

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/05207 A2

(51) International Patent Classification⁷: **G06T**
(21) International Application Number: PCT/IL01/00597
(22) International Filing Date: 28 June 2001 (28.06.2001)
(25) Filing Language: English

(26) Publication Language: English
(30) Priority Data:
60/215,120 29 June 2000 (29.06.2000) US

(71) Applicant (for all designated States except US): **TECHNION RESEARCH AND DEVELOPMENT FOUNDATION LTD.** [IL/IL]; Gutwirth Science Park, Technion City, 32000 Haifa (IL).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SHAMIR, Joseph** [IL/IL]; 27a Chasidey Umot Haolam Street, 32985 Haifa (IL). **HAR, Offer** [IL/IL]; 16 Ben Gurion Street, 44257 Kfar Saba (IL).

(74) Agent: **G. E. EHRLICH (1995) LTD.**; 28 Bezalel Street, 52521 Ramat Gan (IL).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CLASSIFIER FOR IMAGE PROCESSING

(57) Abstract: An image classifier for classifying objects of an image, comprising: an optical and/or electronic and/or hybrid optical-electronic processor for measuring objects to determine whether said objects comprise features useful in classifying said objects, and a conditionality network associated with said optical processor for using conditionality to select said features useful in classifying interactively with measurement outputs of said optical processor, thereby to classify said objects.



WO 02/05207 A2

5

Classifier for Image Processing

Cross-Reference To Related Applications

The present application claims priority from US Provisional Patent Application No. 60/215,120 filed 29th June 2000.

10

Field of the Invention

The present invention relates to a classifier for image processing and more particularly but not exclusively to a classifier for measuring the presence of features within image parts and thereby inferring a classification of the image part.

Background of the Invention

Image recognition is a field of technology still in its infancy and the problem of automatic processing of images is poorly understood. There are a number of ways of processing an image. Depending on which way is used, and

the purpose of the particular image recognition application, the image processing may include some or all of the following steps:

- identifying individual parts within an image,
- breaking down those parts into basic shapes,
- 5 the classification of those basic shapes and parts of the image, and
- attempting to reach an understanding of the image as a whole.

In the case of electronic reading, the units to identify are easy to find, and algorithms exist for distinguishing basic shapes in the letters. Once the letters are identified the original text can be reconstructed.

- 10 The need for an accurate, fast, and reliable object recognition module is encountered in many different applications ranging from postal sorting equipment to home robotics to high-fidelity military applications.

- In the early days of image recognition, the task of developing pattern and object recognition applications fell in the field of Artificial Intelligence (AI) and more specifically in the field of Vision, (Expert) System development. Work in this field resulted in the development of many different approaches and algorithms where the input to the algorithm was the object's image. The image or picture was usually presented as a bit-map and the expected output of the algorithm was the identification of the object. The main methodology behind these algorithms was to use different approaches within image processing in order to find a correlation (or resemblance) between the object in
- 15
- 20

the picture and a class of objects, to which class the algorithm attempts to match the observed object. Although these algorithms were regarded as promising, the modules developed were not particularly satisfactory, and often the main obstacle cited was the computation and processing time required for
5 the classification algorithm.

Outside of AI, object recognition has also evolved in optics, utilizing matched filters. Optical filter approaches started as early as the 1960's with the work of VanderLugh and have evolved over the years. The work, which concentrated on pure optics, faced an obstacle in the form of the lack of
10 computation abilities, mainly due to the limitations of optical processing and computing. The lack of optical computational ability led to a standstill in the development of optical based object recognition modules.

The problem faced within the optical field in dealing with and exploring object recognition led to the development of various hybrid systems which
15 make use of computer based algorithms to enhance optical based object recognition schemes. The development and research in the field of hybrid systems has been initiated only in the last 15-20 years and has allowed for an improvement in the functionality of object recognition modules.

Optical object recognition began with the introduction of the Optical
20 Correlator by Vanderlugt in the early 1960's. The recognition scheme was based on a Matched Filter (MF) and involved measuring the response of the input image to a specific spatial filter. The use of matched filters for signal recognition and classification has been in wide use in communication theory,

however, the optical Recognition and Classification problem is different both in its nature and in its goal.

The matched filter technique is optimal for detecting a known signal that is corrupted by additive White Gaussian Noise, which is the case with a communication signal. However additive white Gaussian noise is not generally a problem in optics. When exploring the optical Recognition and Classification problem one encounters the following:

Usually the objective of an optical correlator is not to identify the existence of a specific signal (image) but to discriminate between several different signals (images) which may be similar in many aspects.

The noise in an optical system is not additive, and does not possess a White Gaussian distribution.

The optical signals (images) which the correlator is required to recognize are derived from three-dimensional objects, and not a one-dimensional signal as in communication theory. Thus signals (images) which are of the same type may appear differently due, for example, to in plane rotation and out of plane tilt of the 3D object.

The development of Optical Object Recognition systems has evolved in several directions since it was first introduced. The first research direction was to improve the capabilities of the optical correlators, which was done solely through the field of optics. A second direction involved the development of better spatial filters, starting with the development of the Spatial Light

Modulator (SLM) including research for an optimal filter for the SLM to display. Lastly, a lot of research was devoted toward the utilization of computer based algorithm in concert with the optical systems to provide for a better utilization of the optical capabilities.

5 The original correlator introduced by VanderLugt is known as the 4F correlator. The correlator's name is simply derived from its physical length which is four times the lens' focal length. The basic architecture of the correlator 10 is shown in Figure-1 below which shows an input plane 12, a first lens 14, a fourier plane 16, a second lens 18 and a correlation plane 20.

10 The simple architecture of the correlator is misleading, and when used two main problems are encountered. First, the spatial filter used in the Fourier plane is holographic in nature and therefore the correlation data rides on a carrier frequency, this means that the second part of the correlator should be tilted from the system's main optical axis. Second, the Fourier transform
15 created by a single converging lens leads to a high demagnification of the Fourier transform and therefore to a loss of resolution.

 The required off-axis measurement of the correlation can be overcome by using a low carrier frequency. However, the use of a low carrier frequency puts a strong limitation on the signal's Space-Bandwidth Product (SBP), which
20 is derived from Nyquist's criteria. The use of Phase Only Filters (POF), can overcome this limitation of the original 4F architecture. However, due to technical limitations, most POFs are implemented using amplitude coding, thereby losing their main advantage. Several attempts to overcome this

limitation where made by implementing the POF using Kinoforms. The technological advances in the development of such elements including the phase modulator SLM further help to overcome this basic limitation of the 4F correlator.

- 5 The high demagnification of the Fourier transform causes a high resolution of the information storage in the Fourier plane. Filtering such high-resolution information requires the use of a very high resolution SLM, a wide dynamic range, and a stringent alignment tolerance.

 The stringent alignment tolerance is met by the Joint Transform
10 Correlator (JTC) as shown in Figure-2.

 In the JTC architecture no Synthetic Filter (SF) is required. Instead, two image subsystems, a reference image subsystem 22 and a main image subsystem 24 are placed side by side. Each subsystem has a lens, 30 and 32 respectively, and output planes are a Fourier plane 34 in the first case and a
15 correlation plane 36 in the second.

 The output of the first plane is taken from the Fourier plane 34, by CCD 38 and projected onto the SLM 28, where it forms a filter onto which the reference image or a derivation thereof is projected as a filter, as will be explained in more detail below. The correlation plane thus receives the result
20 of filtering the main image through the Fourier based derivation of the reference image. The result is detected by CCD 40.

The main advantage of the JTC approach is that there is no need to prepare a spatial filter ahead of time and the filter may be constructed optically when required. Another advantage of the JTC's architecture is its compatibility with electronic computers, basically it is possible to connect a computer
5 between CCD1 38 and the SLM, and thereby apply computerized control to the spatial filter created.

The idea behind the design of the JTC is to use a reference image identical to the input image, so as to give a positive identification of the input image by means of the correlation. Use of the JTC does not necessarily involve
10 computing and creating a specific spatial filter.

The JTC architecture suffers from a further reduction of the SBP due to the partitioning of the input plane to hold both the image to be correlated and the reference image. JTC performance is also limited by the resolution of the SLM and the CCDs. The JTC's architecture as shown in Figure-2 suffers from
15 difficulty in integrating a complex Synthetic Reference Function (SRF).

Reference is now made to Fig. 3, which shows a mathematical equivalent, according to first order optics, of the arrangement of Fig. 1.

In Fig. 2, a so-called 2F correlator, due to its length being twice that of the lenses' focal length, comprises an input plane 42, a Fourier plane 44 and a
20 correlation plane 46. The Fourier plane is surrounded by lenses on either side. An important advantage of the 2F correlator over the 4F correlator is that it can be modified to match the magnification of the SLM's pixel size. The ability to

be modified in this way allows overcoming the limitation posed by the high resolution of the information in the Fourier plane encountered in the 4F correlator.

The use of spatial correlators for the process of pattern classification has lead to research in the field of how to design an optimal Matched Spatial Filter (MSF). The research was initiated in order to improve the results of the matched filter correlators in the process of object classification. The following discussion is mainly based on a tutorial survey by B.V.K. Vijaya Kumar.

The traditional MSF as borrowed from classical communication theory suffers due to three main properties:

1. The MSFs are very sensitive to small changes in the reference signal. In the optical case, since the object's image is highly influenced by in-plane rotation and out-of-plane tilt, the property restricts the usefulness of the MSF.
2. The MSF are light - inefficient because their magnitude response is usually much smaller than one at most frequencies. This magnitude response problem can be overcome by the use of POF. However, at this time the technology does not permit the extensive use of such filters with a high resolution.

Most SLMs allow for amplitude filtering and do not support the complex filter usually required for the optimal MSF.

In order to proceed with a discussion on the design of matched filters, it is necessary to begin with a presentation of the mathematical equivalence of the manner by which the optical correlator computes the correlation, followed by a consideration of the Synthetic Discrimination Function (SDF), and a

5 presentation of some procedures by which the SDF could be computed to enable the optical correlator to provide the optimal results.

Figures 1 to 3 each show different types of Fourier correlators, and all three types of Fourier Plane correlators presented in Figures 1-3 are equivalent mathematically. The basic idea is that these correlators compute the correlation

10 between a reference image that is placed at the input plane and a function (the Spatial Filter) that is placed at the Fourier plane, while the result of the correlation is presented at the output plane.

The output function computed by the Fourier Plane Correlators is given by:

$$\begin{aligned}
 c(x,y) &= \iint F(x_f, y_f) H^*(x_f, y_f) \text{Exp}\{j2\pi(x_f x + y_f y)\} dx_f dy_f \\
 (1) \qquad &= \iint f(x + x_f, y + y_f) h^*(x, y) dx dy \\
 &= \text{Corr}(h(x, y), f(x, y))
 \end{aligned}$$

where:

$f(x, y)$ is the input signal, the image placed at the input plane.

$F(x, y)$ is the Fourier transform of the input image.

$h(x, y)$ is the image of the spatial filter.

$H(x, y)$ is the Fourier transform of the spatial filter image.

(x_f, y_f) denote the coordinates in the Fourier plane.

The output function $c(x, y)$ represents the cross-correlation between the input image, which in our case are the image of the object to be classified, and
 5 the spatial filter's image.

In the design of a Fourier correlator an objective of image processing is to select a spatial filter, $h(x, y)$, whose cross correlation response to all of the images originating from the same class of objects is the same. Obviously, such a task is impossible, instead Hestner and Casasent required that only the values
 10 at the origin of the cross correlation would be the same.

$$(2) \quad \text{Corr}(h(x, y), f_i(x, y)) \Big|_{x=0, y=0} = \text{const}, \quad i = 1, 2, \dots$$

Such a choice of a spatial filter $h(x, y)$, preferably yields the same output peak for all images of the specific object class which had participated in an initial training set. Furthermore, it is expected that the output for an object's
 15 image that is not a part of the training set may be in the vicinity of such a constant.

In order to allow for a simple generation of the spatial filter $h(x, y)$, Hester and Casasent had assumed that the filter could be presented as a linear combination of the specific filters required for each of the images of the
 20 specific object participating in the training set.

$$(3) \quad H(x, y) = \sum_{i=1}^N \alpha_i F_i(x, y)$$

The constants α_i are chosen so the equality of the cross correlation output among all the images belonging to the same class is achieved. The construction of the spatial filter may be achieved by multiple exposure of a film
 5 to the relevant Fourier transform where the exposure time is a function of α_i .

Early computer simulation of the above spatial filter construction methodology has found it to be very useful. In most initial studies out-of-plane tilt was used and the filter still provided good correlation results.

Riggins and Butler showed that it is possible to implement the SDFs
 10 using standard methods of computer generated holograms. However, this approach faced several basic problems: First the computer design did not consider noise that appears in the input plane, thus once noise was encountered (in almost every real experiment) the response of the filter had degraded and therefore the result of the classifier had also degraded. Another problem was
 15 the fact that the spatial filter was a linear combination of the training images thus the response was always a cross correlation and not the auto-correlation expected from a holographic filter. This yielded a situation where the peak of the correlation was not always at the origin of the output plane.

The design of spatial filters for optical classifiers has evolved in several
 20 directions, and a variety of different algorithms have been introduced for the construction of such spatial filters. The main goal in the design of the spatial filters has been to reach an equal correlation peak for all images of objects

belonging to the same class, and a very low correlation peak for images of objects that do not belong to other classes. Some of these methods include:

1. Kallman's Composite Filter Design - aimed at the design of distortion invariant correlation filters.
- 5 2. Circular Harmonic Expansion Based Filters - which are important techniques in the design of distortion invariant correlation filters.
3. Associative Memory based techniques.
4. Lock and Tumble Filters - These filters are based on ideas
10 similar to those of circular harmonic filters and prove to be invariant for in-plane rotation of both the filter and the image.
5. Quadratic Filters - These possess the quality of both distortion invariance and clutter rejection that make them more tolerable to input noise
- 15 6. Linear Phase Coefficient Composite Filters

Other computer based algorithms have been designed and explored, and some of these will be discussed below in connection with Hybrid Optical Systems.

Pattern recognition has evolved as part of the field of Artificial
20 Intelligence, in parallel to its evolution in optics. The field of pattern recognition has been explored and covered by many different papers, surveys,

and books which provide a lot of information on the subject. Following is a discussion about some of the different approaches for computer based pattern classification algorithms.

Through the field of AI several different approaches to pattern
5 classification were investigated. These approaches differ both in the specific classification algorithm and in the problem representation scheme.

The traditional AI approach has been to divide the pattern classification problem into two sub-problems. The first deals with feature extraction, while the second is a decision-making problem. This is probably the most common
10 approach used by the AI community in pattern recognition since it makes use of already present decision making algorithms.

Another approach used by the AI community is to devise specific algorithms that perform both tasks of feature extraction and decision making simultaneously. The simultaneous approaches usually make use of specific
15 tailored mathematical models.

The initial and main goal of AI was to develop computer systems and algorithms that would possess and utilize intelligence in their inference process. When comparing human (intelligent) inference to that of simple computer algorithms we encounter the main difference to be found in the decision
20 making process itself. While humans make intelligent decisions using partial data about the problem, utilizing their knowledge and experience, this is not the case for simple computer algorithms. Researchers in the field of AI set to

explore and develop decision-making algorithms that are intended to behave in a similar manner to that of the human mind. These algorithms deal with making decision under uncertainty.

In the following subsections I present a short overview of some of these
5 decision-making techniques. It should be noted that the presentation in this background section is partial and there are many other algorithms and approaches that are used by the AI community and are not be presented herein.

The most basic decision making paradigm utilized is a Decision-Tree.
The decision tree lies on the boundary between conventional methods and AI.

10 A decision tree is described by a set of questions, "if ... then..." statements, ordered according to some logic about the problem. When the question sequence is followed it allows the problem solver to converge to the correct decision. The decision making process is called a decision tree since, when presented in a diagram, the flow through the problem solving process is
15 presented as a tree. In the decision tree every node (that is not a leaf) presents a question about the object, and every leaf presents the class, i.e. the decision which would be reached. Reference is now made to Figure 4, which represents a simple decision tree. The decision tree comprises nodes and leaves, and starting at the top it is possible to follow the tree downwards via any one of a
20 number of routes to arrive at any given node.

Decision tree based problem-solving schemes seems simple and at first it might seem that decision trees have nothing to do with AI. However, this is

not the case and the main task which the AI community has faced was the process of constructing the decision trees in a manner that may lead for the minimal number of questions to be asked prior to making the decision. Much study is being devoted to the development of algorithms for the induction of
5 decision trees. Such algorithms are expected to receive the definition of the final results, the questions which can be asked, and a training set which presents how each decision is characterized; and from such data to be able to produce the topology of an efficient decision tree.

An Algorithm for the induction of such decision trees has been
10 presented by Quinlan. The main problem with decision trees and real life problem is the fact that each question requires a decisive answer and the algorithm could not cope with probabilistic answers and could not back track if new data had revealed that an error had been made when answering a previous question. Other work by Change et al, Utgoff and Quinlan presents some
15 answers to these problems. New work in the field of pattern recognition and classification utilizing decision trees is presented by Jun et al and Amit et al.

Despite the new work and the logical flow of the problem solving procedure, Decision Trees are not currently widely used in pattern recognition and classification algorithms, mainly due to the difficulty in dealing with
20 uncertainty in the answer to each of the questions. Generally the trees require yes or no answers to the questions, and are not suitable when the answers can only be couched in probabilistic terms. Furthermore the standard decision tree

is not suitable when answers to questions only lead to decisions with a certain degree of probability.

The Bayesian decision-making methodology has been known since 1600, when Bayes first presented what was later to be known as Bayes' Theorem.

- 5 The main advantage in Bayes' theorem is that it allows making "logical decision" when only partial data is known about the situation. However, it is required for the algorithm to know the entire conditional probability of the problem. For example consider Bayes' formula:

$$(4) \quad P(H|A) = \frac{P(A|H)P(H)}{P(A)}$$

Where:

- $P(H|A)$ - is the probability of making decision H, given a known observation A, or as it is usually presented as the probability of hypothesis H being correct given A.
- 5 • The other parameters represent knowledge of the problem domain, and include
 - $P(A|H)$ - is the probability of A given hypothesis H.
 - $P(A)$ - is the probability of A with no restriction of the hypothesis
- 10 • $P(H)$ - is the a priori probability of the hypothesis H.

The field of Bayesian decision making under uncertainty had been investigated and researched for a long time by many different researches.

Again once one learns the mathematical foundation of Bayes' theorem, the main problem faced by him is to represent the knowledge, i.e. the Bayesian network in a manner that allows it to be solved by a computer. A lot of work and research has been devoted to investigating this basic problem.

The use of the Bayesian decision-making technique has not been restricted to the main decision making (classification) process. It has also been put to use in making the decision as to what questions to ask at each time or basically which features should be extracted to allow the problem solver to converge to the correct answer. Marefat and Ji present an algorithm where the

Bayesian method is utilized for recognition and feature extraction of 3D objects in CAD solid models. Wong and Chang present a Bayesian based algorithm for the recognition of Chinese characters, through the work they present the classification problem as a Bayesian problem and show an off-line algorithm
5 that can be used for classification.

One of the main drawbacks in the Bayesian decision-making scheme is the need to have a probabilistic understanding of the underlying problem. Such an understanding serves as the heart of the decision making inference engine and is represented by $P(A|H)$, $P(A)$, and $P(H)$. The problem usually
10 encountered in the development of real decision-making systems is that there is a lack of accurate data about the problem and therefore the conditional probability is not accurately known and the Bayesian method cannot be used.

Fuzzy logic techniques try to overcome the need to know the accurate conditional probability function by employing a fuzzy decision rule that
15 substitutes Bayes' Theorem. In the last few years a lot of work has been devoted to the research of fuzzy logic decision making. Fuzzy logic research has resulted in the development of a wide mathematical foundation for fuzzy decision making which is very similar to the probability theory.

Basically, if one is to use the Fuzzy rules as the conditional probability
20 and at the same time utilize a Bayesian inference engine it is expected that both the Fuzzy decision maker and the Bayesian one may result in the same problem solving scheme.

Abe and Thawonmas present a fuzzy classifier that allows for ellipsoidal uncertainty regions. The classifier presented yields very good classification results given the fact that the actual probabilities were not known.

Another problem on which Fuzzy logic has been put to use is the
5 decision about the probability of detecting an image or a feature in an image.
The work relates to the feature extraction process.

At first Neural networks were presented as a computer problem-solving scheme that simulates human brain activities. As Neural networks were introduced to AI many attempts have been made to use them for decision
10 making and specifically for pattern recognition. Taver presents a long review of the use of these techniques.

The main idea behind Neural networks was to build a net composed of simple elements, perceptrons, which work in concert to yield a solution. Each perceptron receives any number of inputs and generates an output that is
15 usually a binary function. Figure-5 presents a single perceptron having such a series of inputs, leading to a function and from which is produced an output.

The transformation between the input values and the binary output value is usually made by a non-linear function $f(\cdot)$. The Neural network is constructed by connecting many perceptrons together where those in the first
20 levels produce the input for those in the higher levels.

Many different algorithms and research topics deal with the training of a neural network, which is actually the tuning of the parameters w_i .

One of the more interesting approaches utilizing Neural networks in pattern classification and recognition is the use of Associative Memory. The idea in this problem-solving scheme is that the algorithm learns a set of classes and basically given an input image (or a sub-set of an image), finds the class
5 that is closest to it among all the base classes. Basically the results of the algorithm are similar to the matched correlator only that instead of using a matched filter with which the correlation is computed the system utilizes the network to perform a similar task.

One of the more common human problem solving techniques is to guess
10 a solution, test it and, according to the test results, update the guess according to any test error, until the guessed solution is close enough to the required answer, i.e. the error is small enough. In order to converge to a solution humans make use of logic to guide the guessing scheme, that is we actually search the solution space in accordance with some search strategy (the logic),
15 until an acceptable solution is found.

The same problem-solving scheme had evolved in the field of computer problem solving. In the field of AI, many different search algorithms have been developed and again the goal of these algorithms has been to scan the solution space of a problem and converge in the fastest possible manner to a good
20 solution.

Different search techniques have also been employed in pattern classification problems, and such search techniques may be used to guide two different actions in the classification process.

One such technique involves guiding the feature selection, or the tests that may be conducted on the pattern, to allow classification. The idea is similar in nature to the construction of a decision tree where we need to decide at each node which feature to measure. The sole difference between a search and a
5 decision tree is that in a search the decisions are made in real time, while in a decision tree the entire search is prepared prior to the initiation of the algorithm.

A second technique involves guessing a solution and trying to correct the guess following a comparison between the guessed pattern and reference
10 pattern, which comparison the algorithm tries to classify.

Algorithms which follow the first approach presented include most heuristic searches such as: A*, RTA*, LRTA*, which are presented by Korf. Other types of algorithm which can be utilized are Bi-directional and Moving Target Searches, the details of which are known to the skilled person. The main
15 methodology that is followed is to construct in real time the shortest path (where the path is the set of measured features) which leads from an unknown pattern to a classified one.

Additional approaches that have been attempted use Genetic algorithms for the feature selection process. Such approaches that have been attempted
20 have been proven to be effective specifically when the number of features is very large and some statistical correlation between the features exists.

The second technique referred to above, namely that of guessing a solution and testing it to converge on a result, is simpler to comprehend. Implementations of the technique may for example utilize Genetic Algorithms, Simulated annealing, and other solution oriented search techniques. The

5 technique involves guessing a solution and trying to correct it according to the differences obtained from comparing the guess with the reference pattern.

The above listed decision-making techniques present most of the methodologies used by the AI community in the development of pattern recognition algorithms. Many other approaches have been investigated ranging

10 from Reinforcement Learning through the combination of different decision making methodologies to allow for an accurate and fast convergence.

All the decision-making algorithms considered above require, as an input, a set of properties of the pattern or shape to be classified. These properties are referred to herein as features, and the features can be used to

15 describe the pattern or shape in the solution space. One of the more complicated problems in computer based pattern classification and recognition problems is to measure the properties that are present in the pattern or shape to be classified or, more precisely to extract the features.

Feature extraction using digital computer-based classifiers is extremely

20 computationally intensive.

The first problem encountered is the image segmentation task. In image segmentation, the computer is required to separate the pattern to be classified

from the background scenery and other objects and patterns which may be a part of the image. Many different approaches to image segmentation have been explored through the years and currently there is no one closed solution that accomplishes the task of flawless image segmentation. The main reason is that

5 a correct image segmentation scheme should not restrict itself to geometric properties, nor to photometric properties, but should also rely on the content of the image presented. However, such image segmentation yields some ambiguity since the content of the image may be known only after the pattern has been classified, and segmentation is a requirement for classification. So, as

10 could be seen a closed-loop may be encountered and the only way out is using search techniques or hypothesis testing. In such a case a preferred approach is to assume a solution, or a part of a solution, and try to solve the problem given the hypothesis. Once a solution is reached the hypothesis is re-examined. The solution thus comprises a guided search of the problem domain.

15 The second problem that is encountered in computer based feature extraction is the amount of computation needed for the process. Assuming that the image presented contains a single pattern, i.e. ignoring the segmentation issue, it is possible to perform a frequency domain filtering technique in order to extract a feature. Such a procedure requires the performance of several 2D

20 Fourier transforms, and matrix operations. Such transforms and operations are computationally intensive. Furthermore, with a digital computer it is not possible to apply a real Fourier transform but rather a Discrete Transform, which has somewhat different properties.

Considerable efforts have been devoted to the selection of features that could be simply and confidently measured by a digital computer from a bitmap image of the pattern.

Another important issue in real pattern classification problems is the
5 decision about which feature conveys the most information and, therefore, should be measured to provide for a fast convergence of the classification process. The amount of information residing in a measurement has long been associated with the entropy of the signal. Basically, the higher the entropy the more information will be gained from measuring it.

10 An important point that is considered in feature selection and classifier design, is the amount of data required for a reliable classification.

Other problems and difficulties exists in the process of computer based feature extraction techniques, these include several topics such as attention focusing, i.e. to what part of the image to focus the attention when trying to
15 locate an object or to extract a feature. A Bayesian based solution for this task is known to the skilled person.

The exploration of computer based pattern recognition algorithm has not been confined only to known AI methodologies such as those described above, and efforts have been devoted towards the development of specific methods for
20 pattern classification and recognition.

Most of these concepts and methods utilize some form of correlation computation, although in some cases this form of correlation computation is

hidden behind a complicated mathematical model that compares different features of the object. In many cases these computations are actually the computation of an inner product.

5 Sclaroff and Pentland present a modal matching correspondence and recognition algorithm that finds a correspondence (the correlation) between an object model and an image thereof. The classification is reached by finding the model that is least deformed by matching to the reference object. This is actually the computation of a correlation and the choice of the closest object. Of course the computation is not performed as a direct correlation between the
10 images but the idea is still similar, and makes use of the computational advantages of the computer in mathematical computation and not only bitmap matching.

Ben Arie and Wang present another algorithm that performs object classification. The algorithm presented makes use of Affine correspondence in
15 the frequency domain. Again the basic idea behind the algorithm is to find a correlation between the object's image presented and basic properties of the object which are known. In this case those properties (features) are all in the frequency domain, and the correlation is computed by finding an Affine transformation between the requested feature and the property of the object at
20 hand.

He and Kundu present an algorithm that assumes that objects can be described using a Markov model. The algorithm utilizes a hidden Markov model and uses it for the classification process. Again the method relies on the

idea of measuring the correlation between the object and some reference set of objects, but utilizing a different methodology in order to converge to the solution.

The above-described systems are based on digital computers. Parallel
5 development of optically based systems has encountered several difficulties that have forced researchers to develop systems that do not solely rely on optical processing. This research led to the development of Hybrid optical systems, these systems make use of both optical and electronic processing utilizing the best of each. That is every component is assigned a task for which
10 it can provide a better solution:

Optical components are very good at performing massive computation such as Correlation computation, Fourier analysis, etc, since these can be performed at the speed of light on an entire image or image part, relying on the properties of wave propagation and those of the optical system.

15 Electronic computers are used for the result analysis of the optical measurements and for the control of the entire system including the optical system.

A long survey of hybrid optical systems can be found in IEEE Computer magazine issue 2, 1998.

20 Much effort in the field of Hybrid optical systems has been devoted to the development of optical classifiers. The reason is the fact that the optical measuring and computational speeds are far ahead of those of electronic

computers and therefore it is possible to overcome some of the problems which are encountered in computer based pattern recognition systems.

Since the optical components in such a Hybrid classifier are utilized mainly for the measurement of the correlation, many algorithms have been developed for computing the optimal set of filters to be used. Such algorithms include the algorithm known to the skilled person as Projection on Constraint Sets (POCS)

Many other algorithms and systems have been developed as Hybrid optical systems and more specifically as hybrid classifiers. Nevertheless, the current state of the hybrid optical classifier is unsatisfactory in terms of the time taken to converge on a classification decision, and particularly the ability to cope with uncertainty.

Summary of the Invention

According to a first aspect of the present invention there is thus provided an image classifier for classifying objects of an image, comprising:

a feature measurer for measuring objects to determine whether said objects comprise features useful in classifying said objects, and

a conditionality network associated with said feature measurer for using conditionality to select said features useful in classifying interactively with measurement outputs of said feature measurer, thereby to classify said objects.

Preferably, said feature measurer comprises an optical processor.

Alternatively, said feature measurer comprises a hybrid electro-optical processor.

Alternatively, said feature measurer comprises a digital processor.

5 Preferably, said features useful in classifying are arranged to form a classification feature language.

A preferred embodiment further comprises a feature builder for building a set of features useful in classifying into a classification feature language.

Preferably, said conditionality network comprises a Bayesian network.

10 Preferably, said Bayesian network is controllable to use back propagation of said measurement outputs to select a next feature for measurement.

Preferably, said conditionality network comprises a decision tree.

15 Preferably, said decision tree is controllable to use back propagation of said measurement outputs to select a next feature for measurement.

Preferably, said feature measurer comprises a $4f$ optical correlator, connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

20 Preferably, said feature measurer comprises a $2f$ optical correlator, connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

Preferably, said feature measurer comprises a JTC, connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

Preferably, said feature measurer comprises an electronic correlator,
5 connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

Preferably, said feature measurer further comprises a Fourier transform unit connected to said correlator, which may be an optical correlator, for forming a Fourier transform of said object.

10 Preferably, said Bayesian network comprises an inference engine.

Preferably, said inference engine comprises electronic circuitry.

Preferably, said inference engine is connected to said correlator, and comprises parameter control functionality to control the features useful for classification to be measured by the optical correlator.

15 Preferably, said electronic circuitry further comprises a classifier to conduct classification computations on the output of said correlator, thereby to perform image classification.

Preferably, said electronic circuitry comprises digital electronic circuitry.

20 Preferably, said inference engine comprises a Bayesian network.

Preferably, said inference engine comprises a feature selector for selecting an image feature for measurement by said optical correlator.

A preferred embodiment further comprises a feature register for holding a plurality of features for selection.

5 Preferably, said plurality of features comprising any of the following:

symmetry in the y axis, symmetry in the x axis, rotational symmetry, lines; geometrical shapes; corners; line crossings; energy primarily concentrated at low frequencies, and energy primarily concentrated at high frequencies. More generally, features for detection may include any feature of
10 an object that can be measured using the techniques described herein.

Preferably, the above described feature of rotational symmetry includes any rotational symmetry having an order number (that is the number of lines of symmetry) between 2 and infinity.

In particular, rotational symmetries having order numbers of 2, 3, 4, 6,
15 10, and infinity respectively are especially favored.

Preferably, the feature selector comprises a classification probability estimator for estimating a probability of successful classification of said object given recognition of a given feature, said feature selector being operable to use said probability to select a feature to present said feature to said optical
20 correlator.

Preferably, the optical feature extractor comprises a first imaging device located at the output of said Fourier transform unit for forming an image of said Fourier transform of said feature.

A preferred embodiment further comprises a duplicator connected
5 between the output of the Fourier transform unit and the input of the optical correlator, said duplicator being controllable from said inference engine, said duplicator being operable to generate a spatial filter for use in correlation of said image.

Preferably, said duplicator is operable to use said Fourier transform
10 output as an input to said generation of said filter.

A preferred embodiment receives an instruction from said inference engine to measure a given symmetric property and generates said filter by duplicating said Fourier transform, thereby to form a filter having said given symmetric property.

15 Preferably, said duplicator is operable to generate an arbitrary filter.

Preferably, said arbitrary filter is selectable by said inference engine from any of a low pass filter, a selection of band pass filters and a high pass filter.

Alternatively or additionally, said arbitrary filter is selectable from a low
20 pass filter, a selection of band pass filters, a high pass filter, circular sine filters, circular cosine filters, sinc filters and circular sinc filters.

In a preferred embodiment, the prestored set of filters comprises at least three each of said bandpass filters, said circular sine filters, said circular cosine filters, said sinc filters and said circular sinc filters respectively.

Preferably, duplicator is operable to generate a filter in accordance with
5 input from said inference engine.

Preferably, said duplicator is connected to display said spatial filter in said correlator.

A preferred embodiment further comprises a second imaging device at the output of said correlator, for forming an image of a response of said feature
10 to said spatial filter.

A preferred embodiment further comprises an evaluator connected to said second imaging device for evaluating said imaged response and ascribing a value thereto, said value being transferable to said inference engine.

Preferably, said features are object dependent features.

15 A preferred embodiment further includes object-dependent features which are any of geometric features, relational features and frequency features.

Additionally or alternatively, said geometric features are any ones of a group comprising features containing symmetry and features containing internal primitive features.

20 A preferred embodiment further comprises the ability to identify an internal primitive feature within a feature by setting said duplicator to generate

a spatial filter of a suspected internal feature and measuring a correlation of said feature with said spatial filter.

In a preferred embodiment the feature extractor further comprises a spatial angle adjustment unit for rotating said Fourier transform to produce a plurality of spatial correlations with said filter.

A preferred embodiment is operable to identify an internal primitive feature within a feature by setting said duplicator to select from a prestored set of spatial filters at least one spatial filter of an internal feature to be tested and measuring a correlation of said feature with said spatial filter.

Preferably, said feature extractor further comprises a spatial angle adjustment unit for rotating said Fourier transform to produce a plurality of spatial correlations with said filter.

A preferred embodiment is operable to identify a geometric feature by setting said duplicator to generate a spatial filter of a synthetic image related to said geometric feature and measuring a correlation of said geometric feature with said spatial filter.

A further preferred embodiment is operable to test said geometric feature for rotational symmetry by setting said duplicator to form said spatial filter by taking a rotational segment of said Fourier transform and duplicating said rotational segment to synthetically complete said transform.

A preferred embodiment is operable to test said geometric feature for axial symmetry by setting said duplicator to form said spatial filter by taking an

axial segment of said Fourier transform and duplicating said axial segment to synthetically complete said transform.

A preferred embodiment is operable to identify frequency features of an object by setting said duplicator to produce a high pass filter and a low pass
5 filter, each for respective intensity measurements of said feature, and calculating a ratio between said high frequency and said low frequency measurements.

In a preferred embodiment, said duplicator is operable to produce at least one of a high pass, a low pass and a band pass filter, and to output said
10 filter to said correlator for correlation with said object, said correlator further comprising an evaluator for comparing said correlated output with an uncorrelated output to ascribe a value to said correlation.

A preferred embodiment is operable to identify a relational feature by separately identifying two features, determining a spatial relationship
15 therebetween and applying a best fit relational description to said determined spatial relationship.

A preferred embodiment comprises a picture dependence analyzer for determining object features that are specific to a current orientation of an input camera to said object, thereby to screen out said object features.

20 Preferably, said inference engine is operable to control said duplicator to generate a series of filters for correlating with each object, said inference

engine comprising an analyzer for using said series of correlations to determine a most likely object.

Preferably, said inference engine uses a Bayesian classification to select said series of filters to generate for correlating with each given object.

5 Preferably, the Bayesian classifier is constructed to divide classification uncertainty of said objects into complementary sets such that one of said sets can be substantially ruled out after each correlation.

Preferably, said optical processor comprises a spatial light modulator electronically controllable to produce filters.

10 According to a second aspect of the present invention there is provided an optical feature extractor comprising an input for receiving a visual object, a part extractor for extracting a part of said visual object and building a filter from said part, and a correlator for correlating between said input visual object and said filter, thereby to determine the presence of a feature in said visual
15 object.

Preferably, said feature is symmetry and said part extractor is operable to duplicate said part at least once to construct said filter.

Preferably, said feature is n^{th} degree symmetry, said part is an n^{th} part of said object and said part extractor is operable to duplicate said part n times to
20 construct said filter.

Preferably, the part extractor is electronically controllable.

According to a third aspect of the present invention there is provided an optical feature classification decision mechanism comprising:

a Bayesian network linking a first layer of nodes and a final layer of classes via probability links, at least said first layer of nodes representing measurable features of visual objects and said classes representing potential classification groups of said visual objects, said probability links comprising probability numbers representing potential classifications given feature measurement results,

said mechanism being operable to use said probabilities to guide a visual object measurement and classification process.

A preferred embodiment is operable to receive a result of a first measurement and to use said probabilities to determine whether to request a second measurement or to output a classification decision.

A preferred embodiment further comprises:

an error calculator for calculating the possibility that a present classification result is erroneous, and

a thresholder for comparing said calculated possibility with a predetermined threshold, thereby to decide whether to output said classification decision or whether to request said further measurement.

According to a fourth aspect of the present invention there is provided a method of classifying an object taken from an image, the method comprising:

optically processing said object to extract a filter,
comparing said object with said filter,
using Bayesian networking of a predetermined set of features and a
result of said comparison, selecting either one of a feature classification result
5 and a next filter for a next comparison, and
continually comparing and selecting until a classification result is
reached.

Brief Description of the Drawings

10 For a better understanding of the invention and to show how the same
may be carried into effect, reference will now be made, purely by way of
example, to the accompanying drawings, in which:

Fig. 1 is a simplified schematic diagram of a prior art 4F correlator,

Fig. 2 is a simplified diagram of a prior art joint transform correlator
15 (JTC),

Fig. 3 is a simplified schematic diagram of a 2F correlator,

Fig. 4 is a simplified diagram of a decision tree useful for classification
between five classes,

Fig. 5 is simplified diagram of a perceptron of a neural network,

Fig. 6 is an overall block diagram of a classifier according to a first preferred embodiment of the present invention,

Fig. 7 is a simplified block diagram of the optical feature extractor of Fig. 6,

5 Fig. 8 is a schematic diagram of the optical feature extractor of Fig. 7,

Figs. 9 – 11 are diagrams showing the construction of a simplified Bayesian network,

Fig. 12 shows a set of five objects useful for training and testing embodiments of the present invention,

10 Figs. 13-15 are graphs showing experimental results obtained with an embodiment of the present invention and showing a number of features that needed to be measured before convergence on the result,

Fig. 16 is a schematic representation of a multiple layer Bayesian network for use in embodiments of the present invention.

15

Description of the Preferred Embodiments

Reference has been made to Figs. 1 to 5 in the background and the content thereof will not be considered again.

Reference is now made to Fig. 6, which is a simplified block diagram
20 showing a hybrid classifier according to a first embodiment of the present invention. In Fig. 6, a hybrid classifier 60 receives as an input an object to be

classified 62. The object 62 is preferably the output of a segmentation process and is preferably a relatively low level geometric feature. The hybrid classifier itself comprises two parts, an optical feature extractor 64 which comprises optical components arranged as will be described in more detail below and
5 whose aim is to measure features of the object to be classified. The hybrid classifier 60 further comprises a digital inference engine 66 which is a digital processing device able to receive outputs from the optical feature extractor regarding features present or absent and use the received input to do two things, firstly to decide on a next feature to measure for optimal convergence on a
10 decision, and secondly to reach a decision as to the classification of the object 62. As will be explained in more detail below the inference engine 66 selects features according to estimated probabilities for reaching a rapid decision from the current measurement given the previous measurements. A classification output 68 is produced indicating a class of objects to which the present object
15 62 has been inferred to belong.

Reference is now made to Fig. 7, which is a simplified block diagram showing in more detail the optical feature extractor 64 of Fig. 6. The optical feature extractor 64 preferably comprises a Fourier transform system 70 for optically carrying out a Fourier transform on an input object, which transform
20 is detected by a first CCD 72. The optical feature extractor 64 further comprises an optical correlator 74, whose structure will be discussed in greater detail below. The optical feature extractor is able to produce a Fourier transform of an input object in the Fourier transform system 70 and is able

within the correlator to produce a correlation result of the input object with any filter that may be used as an input into the correlator. As will be discussed below, sources for such filters include the Fourier transform system and the inference engine. The correlation output is processed by computer 78, either to
5 make a classification decision or to choose a new feature to measure.

Reference is now made to Fig. 8, which is a simplified schematic diagram showing the optical processor of Fig. 7 in greater detail. Parts that are identical to those shown above are given the same reference numerals and are not referred to again except as necessary for an understanding of the present
10 embodiment. An input signal bearing a pattern or object to be classified firstly arrives at a beamsplitter 80, which splits the input signal between the Fourier transform system 70 and the correlator 74. The Fourier transform system comprises a lens 82 and a Fourier plane 84.

At the other output of the beamsplitter, a lens 86 and mirror 88 direct the
15 signal to the correlator 74, which comprises a spatial light modulator (SLM) 90. The SLM 90 is electronically controllable to hold a filter. The SLM 90 is followed by a system of two lenses 92 and 94, the second being placed in what is known as the correlation plane.

Between the first CCD detector 72 and the SLM 90 is located a
20 duplicator. The duplicator is preferably an electronic device for modifying a Fourier output to form a filter. More generally the duplicator is able to generate any desired filter under electronic control. Specifically, the duplicator is able to take the Fourier analysis output of part of the input signal and

duplicate it to form a filter covering a whole signal comprising symmetry. The SLM 90 may then perform a correlation between the real image and the synthetic image as defined by the filter, thereby to determine whether the object represented by the signal is symmetrical or not. As will be described in more
5 detail below, there are numerous types of symmetry that can be tested, and the presence or absence of one or more of such types of symmetry can be used to classify the object. Aside from symmetry there are other features that may be considered for testing, as will be described below and such filters are also preferably provided via the duplicator, although generally without reference to
10 the generated Fourier transform.

The concept of features is used throughout the *Artificial Intelligence* (AI) literature, and definitions vary. Although these variations are minor and their influence on the classification procedure is negligible, we shall use the following definition:

15 A *feature* is an atomic piece of information that is used in order to describe the state/object.

The basic idea, in the above definition, is to use simple/primitive features that describe the object. The use of primitive features provides an ability to reuse the feature in different classification problems without the need
20 to rebuild a new feature extraction process, (that is to say the features may be individual components of a Generic Classification Language). For example consider the manner by which humans remember a person, we may pick several features such as {height, weight, hair color, eye color, etc.} and give

each of these a value from a predefined range, these values actually representing the person. That is we are actually defining the person's location in a feature space where knowing his location (in the feature space) is equivalent to identifying him.

5 The features actually define a language that is used for the description of the objects by the classifier and later on for their classification. Through this project we shall use several different features which are used to describe the visualization of an object.

 There are different types of optical features that could be used for the
10 identification of the objects. These features may be categorized as follows:

 1. Object dependent – object dependent features depend solely on the object to be classified, and are not effected by the surroundings, nor by the manner the object was photographed. In the embodiments herein, treatment is mainly confined to object-dependent features, however this is by no means
15 intended as a restriction on the invention, which extends to any feature that is optically measurable. Object dependent features may be considered to include geometric features, relation features, and frequency features, although frequency features may be considered to be affected by overall scaling.

20 2. Surrounding dependent – Surrounding dependent features depend on the surroundings and the background of the object, for example a ship is presented against a blue background which is the sea.

3. Picture dependent - The picture dependent features are features that depend on the relative orientation of the camera v. the object. These features do not help in the course of the classification and it is in fact useful to filter out such features. The problem is that in many cases

5 picture dependent features (if not ignored) distort the understanding of other features and lead to a wrong classification.

Geometric features are the basic features that describe an object. The Geometric features to be used throughout this work include:

1. **Is symmetric in "Y"**. These objects agree with the following, For

10 some value y_0 the following equality holds

$$(5) \quad I(x, y_0 - y) = I(x, y_0 + y),$$

where $I(x, y)$ is the intensity (gray level) of the pattern.

2. **Is symmetric in "X"**. These objects agree with the following, For

some value x_0 the following equality holds

$$15 \quad (6) \quad I(x_0 - x, y) = I(x_0 + x, y).$$

3. **Has rotational symmetry of order k** . These objects agree with the following equality.

$$(7) \quad I(r, \theta) = I(r, \theta + \Delta\theta),$$

where $\Delta\theta = \frac{2\pi}{k}$, and $I(r, \theta)$ is the intensity (gray level) of the pattern

20 in polar coordinates.

4. **Contains a primitive** - The objects contain a pattern of a primitive object, for example, within the object we can locate a circle, or within the object we can locate a triangle. The present embodiments preferably use different primitives in order to describe as wide a range of objects as possible.

The manner in which these features are extracted is as follows:

The first three types of features: {**Is symmetric in “Y”, Is symmetric in “X”, Has rotational symmetry of order k** } may be extracted by measuring the correlation between the original image and a synthetic image reproduced from the original image. For example we can produce a k order symmetric Fourier transform by duplicating a section of the original Fourier transform, then by measuring the correlation between the original image and the synthetic transform we can infer about the symmetry features of the object.

The forth type of feature *Contains a primitive* is measured by measuring the correlation between the original image and an image of the primitive. For example we can produce via the *Duplicator* the Fourier transform of a circle and then measure the correlation. Considering the correlation peaks and strength, it is possible to infer whether there is a circle inside the original object.

As may be understood the extraction of the above features is preferably done by measuring the correlation of the image with that of a synthetic image produced by the *Duplicator*. The Duplicator creates a synthetic Fourier

transform of an object given some part of the real object's image for example for measuring a *symmetry feature*, or a totally synthetic Fourier transform for measuring a *containment feature*.

The frequency features represent information about the ratio between the
 5 object's energy in a specified range of frequencies. We can define the following:

A pattern has sufficient energy in the frequency range $[\rho_1, \rho_2]$, if the following holds:

$$(8) \quad \frac{\int_0^{2\pi} d\varphi \int_{\rho_1}^{\rho_2} |\mathfrak{F}_{\rho,\varphi} \{I(r, \theta)\}|^2 d\rho}{\int_0^{2\pi} d\varphi \int_0^{\infty} |\mathfrak{F}_{\rho,\varphi} \{I(r, \theta)\}|^2 d\rho} > Th$$

10 Where Th is some arbitrary threshold, and $\mathfrak{F}_{\rho,\varphi}$ is the Fourier transform in a Polar coordinate system.

Specifically we may define the High-pass and Low-pass features of an object in the following manner:

1. The pattern is mainly low-pass. In this case we state that most of
 15 the pattern's energy is concentrated in low frequencies. That is to say

$$(9) \quad \frac{\int_0^{2\pi} d\varphi \int_0^{\rho_{LP}} |\mathfrak{F}_{\rho,\varphi} \{I(r, \theta)\}|^2 d\rho}{\int_0^{2\pi} d\varphi \int_0^{\infty} |\mathfrak{F}_{\rho,\varphi} \{I(r, \theta)\}|^2 d\rho} > Th.$$

Where ρ_{LP} is the highest frequency that is still considered as a low frequency, Th is some arbitrary threshold, and $\mathfrak{F}_{\rho,\varphi}$ is the Fourier transform in a Polar coordinate system.

2. The pattern is mainly high-pass. In this case we state that most of
 5 the pattern's energy is concentrated in high frequencies. That is to say

$$(10) \quad \frac{\int_0^{2\pi} d\varphi \int_{\rho_{HP}}^{\infty} |\mathfrak{F}_{\rho,\varphi}\{I(r,\theta)\}|^2 d\rho}{\int_0^{2\pi} d\varphi \int_0^{\infty} |\mathfrak{F}_{\rho,\varphi}\{I(r,\theta)\}|^2 d\rho} > Th.$$

Where ρ_{HP} is the lowest frequency that is still considered as a high frequency, Th is an arbitrary threshold, and $\mathfrak{F}_{\rho,\varphi}$ is the Fourier transform in a Polar coordinate system.

- 10 3. It is possible to define any border frequency ρ_{HP} as a feature.

The main difficulty encountered with the above features is that in order to measure them correctly it is necessary to be able to separate the object from the surroundings i.e. perform some type of segmentation. Segmentation is a computationally difficult task, especially if it is attempted via optical means. In
 15 this case the size of the object v. the entire picture influences the measurement, i.e. the features become picture dependent and this is what we are trying to avoid.

The measurement of the features is done by measuring the band-pass and whole energy of the object's image and computing the ratio between them.

This is done by creating a low/high pass filter via the *Duplicator* and measuring the response of the image.

Relational features refer to the relation between primitives that are present in the image. An example for such primitives could be: {"Is there a
5 triangle above a square?", "Are there at least two triangles in the image?", "A circle, a triangle and a rectangle should form a 60-60-60 triangle", etc.} As could be expected the measurement of these features is a very complicated task, since such a "vocabulary" of relations is very large. That is to say we need to create a complex language in order to describe objects using such features.
10 Furthermore, the process of feature extraction becomes more complicated.

The manner in which such features may be measured is by finding some correspondence between the response of the image to different primitives. By that we may find the spatial relationship between the primitives. The following embodiments do not explicitly consider relational features except where the
15 spatial relationship is a rotation.

Background features are not features of the object itself but refer directly to its surrounding. When humans identify objects they take information from the object's surroundings. For example we know that a ship could not be in the air or that a tank could not be at sea. That is, by understanding the surroundings
20 of the object it is possible to infer more about the object itself. An example of such an inference could be an object photographed on a very steep hill. If one knows that a tank could not travel on such an incline one could infer that the object is not a tank.

The following are examples of a set of features that an embodiment of the present invention may be set to measure.

- a) Rotation symmetry of various orders.
- b) Axial symmetry:
 - 5 i) Symmetry in Y.
 - ii) Symmetry in X.
- c) Contains a triangle: Right angle triangles, Equal sided triangles.
- d) Contains a square.
- 10 e) Contains a rectangle.
- f) Contains a pentagon.
- g) Contains a circle.
- h) Contains an ellipse - Different types of ellipses may be tested where the difference used in testing may be the
15 eccentricity of the ellipse.
- i) Any other feature that can be defined and recognized by optical or/and digital means.

2. Frequency Features - These features will be used throughout the classification process, the point that should be taken into

account is that the scale of the object with respect to the entire image highly influences these features.

3. Relational features – As mentioned above, relational features are not discussed specifically in the present embodiments.

5 4. Background Features – Background features are not discussed explicitly in the present embodiments except to say that the background colors may be taken as additional cues for the inference engine.

10 5. Arbitrary Spatial Filters - a set of predetermined spatial filters may be made available to the system so that the system may check the response of the object to filters that do not depend on the object.

15 6. Polarization – Sometimes the polarization of the light arriving from an object or its background may carry characteristic information. Polarization features are not discussed explicitly in the present embodiments except to say that they may be taken as additional cues for the inference engine

Feature measurement requires an object to be input to the optical feature extractor, which object is now measured for the presence of one or the other of the features that are being considered.

20 Given an object as input to the system, an image of the object may be represented by the function $f(x,y)$. Now assuming that the object has a rotation symmetry of α degrees, then, if the object is rotated by an angle of α

there should not be any difference in its appearance. In this case the following equation should hold:

$$(11) \quad \begin{aligned} f(x, y) &= f(x', y') \\ \begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \end{aligned}$$

The symmetry is for any numbers of turns in an angle of α degrees,

5 therefore

$$(12) \quad \forall n \in \mathbb{N} \exists \begin{cases} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}^n \begin{bmatrix} x \\ y \end{bmatrix} \\ f(x, y) = f(x', y') \end{cases}$$

Remembering the properties of the Fourier transform:

$$(13) \quad \begin{aligned} \mathfrak{F}\{f(\bar{x})\} &= F(\bar{u}) \quad \bar{x} = (x, y), \bar{u} = (u, v) \\ \mathfrak{F}\{f(\underline{A}\bar{x})\} &= \frac{1}{|\underline{A}|} F(\underline{A}^T \bar{u}) \end{aligned}$$

we observe that if the matrix $\underline{A}$ is a rotation matrix than a rotation in the
10 real space causes a rotation in the opposite direction in the Fourier space.

Therefore, if the object has a rotation symmetry of α degrees its Fourier transform will also have a rotation symmetry of α degrees.

Using this property of the Fourier transform the rotation symmetry feature may be measured in the following manner:

15 1. The object's Fourier transform may be captured using the Fourier transform system.

2. Via the *duplicator* a section of α degrees of the picture may be duplicated $n * \left(\frac{360}{\alpha} \right)$ times in order to form a complete Fourier transform.
3. The synthetic Fourier transform may be displayed on the SLM.
4. The correlation between the original object and its “symmetric” Fourier transform may then be measured.
5. The existence of the rotation symmetry feature will be decided according to the correlation value.

Axial symmetry features may be measured in a similar manner to the Rotation symmetry features. Via the Duplicator one may create a synthetic Fourier transform of an object possessing the requested symmetry and measure the correlation between the original object and the symmetric object.

Assume that the function $f(x, y)$ has an axial symmetry. For the sake of convenience one may assume the function to be symmetric in X. That is to say the function has the following property:

$$(14) \quad f(x_0 + x, y) = f(x_0 - x, y)$$

For reason of convenience one may further assume that the symmetry is around the point $x = x_0$.

Now remembering the properties of the Fourier transform and the fact that the function $f(x, y)$ is a real function, the following holds:

$$(15) \quad \begin{aligned} \mathfrak{F}\{f(x, y)\} &= f(u, v) & |\mathfrak{F}\{f(x + x_0, y)\}| &= |f(u, v)| \\ f(x_0 + x, y) &= f(x_0 - x, y) & |f(u, v)| &= |f(-u, v)| \end{aligned}$$

Therefore, once again the presence of the feature may be measured using the following steps:

The object Fourier transform may be captured using the Fourier
5 transform system.

a synthetic Fourier transform is preferably created via the duplicator.
The synthetic transform possesses the symmetry around the requested axis X,
or Y respectively.

The synthetic Fourier transform is then preferably displayed on the
10 SLM.

The correlation between the original object and its “symmetric” Fourier
transform is now measured.

The existence of the axis symmetry feature may be decided according to
the correlation value.

15 The containment features are features that indicate whether a primitive
object is contained within the object to be identified. As may be appreciated by
the skilled person, the Fourier transform is a Linear Operator, therefore:

$$(16) \quad \mathfrak{F}\{f_1(x, y) + f_2(x, y)\} = \mathfrak{F}\{f_1(x, y)\} + \mathfrak{F}\{f_2(x, y)\} = f_1(u, v) + f_2(u, v)$$

Now assuming $f(x,y)$ to be an image of an object that may be presented as a sum of two positive images:

$$(17) \quad f(x,y) = f_1(x,y) + \tilde{f}(x,y)$$

Now assume that $f_1(x,y)$ is an image of a known primitive. If this is the case then the correlation between the image $f(x,y)$ and that of the primitive $f_1(x,y)$ will have a high positive value.

The containment features may be measured as follows:

First of all, a set of primitive objects and corresponding Fourier transforms are preferably set up. The Fourier transform is preferably rotated several times in order to allow for some tilt and rotation of the primitive within the original image.

Preferably the synthetic Fourier transform of the primitive is sent, using the duplicator, to the SLM.

The correlation between the original object and the "symmetric" Fourier transform is measured.

The existence or otherwise of the primitive in the original image is then decided according to the correlation value.

Frequency features depend on the amount of energy the object has in low frequencies v. the energy in possesses in high frequencies.

Frequency features may be measured in the following manner:

Using the duplicator, it is possible to create High pass, Low pass, and Band pass, spatial filters.

The spatial filters are preferably displayed on the SLM.

The energy of the reconstructed filtered object is measured and
5 compared to the energy of the unfiltered object.

The existence of the requested frequency feature will be determined according to the ratio between the correlation value of the object with itself and the BP-filtered object.

The measurement of arbitrary features may be carried out by displaying
10 some arbitrary (predefined) spatial filter on the SLM and measuring the response of the object to the filter.

The arbitrary features may be measured in the following manner:

Arbitrary spatial filters are created, preferably using the duplicator.

The spatial filters are presented on the SLM.

15 The energy of the reconstructed filtered object is measured and compared to the energy of the unfiltered object.

It will be noted that the outputs of the correlations need not be binary results. Rather numerical values may be ascribed to correlations and feature recognition can be probabilistic.

20 The classification algorithm that is utilized in the present embodiment is based on a Bayesian inference engine. The algorithm preferably makes use of a

Bayesian network that is constructed automatically from a training set presented to the system.

The construction of the Bayesian network via the training set that is presented to the specific optical system provides the algorithm with the ability
5 to be adaptive to both the classes that have to be classified, and to the specific optical system. This allows the algorithm to provide a robust classification solution to a wide variety of classification problems using optical systems whose fidelity might vary.

The algorithm implemented in this work preferably comprises the
10 following steps:

1. Measure the response of the training set to all the different spatial filters (features).
2. Learn the conditional probabilities of the features, and construct a Bayesian network.
- 15 3. Classify an object using the following procedure:
 - a) Decide upon a feature to be measured.
 - b) Measure the feature.
 - c) Compute the posteriori probabilities.
 - d) Decide on the final classification, or return to (a) if
20 a decisive classification is not reached.

In the following section a short overview of Bayesian networks is provided together with a specific description of the network chosen to be implemented in the present embodiments.

Bayesian networks are a tool that enables a representation of the relationship between cause and effect, thereby to use such a relationship in a reasoning process. Herein, a short discussion about Causes, Effects, and the manner in which they are presented by a Bayesian network is provided.

The relationship between cause and effect is true in object recognition. When one builds a classifier one actually develops a theory that comprises an understanding of the relationship between the observed features (the effect) and the observed pattern (the cause).

The main idea in *cause* and *effect* is that the order of appearance is well defined. At first the cause exists and only afterwards the effect takes place. Stating this in words yields: “If *A* is the cause of *B* then *B* occurs later than *A*.”

In First Order Logic (FOL) or any other predicate logic “ $\alpha \rightarrow \beta$ is equivalent to $(\neg\beta) \rightarrow (\neg\alpha)$ ”, so there is no directionality to model causality. This problem of modeling causality has been partially overcome by rule based systems where the interpretation of the “if __ then __” statements is unidirectional. That is “if *a* then *b*” does not imply “if not *b* then not *a*”. Semantic nets also incorporate directionality by using directed graphs. This suggests that some sort of directed graph structure could be used to describe the relationship between the cause and the effects.

Conditional probability is very suggestive of causality, although if one examines Bayes' theorem the relationship to causality is temporarily lost. The expression $P(B|A)$ can be read as the probability of an effect, **B**, given a cause,

A. If we now apply Bayes' theorem we can compute $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

- 5 Using the theorem we actually compute the probability of the cause **A** given that we have observed the effect **B**. In explaining the computation we have actually imposed a directed graph structure over the event space. We had two events **A** and **B** and using the graph we have implied that **A** causes **B**.

The above is the main idea behind Bayesian networks:

- 10 1. We consider the entire space of events containing both the *causes* (the objects we want to recognize) and the *effects* (the observed features).
2. We apply a directed graph leading from the *causes* to the *effects*.
3. We learn the conditional probabilities of the *effects* given the
- 15 *causes*.
4. Finally, we use this information to infer the *cause* given the observed *effects*.

- As presented at the end of the previous section a Bayesian network has a structural part reflecting the causal relationship between two random variables,
- 20 in our case the *causes* (the objects to be recognized) and the *effects* (the observed features). Another part of the Bayesian network is the conditional

probability associated between the causes and the effects. In the Bayesian network we consider both the causes, and the effects as random variables. The user of the Bayesian network specifies each of the observed features and via the relationships represented by the network one may calculate the probability of
5 each of the causes. Such a calculation is performed using Bayes' theorem.

Bayesian networks have certain advantages and disadvantages:

1. Advantages:

- a) There is a solid theoretical basis.
- b) Results are given in terms of probabilities, therefore
10 they have a readily understood meaning.

2. Disadvantages:

- a) In the general case, the only known algorithm
requires a large amount of data and calculation. However, it is the
belief of the present inventor that this disadvantage does not
15 apply to object recognition and pattern classification.

- b) A Bayesian network may be the wrong tool for
solving the problem. It is well known that not all problems are
Bayesian in nature. If one tries to apply a Bayesian solution to a
non-Bayesian problem one will probably get a meaningless result.
20 For example if one tries to solve the radar detection problem
using a Bayesian model he will probably fail. Once again this is

not the case in object recognition since we can define the problem as a Bayesian one.

- c) The Bayesian system gives numerical values as output. Numerical values may give a false sense of accuracy, if
- 5 the initial probabilities on which they are inevitably based are inaccurate. This is a problem that would be considered in the stage of interpreting the results of the calculation.

When considering human reasoning one finds that it is heavily dependent on causality. Humans may consider a given effect by searching for

10 other effects until reaching a conclusion that could represents the cause of these effects. Such a method of reasoning might fail if one is not capable of grasping the numerical aspects of uncertainty. Many AI reasoning methods consider all mathematical and numerical aspects of the problem, but usually fail to grasp the causality. The Bayesian model incorporates both the causality and the

15 mathematical aspects of the problem in a single reasoning model. Such may be achieved by using two mathematical tools, probability, which is the main mathematical tool which deals with uncertainty, and the Graph theory, or more precisely the *directed acyclic graph* (DAG), which is a tool with which causally related events may be represented.

20 The combination of these two tools leads to the creation of a Bayesian network.

Considering the following example:

In a large department store a burglar alarm has been installed. The Security Company decides to build an expert system that will help decide whether a break-in is in progress. The observer is a watch officer who listens to the alarm. He produces the input that is his belief that he has heard
5 the alarm. Given this input the system has to decide whether a break-in is in progress.

When one starts to consider the possible chain of events one may reach the following possibilities:

- A burglar has broken in $\rightarrow$ the alarm goes off $\rightarrow$ The Watch
10 officer hears the alarm.

- A malfunction occurs $\rightarrow$ the alarm goes off $\rightarrow$ The Watch officer hears the alarm.

- Nothing happens $\rightarrow$ the alarm does not go off $\rightarrow$ The watch officer thinks he had heard the alarm.

15 The causality graph of the above rules is presented in Figure-9.

From observing the causality network presented in Figure-9, one may conclude the following:

1. The graph is a directed acyclic graph (DAG).
2. There are three basic causes: A break-in, a malfunction, and
20 No-event.

There are four different effects: Nothing heard, alarm heard, alarm goes off, and nothing occurs.

1. Two of the effects, namely: Alarm goes off, and nothing occurs, are also the causes of the two observations: nothing heard, and alarm
5 heard.

It is obvious that not all of the possibilities are presented in the graph but the graph still presents all the rules specified in the example.

Now in order to be able to infer the basic cause from the observation of the watch officer, one needs to know the probabilities associated with each of
10 the edges in the graph. It may be obvious that both the sum of the edges leading to a node and the sum of the edges exiting a node are positive numbers which are not bounded by unity, (considering conditional probability). Nevertheless it should be clear that the value of each of the edges is bounded by unity.

If one is to reconsider the dependency network presented in Figure-9,
15 one may find that there is a problem with the chain of events, namely that there is no consideration of the fact that two of the events, the break-in, and malfunction, can occur simultaneously. Although this event has not been presented in the basic rules at the beginning of the example it should have been added as was the event "Nothing happens". If these are not introduced the
20 problem representation will be incomplete and the calculated results will be misleading.

The complete dependency graph, including the above-mentioned additions, is presented in Figure-10.

Considering the modified dependency graph one may still see that there is a missing event and that is simply that the alarm goes off and the watch officer does not hear it. Let us assume that the probability of such an event is negligible and therefore it is omitted from the graph.

Now, if one is to compute the probabilities associated with each of the edges, one may solve the mathematical problem and, given a probability that the alarm has been heard, compute the probability that a break-in took place.

Reference is now made to Fig. 11 which is a simplified diagram of a two-level Bayesian network, in which a series of features, such as the features that may be measured in the image classification process, are associated using probabilistic connections with a series of classes that are intended to classify the objects. The present embodiments use a two level Bayesian network to represent the knowledge required for the classification process, although the invention is of course in no such way limited and any number of levels may be used.

The limitation posed by restriction to a two-level knowledge representation scheme is that the features used throughout the classification are required to be statistically independent. Such a requirement poses a very heavy restriction on the choice of the features that form the Classification Language to be used. However, a choice of a multi level Bayesian network as the

knowledge representation scheme may be expected to complicate the inference process and may additionally require a considerable level of preprocessing computations in order to simplify the network to allow it to be solved easily (fast) by a computer.

- 5 After choosing the knowledge representation scheme to be a two level Bayesian network as represented in Figure-11, the first step in the construction of the classifier is to learn the conditional distribution of the features given the different classes of objects.

 The learning of the conditional distribution is preferably achieved by
10 presenting the hybrid classifier with a training set. The training set preferably contains several different observations of objects from each class and the classifier measures each of the features that can be used in the classification process. That is, for each observation of an object the classifier measures all the features composing the Generic Classification Language.

- 15 After the measurement process is completed the classifier preprocess the knowledge acquired from the training set in order to compute the following:

1. The a-priori probability for each class.
2. The distribution of each feature.
3. The conditional distribution of a feature given any of the
20 classes.
4. The entropy of each feature.

5. The conditional entropy of each feature.

These parameters are then used by the classifier throughout the classification process for the following tasks:

1. Choosing which feature may be measured at any given
5 time.
2. Computing the posterior probability that the object belongs
to a specific class.

The a-priori probability of the classes is a parameter that is used in the initializing phase of the Bayesian classification process. Basically, this
10 parameter influences the following:

1. The initial probabilities - which influence all the calculated
probabilities.
2. The feature to be measured selection.
3. The rate at which the algorithm converges.

15 Preferably, the training set is selected from the entire set of objects in a manner that the probabilities of the occurrence of any given class remain unchanged. Theoretically the initial probability of occurrence of a class, as long as it is not zero or unity need not effect the final classification result, but rather simply require further features to be measured, thereby extending the time that
20 is required for the classification to take place.

The a-priori probability of the classes is calculated by summing the occurrences of each class in the training set and dividing the sum by the total number of elements in the training set. Therefore, the a priori probability to encounter class j is given by:

$$5 \quad (18) \quad P_0(c_j) = \frac{\sum_{i=1}^N (E_i = c_j)}{N}$$

Where:

- N - is the number of elements in the training set.
- E_i - is the i^{th} element in the training set.

When learning the distribution of the features one has to obtain several different parameters.

One such parameter is the distribution of a feature - The distribution is used in order to compute the probability of measuring a specific value for a feature. This computation is essential for the use of Bayes' theorem.

Another such parameter is the conditional distribution of a feature given any of the classes - This is important for both the computation of the probability of a feature given a hypothesis (an assumed class), and the computation of the information residing in a feature, i.e. its entropy.

A further parameter is the conditional entropy of a feature - The conditional entropy of a feature is computed directly from the conditional distribution of a feature.

Yet another parameter is the entropy of a feature - The entropy of a feature is computed directly from the distribution of a feature. Note that the entropy of a feature changes as the classification process is performed this is due to the fact that the probabilities of the classes change and therefore the probability distribution of the features also change.

A parameter which is saved by the algorithm is the conditional probability of the features given the classes. This parameter is saved in an $N \times M$

matrix where each element is a function representing the features distribution, (N is the number of classes while M is the number of features).

All the other parameters are computed directly from the conditional distribution of the features.

- 5 The conditional distribution of the features is represented by a distribution function. The distribution function is preferably modeled by a Gaussian distribution convolved with a chain of Delta functions each representing a single occurrence of the feature, as follows:

$$(19) \quad f(F=x|c) = C_c \sum_{i=1}^N \text{Exp} \left\{ -\frac{(x_i - x)^2}{2\sigma_c^2} \right\}$$

- 10 wherein:

- x_i - is the value of the feature as measured in its Ith occurrence.
- σ_c - is the standard deviation of the feature's measurement. This parameter is needed due to errors of measurement and fluctuations in the measurement process, which are caused by the optical system, the photograph of the object, etc.

- c - is the class for which the distribution of the feature is represented.
- C_c - is a normalizing constant.

The probability of measuring a feature with a value of x independent of the class is computed directly by:

$$(20) \quad f_0(F=x) = \sum_{c \in C} P_0(c) f(F=x|c)$$

wherein:

- $P_0(c)$ is the a-priori probability of class c .

Throughout the classification process there is a need to update the distribution function of the features. Updating is required for the computation of the entropy and the feature selection process. The updated distribution function of the features may be computed as follows:

$$(21) \quad f(F=x) = \sum_{c \in C} P(c) f(F=x|c)$$

wherein

- $P(c)$ is the probability of class c .
- x is the value of the feature F for which the computation is done.

The entropy of a feature is computed from the features probability distribution function. The entropy is defined by:

$$(22) \quad I(x) = - \int_{\Sigma} f(F=x) \text{Log}(f(F=x)) dx$$

Such a computation is difficult to perform analytically by a computer, therefore the computation is preferably performed by dividing the space of the feature, Σ , into bins and using the following formula:

$$(23) \quad I(x) = - \sum_{i \in \Sigma} P(i-1 \leq x < i) \text{Log}(P(i-1 \leq x < i))$$

Where $P(i-1 \leq x < i)$ is the probability that the feature will be in the bin that is between $i-1$ and i . The value chosen is the value of the probability distribution function in the center of the bin.

$$5 \quad (24) \quad P(i-1 \leq x < i) \equiv f(F = i - \frac{1}{2})$$

Such a form of computation is actually a simple form of numerical integration, and may be regarded as sufficiently accurate for the classification process.

Using the above definitions the algorithm computes the following:

- 10
- The entropy of a feature given a class:

$$(25) \quad I(x|c) = - \sum_{i \in \Sigma} f(F = i - \frac{1}{2}|c) \text{Log}(f(F = i - \frac{1}{2}|c))$$

- The internal entropy of a feature:

$$(26) \quad I_I(x) = \sum_{c \in C} P(c) I(x|c)$$

15 It is noted that the internal entropy of a feature depends on the current probabilities of the classes. The internal entropy parameter allows for the dynamic adaptation of the algorithm to the observed object.

- The external entropy of a feature is given by:

(27)

$$I(x) = - \sum_{i \in \Sigma} \left(\sum_{c \in C} P(c) f(F = i - \frac{1}{2} | c) \right) \text{Log} \left(\sum_{c \in C} P(c) f(F = i - \frac{1}{2} | c) \right)$$

It is noted that the external entropy or just entropy of a feature depends on the current probability of the classes.

5 The classifier is preferably dynamically adaptive in respect to the observed class. Thus it is neither desirable nor possible to a-priori define the sequence by which the features would be measured. Therefore, the decision on which feature to measure next, is preferably made in real-time during the classification procedure.

10 In order to converge as fast as possible to the correct classification, it is desirable to measure the feature whose determination of presence or absence produces as much relevant information as possible.

Remembering the properties of the entropy function one can infer the following:

15 • If the features observations are concentrated near a specific value the entropy of the distribution will be low. More specifically if the features distribution function can be presented by Dirack's Delta function the entropy will be zero.

 • If the features observations are scattered, thereby yielding a
20 distribution function that is not concentrated, the entropy will be high.
The highest value of the entropy is encountered when the distribution

function is uniform. Therefore, the feature that conveys the most information may be selected by taking the feature with the highest ratio between its external entropy and internal entropy. That is to say that we would like to find a feature that changes over the classes but is constant within each class. Let us define the following:

1. The probability of a feature is

$$(28) \quad P(f_i) = \sum_{j=1}^N P(C_j) P(f_i | C_j)$$

2. The external entropy of a feature (or just entropy) is:

(29)

$$10 \quad I(x) = - \sum_{i \in \mathbb{Z}} \left(\sum_{c \in \mathcal{C}} P(c) f(F = i - \frac{1}{2} | c) \right) \text{Log} \left(\sum_{c \in \mathcal{C}} P(c) f(F = i - \frac{1}{2} | c) \right)$$

3. The internal entropy of the I^{th} feature over the J^{th} class is

$I(f_i | C_j)$, and is computed during the knowledge preprocessing stage. The

total internal entropy of a feature is calculated by

$$(30) \quad I_I(f_i) = \sum_{j=1}^N P(C_j) I(f_i | C_j)$$

15 4. The information measure of a feature by

$$(31) \quad Inf(f_i) = \frac{I(f_i)}{I_I(f_i)}.$$

Using the above definitions it is possible to select for measurement those features whose information level is maximal $\max_{f_i} \arg \{ Inf(f_i) \}$.

Given the measurement of a new feature f_i , the posteriori probabilities of the classes may be updated. The computation of the new probabilities is performed according to Bayes' rule. This is performed as follows:

1. The current probability of the J^{th} class is $P(C_j)$.

2. The a priori probability of the I^{th} feature is $P(f_i)$.

3. The conditional probability of the I^{th} feature given the J^{th} class is $P(f_i|C_j)$.

4. For each class one may compute

$$(32) \quad P^+(C_j) = P(C_j|f_i) = \frac{P(f_i|C_j)P(f_i)}{P(C_j)}.$$

5. One may now choose the class with the highest probability

$$Ans = \max_{c_j} \arg \{ P^+(C_j) \}.$$

The probability that the above class selection was an error may be calculated as $P_e = 1 - \max\{P^+(C_j)\}$. If the probability of error is below a predefined threshold the classification may be considered a result, if this is not the case then classification is continued with another feature.

A prototype apparatus according to an embodiment of the present invention was constructed and trained and then used to classify the different shaped objects shown in Fig. 12. The results are shown graphically in Figs. 13, 14 and 15 which respectively show the suggested classifications achieved by the embodiment after different numbers of feature tests. Specifically, Fig. 13 shows attempts to classify a square, and it is seen that after three measurements the results are unambiguous. Fig. 14 shows attempts to classify a triangle and, again, it is seen that three measurements are sufficient. Fig. 15 shows attempts to classify a star of David, and again reliable classification is achieved after three measurements.

In the above experiment, the classifier always began by measuring for the feature of "Has a 72° Rotational symmetry" and then proceeded to a new feature depending on the result. The choice of measurement reflects the fact that the above-mentioned feature is the most discriminatory between the object classes considered.

The second feature to be measured may now depend on the class of the test object that is assumed by the classifier as a result of updating the posteriori probabilities given the first measurement. The process continues until a classification decision is reached, which is to say that a classification is made with an error probability that is below a predetermined threshold.

The mean number of features that was required by the classifier in order to reach the correct classification of the objects in the above experiment was 2.3474.

Reference is now made to Fig. 16 which is a simplified diagram
5 showing a Bayesian network for use with the present invention. In the embodiment of Fig. 16 a series of measurable features are arranged in two layers with interconnecting probabilities therebetween. The use of a second layer of features allows the Bayesian network to better represent interrelationship probabilities between the features and therefore to lead to
10 faster and more reliable convergence.

The above-described embodiments thus comprise a hybrid optical classifier that makes use of a Bayesian network inference engine and an optical feature extractor.

The correct usage of the Bayesian inference engine has been found to
15 provide a robust classifier that can adapt to different types of optical systems even optical system with a very low fidelity.

The optical system in the present embodiments is similar to the most basic 4F correlator introduced by VanderLugt in the early 1960s. However, the use of an SLM that is fully controllable by the computer allows the system to
20 display different spatial filters that are generated in real time. The controllable SLM allows measuring of features related to the specific object presented. (Example of such features are the rotation symmetry features).

The embodiments provide the use of real-time designed filters to measure a specific feature that depends on the object to be classified and can not be generated ahead of time.

5 The classification algorithm used is not new and the use of Bayesian networks in the field of decision making under uncertainty is known. Furthermore, the feature selection process used, which measures the information carried by each feature and chooses the most informing feature, is also known. However, the feature information measure used in this project has been tuned for the classification of visual objects.

10 When examining different feature information measures used, the current art often uses cost functions, which consider both the information conveyed by a feature and the cost that will be required in order to measure the feature. In our case the application of a similar measurement process to all features renders such a cost feature redundant.

15 Furthermore, the feature selection process used in this work recalculates both the internal and external entropy of the features prior to every decision, while most other algorithms only recalculate the external information. This recalculation provides for a higher degree of adaptability in the feature selection process yielding a faster conversion to the correct classification.

20 The use of features in the way described permits a Generic Classification Language. The development of a Generic Classification Language that does not depend on the optical system, nor on the object to be

classified, allows the development of any type of classifier and its usage for all types of purposes without having to take into account its specific operating environment or purpose. This capability is provided since the actual selection of the features to be used is decided in real time by the classifier without
5 effecting the classification speed nor the correctness of the classification.

The classification features used herein are not applicable only to the classifier herein described but are also applicable to other classifiers and other optical systems. The generic nature of the classification features allows the development of further optical analyzing tools for object classification, without
10 the need to specifically tailor the tools at the development stage to the specific classification task.

It is particularly pointed out that part or all of the classifiers described herein in the preferred embodiments may be used as training devices for more general use of the recognition system.

15 As mentioned above, the classification features are of generic application to image recognition, and training of the system allows for refining of the feature set. The features in the feature set may thus be built up into a generic language for feature classification, individual shapes and other features comprising the vocabulary of the language.

20 It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the

invention that are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described
5 hereinabove. Rather the scope of the present invention is defined by the appended claims and includes both combinations and subcombinations of the various features described hereinabove as well as variations and modifications thereof which would occur to persons skilled in the art upon reading the foregoing description.

Claims

1. An image classifier for classifying objects of an image, comprising:

a feature measurer for measuring objects to determine whether said objects comprise features useful in classifying said objects, and

a conditionality network associated with said feature measurer for using conditionality to select said features useful in classifying interactively with measurement outputs of said feature measurer, thereby to classify said objects.

2. An image classifier according to claim 1, wherein said feature measurer comprises an optical processor.

3. An image classifier according to claim 1, wherein said feature measurer comprises a hybrid electro-optical processor.

4. An image classifier according to claim 1, wherein said feature measurer comprises a digital processor.

5. An image classifier according to claim 1, wherein said features useful in classifying are arranged to form a classification feature language.

6. An image classifier according to claim 1 further comprising a feature builder for building a set of features useful in classifying into a classification feature language.

7. An image classifier according to claim 1, said conditionality network comprising a Bayesian network.

8. An image classifier according to claim 7, said Bayesian network being controllable to use back propagation of said measurement outputs to select a next feature for measurement.

9. An image classifier according to claim 1, said conditionality network comprising a decision tree.

10. An image classifier according to claim 9, said decision tree being controllable to use back propagation of said measurement outputs to select a next feature for measurement.

11. An image classifier according to claim 1, said feature measurer comprising a 4f optical correlator, connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

12. An image classifier according to claim 1, said feature measurer comprising a 2f optical correlator, connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

13. An image classifier according to claim 1, said feature measurer comprising a JTC, connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

14. An image classifier according to claim 1, said feature measurer comprising an electronic correlator, connected to a Fourier transform unit, for performing correlation on said object against a filter representing said feature.

15. An image classifier according to claim 11, said feature measurer further comprising a Fourier transform unit connected to said optical correlator for forming a Fourier transform of said object.

16. An image classifier according to claim 15, wherein said Bayesian network comprises an inference engine.

17. An image classifier according to claim 15, wherein said inference engine comprises electronic circuitry.

18. An image classifier according to claim 17, said inference engine, being connected to said correlator, comprising parameter control functionality to control the features useful for classification to be measured by the optical correlator.

19. An image classifier according to claim 17, said electronic circuitry further comprising a classifier to conduct classification computations on the output of said correlator, thereby to perform image classification.

20. An image classifier according to claim 17, wherein said electronic circuitry comprises digital electronic circuitry.

21. An image classifier according to claim 1, wherein said inference engine comprises a Bayesian network.

22. An image classifier according to claim 1, wherein said inference engine comprises a feature selector for selecting an image feature for measurement by said optical correlator.

23. An image classifier according to claim 22, comprising a feature register for holding a plurality of features for selection.

24. An image classifier according to claim 23, said plurality of features comprising any set of the following:

symmetry in the y axis,

symmetry in the x axis,

rotational symmetry,

lines;

geometrical shapes;

corners;

line crossings;

energy primarily concentrated at low frequencies, and

energy primarily concentrated at high frequencies

25. An image classifier according to claim 24, said rotational symmetry comprising any one of a group of rotational symmetries having order numbers between 2 and infinity.

26. An image classifier according to claim 25, said group of rotational symmetries comprising rotational symmetries having order numbers of 2, 3, 4, 6, 10, and infinity respectively.

27. An image classifier according to claim 22, said feature selector comprising a classification probability estimator for estimating a probability of successful classification of said object given recognition of a given feature, said feature selector being operable to use said probability to select a feature to present said feature to said optical correlator.

28. An image classifier according to claim 1, said optical feature extractor comprising a first imaging device located at the output of said Fourier transform unit for forming an image of said Fourier transform of said feature.

29. An image classifier according to claim 1, further comprising a duplicator connected between the output of the Fourier transform unit and the

input of the optical correlator, said duplicator being controllable from said inference engine, said duplicator being operable to generate a spatial filter for use in correlation of said image.

30. An image classifier according to claim 29, said duplicator being operable to use said Fourier transform output as an input to said generation of said filter.

31. An image classifier according to claim 30, operable to receive an instruction from said inference engine to measure a given symmetric property and to generate said filter by duplicating said Fourier transform, thereby to form a filter having said given symmetric property.

32. An image classifier according to claim 29, said duplicator being operable to generate an arbitrary filter.

33. An image classifier according to claim 32, said arbitrary filter being selectable by said inference engine from a predetermined set of frequency filters comprising a low pass filter, a selection of band pass filters and a high pass filter.

34. An image classifier according to claim 32, said arbitrary filter being selectable from a prestored set of filters comprising circular sine filters, circular cosine filters, sinc filters and circular sinc filters.

35. An image classifier according to claim 34, said prestored set of filters comprising at least three each of said circular sine filters, circular cosine filters, sinc filters and circular sinc filters respectively.

36. An image classifier according to claim 32, said arbitrary filter being selectable by said inference engine from a predetermined set of frequency filters comprising a low pass filter, a selection of band pass filters, a high pass filter, circular sine filters, circular cosine filters, sinc filters and circular sinc filters.

37. An image classifier according to claim 36, said prestored set of filters comprising at least three each of said bandpass filters, said circular sine filters, said circular cosine filters, said sinc filters and said circular sinc filters respectively.

38. An image classifier according to claim 29, said duplicator being operable to generate a filter in accordance with input from said inference engine.

39. An image classifier according to claim 29, said duplicator being connected to display said spatial filter in said correlator.

40. An image classifier according to claim 29, further comprising a second imaging device at the output of said correlator, for forming an image of a response of said feature to said spatial filter.

41. An image classifier according to claim 40, comprising an evaluator connected to said second imaging device for evaluating said imaged response and ascribing a value thereto, said value being transferable to said inference engine.

42. An image classifier according to claim 1, said features being object dependent features.

43. An image classifier according to claim 42, said object-dependent features being any ones of a group comprising geometric features, relational features and frequency features.

44. An image classifier according to claim 43, said geometric features being any ones of a group comprising features containing symmetry and features containing internal primitive features.

45. An image classifier according to claim 29, operable to identify an internal primitive feature within a feature by setting said duplicator to generate a spatial filter of a suspected internal feature and measuring a correlation of said feature with said spatial filter.

46. An image classifier according to claim 45, said feature extractor further comprising a spatial angle adjustment unit for rotating said Fourier transform to produce a plurality of spatial correlations with said filter.

47. An image classifier according to claim 29, operable to identify an internal primitive feature within a feature by setting said duplicator to select from a prestored set of spatial filters at least one spatial filter of an internal

feature to be tested and measuring a correlation of said feature with said spatial filter.

48. An image classifier according to claim 47, said feature extractor further comprising a spatial angle adjustment unit for rotating said Fourier transform to produce a plurality of spatial correlations with said filter.

49. An image classifier according to claim 29, operable to identify a geometric feature by setting said duplicator to generate a spatial filter of a synthetic image related to said geometric feature and measuring a correlation of said geometric feature with said spatial filter.

50. An image classifier according to claim 49, operable to test said geometric feature for rotational symmetry by setting said duplicator to form said spatial filter by taking a rotational segment of said Fourier transform and duplicating said rotational segment to synthetically complete said transform.

51. An image classifier according to claim 49, operable to test said geometric feature for axial symmetry by setting said duplicator to form said spatial filter by taking an axial segment of said Fourier transform and duplicating said axial segment to synthetically complete said transform.

52. An image classifier according to claim 29, operable to identify frequency features of an object by setting said duplicator to produce a high pass filter and a low pass filter, each for respective intensity measurements of said feature, and calculating a ratio between said high frequency and said low frequency measurements.

53. An image classifier according to claim 29, said duplicator operable to produce at least one of a high pass, a low pass and a band pass filter, and to output said filter to said correlator for correlation with said object, said correlator further comprising an evaluator for comparing said correlated output with an uncorrelated output to ascribe a value to said correlation.

54. An image classifier according to claim 43, operable to identify a relational feature by separately identifying two features, determining a spatial relationship therebetween and applying a best fit relational description to said determined spatial relationship.

55. An image classifier according to claim 1, comprising a picture dependence analyzer for determining object features that are specific to a

current orientation of an input camera to said object, thereby to screen out said object features.

56. An image classifier according to claim 29, said inference engine being operable to control said duplicator to generate a series of filters for correlating with each object, said inference engine comprising an analyzer for using said series of correlations to determine a most likely object.

57. An image classifier according to claim 56, said inference engine using Bayesian classification to select said series of filters to generate for correlating with each given object.

58. An image classifier according to claim 57, said Bayesian classifier being constructed to divide classification uncertainty of said objects into complementary sets such that one of said sets can be substantially ruled out after each correlation.

59. An image classifier according to claim 1, said optical processor comprising a spatial light modulator electronically controllable to produce filters.

60. An optical feature extractor comprising an input for receiving a visual object, a part extractor for extracting a part of said visual object and building a filter from said part, and a correlator for correlating between said input visual object and said filter, thereby to determine the presence of a feature in said visual object.

61. An optical feature extractor according to claim 60, wherein said feature is symmetry and said part extractor is operable to duplicate said part at least once to construct said filter.

62. An optical feature extractor according to claim 61, wherein said feature is n^{th} degree symmetry, said part is an n^{th} part of said object and said part extractor is operable to duplicate said part n times to construct said filter.

63. An optical feature extractor according to claim 61, said part extractor being electronically controllable.

64. An optical feature classification decision mechanism comprising:
a Bayesian network linking a first layer of nodes and a final layer of classes via probability links, at least said first layer of nodes representing measurable features of visual objects and said classes representing potential

classification groups of said visual objects, said probability links comprising probability numbers representing potential classifications given feature measurement results,

said mechanism being operable to use said probabilities to guide a visual object measurement and classification process.

65. An optical feature classification decision mechanism according to claim 64, operable to receive a result of a first measurement and to use said probabilities to determine whether to request a second measurement or to output a classification decision.

66. An optical feature classification decision mechanism according to claim 65, further comprising:

an error calculator for calculating the possibility that a present classification result is erroneous, and

a thresholder for comparing said calculated possibility with a predetermined threshold, thereby to decide whether to output said classification decision or whether to request said further measurement.

67. A method of classifying an object taken from an image, the method comprising:

optically processing said object to extract a filter,
comparing said object with said filter,
using Bayesian networking of a predetermined set of features and a
result of said comparison, selecting either one of a feature classification result
and a next filter for a next comparison, and
continually comparing and selecting until a classification result is
reached.

1/8

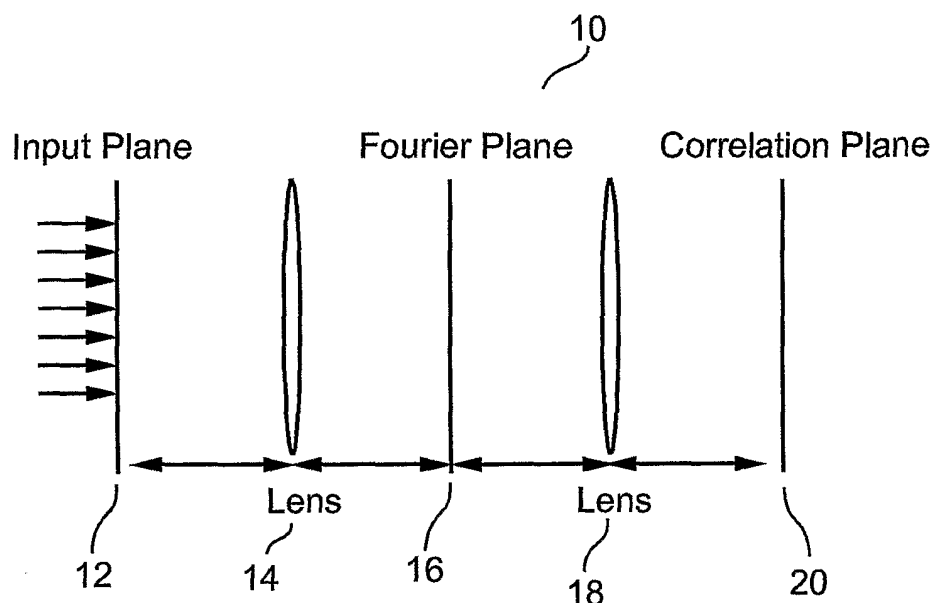


Fig. 1

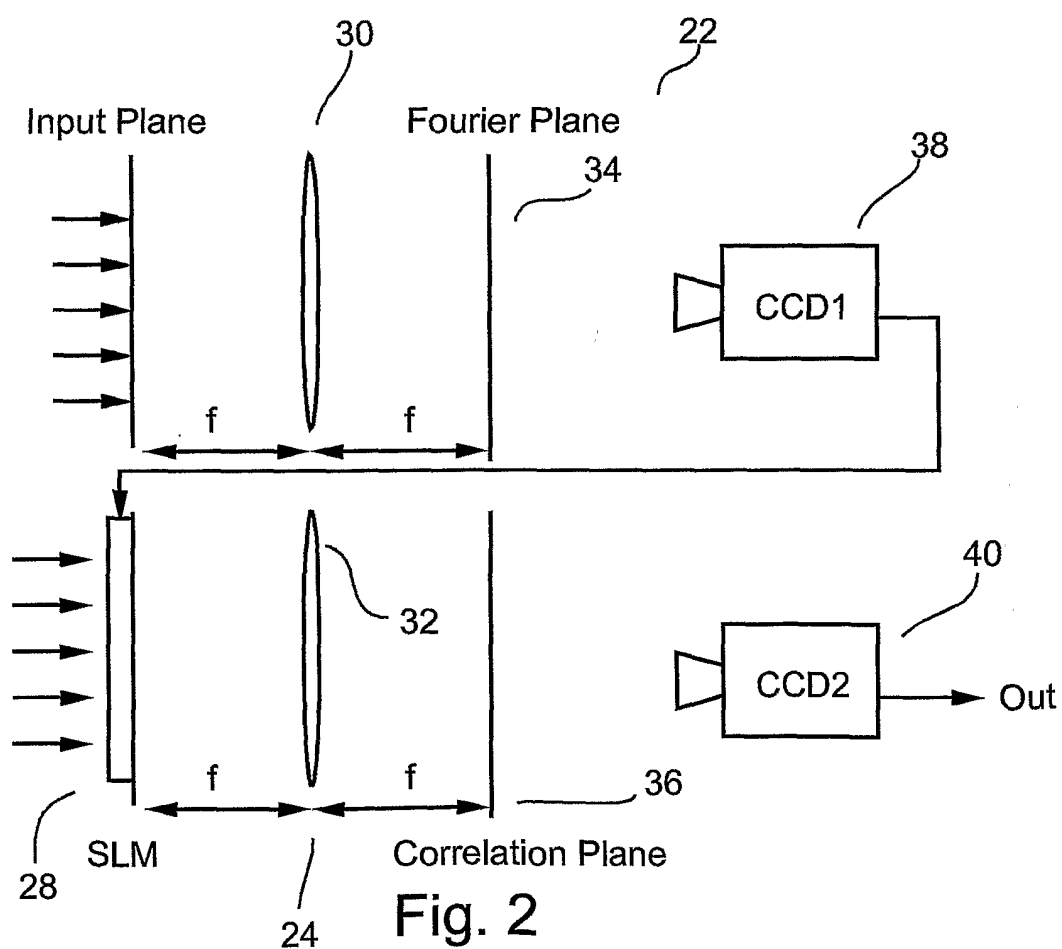
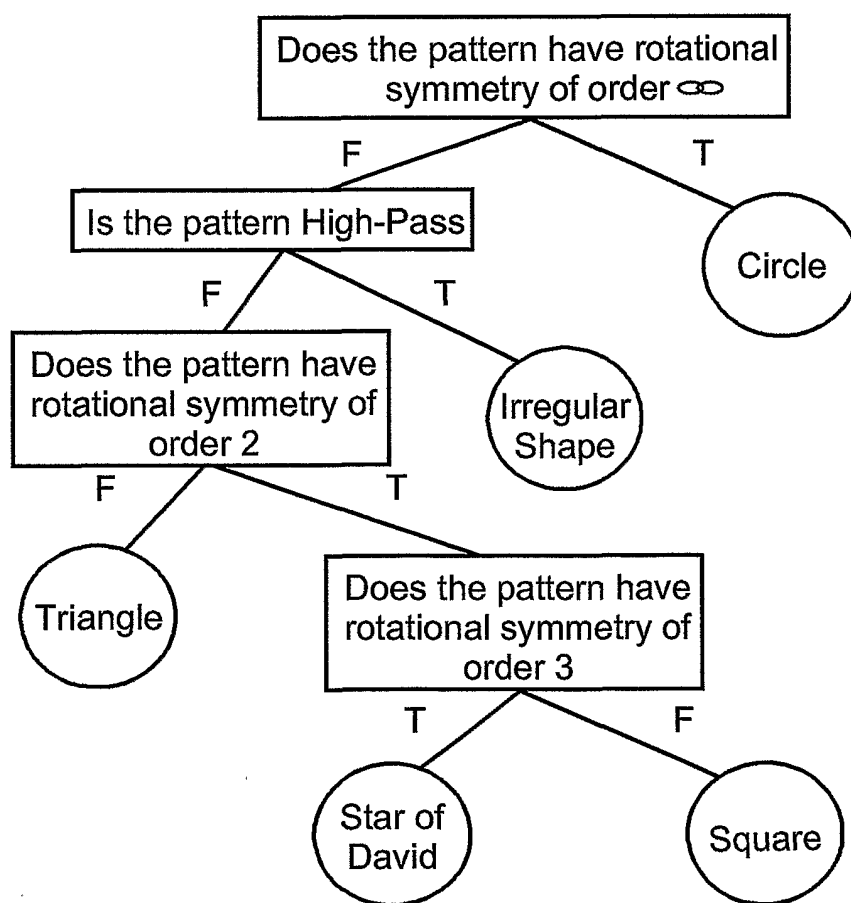
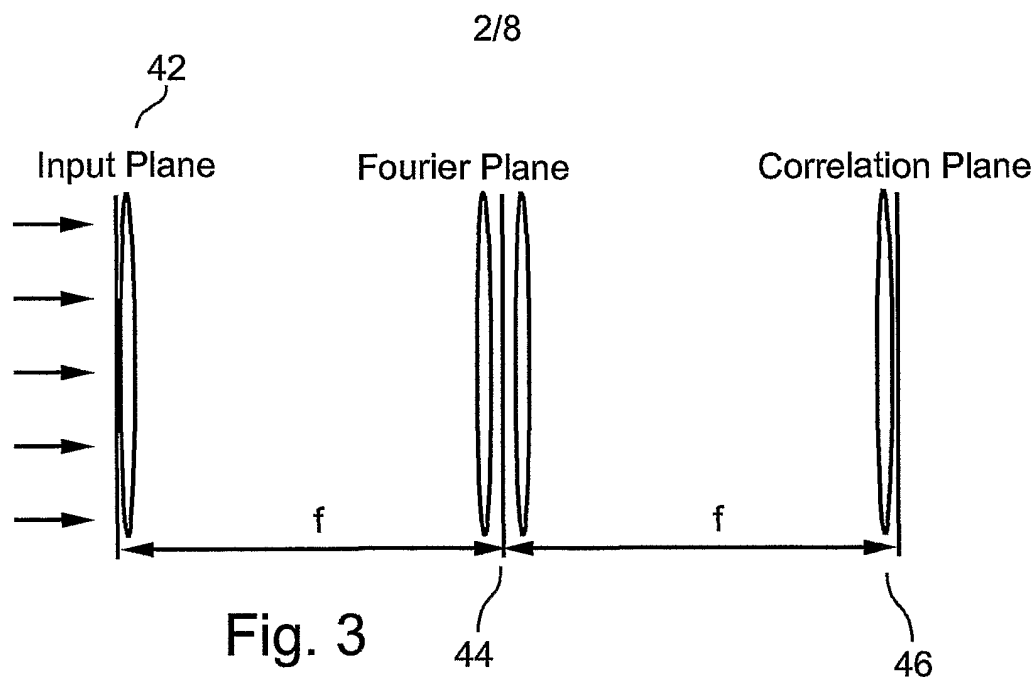


Fig. 2

**Fig. 4**

3/8

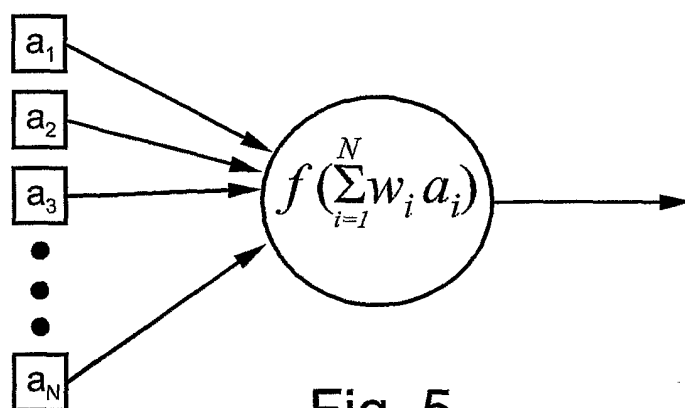


Fig. 5

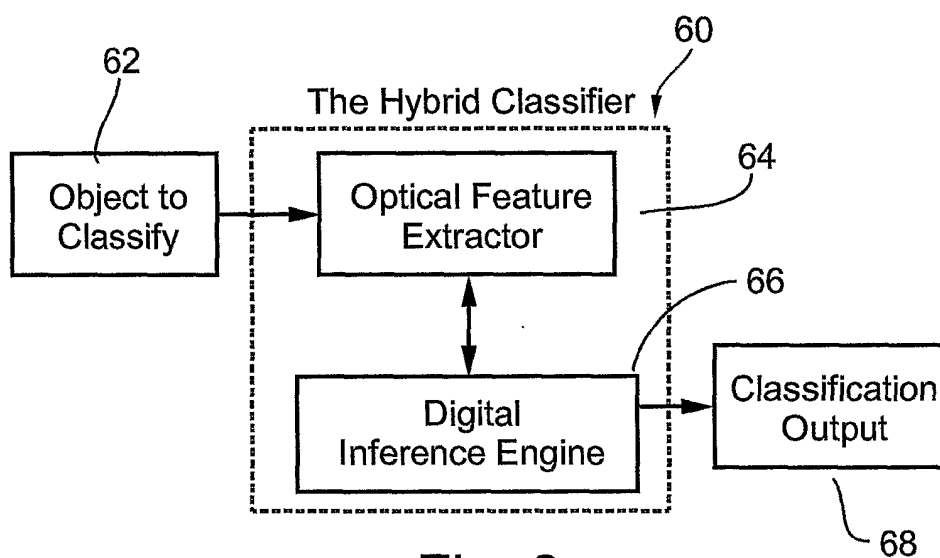


Fig. 6

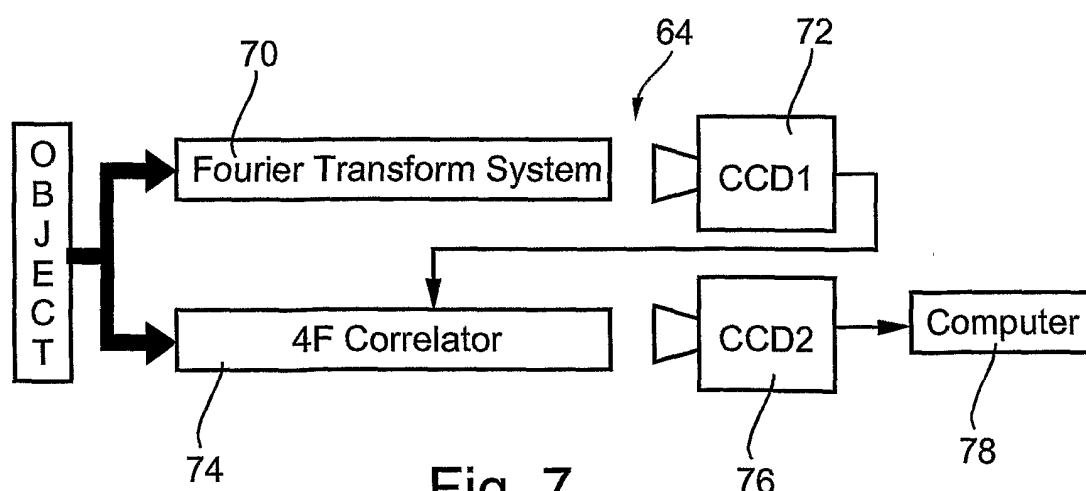


Fig. 7

4/8

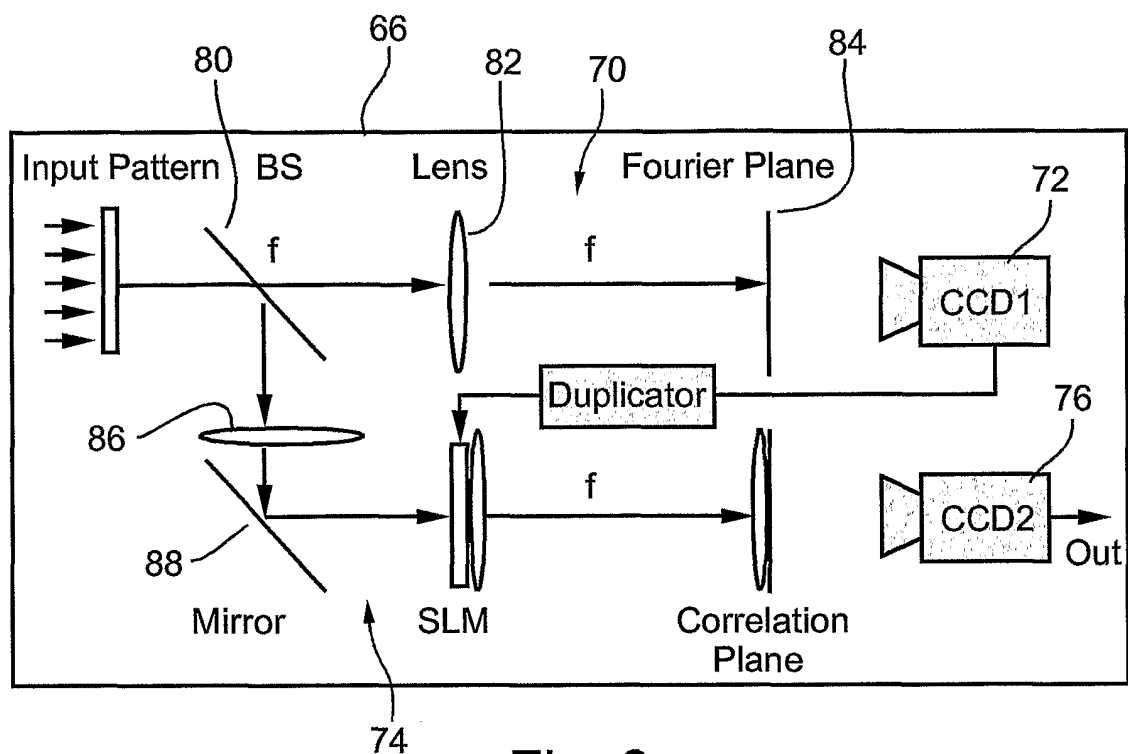


Fig. 8

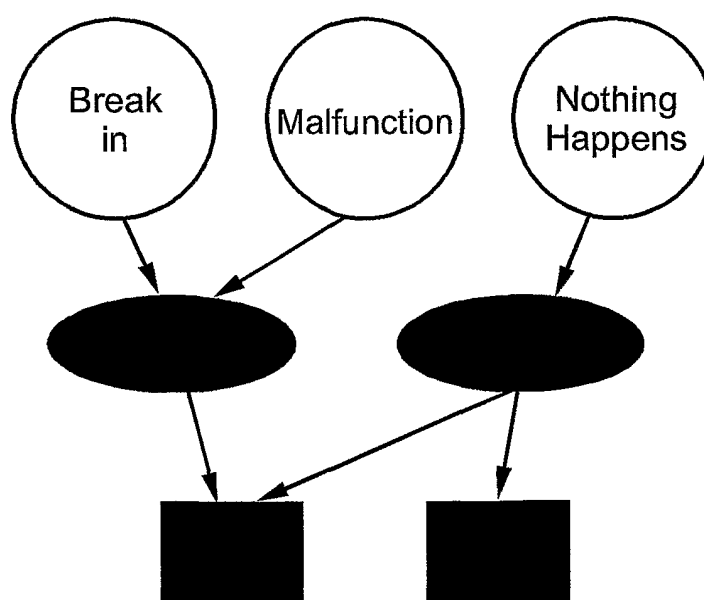


Fig. 9

5/8

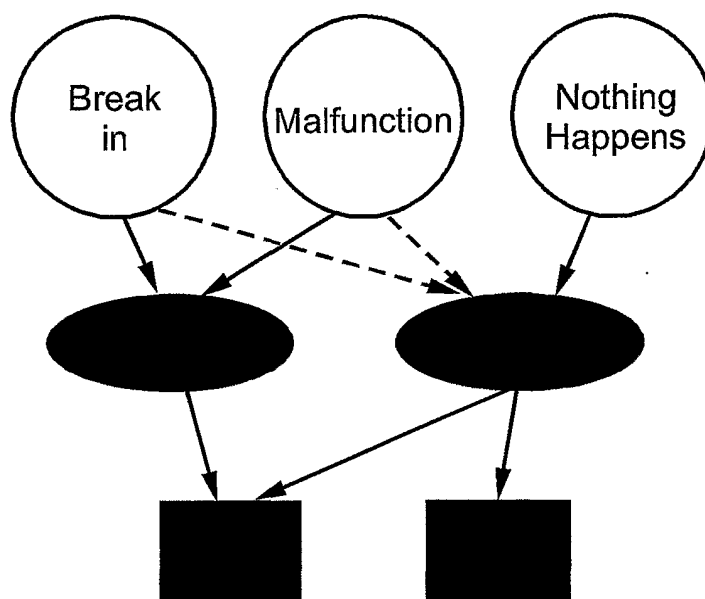


Fig. 10

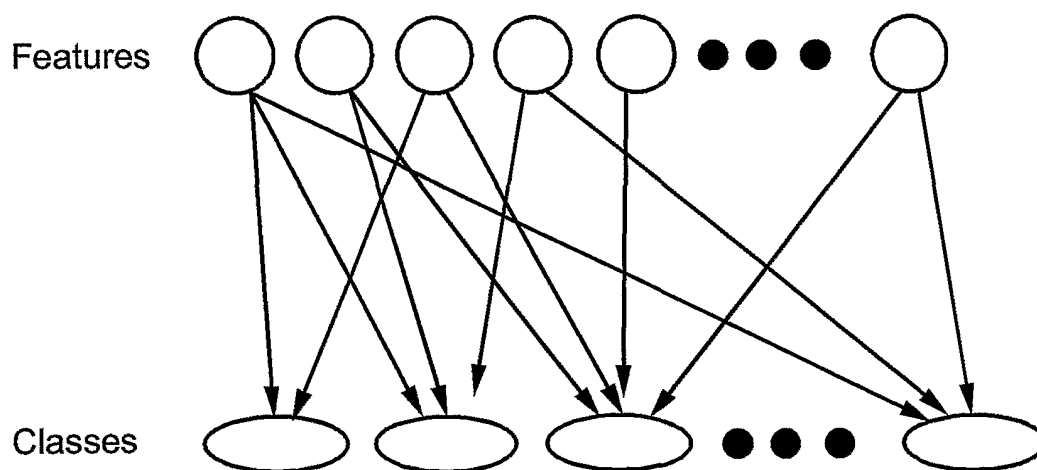


Fig. 11

6/8

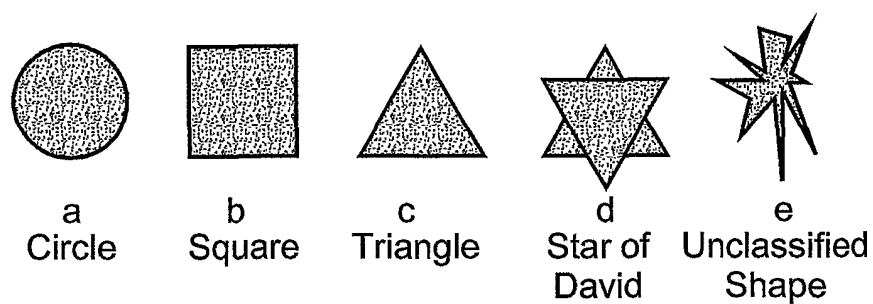


Fig. 12

Probability Distribution Over the Classes

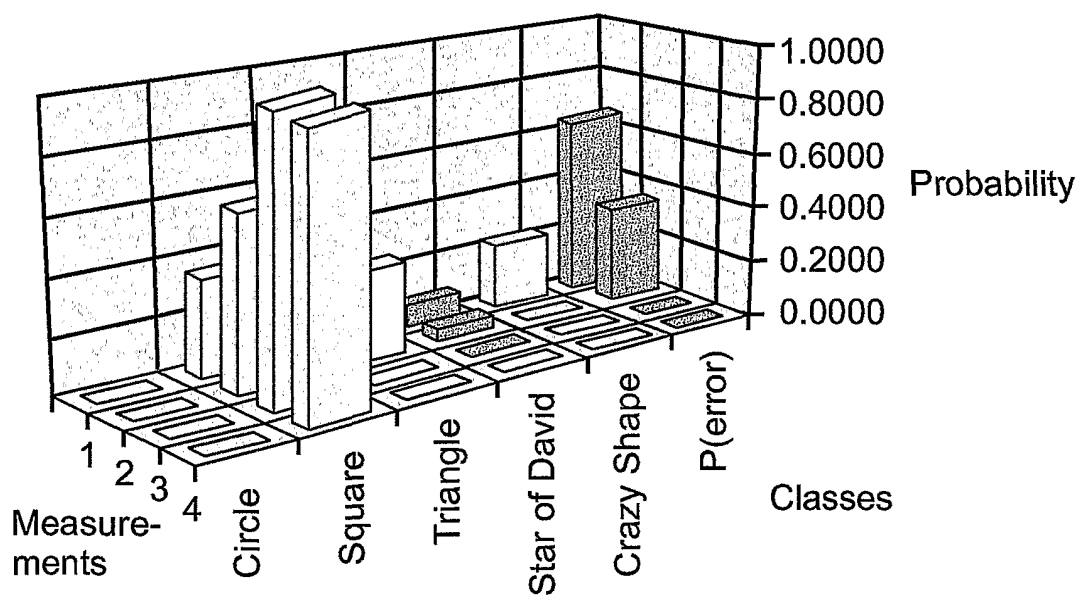


Fig. 13

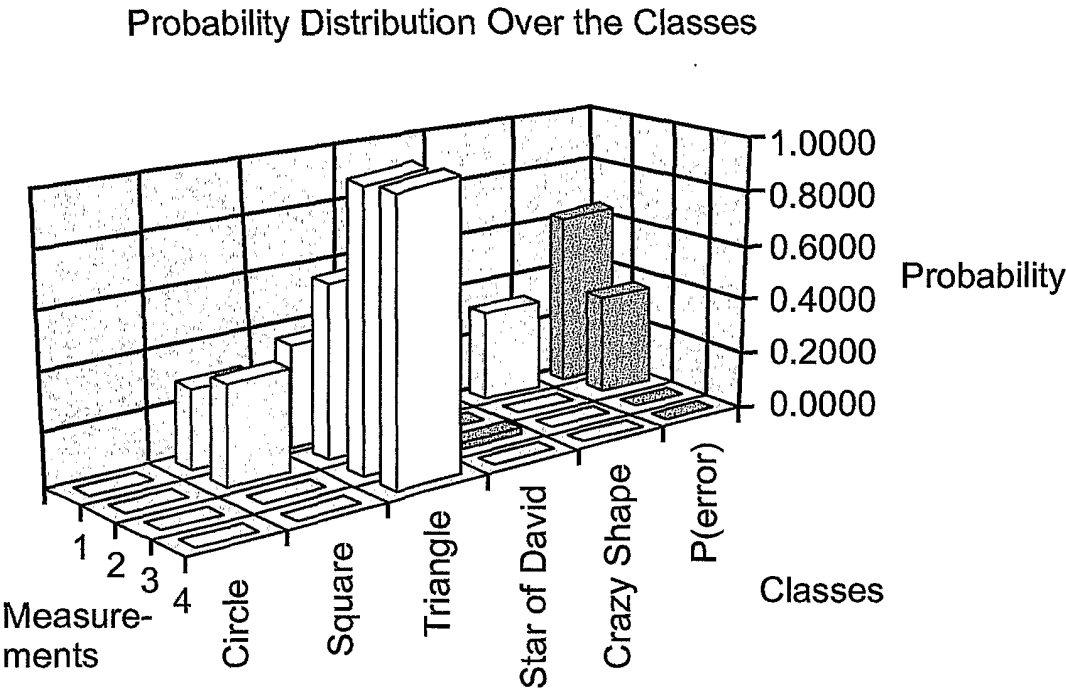


Fig. 14

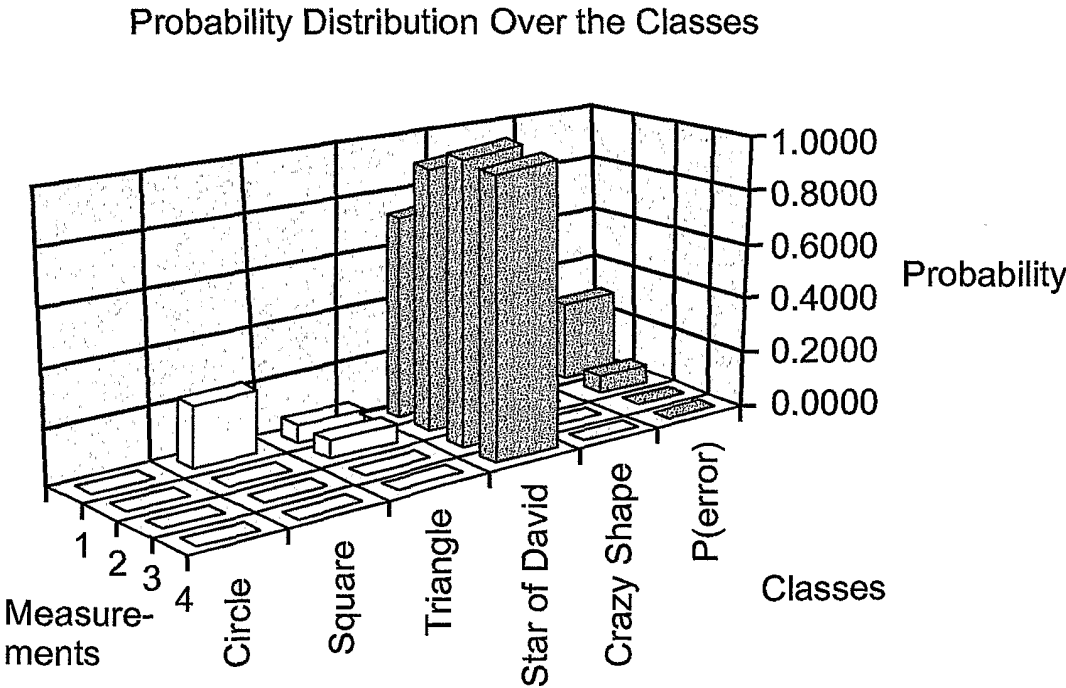


Fig. 15

8/8

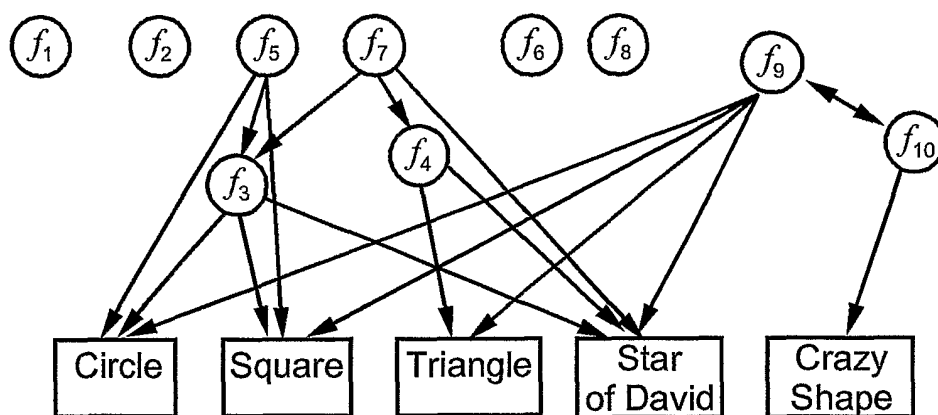


Fig. 16