lie in the same subspace, whereas \( s_v \) lies in another subspace. We can thus say that the similarity between \( s_0, s_h, \) and \( s_v \) is greater than the similarity between \( s_0, s_h, \) and \( s_v \).

The purpose of this paper is to show how to construct a feature-extraction unit that can use the regularity of the space of input signals. We show that such a unit computes two essentially equal feature values \( f(s_0) \) and \( f(s_v) \) for two essentially equal signals \( s_0 \) and \( s_v \). We demonstrate also that the unit can recognize all signals in a given invariant subspace once it has seen one member of it. In the edge-detection example mentioned above, we could say that such a unit can recognize a whole class of patterns, and the class of edgelike gray-value distributions. Furthermore, the system can recognize all edges once it has encountered one special example of an edge.

In Section 2 we develop the concept of a symmetrical signal space. We then review some important results from the theory of compact groups. These results then are used to get an overview of all linear feature-extraction units on...
signals are gray-value distributions and the regularities or symmetric signal spaces, and we shall derive some important properties of these feature-extraction units. The general theory is illustrated here with two examples in which the signals are gray-value distributions and the regularities or symmetries are defined with the help of the group of 2-D or three-dimensional (3-D) rotations.

2. DESCRIPTION OF THE REGULARITIES IN THE SIGNAL SPACE

From the edge-detection example in the Introduction we can see that we would like to introduce some type of structure in the space of all possible input signals. In this section we shall introduce an input pattern that allows us to divide the space of all possible input patterns into subspaces of essentially the same patterns.

When we develop our model of a regular or a symmetrical signal space it might be useful to have a concrete example in mind against which we can compare the general definitions and constructs. We shall use mainly the familiar 2-D edge-detection problem as such an example. The receptive field of the feature-extraction unit consists now of a number of detectors located on a disk. If \( x \) denotes a point on this disk, then we denote by \( s(x) \) the intensity measured by the detector located at \( x \). The input to the unit, the signal \( s \), is in this case the light intensity measured by the sensors on the disk. Since the detectors are located on a disk, we can assume that \( s \) is a function defined on some disk. For notational simplicity we can assume that this disk has a radius of 1, and we assume further that the function \( s \) is square integrable. We denote the unit disk by \( D \), and we assume that our signal space is the space of all square-integrable functions on the unit disk. This space is denoted by \( L_2(D) \). Two typical step edges from this space are shown in Fig. 1.

From functional analysis we know that this signal space forms a Hilbert space; i.e., there is a scalar product \( (\cdot, \cdot) \) with which we can measure the length of a signal and the cosine of the angle between two signals (for an exact definition of a Hilbert space the reader is referred to Appendix A and the literature; see, for example, Refs. 1 and 2). These two values are given by \( ||s||^2 = (s, s) \) and \( ||s_1|| ||s_2|| \cos \theta(s_1, s_2) = (s_1, s_2) \). Furthermore, we can compute the distance between two signals as \( ||s_1 - s_2|| \). This is about all that we can say about the signals as elements of the Hilbert space. In our case these measurements can be quite misleading, as Fig. 1 shows. As elements of the Hilbert space these two signals are orthogonal, i.e., as uncorrelated or different as possible. At a higher level of understanding they are, however, nearly identical: they are only rotated against each other. We thus must add a higher level of similarity to our basic, low-level Hilbert-space model.

In the edge-detection problem we introduce such a similarity by defining two gray-value distributions to be essentially equal if they are rotated versions of each other. We thus have a higher-level symmetry or a regularity defined by the set of all rotations. In this section we show how to build a model that describes such a regular signal space. We then investigate what can be said about feature-extraction units that can use these regularities.

We begin by introducing some basic concepts from group theory that will be of fundamental importance below.

Definition 1

1. A transformation of a space \( X \) is a continuous, linear, one-to-one mapping of \( X \) onto itself. The set of all transformations of \( X \) is a group under the usual composition of mappings, and it is denoted by \( GL(X) \).

2. A regular or a symmetrical signal space is a triple \((H, G, T)\) of a Hilbert space \( H \), a compact group \( G \), and a mapping \( T: G \rightarrow GL(H); \ g \mapsto T(g) \) from the group \( G \) into the group of transformations of \( H \). \( T \) preserves also the group structure; i.e., we have \( T(g_1 g_2) = T(g_1)T(g_2) \) for all \( g_1, g_2 \in G \). The Hilbert space \( H \) is called the signal space, the group \( G \) is called the symmetry group of the signal space, and the mapping \( T \) is called its representation.

3. If \( g \in G \) is a fixed group element and \( s \in H \) is a fixed signal, then \( T(g)s \) is the transformed version of \( s \). Sometimes we shall also write \( s^g \) instead of \( T(g)s \).

In our edge-detection example these definitions translate as follows: The Hilbert space \( H \) is the signal space \( L_2(D) \) of square-integrable intensity distributions on the unit disk. If \( s \in H \) is a given intensity distribution and \( g \) is a 2-D rotation, then we can map this signal \( s(x, y) \) onto the rotated intensity distribution \( s[g^{-1}(x, y)] \). Now we select a fixed rotation \( g \). The mapping \( T(g) \) that maps a function \( s \in H \) to its rotated version is obviously a linear mapping from \( H \) into \( H \). The symmetry group \( G \) is in this example the group of 2-D rotations, denoted by \( SO(2) \). The mapping \( T \) maps the rotation \( g \) to the rotation operator \( T(g) \), where \( T(g)s(x, y) = s[g^{-1}(x, y)] \). Note that \( T(g_1 g_2) = T(g_1)T(g_2) \) for all \( g_1, g_2 \in G \).

For the definition of a compact group and groups in general the reader is referred to Appendices A and B and the literature (see, for example, Refs. 3 and 4). Among the possible symmetrical signal spaces there are, of course, also uninteresting ones. As an example, take \( L_2(D) \) as the Hilbert space and \( G = SO(2) \) as described above. As representation, select the mapping \( g \mapsto id \), where \( id \) is the identity in \( GL(H) \). For every signal \( s \in H \) and all \( g \in G \) we have thus \( T(g)s = s \); the only transformed version of the signal is the signal itself. This example demonstrates that the mapping \( T \) is responsible for the power of the constraints that the symmetry group has on the signal space: In the case in which all group elements are mapped to one element, the identity, we have...
no constraint at all, and the pattern classes consist of one element only. If the mapping \( T \) is such that for a fixed \( s \) the set \(|T(g)sg|g \in G|\) is a large set of signals, then the constraints imposed by the group are powerful.

We now define exactly what we mean when we speak of essentially the same signals. If \((H, G, T)\) is a symmetrical signal space, then we define two signals \( s_1, s_2 \in H \) as \( G \)-equivalent if there is a \( g \in G \) such that \( T(g) s_1 = s_2 \); in this case we write \( s_2 = s_1^g \) or \( s_1 = s_2^g \). If \( p \) is an element in \( H \), then we denote the set of all signals \( s \in H \) that are equivalent to \( p \) by \( p^G = \{ s \in H | s = p^g, g \in G \} \). We call \( p^G \) a pattern class and call \( p \) the prototype pattern of \( p^G \). The set of all pattern classes in \( H \) is denoted here by \( H/G \). We shall also denote a complete set of prototype patterns by \( H/G \), keeping in mind that this is an inexact notation.

From the properties of \( T \) it is clear that \( G \)-equivalence is an equivalence relation and that \( H \) can thus be partitioned into pattern classes. This means that \( H \) is the disjoint union of the different pattern classes; i.e., \( H = \bigcup_{g \in G} p^G \), and if \( s_1^G \cap s_2^G \neq \emptyset \), then \( s_1^G = s_2^G \). These definitions describe the high-level similarity referred to above: two signals represent essentially the same pattern if they lie in the same pattern class, independent of their Hilbert-space relation.

In Section 3 we introduce the definitions and results from the theory of group representations that are used in the rest of the paper. In Section 4 we develop the general theory of feature extraction, and in Section 5 we apply this theory to the edge-detection problem.

### 3. Some Results from the Theory of Compact Groups

We now collect some important results from the theory of compact groups. The interested reader may consult the literature (for example, see Refs. 3, 4, and 7) for a detailed treatment.

Let \( H \) be a finite-dimensional Hilbert space, and let \( T: G \to GL(H) \) be a representation. If we select a fixed basis in \( H \), then we can describe every linear mapping in \( GL(H) \) by a matrix. The representation \( T \) maps thus a group element \( g \) onto a matrix \( T(g) \) such that those matrices satisfy the equation \( \tilde{T}(g_1 g_2) = \tilde{T}(g_1) \tilde{T}(g_2) \) for all group elements \( g_1 \) and \( g_2 \). Now, assume further that \( b = [e_1, \ldots, e_N] \) and \( b' = [e'_1, \ldots, e'_N] \) are two sets of basis elements connected by means of the matrix multiplication \( b' = Ab \). In the coordinate systems defined by these two basis sets the map \( T(g) \) will then be described by the two matrices \( \tilde{T}(g) \) and \( \tilde{T}(g) \), respectively. It is easy to show that the two matrices \( \tilde{T}(g) \) and \( \tilde{T}(g) \) are linked by the equation \( \tilde{T}(g) = A \tilde{T}(g) A^{-1} \). A mapping \( \tilde{T}: G \to GL(n, C) \) as defined above is called a matrix representation of \( G \). For notational simplicity we sometimes ignore the difference between the representation \( T \) and its corresponding matrix representation. In this case we also denote the matrix representation by \( T \).

The next definition introduces some important properties of representations.

**Definition 2**

1. A representation is called a finite representation if \( H \) is finite dimensional.
2. Let \( T \) and \( T' \) be a two finite representations in Hilbert spaces of the same dimension. These two representations define two matrix representations \( \tilde{T} \) and \( \tilde{T}' \). The two representations are called equivalent if there is a matrix \( A \) such that \( \tilde{T}'(g) = A \tilde{T}(g) A^{-1} \) for all \( g \in G \).
3. A finite representation \( T \) of \( G \) is reducible if there is a basis \( \{e_1, \ldots, e_m, e_{m+1}, \ldots, e_N\} \) of \( H \) such that all the maps \( T(g) \) leave the two spaces generated by \( \{e_1, \ldots, e_m\} \) and \( \{e_{m+1}, \ldots, e_N\} \) invariant. A representation is irreducible if it is not reducible.
4. If \( T \) is reducible, then we can find a basis of \( H \) such that the matrices \( \tilde{T}(g) \) with respect to this basis have the form
   \[
   \tilde{T}(g) = \begin{bmatrix} \tilde{T}_1(g) & 0 \\ 0 & \tilde{T}_2(g) \end{bmatrix}
   \]
   with \( m \times m \) matrices \( \tilde{T}_1(g) \) and \((N - m) \times (N - m)\) matrices \( \tilde{T}_2(g) \).
5. We say also that the matrix representation is reducible if there is a matrix \( A \) such that
   \[
   A^{-1} \tilde{T}(g) A = \begin{bmatrix} \tilde{T}_1(g) & 0 \\ 0 & \tilde{T}_2(g) \end{bmatrix}
   \]
6. If \( T \) is reducible and \( T_1 \) and \( T_2 \) are defined as above, then \( T \) is the direct sum of \( T_1 \) and \( T_2 \).
7. If the representation preserves the scalar product in the Hilbert space, then the representation is unitary. In this case we have, for all \( g \in G \) and all \( s_1, s_2 \in H \), \( \langle s_1, s_2 \rangle = \langle T(g) s_1, T(g) s_2 \rangle \).

In most cases of interest to us we can decompose a given representation \( T \) into a sum of irreducible representations \( T^k \). Consequently we can decompose the underlying Hilbert space \( H \) into a sum of smaller subspaces \( H^1, H^2, \ldots \) such that all \( H^k \) are invariant under all mappings \( T(g) \). These subspaces are also minimal in the sense that they do not contain smaller invariant subspaces. \( T^k: G \to GL(H^k) \) is then an irreducible representation of the group \( G \) on the Hilbert space \( H^k \). We shall sometimes refer to \( H^k \) as the subspace belonging to \( T^k \).

It is easy to show that equivalent representations define an equivalence relation on the space of representations, and it is therefore natural to select, from each class of equivalent representations, one member with convenient properties. Some ways of doing this are suggested in the next few theorems, which collect some important results from the theory of compact groups.

**Theorem 1**

Any finite-dimensional, unitary representation is completely reducible; i.e., either it is irreducible or it is the direct sum of irreducible representations.

**Theorem 2**

Let \( T \) be a finite-dimensional representation of a compact group \( G \); then \( T \) is equivalent to a unitary representation; i.e., there is a fixed matrix \( A \) such that \( A \tilde{T}(g) A^{-1} \) is a unitary representation.

**Theorem 3**

All irreducible, unitary representations of a compact group are finite dimensional.

**Theorem 4**

Every irreducible, finite-dimensional representation of a commutative group \( G \) is one dimensional (1-D).
The simplest illustration of these results is provided by the symmetrical signal space \( (H, G, T) \) in which \( H \) is the space of all square-integrable functions on the unit circle, \( G = SO(2) \), and \( T \) is the rotation operator \( T(g) s(\phi) = T(\phi) s(\phi) = s(\phi - \psi) \) (\( \psi \) is the rotation angle of the rotation \( g \)). An element \( s \in H \) can be decomposed into a Fourier series \( s(\phi) = \sum c_n \exp(2\pi i n \phi) \); i.e., \( H \) is the direct sum of the one-dimensional subspaces \( H^0 \) consisting of the complex multiples of \( \exp(2\pi i n \phi) \). The representation \( T \) can be reduced to the sum of 1-D, irreducible representations \( T^n \) defined by \( T^n(s(\phi)) = T^n(\phi)[\sum c_n \exp(2\pi i n \phi)] = \sum n \cdot k \cdot c_n \exp(2\pi i n \phi) + c_{\phi} \exp(2\pi ik(\phi - \phi')) \). The (1-D) matrix representation is given by the matrix entry \( t_{nk}(g) \) defined by \( g \mapsto \exp(-2\pi i k \phi) \).

The next theorem, known as the Peter-Weyl theorem, is one of the most important results in the theory of compact groups. It describes how a function on the group can be decomposed into a series of simple basis functions. It is also possible to say that the theorem describes a kind of Fourier expansion of functions on compact groups.

**Theorem 5**

Let \( T^k = 1, 2, \ldots \) be a complete set of inequivalent, irreducible, unitary representations of a compact group \( G \), let \( d_k \) be the dimension of \( T^k \), and let \( t_{nmk}(g) \) be the matrix elements of the transformation \( T^k \); then the functions

\[
(d_k)^{1/2} t_{nmk}(g)
\]

with \((k = 1, 2, \ldots) (1 \leq n, m \leq d_k)\) form a complete orthonormal system of square-integrable functions on \( G \) with respect to the inner product based on the invariant integration on \( G \).

We can reformulate the result as follows: each function \( s(g) \) can be written as a series:

\[
s(g) = \sum_{k=1}^{d_k} \sum_{n,m=1}^{d_k} \alpha_{nm}^k t_{nmk}(g),
\]

with

\[
\alpha_{nm}^k = (d_k)^{-1} \int_G s(g)t_{nmk}(g) dg.
\]

**4. APPLICATION TO FEATURE EXTRACTION**

We can now apply the results presented so far to investigate feature-extraction units. For this purpose we assume that the signal space of our unit is a symmetrical signal space \( (H, G, T) \) as defined in definition 1. In the discussion of this definition we noted that \( GT \) equivalence is an equivalence relation, and we partitioned \( H \) into the equivalence classes \( H/G \). As a result we can now decompose an arbitrary signal \( s \in H \) into a component depending on \( G \) and one depending on the prototype; we write \( s = s_1(g, p) \). We now define a feature.

**Definition 3**

1. A feature is a complex-valued function \( f \) on \( H \). If we write \( f(s) = f[s(g,p)] = f(g,p) \), then we also assume that \( f \), as a function of \( g \), is square integrable on \( G \).

2. A linear feature is a feature that is also a linear map; i.e., it satisfies the condition \( f(c_1s_1 + c_2s_2) = c_1f(s_1) + c_2f(s_2) \) for all signals \( s_1, s_2 \in H \) and all complex constants \( c_1, c_2 \).

With the help of the Peter–Weyl theorem we can now give a simple description of a features: if we keep the prototype signal \( p \in H/G \) fixed, then \( f \) is a function of \( g \), and we can use the Peter–Weyl theorem to get a series expansion:

\[
f(g, p) = \sum_{n, m} \alpha_{nm}^k(p) t_{nmk}(g).
\]

This expression describes how \( f \) depends on \( p \) and \( g \); we could also say that we have separated the dependencies of \( f \) on \( p \) and \( g \).

Using this expression for the feature \( f \) and the orthogonality of the functions \( t_{nmk}(g) \), we can apply Haar integration to compute the mean value of \( f \) over one pattern class or the correlation between the signals in two pattern classes. For the mean we get \( \int_G f(g, p) dg = \alpha_0(p) \), where \( \alpha_0 \) is the coefficient belonging to the trivial, 1-D representation defined by \( t(g) = 1 \) for all \( g \in G \). For the correlation we get

\[
\int_G f(g, p_1)f(g, p_2) dg = \int_G \left[ \sum_{n,m} \alpha_{nm}^k(p_1) t_{nmk}(g) \right] \times \left[ \sum_{n,m} \alpha_{nm}^k(p_2) t_{nmk}(g) \right] dg
\]

\[
= \sum_{n,m} \sum_{l,j} \sum_{i,j} \alpha_{nm}^k(p_1) \alpha_{lm}^k(p_2) \times \int_G t_{lj}(g) t_{nmk}(g) dg
\]

\[
= \sum_{n,m} \sum_{l,j} \alpha_{nm}^k(p_1) \alpha_{nm}^k(p_2) d_h^{-1}.
\]

For our purposes these features are too general, and in the rest of this section we therefore consider only linear features and simple functions of linear features.

Assume now that \( (H, G, T) \) is a symmetrical signal space and that the representation \( T \) is the direct sum of a complete set of inequivalent, unitary, irreducible representations \( T^k \) of \( G \). We denote the Hilbert space belonging to the irreducible, unitary representation \( T^k \) by \( H^k \). The dimension of this space is denoted by \( d_k \), and the basis elements of \( H^k \) are signals \( e_n^k \). The matrix entries of \( T^k \) with respect to this basis are the functions \( t_{nmh}(g) \). For a fixed index \( k \) we denote the matrix of size \( d_k \times d_k \) and elements \( t_{nmh}(g) \) by \( T^k(g) \).

For simplicity we shall also assume that \( H \) is the direct sum of all the subspaces \( H^k \). A basis of \( H \) is thus given by the elements \( e_n^k \).

Using the fact that the elements \( e_n^k \) form a basis of \( H \), we find for a prototype signal \( p \) the following expansion:

\[
p = \sum_{k} \sum_{n=1}^{d_k} \beta_n^k e_n^k.
\]

Using the scalar product in the Hilbert space \( H \), we can compute the coefficients \( \beta_n^k \) as usual by \( \beta_n^k = \langle p, e_n^k \rangle \). The subspaces \( H^k \) are invariant under the group operations, and the transformation of the basis elements is described by the
matrices \( T^k(g) = [t_{nm}^k(g)] \). For an arbitrary signal \( s = s(g, p) = p^x \), we therefore find the expansion

\[
s(p^x) = \sum_k \sum_{nm} \beta^x_{mn} t_{nm}^k(g) e^x_n.
\]  

From this expansion and the linearity of the feature function \( f \), we get the following equation for the value of the linear feature \( f \):

\[
f(s) = f(p^x) = \sum_k \sum_{nm} \beta^x_{mn} t_{nm}^k(g) f(e^x_n).
\]

This equation shows how \( f \) depends on the prototype signal \( p \), the group element \( g \), and the values of \( f \) at the basis elements \( e^x_n \). We see also that a linear feature function \( f \) is defined completely by its values at the basis elements \( e^x_n \). Among the linear features there is a set of especially simple ones, which are defined below.

**Definition 4**

1. A linear feature that has a value of 1 at one fixed basis element \( e^x_n \) and is 0 for all other basis elements is called a basic linear feature. It is denoted here by \( f_n^x \).
2. A basic feature vector \( f^x \) is a map \( f^x : H \rightarrow \mathbb{C}^k \) of the form \( s \mapsto f^x(s) = [f^x_1(s), \ldots, f^x_k(s)] \), where the \( f^x_k \) are the basic features belonging to the subspace \( H^k \) of the representations \( T^k \).

Equation (7) shows that all the linear features are linear combinations of basic linear features, and in the rest of this paper we therefore restrict ourselves to the study of basic linear features.

For the prototype signal \( p \) we get, from Eqs. (5) and (7),

\[
f(p) = \sum_k \sum_{mn} \beta^x_{mn} t_{nm}^k(g) f^x(e^x_n),
\]

and, if \( f \) is a basic linear feature, say, \( f^x \), then we have

\[
f^x(p) = \beta^x f^x.
\]

If \( s \) is a signal that is \( GT \) equivalent to \( p^x = p^y \), then we find, for the feature value [see Eq. (7)],

\[
f^y(s) = f^x(p^y) = \sum_k \sum_{mn} \beta^y_{mn} t_{nm}^k(g) f^x(e^x_n)
\]

\[
= \sum_m \beta^y_m t^x_m(g) \sum_k f^x(p) t_{nm}^k(g).
\]

Using the basic feature vector notation, we then find the following theorem.

**Theorem 6**

First,

\[
f^x(s) = f^x(p^x) = T^x(g) f^x(p).
\]

Second, \( T^x \) was a unitary representation, and therefore we have, for all \( g \in G \),

\[
\|f^x(p)\| = \|f^x(p^x)\|.
\]

The magnitude of the feature vector is thus invariant under the transformation \( p \mapsto p^x \).

The previous results show that we can treat the basic feature vectors \( f^1 \) and \( f^2 \) separately if \( k_1 \neq k_2 \) but that we must consider simultaneously all basic linear features belonging to the same representation. Equation (10) is, of course, valid only for basic feature vectors; if the different basis elements have different feature values, then we do not in general have the same magnitude for all feature vectors of \( GT \)-equivalent signals.

The results in theorem 6 suggest the following procedure for solving the pattern-recognition problem of detecting a certain class of signals.

Assume that we know that our signal space is symmetric, and assume further that we know the symmetry group of our problem and the irreducible, unitary representations of this group. In the learning phase we present to the system one prototype signal \( p \). The feature-extraction unit computes from this signal the feature vector \( f^x(p) = [f^x_1(p), \ldots, f^x_k(p)] \) = \( (\langle p, e_1^x \rangle, \ldots, \langle p, e_k^x \rangle) \), where the \( e_i^x \) form a basis of the subspace \( H^k \) belonging to the irreducible, unitary representation \( T^x \). The next time that the unit receives an unknown signal \( s \) at its input, it computes \( f^x(s) = [f^x_1(s), \ldots, f^x_k(s)] = (\langle s, e_1^x \rangle, \ldots, \langle s, e_k^x \rangle) \). It then compares \( \|f^x(p)\| \) and \( \|f^x(s)\| \), and, if these two values are different, then it concludes that the two signals do not belong to the same class. If they are more or less equal, then we can conclude that they might be equivalent, i.e., that there might be a group element \( g \in G \) such that \( s = p^x \). However, it is not guaranteed that they are similar. If it is necessary to find the transformation \( g \) also, then we might try to recover \( g \) from the matrix equation \( f^x(s) = T^x(g) f^x(p) \). However, this is not always possible, as we shall see in Section 5.

The Hilbert space \( H \) is in practically all cases infinite dimensional, i.e., there are infinitely many basic feature vectors \( f^x \). However, in a real application we can compute only finitely many feature values \( f^x_k(s) = (s, e^x_k) \). Therefore we must select a finite number of these basic features. This problem cannot be solved by using symmetry considerations only; instead we must take into account the problem at hand. As one example of how we can select suitable feature vectors, we consider again the problem in which we want to recognize transformed versions of a prototype signal \( p \). The feature vector \( f^x(p) \) describes in this case the projection of \( p \) into the subspace \( H^k \), and \( \|f^x(p)\| \) is a measure of the portion of \( p \) that lies in \( H^k \). Therefore it seems to be reasonable to choose the subspace \( H^k \) such that a maximal part of \( p \) lies in \( H^k \); i.e., we select \( k \) such that \( \|f^x(p)\| \) is maximal. This is a direct generalization of the optimal basis-function approach to edge detection developed by Hummel and Zucker. The theory also generalizes the pattern-recognition strategies based on Fourier–Mellin transforms (see, for example, Refs. 10 and 11).

**5. GROUP-THEORETICAL DESIGN OF FILTER FUNCTIONS**

In this section we shall use the abstract results of the preceding sections to get an overview of the symmetrical signal spaces, and we shall also discuss methods of constructing filter functions.

Consider now a fixed signal space \( H \) and a fixed (compact) symmetry group \( G \). If \( (H, G, T) \) is a symmetrical signal space, then we know from the previous theorems that \( T \) is equivalent to the direct sum of finite-dimensional, irreducible representations \( T^x \). This gives all the possible spaces \( (H, G, T) \). As is shown above, the properties of \( T \) describe
the strength of the symmetry constraints: if \( T(g) \) is the identity mapping for all \( g \in G \), then we impose no group-theoretical constraint, since every signal \( s \in H \) is \( GT \) equivalent only to itself. On the other hand, if \( T \) is the direct sum of a complete set of irreducible representations of \( G \), then \( T \) imposes the strongest constraint possible.

Let us now return to our 2-D edge-detection problem. The signal space is in this case the space of all square-integrable functions defined on the unit disk \( \mathbb{D} \): \( H = L^2(\mathbb{D}) \). The symmetry group of our problem is the group of 2-D rotations, which means that we want to detect the pattern independent of its orientation in the image. This group is commutative, and the irreducible, unitary representations are therefore all 1-D (see theorem 4). If \( \psi \) is the rotation angle of the rotation \( g \), then the matrix entry \( t_{11}(g) \) belonging to the irreducible representation \( T^k \) is the function \( \exp(-2\pi ik\psi) \). The Peter–Weyl theorem states that, in this case, functions \( t_{11}(g) = \exp(-2\pi ik\psi) \) form a complete function set on the unit circle. Thus the Peter–Weyl theorem is a generalization of the Fourier expansion of periodic functions on a finite interval. The 1-D subspaces \( H^k \) of \( H \) belonging to \( T^k \) are spanned by functions of the form \( W_k(r)\exp(2\pi ik\phi) \) with a radial weight function \( W_k \). These are the so-called rotation-invariant operators investigated in a number of papers (see, for example, Refs. 12 and 13).

Now suppose that \( T^k \) is an irreducible representation of \( SO(2) \); then \( T^k \) transforms a signal \( s \) by multiplying the \( k \)-th component in the Fourier expansion by the factor \( \exp(-2\pi ik\psi) \). If the signal \( s = s(r, \phi) \) has the Fourier decomposition \( s(r, \phi) = \sum \omega_k(r)\exp(2\pi ik\phi) \), then the transformed pattern \( T^k(g)s \) becomes

\[
T^k(g)s = T^k(\psi)s(r, \phi) = \sum \omega_k(r)\exp(2\pi ik\phi) + w_k(r)\exp[2\pi ik(\phi - \psi)]. \tag{11}
\]

From this equation we infer that two signals are \( GT^k \) equivalent if their Fourier coefficients are equal for all \( l \neq k \) and if the Fourier coefficients for the index \( k \) have the same magnitude. An arbitrary representation is the direct sum of irreducible, 1-D representations, and two signals are equivalent if their Fourier coefficients are equal for all indices that are not involved in the representation. For all indices involved in the decomposition of the representation, the Fourier coefficients have equal magnitudes. The number of Fourier components involved in the representation \( T \) describes how powerful the symmetry constraint imposed by the representation \( T \).

The scalar product in \( H \) is given by the integral

\[
\langle s_1(r, \phi), s_2(r, \phi) \rangle = \pi^{-1} \int_0^{2\pi} \int_0^1 s_1(r, \phi)s_2(r, \phi) r d\phi dr.
\]

If \( W_k(r)\exp(-2\pi ik\phi) \) is a fixed basis element \( e_k \), and if the signal \( s \) has the Fourier decomposition \( s(r, \phi) = \sum \omega_k(r)\exp(2\pi ik\phi) \), then the computed feature value becomes \( \langle s, e_k \rangle = \int_0^1 \int_0^{2\pi} W_k(r)\omega_k(r) r dr \). Feature extraction thus amounts to a convolution of the signal \( s \) with the filter function \( W_k(r)\exp(2\pi ik\phi) \) or to a Fourier decomposition of the signal in polar coordinates.

Equation (11) shows also that two different group elements \( g_1 \neq g_2 \) can produce the same transformation \( T^k(g_1) = T^k(g_2) \); we need only to take as \( g_1 \) the rotation with an angle \( \psi \) and to take as \( g_2 \) the rotation with an angle \( \psi + 2\pi/k \). Therefore, in general, we cannot recover the group element \( g \) from the feature values.

In the next example we shall study the same problem as in the previous example but for 3-D images. The signal space is in this case the space of square-integrable functions on the unit ball, and the symmetry group is the group 3-D rotations \( SO(3) \). This symmetry group is unfortunately no longer commutative. Therefore we cannot apply theorem 4, and we must consider representations of dimensions greater than 1. The representation \( T^k \) of \( SO(3) \) has now the dimension \( 2k + 1 \), and the basis elements in the spaces \( H^k \) are the surface harmonics \( P_k^m(\cos \theta)\exp(2\pi im\phi) \) (\(-k \leq m \leq k\)) where \((r, \theta, \phi)\) are the polar coordinates in three-dimensional spaces and \( P_k^m \) are the associated Legendre polynomials. The basic features are computed now by convolving the signal with the filter functions \( W_k^m(r)P_k^m(\cos \theta)\exp(2\pi im\phi) \) with radial weight functions \( W_k^m(r) \). For a detailed study of these functions and their properties, the reader is referred to the literature (see, for example, Refs. 14 and 15). For an application to filter design see also Refs. 13 and 16.

6. FURTHER APPLICATIONS

The feature-extraction problem was the starting point for the development of the group-theoretical model presented in this paper. The model is, however, so general that it seems to be useful also in other areas of image processing and pattern recognition. We mention here only a few areas in which such a strategy might be useful.

The problem of detecting a pattern independent of its orientation in an image is the simplest example in which group-theoretical methods are useful in image processing. Its natural generalization is connected with the analysis of camera motion. This can be seen as follows. Describe the image that is detected by a camera placed at a fixed position in space by the function \( s \). If the camera is rotated around its optical axis, then we get after the rotation \( g \) the image \( s' \), where \( g \) is a 2-D rotation as in the edge-detection example. If we now allow arbitrary Euclidian motions \( g \), then in the same way we get a transformed image \( s'' \). It can be shown that the investigation of this problem leads to the study of the group \( SL(2, C) \), the group of \( 2 \times 2 \) matrices with complex entries and determinant 1. This is the simplest noncommutative, noncompact group, and it is known that its irreducible representation is no longer finite dimensional.5,6,7

As another example, consider image coding. Here the problem is to compress the information content in an image. One approach to solving this problem is to partition an image into small segments. The gray-value distribution in such a segment then is described by a series. The sender then transmits the coefficients in this series and not the original gray values in this segment. The problem here is, of course, which function set should be used in the series expansion.

One approach might be as follows. Assume that we want to code the signals in one class \( p^0 \). Assume further that \( f \) is an arbitrary basis function with \( \|f\| = 1 \). We measure the quality of \( f \) by the mean-squared error \( M(f) = \int_0^1 \int_0^1 (f, p^0)^2 f^2 dg \). A good basis function is thus a function that has a low mean-squared error. This value can be computed with the help of the Peter–Weyl theorem as follows.

We first use the properties of the scalar product to get the
following expressions for the squared error and the mean-squared error:

\[ M(f) = \int_{G} |p^g| - |f, p^g|^2 \, dg = \int_{G} |p^g|^2 - |f, p^g|^2 \, dg. \]

We now use the expansion for \( p \) and \( p^g \) as in Eqs. (5) and (6) to find

\[ \int_{G} |f, p^g|^2 \, dg = \sum_{m n} \sum_{i j} \beta_{m n} \tau_{n m} (g) \delta_{i j} (g) \langle e_{m}, f \rangle \langle e_{n}, f \rangle \, dg \]

\[ = \sum_{m n} \sum_{i j} \beta_{m n} \tau_{n m} (g) \delta_{i j} (g) \langle e_{m}, f \rangle \langle e_{n}, f \rangle \, dg \]

\[ = \sum_{i j} \delta_{ij} \sum_{m} \beta_{m} \tau_{m} (g) \delta_{i j} (g) \langle e_{m}, f \rangle \langle e_{i}, f \rangle \, dg. \]

The first part of the sum describes the part of the prototype pattern that lies in the subspace \( H^k \) spanned by the \( e_1, e_2, \ldots, e_d \). The second sum describes the part of the filter function that lies in \( H^k \). We conclude that it is best to select the basis function \( f = e_k \) such that the space \( H^k \) contains the largest proportion of \( p \). All basis functions \( e_k \) in this space have the same importance.

Our last example is motivated by the optimal basis function approach to filter design, and by some problems in the study of neural networks. In both cases it turns out that we must find the solutions of an eigenvalue problem of the form

\[ \int_{X} k(\xi, \eta) f(\xi) \, d\mu(\xi) = \lambda f(\eta). \]  

Here \( X \) is a set, \( k \) is the kernel of the equation, the integral is a Lebesgue integral, \( \lambda \) is a complex constant, the eigenvalue, and \( f \) is the eigenfunction. In the case of the neural networks the solutions would describe the stable states of the network, and in the filter design method the solutions are the required filter functions. We want to show that we can deduce a number of properties of the solutions from only some general properties of the kernel and the integral. We make the following assumptions:

- The group elements \( g \) are one-to-one mappings from \( X \) onto \( X \), and the integral is invariant under \( G \); i.e., \( \int_{X} f(\xi) \, d\mu(\xi) = \int_{X} f(g(\xi)) \, d\mu(\xi) \) for all functions \( f \) and all group elements \( g \in G \).
- The kernel is symmetrical in the sense that

\[ \int_{X} k(\xi, \eta) \, d\mu(\xi) = \int_{X} k(\xi, \eta) \, d\mu(\eta) \]

for all \( g \in G \) and all \( f \).

We now define the transformation \( T \) as \( T(g)f(\xi) = f(\xi) = f(g^{-1}(\xi)), \) and we finally define (the unknown) Hilbert space \( H \) as the space of the eigenfunctions belonging to a fixed eigenvalue \( \lambda \) in Eq. (12). We then have a symmetrical signal space \( (H, G, T) \), since \( f^g \) is an element of \( H \) for all \( f \in H \) and all \( g \in G \), as can be seen by applying the properties of the kernel and the integral introduced above. For a large number of groups \( G \), the representation spaces \( H \) and many of their properties are known. Since \( H \) is related to \( G \) by a representation, we can now conclude that \( H \) must be one of these known Hilbert spaces. If, for example, \( G \) is a compact, Abelian group, then we find immediately from the representation-theoretical theorems that \( T \) is equivalent to the direct sum of 1-D unitary, irreducible representations. In the case in which \( X \) is the unit circle or the unit disk and \( G \) is a subgroup of \( SO(2) \) we find again as basic solutions the complex exponentials. In the 3-D case in which the symmetry group is given by a subgroup of \( SO(3) \) we find the surface harmonics as solutions.

This result can also be reformulated in the following way: Assume that \( H' \) is the Hilbert space \( L_0(X) \) and that \( G \) and \( T \) are defined as above; then the mapping \( K \) defined as \( K(f)(\eta) = \int_{X} k(\xi, \eta) f(\xi) \, d\mu(\xi) \) is an operator from \( H' \) to \( H' \). Under the conditions formulated above we find that \( K \) commutes with the action of \( G \), i.e., \( K(f^g) = [K(f)]^g \) for all \( f \) and all \( g \). If we now define \( H \) as the space of all eigenfunctions belonging to the fixed eigenvalue \( \lambda \), then we find that \( (H, G, T) \) is the same symmetrical signal space as is described above. This new formulation of the integral equation problem has, however, the advantage that we can now study arbitrary operators \( K \) that commute with the group operation. In the case of the rotation groups \( SO(2) \) and \( SO(3) \), we find that the complex exponentials and the surface harmonics are eigenfunctions of operators that commute with the actions of \( SO(2) \) and \( SO(3) \), respectively. They are especially eigenfunctions of the Fourier transform and the Laplacian. This is a group-theoretical explanation of why the filter functions obtained by the optimal basis-function approach have a number of other nice properties, such as Fourier-transform invariance.

These examples may demonstrate the power of the group-theoretical method. The method allows us to find results that we could obtain otherwise only by means of lengthy computations, and it gives an explanation of otherwise unrelated properties. We shall not go into further detail here, but the interested reader can turn to the literature.

**APPENDIX A: GROUPS AND HILBERT SPACES**

In this appendix we summarize some notations, definitions, and basic properties of groups and Hilbert spaces.

If \( A \) and \( B \) are sets and \( f \) is a mapping from \( A \) into \( B \), then we denote this by \( f: A \rightarrow B; a \leftrightarrow b = f(a) \). By \( R \) and \( C \) we denote the real line and the complex numbers, respectively. \( R^n \) and \( C^n \) are the spaces of \( n \)-dimensional real and complex vectors, respectively. The unit circle \( S^1 \) is denoted by \( e \) and the unit disk; i.e., the unit circle together with its interior is denoted by \( D \). The conjugate complex value of the complex number \( z \) is denoted by \( \bar{z} \).

We now define the algebraic concept of a group.

**Definition 5**

A nonempty set \( G \) is called a group if a product \( \circ: G \times G \rightarrow G; (g_1, g_2) \rightarrow g_1 \circ g_2 \) is defined for every two elements \( g_1 \) and \( g_2 \) in \( G \), for which the following conditions hold:

- \( (g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3) \). We say that the product is associative.
- There is a unique element \( e \in G \) such that \( e \circ g = g \circ e = g \) for all \( g \in G \). \( e \) is called the identity element.
- For every element \( g \in G \) there is a unique element \( g^{-1} \) such that \( g^{-1} \circ g = g \circ g^{-1} = e \). \( g^{-1} \) is called the inverse of \( g \).
A group is called commutative or Abelian if \( g_1 \circ g_2 = g_2 \circ g_1 \) for all \( g_1, g_2 \in G \).

**Definition 6**

1. Assume that \( H \) is a (complex) vector space. A scalar product is a mapping \( \langle \cdot, \cdot \rangle \) with the following properties:
   - \( \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \) for all \( x, y, z \in H \).
   - \( \langle \alpha \cdot x, y \rangle = \alpha \cdot \langle x, y \rangle \) for all \( x \in H \) and all \( \alpha \in \mathbb{C} \).
   - \( \langle x, y \rangle = \langle y, x \rangle \) for all \( x, y \in H \).
2. A vector space together with a scalar product is called an inner product space or a pre-Hilbert space.

**Theorem 7**

If \( H \) is a pre-Hilbert space, then \( \| x \| = \langle x, x \rangle \) defines a norm on \( H \).

**Definition 7**

1. A normed space is complete if every Cauchy sequence converges in this space.

2. A Hilbert space is a pre-Hilbert space that is complete in the norm defined by its scalar product.

**APPENDIX B: TOPOLOGICAL GROUPS, REPRESENTATIONS, AND THE HAAR INTEGRAL**

The definitions and the theorem below introduce the concept of a topological group, representation of such groups, and invariant integration on a group.

**Definition 8**

1. If the set \( G \) underlying the group is also a topological space, then \( G \) is a topological group if the product \( (g_1, g_2) \rightarrow g_1 \circ g_2 \) and the inversion \( g \rightarrow g^{-1} \) are continuous maps.

2. A topological group is called a compact group if \( G \) as a topological space is compact, i.e., if any open covering of \( G \) has a finite subcovering.

**Definition 9**

Assume that \( G \) is a topological group, \( H \) is a Hilbert space, and \( GL(H) \) is the space of continuous, linear, one-to-one mappings of \( H \) into itself. A representation of \( G \) is a map \( T: G \rightarrow GL(H); g \mapsto T(g) \) that satisfies the following conditions:

- \( T(e) = id \) (\( e \) is the identity in \( G \) and \( id \) is the identity function on \( H \)).
- \( T(g_1 g_2) = T(g_1) T(g_2) \) for all \( g_1, g_2 \in G \).
- \( (x, g) \mapsto T(g)x \) is a continuous mapping from \( X \times G \) onto \( X \).

**Theorem 8**

For compact groups (and the slightly more general locally compact groups) it is possible to construct a positive, right-invariant, and normed integral, i.e., an integral that satisfies the following three conditions:

1. \( \int f(g)dg \geq 0 \) if \( f(g) \geq 0 \) for all \( g \in G \).
2. \( \int f(g_1)dg = \int f(g)dg \) for all \( g_1 \in G \).
3. \( \int 1dg = 1 \).

**Definition 10**

The integral described above is unique, and it is called the Haar integral of the group \( G \). It is also possible to define a left-invariant integral of a group, and it can be shown that those integrals are identical for compact groups. Integration over a group is an averaging operation in which all elements in the group are equally likely. This is a direct consequence of the fact that the integral is normed and invariant.

**ACKNOWLEDGMENT**

The research reported in this paper was supported by a grant from the Swedish National Board for Technical Development.

**REFERENCES**