ECSE-626
Statistical Computer Vision

Bayesian Object Recognition

Object Recognition Problem

- Object recognition has been one of central problems addressed in field of computer vision.
- Most common approach is to acquire database of known objects off-line during training.
- On-line, sensor measurements (e.g. from a camera) of an unknown object are acquired.

Inverse Problem

- Given a set of sensor measurements of an object in a scene (e.g. from a camera), determine the object in the database that generated those measurements.

Object Recognition Problem

Generate algorithm to get a computer recognize an object placed in front of it *in any orientation* from a single 2D image.

What is this?

Database of objects

Which 3D object in the database gave rise to this image?

Object Recognition Problem

Is this problem *well-posed*?

Recognition - Ill-posed Problem

This problem is not well posed primarily due to the violation of the uniqueness constraint.

Uniqueness: Several objects can give rise to the same measurement.

Recognition - Ill-posed Problem

What is this?

Viewpoints where impossible to pick one solution!

Object Recognition Problem

- As a result, it is not possible to identify an unknown object uniquely and several solutions may be equally viable.
- Imposing a closed-world assumption (i.e. only database objects can exist) helps to constrain the solution space.
- However, requiring a single object identity to be chosen can lead to instability in the solution – violation of condition of continuity.

Object Recognition Problem

- In addition, the solution process can be ill-conditioned (i.e. in terms of robustness against noise).
- Experimental uncertainty gives rise to uncertain measurements.

Object Recognition Solutions

- In general, shortcomings in solutions to the problem of object recognition can be categorized as follows:

Object Recognition Solutions

Deterministic strategies:

- Uncertainties are not explicitly represented in the solution.
- Solutions return a single object label from a set of measurements, even with the possibility of ambiguous or erroneous results.
- No way to assess the validity of the result.

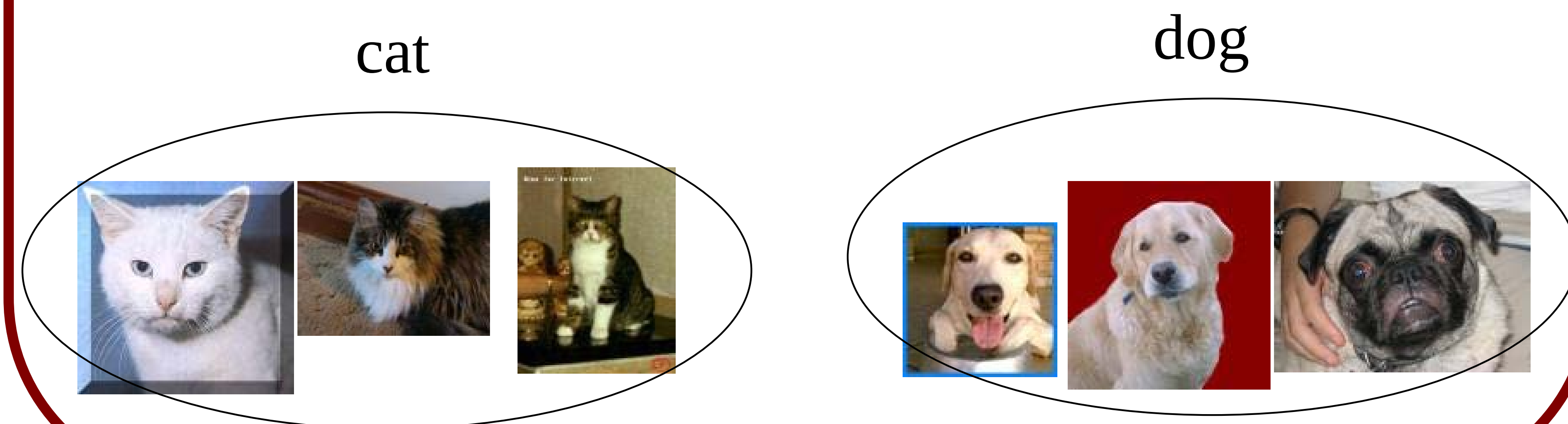
Object Recognition Solutions

Static Approaches:

- Static approaches focus on the problem of identifying an object based on data acquired from a single viewpoint in a scene.
- Usually results based on a single image are not sufficient.

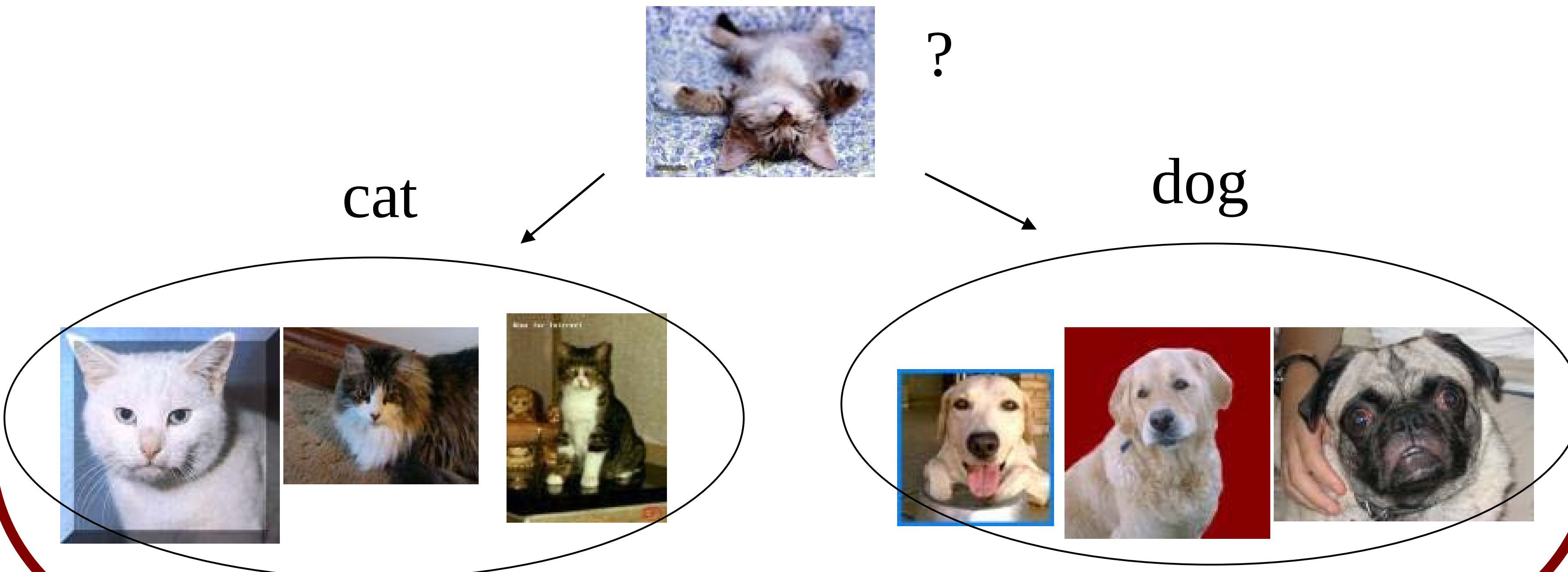
Classification

- A more general problem is that of classification: Camera measurements of objects of different classes are acquired off-line during training.



Classification

- Given a new image, which class does this new object belong to?



Appearance-based Techniques

- Appearance-based techniques were developed for the problem of object recognition.
- These techniques (Nayar & Murase) became popular due to their:
  - simplicity
  - fast indexing time
  - Accuracy in constrained environments
  - Relative insensitivity to small perturbations in the image

Appearance-based Techniques

- Off-line, images of each object in the database are acquired.
- A lower dimensional subspace is build based on the entire set of raw images acquired using Principal Components Analysis (PCA).
- The lower dimensional space in which the images are represented is referred to as an *appearance manifold*.

Appearance-based Techniques

- Recognition or indexing is based on projecting images acquired on-line onto the manifold and finding the closest stored image.
- The major drawback is that tight control over the image formation parameters (e.g. lighting, background) has to be enforced to ensure repeatable appearance.

Application to Robotic Vision Tasks

- One of the first application areas for appearance based recognition was in the area of industrial robotics.
- Many difficult problems in robotics are simplified when effective vision processes are available.

Application to Robotic Vision Tasks

- Some of these problems include:
  - visual position control (visual servoing)
  - visual object tracking (for controlling moving objects)
  - object inspection and recognition
- The appearance-based approaches of Nayar et al were effective in the controlled, real-time, conditions of robotics.

(S. K. Nayar, S. A. Nene, and H. Murase, "Subspace Methods for Robot Vision," CUCS-06-95, Technical Report, Department of Computer Science, Columbia University, New York, February 1995)

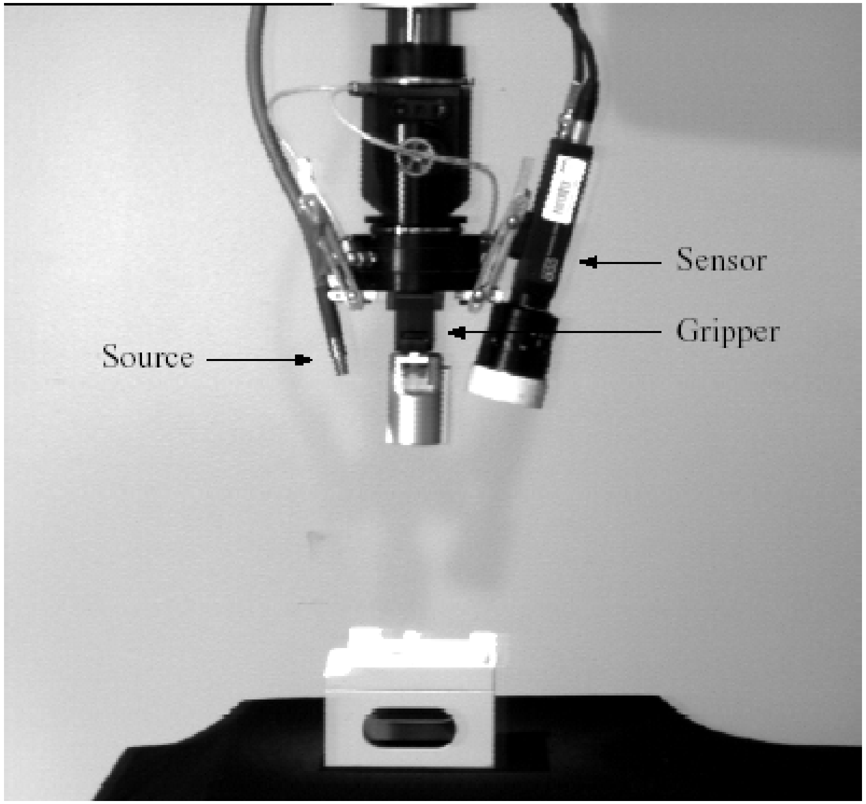


Figure 1: The hand-eye system used for visual positioning, tracking and inspection. The end-effector includes a gripper and an image sensor. In applications where ambient lighting is not diffuse, a light source may be mounted on the end-effector to achieve illumination invariance.

Application to Robotic Vision Tasks

- Nayar et. al. acquire images of an object over a sampling of positions in the *workspace* of a robotic manipulator, where the camera is fixed to the manipulator.
- A low-dimensional eigenspace is created from the principal components of all of these images.

Application to Robotic Vision Tasks

- Since the images of the object taken from neighboring points in the manipulator workspace tend to be highly correlated, they will tend to have neighboring representations in the eigenspace as well.
- Thus there is usually a smooth relationship between spatial coordinates in the robot's manipulator workspace and the eigenspace coordinates.

Application to Robotic Vision Tasks

- This relationship between the visual eigenspace and the manipulator workspace is very useful for *visual servoing* tasks.
- The relationship is easily computed.
- No feature extraction is needed, nor does the camera need to be calibrated.

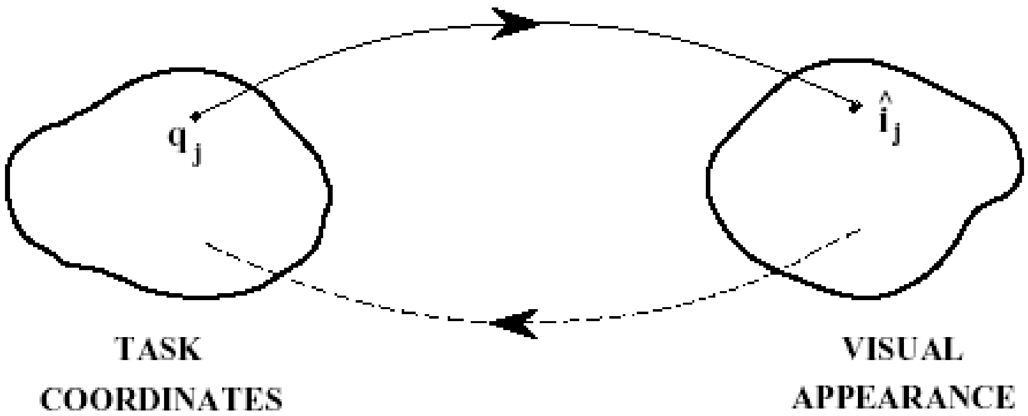
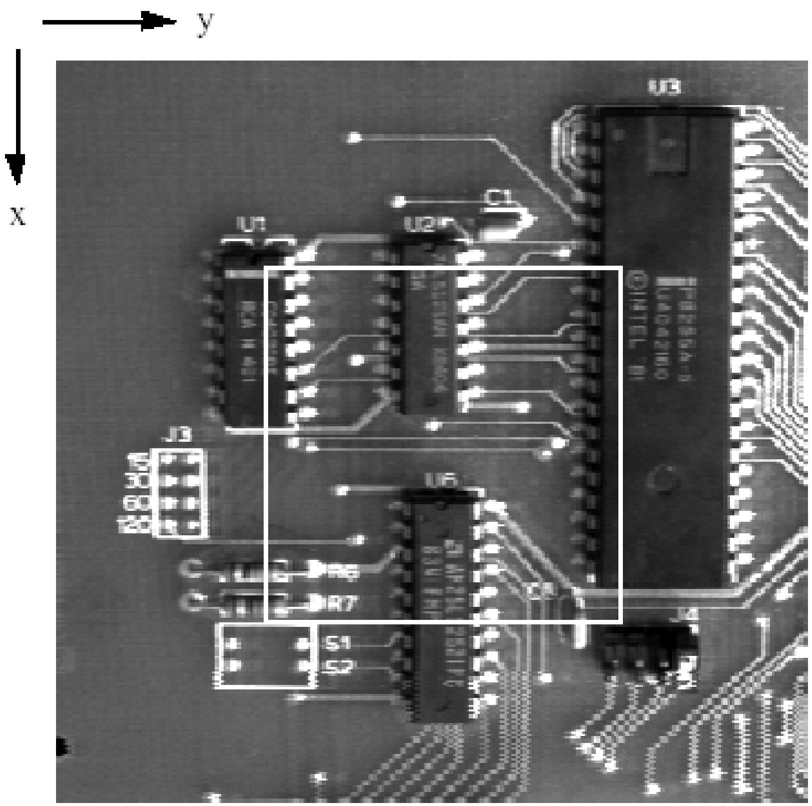


Figure 3: Estimation of the task coordinates (for instance, robot's displacement from its desired position in servoing) is based on the observation that each task coordinate can be forced to produces a unique visual appearance.

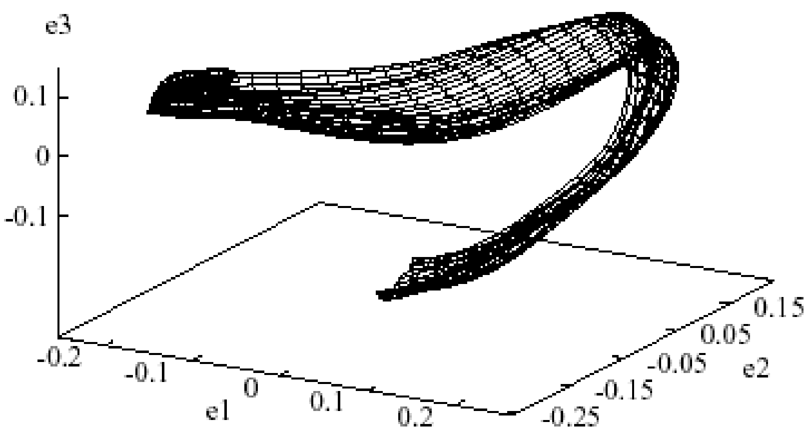
Visual Positioning Tasks

Every position of the camera gives a different image, which then defines a point in the eigenspace. Thus, every camera position corresponds to a point in the eigenspace (and vice versa).



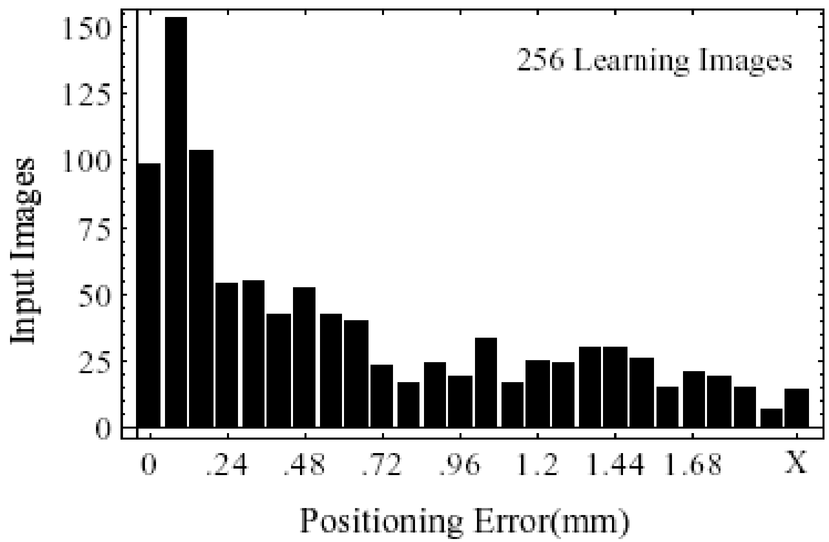
(a)

Appearance manifold

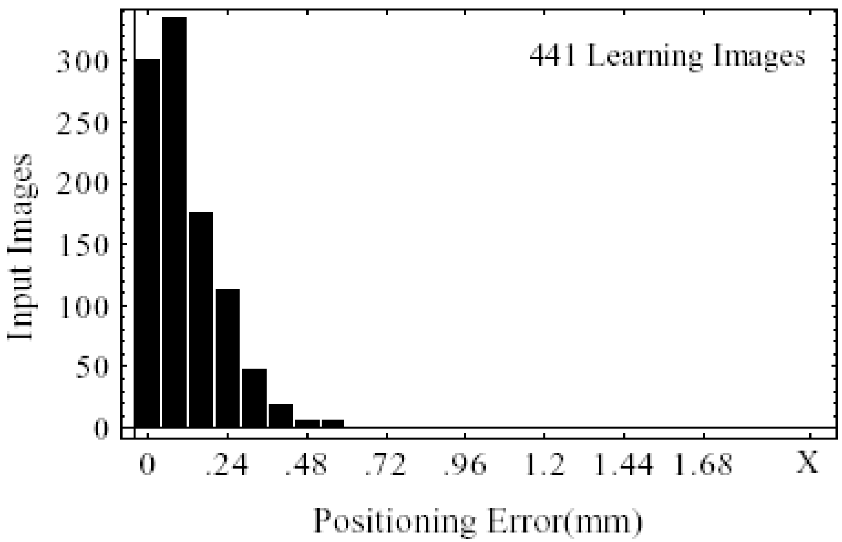


(b)

The applications use a 15-dimensional eigenspace - only a 3-D subspace is shown here.



(c)



(d)

Figure 6: Visual positioning experiment: printed circuit board. (a) Image window used for learning and positioning. (b) Parametric eigenspace representation of visual workspace displayed in 3-D. Displacements are in two dimensions (x and y). Histograms of absolute positioning error (in mm) for (c) 256 learning images and (d) 441 learning images.

Inspection Tasks

- Inspection of objects can be easily implemented with appearance-based eigenmethods.
- Nayar et. al. describe a system which passes a camera in a known trajectory over a good object (one with no defects).
- The images obtained over this trajectory are projected onto a low dimensional eigenspace and the eigenspace trajectory is stored.

Inspection Tasks

- Then, an object to be inspected is presented and the camera passes over it in the same trajectory.
- The images are again projected onto the eigenspace, and the distances between points in the inspected object eigenspace trajectory and points in the model object trajectory are computed.

Inspection Tasks

- A defect in the object under inspected will show up as a large distance in the eigenspace between the two trajectories at the location of the defect.

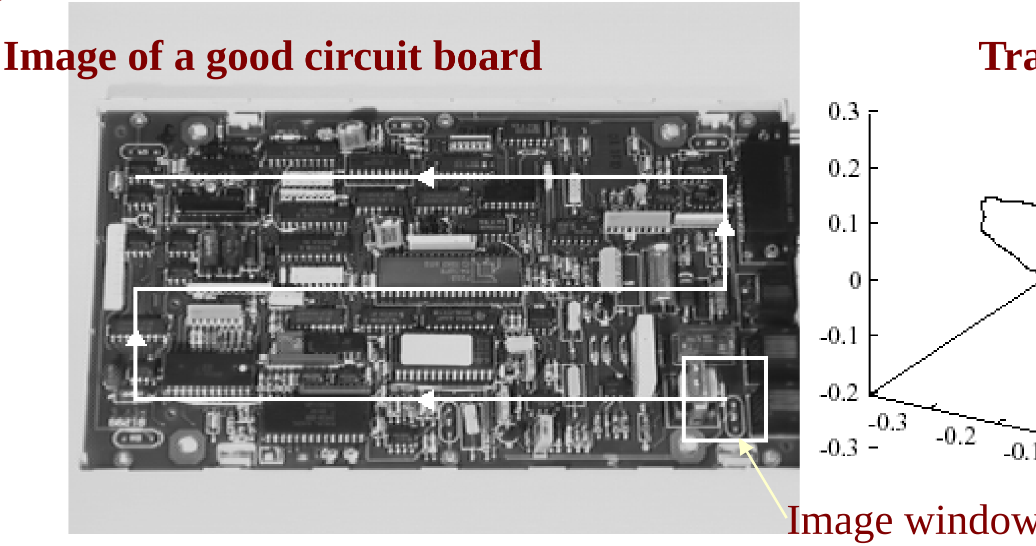
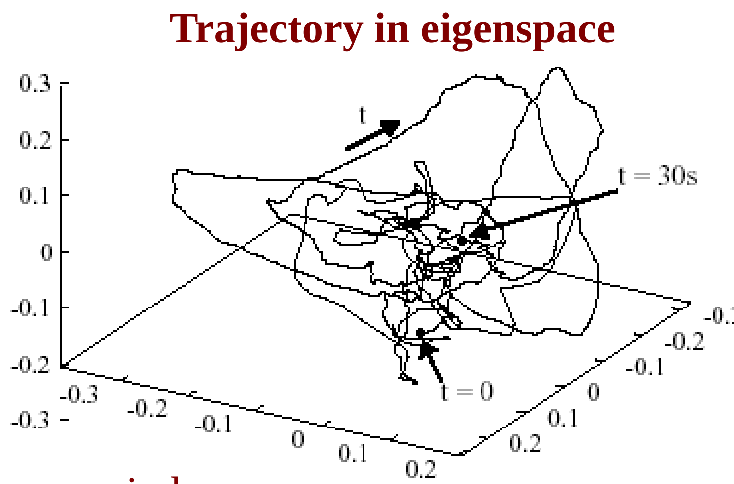


Image of a good circuit board
Image window


Trajectory in eigenspace
t
t = 30s
t = 0

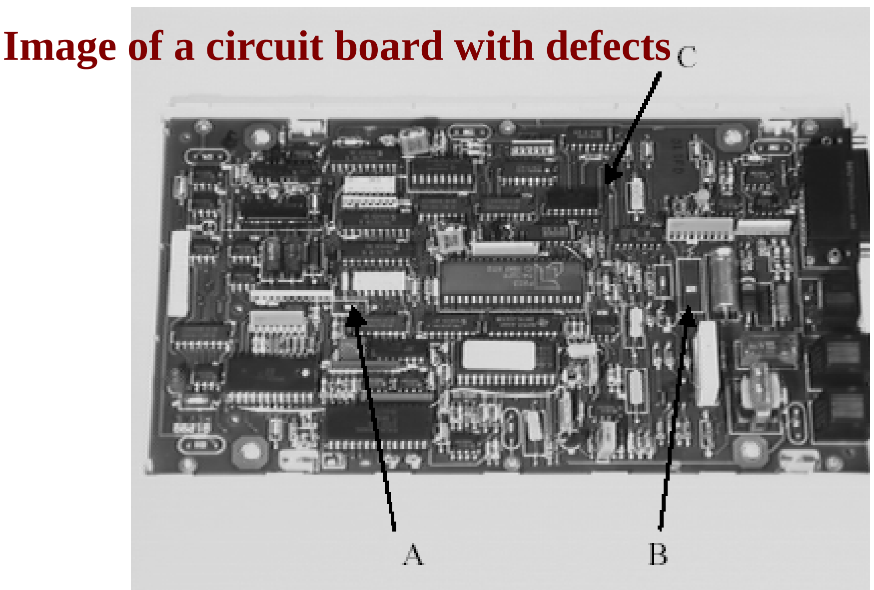
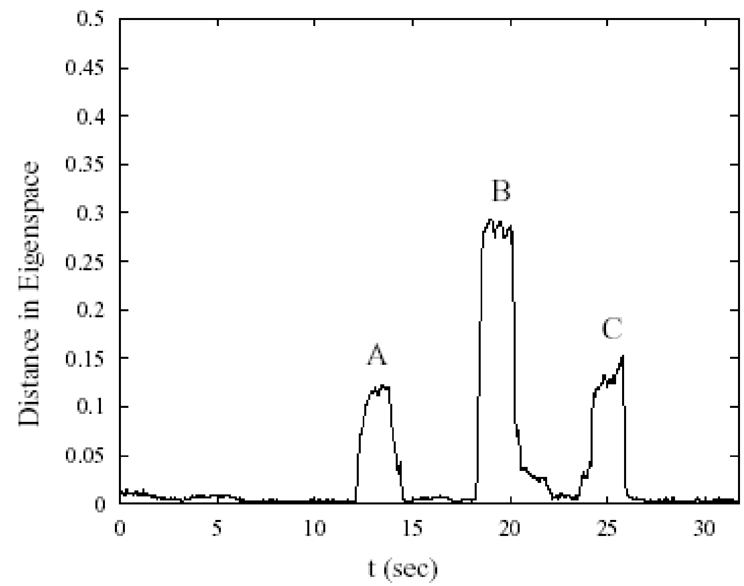


Image of a circuit board with defects
A
B
C

(c)


Distance in Eigenspace
t (sec)
A
B
C

(d)

Object Recognition

- Many approaches were developed for the ill-posed problem of object recognition. Most of the approaches were:
  - Not general, easily portable
  - Deterministic
  - Static
- Most approaches seek a single object label.

Object Recognition

- We examine a probabilistic solution to the recognition problem:
  - Sources of uncertainty explicitly represented by probability density functions.
  - Mathematically sound recipe for combining information – don't need to impose heuristics.
  - Subjective priors can be factored – their effect reduced with data collected.
  - Qualification of recognition result.

Bayesian Recognition Problem

- Recognition problem: Given a set of measurements of an unknown object, compute the *degree of likelihood* of it matching each of the objects in a stored database.
- Define the set of K possible object hypotheses:

$$\{O_i\} \mid i = 1..K$$

Bayesian Recognition Problem

- Posterior beliefs over entire set of K object hypotheses, given a set of measurements $\mathbf{d}$ in a scene (i.e. from a camera, laser rangefinder), is represented by a posterior, discrete, conditional probability density function:

Bayesian Recognition Problem

- Posterior beliefs over entire set of K object hypotheses, given a set of measurements $\mathbf{d}$ in a scene (i.e. from a camera, laser rangefinder), is represented by a posterior, discrete (conditional) probability density function:

$$p(O_i \mid \mathbf{d}) \mid_{i = 1..K}$$

Feature-based Recognition

- Consider the problem where objects are represented by models (e.g. parametric models, features) inferred from the measurements.
- Inference still takes the form of describing the belief in a set of object labels.
- There is another level of inference – involving a mapping from model space M and the space of all possible labels O.

Object Space

- O space can be visualized as a separate space, where each point is assigned a particular label i - object corresponding to the label is O_i.
- Alternately, O space can be thought of as simply a subspace of the set of all possible model parameters/features M – object labels can be linked to a particular set of model parameters/features.

Object Space

- For the particular case of object recognition, *O* space is discrete, with each point in *O* denoting a particular object in the database.

Inference Chain

- The general inference chain for this type of inverse problem can be represented by:

$$\mathbf{d} \rightarrow \mathbf{m} \rightarrow O_i$$

- This implies a level of inference from raw data, **d**, to model parameters, **m**.
- Another level of inference takes us from parameters **m** to label O_i.

Sources of Information

- Let's examine each of the sources of information available and represent each as probability density functions…

Forward Problem

- We first examine the forward problem – information from physical theory.
- For each object label, O_i, the forward problem consists of predicting the values of the observations.
- In this context, we have an additional level of inference so the forward problem consists of predicting the values of the observed parameters, **m**, for each model label, O_i.

Forward Problem

- In the case of parametric shape recognition, a measurement produces a vector description of object features.
- The theory is rarely exact due to modelization uncertainties. This information is represented by a conditional probability density function,

representing how the observed parameters **m** vary given a particular object, O_i .

Forward Problem

- In the case of parametric shape recognition, a measurement produces a vector description of object features.
- The theory is rarely exact due to modelization uncertainties. This in formation is represented by a conditional probability density function,

$$p(\mathbf{m} \mid O_i)$$

representing how the observed parameters **m** vary given a particular object, O_i .

Forward Problem

- This information is usually obtained during a training or learning phase of the recognition process.
- The general form of the function over the entire space of objects is:

$$p(\mathbf{m} \mid O)$$

Prior Information

- Information from unspecified sources about the kinds of objects which exist in the world is explicitly represented by the probability density function:

Prior Information

- Information from unspecified sources about the kinds of objects which exist in the world is explicitly represented by the probability density function:

$$p(O)$$

- For the particular case of object recognition, this implies *a priori* belief in each of the objects in the database, $p(O_i)$, i.e. a discrete distribution.

Measurement Information

- In the first stage of inference, data extracted from the scene lead to model descriptors. Estimating the parameters of the underlying model that generated the data set is an ill-posed inverse problem.
- The resulting probability density function over models parameters from a set of measurements is represented by:

Measurement Information

- In the first stage of inference, data extracted from the scene lead to model descriptors. Estimating the parameters of the underlying model that generated the data set is an ill-posed inverse problem.
- The resulting probability density function over models parameters from a set of measurements is represented by:

$$p(\mathbf{m} \mid \mathbf{d})$$

Posterior Information

- The goal of the system is to compute the degree of confidence in a set of object labels, O_i, in a predetermined database, given a set of measurements, **d**.

$$p(O\,|\,\mathbf{d}) = \frac{p(\mathbf{d}\,|\,O)\,p(O)}{p(\mathbf{d})}$$

- However, we have an additional level of inference, that inferring model parameters or features from the raw data.

Posterior Information

- As we are interesting in estimating $p(O|\mathbf{d})$, we refer to the model parameter or features, **m**, as nuisance parameters.
- These can be marginalized out as follows:

Posterior Information

- As we are interesting in estimating $p(O|\mathbf{d})$, we refer to the model parameter or features, $\mathbf{m}$, as nuisance parameters.
- These can be marginalized out as follows:

$$p(O \mid \mathbf{d}) = \int_M p(O, \mathbf{m} \mid \mathbf{d}) d\mathbf{m}$$

Posterior Information

- We wish to simplify this further given our sources of information.
- We wish to express the posterior in terms of the measurement information, for example.
- How can we split the joint density function into components that we know?

Posterior Information

- We wish to simplify this further given our sources of information.
- We wish to express the posterior in terms of the measurement information, for example.
- How can we split the joint density function into components that we know?

$$p(O \mid \mathbf{d}) = \int_M p(O \mid \mathbf{m}, \mathbf{d})\, p(\mathbf{m} \mid \mathbf{d}) d\mathbf{m}$$

Posterior Information

- Can this be simplified further?
- We postulate that **m** is a sufficient statistic for O. This simplifies the term:

$$p(O \mid \mathbf{m}, \mathbf{d}) = ?$$

Posterior Information

- Can this be simplified further?
- We postulate that **m** is a sufficient statistic for *O*. This simplifies the term:

$$p(O \mid \mathbf{m}, \mathbf{d}) = p(O \mid \mathbf{m})$$

Posterior Information

Thus the posterior reduces to:

$$p(O \mid \mathbf{d}) = \int_M p(O \mid \mathbf{m})\, p(\mathbf{m} \mid \mathbf{d})\, d\mathbf{m}$$

Given the information that we have, we can invoke Bayes' law further:

Posterior Information

$$p(O \mid \mathbf{d}) = \int_M \frac{p(\mathbf{m} \mid O)\, p(O)\, p(\mathbf{m} \mid \mathbf{d})}{p(\mathbf{m})} d\mathbf{m}$$

$$= p(O) \int_M \frac{p(\mathbf{m} \mid O)\, p(\mathbf{m} \mid \mathbf{d})}{p(\mathbf{m})} d\mathbf{m}.$$

Posterior Information

Should the posterior information from measurements, $p(\mathbf{m}\,|\,\mathbf{d})$, not be available, the solution can be equivalently expressed in terms of the physical theory or forward solution for measurements, $p(\mathbf{d}\,|\,\mathbf{m})$:

Posterior Information

Should the posterior information from measurements, $p(\mathbf{m} \mid \mathbf{d})$, not be available, the solution can be equivalently expressed in terms of the physical theory or forward solution for measurements, $p(\mathbf{d} \mid \mathbf{m})$:

$$p(O \mid \mathbf{d}) = \frac{p(O)}{p(\mathbf{d})} \int_M p(\mathbf{m} \mid O)\, p(\mathbf{d} \mid \mathbf{m})\, d\mathbf{m}$$

where $p(\mathbf{d})$ is a constant of proportionality such that:

$$p(O \mid \mathbf{d}) \propto p(O) \int_M p(\mathbf{m} \mid O)\, p(\mathbf{d} \mid \mathbf{m})\, d\mathbf{m}.$$

Evidence

- The *evidence* for the hypotheses is determined by the $p(\mathbf{d}|O)$ term.
- In this formulation the evidence refers to:

Evidence

- The *evidence* for the hypotheses is determined by the $p(\mathbf{d}|O)$ term.
- In this formulation the evidence refers to:

$$p(\mathbf{d}\,|\,O) = \int_M p(\mathbf{m}\,|\,O)\,p(\mathbf{d}\,|\,\mathbf{m})\,d\mathbf{m}.$$

- The evidence is the Bayesian transportable quantity for comparing alternate models.

Bayesian Solution to Object Recognition

- This equation is a Bayesian solution to the object recognition problem. It's form is general and can be applied to inverse problems where the posterior belief in model hypotheses is desired.
- Rather than a solution in the form of a single object identity, the system produces a probability density function describing the likelihood of the various objects in the database having led to the measurements.

Bayesian Solution to Object Recognition

- Given the ill-posedness of the problem, the importance of this result is that the *existence* of a solution is guaranteed should $p(O|\mathbf{d})$ not be identically null.
- The solution $p(O|\mathbf{d})$ is *unique*.
- In this sense, the ill-posedness of the problem can be considered alleviated.
- The shape of the distribution determines the ill-posedess of the recognition result in the classic sense (e.g. many peaks).

Sequential Recognition

- Most solutions in the literature are *static* – base interpretation on a single data set.
- Difficulty is that, even with strong *a priori* knowledge, there are still ambiguous cases where a “significant” belief in more than one model exists.
- This is further complicated by noise and quantization error inherent to most practical algorithms.

Sequential Recognition

- Ambiguities exist as a result of:
  - Occlusion
  - Projective singularities
  - Errors in parametrization and modeling
- Robustness should increase if decisions could be deferred until a sufficient confidence level is established.

Sequential Recognition

- How do we accumulate evidence for each model hypothesis, O_i, over a sequence of images?
- Evidence is in the form of a conditional probability density function. Let $\mathbf{d}_t$ denote the data acquired at time t.

$$p(O \mid \mathbf{d}_t, \mathbf{d}_{t+1}) = ?$$

Sequential Recognition

- How do we accumulate evidence for each model hypothesis, O_i, over a sequence of images?
- Evidence is in the form of a conditional probability density function. Let $\mathbf{d}_t$ denote the data acquired at time t.

$$p(O \mid \mathbf{d}_t, \mathbf{d}_{t+1}) = \frac{p(\mathbf{d}_t, \mathbf{d}_{t+1} \mid O)\, p(O)}{p(\mathbf{d}_t, \mathbf{d}_{t+1})}$$

Sequential Recognition

- In order to obtain an inexpensive solution to this problem, we make the following assumption:

Data acquired at time t is statistically independent of data sets acquired at other times.

$$p(\mathbf{d}_{t+1} | \mathbf{d}_t) = p(\mathbf{d}_{t+1})$$

Sequential Recognition

- Based on this assumption, information can easily be merged at the level of probabilities by using a recursive Bayesian chaining strategy.

$$p(O \mid \mathbf{d}_t, \mathbf{d}_{t+1}) = \frac{p(\mathbf{d}_t, \mathbf{d}_{t+1} \mid O)\, p(O)}{p(\mathbf{d}_t, \mathbf{d}_{t+1})},$$

$$= \frac{p(\mathbf{d}_t \mid O)\, p(\mathbf{d}_{t+1} \mid O)\, p(O)}{p(\mathbf{d}_t)\, p(\mathbf{d}_{t+1})},$$

$$= \frac{p(\mathbf{d}_{t+1} \mid O)}{p(\mathbf{d}_{t+1})}\, p(O \mid \mathbf{d}_t)$$

Sequential Recognition

- Thus the posterior at time t, $p(O|\mathbf{d}_t)$ can be fed back to the system as the prior at time $t+1$.
- Can we derive what this implies in our context?

Sequential Recognition

$$\frac{p(\mathbf{d}_{t+1} \mid O)}{p(\mathbf{d}_{t+1})} = \frac{p(O \mid \mathbf{d}_{t+1})}{p(O)}$$

In our context, this gives:

$$p(O \mid \mathbf{d}_t, \mathbf{d}_{t+1}) = \frac{p(O \mid \mathbf{d}_t)}{p(O)} p(O \mid \mathbf{d}_{t+1}).$$

We know that:

$$p(O \mid \mathbf{d}_{t+1}) \propto p(O) \int_M p(\mathbf{m} \mid O) p(\mathbf{d}_{t+1} \mid \mathbf{m}) d\mathbf{m}.$$

$$p(O \mid \mathbf{d}_t, \mathbf{d}_{t+1}) \propto \frac{p(O \mid \mathbf{d}_t)}{p(O)} p(O) \int_M p(\mathbf{m} \mid O) p(\mathbf{d}_{t+1} \mid \mathbf{m}) d\mathbf{m},$$

Sequential Recognition

$$p(O \mid \mathbf{d}_t, \mathbf{d}_{t+1}) \propto p(O \mid \mathbf{d}_t) \int_M p(\mathbf{m} \mid O)\, p(\mathbf{d}_{t+1} \mid \mathbf{m})\, d\mathbf{m}$$

Sequential Recognition

- By propagating evidence in this manner, evidence in the true hypothesis should grow over a short number of views, while the confidence in the others should decline.
- By quantifying the level of confidence in the various hypotheses at each stage, an active agent can gather evidence until the composite belief associated with a particular hypothesis exceeds a prescribed level of merit or until a clear winner emerges.

Sequential Recognition