Administrative Information

Lectures: Tue/Thu 2:20-3:40

- One lecture or Two?

Location: room 2311 **OR**
room 1441 (if seminars are being held in 2311)

Instructor: Prof. M. Alex O. Vasilescu

Office Hours: Tue 3:40-5 or by appointment

Office: 1422

URL: www.cs.sunysb.edu/~cse692

Grading:

Project: 50%

- *industry led project ?*
- *www.linkedin.com* – research statement and resume

Presentation: 30%

Class Participation: 20%

Course Reading:

- selected research papers

Course Topics

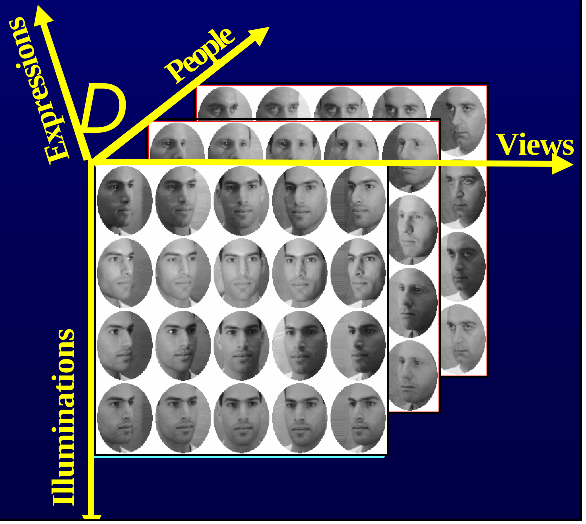
- This course emphasizes the fundamentals of:
  - statistical learning,
  - pattern analysis and recognition

  for computer vision, and computer graphics applications.
- Topics covered include:
  - supervised and unsupervised learning,
  - regression methods,
  - classification methods,
  - basic expansions and regularization,
  - kernel methods,
  - model assessment and selection,
  - model inference and averaging,
  - boosting,
  - neural networks,
  - support vector machines,
  - multilinear (tensor) analysis,
  - nearest-neighbor methods, and
  - unsupervised clustering.

Statistical Learning

- Statistics: the science of collecting, organizing, and interpreting *data.*
  - *Data collection.*
  - *Data analysis - organize & summarize data to bring out main features and clarify their underlying structure.*
  - *Inference and decision theory – extract relevant info from collected data and use it as*

    a guide for further action.



Data Collection

- **Population:** the entire group of individuals that we want information about.
- **Sample**: a *representative* part of the population that we actually examine in order to gather information.
- **Sample size**: number of observations/individuals in a sample.
- **Statistical inference:** to make an inference about a population based on the information contained in a sample.

Definitions

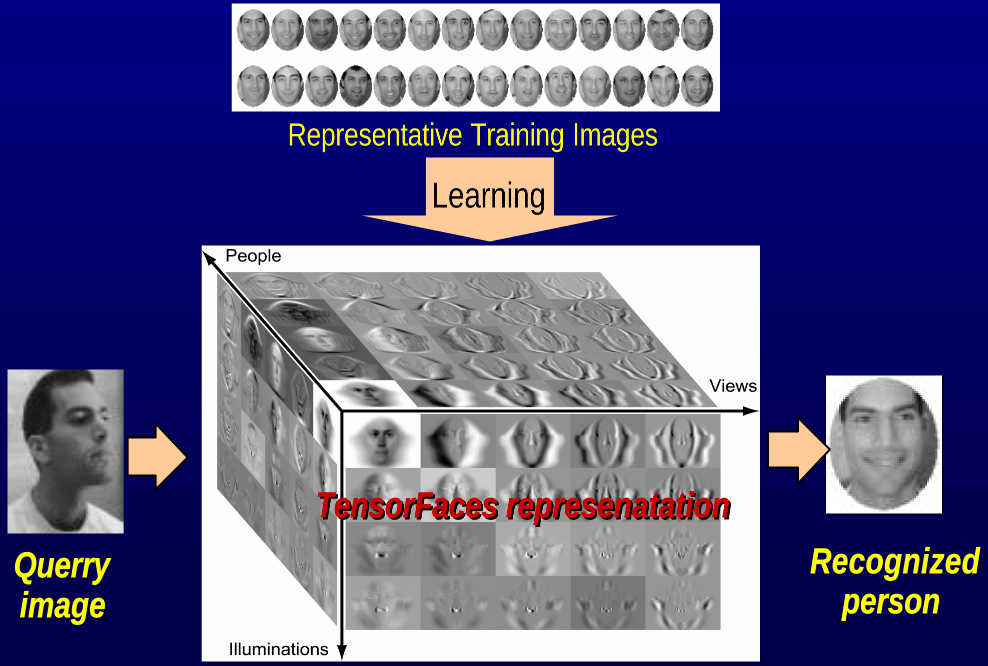
- **Individuals** (people or things) -- objects described by data.
- Individuals on which an experiment is being performed are known as **experimental units, subjects.**
- **Variables**--describe characteristics of an individual.
  - – **Categorical variable** – places an individual into a category such as male/female.
  - – **Quantitative variable** – measures some characteristic of the individual, such as height, or pixel values in an image.

Data Analysis

- Experimental Units: images
- Observed Data: Pixel values in images are directly measurable but rarely of direct interest
- Data Analysis: extracts the relevant information

Data inference

- Incorporate learned "knowledge" in the larger decision making process.



Representative Training Images

Learning

TensorFaces represenatation

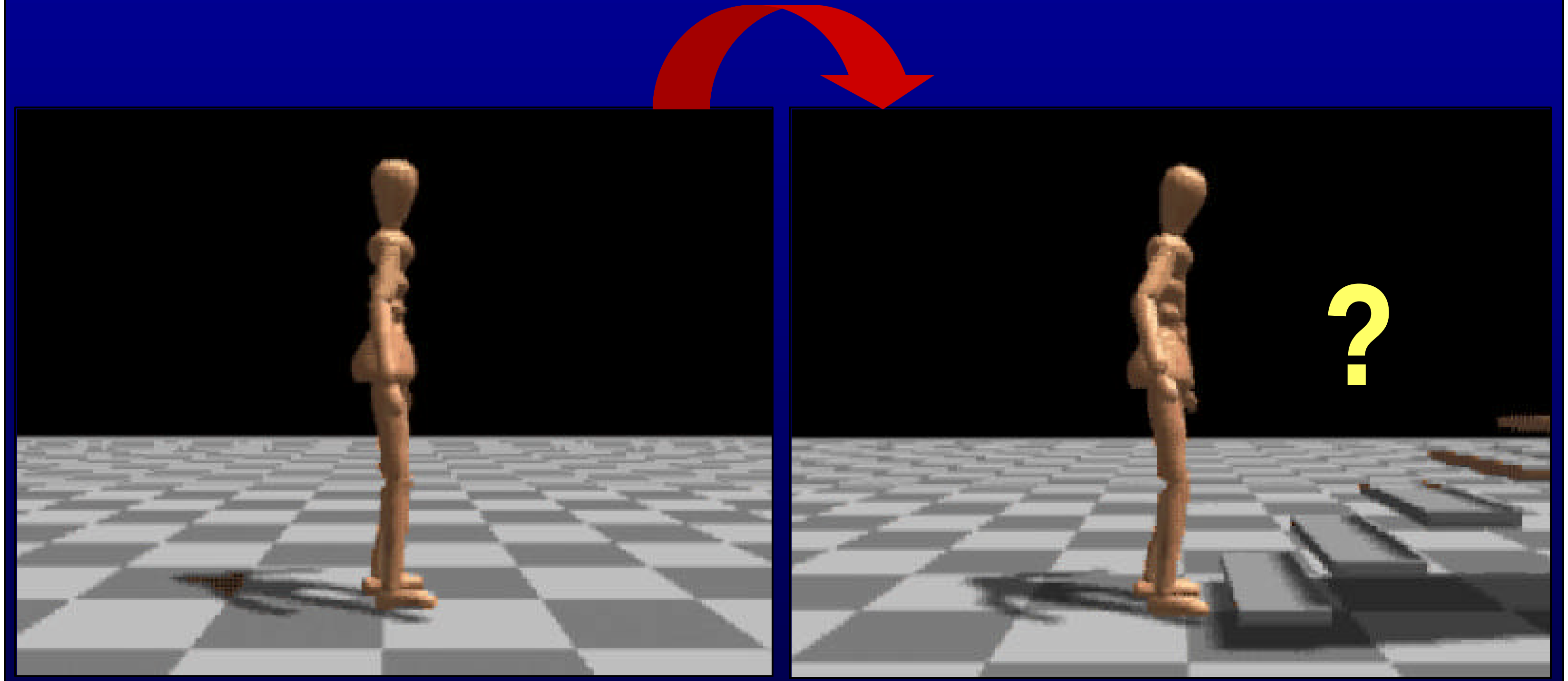
Querry image

Recognized person

Human Motion Signatures

ICPR 2002

Extract the Style Signature → Map Signature to New Action



Style Signatures are learned from motion capture and used for:

- recognition of the individual from actions not all ready in the data base
- synthesis of new actions in the extracted style

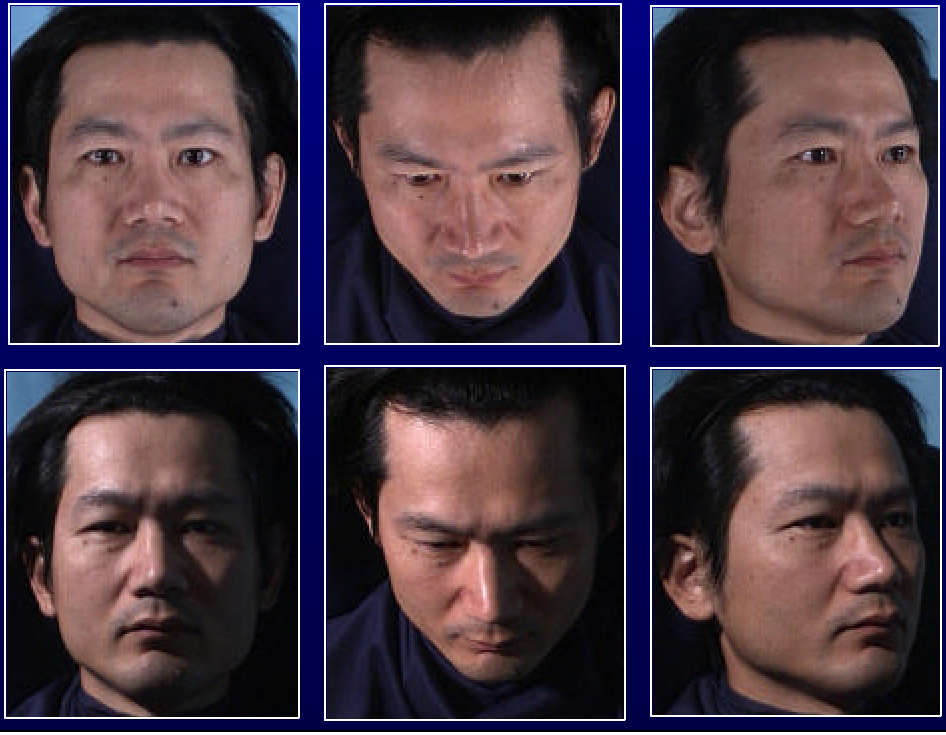
Variables

- **Response Variables** – are directly measurable, they measure the outcome of a study.
  - Pixels are response variables that are directly measurable from an image.

- **Explanatory Variables, Factors** – explain or cause changes in the response variable.
  - Pixel values change with scene geometry, illumination location, camera location which are known as the explanatory variables

Response vs. Explanatory Variables

- Pixels (response variables, directly measurable from data) change with changes in view and illumination, the explanatory variables (not directly measurable but of actual interest).



Causation vs. Association

- Strong **association** measured by **correlation** between two variables does NOT imply **causation.**
- Example of causation:
  - – Increased drinking of alcohol causes a decrease in coordination.
  - – Smoking and Lung Cancer.
- Example of association:
  - – High SAT scores are associated with a high Freshman year GPA.
  - – Smoking and Lung Cancer.

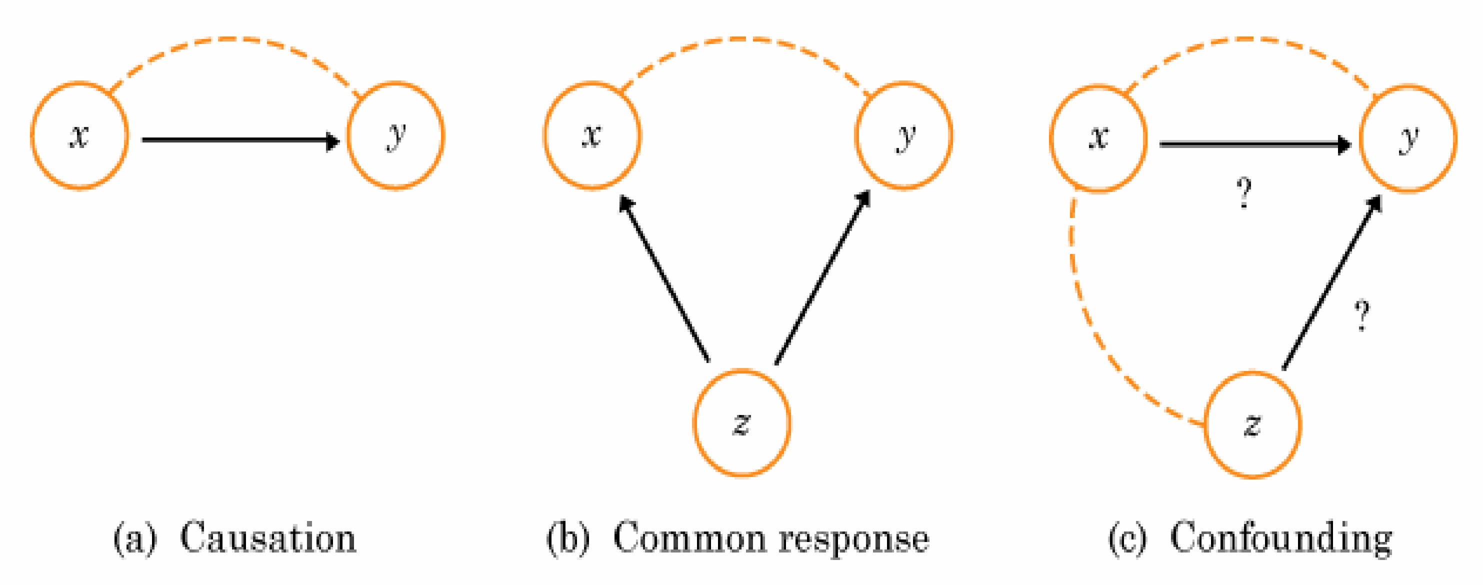
    (Genetic factor that predisposes people to nicotine addiction and lung cancer)

Explaining Association

An association between two variables x and y can reflect many types of relationships

association

causality



(a) Causation

(b) Common response

(c) Confounding

The question of causation

- A strong relationship between two variables does not always mean that changes in one variable causes changes in the other.
- The relationship between two variables is often influenced by other variables which are lurking in the background.
- The best **evidence for causation** comes from *randomized comparative experiments.*

- The observed relationship between two variables may be due to direct causation, common response or confounding.
- ***Common response*** refers to the possibility that a change in a lurking variable is causing changes in both our explanatory variable and our response variable
- ***Confounding*** refers to the possibility that either the change in our explanatory variable is causing changes in the response variable OR that a change in a lurking variable is causing changes in the response variable.

Causation

- Cause-and-effect
- Examples
  - Amount of fertilizer and yield of corn
  - Dosage of a drug and the survival rate of the mice

Common Response

- Lurking variables
- Both x and y change in response to changes in z, the lurking variable
- There may not be direct causal link between x and y.
- Lurking variables can create nonsense correlations!

TV and life expectancy

- For the world’s nations, let *x* be the number of TVs/person and *y* be the average life expectancy;

- A high positive correlation
  - nations with more TV sets have higher life expectancies.
  - Could we lengthen the lives of people in Rwanda by shipping them more TVs?

TV example

- Lurking variable: wealth of the nation
  - Rich nations: more TV sets.
  - Rich nations: longer life expectancies because of better nutrition, clean water, and better health care.

- There is no cause-and-effect tie between TV sets and length of life.

Confounding

- Two variables are confounded when their effects on a response variable are mixed together.
- One explanatory variable may be confounded with other explanatory variables or lurking variables.
- Examples:
  - Religious people live longer.

    (Religious people tend to have healthier habits, less likely to smoke, more likely to exercise. Good habits are confounded

    with the direct effects of attending religious services)

Does smoking cause lung cancer?

- causation: smoking causes lung cancer.
- common response: people who have a genetic predisposition to lung cancer also have a genetic predisposition to smoking.
- confounding: people who drink too much, don't exercise, eat unhealthy foods, etc. are more likely to get lung cancer as a result of their lifestyle. Such people may be more likely to be smokers as well.

The question of causation cont'd

- An observed relationship *can* be used for prediction without worrying about causation as long as the patterns found in past data also hold true in data we collect in the future.
- Randomized comparative experiment is the best approach in establishing causation. But, we *can* also establish causation in the absence of a randomized comparative experiment, as long as a few things hold true.

Establishing causation without a randomized comparative experiment

- The association is strong.
- The association is consistent. If we see a strong relationship between the same two variables in study after study on *different* people, then this makes establishing cause-and-effect easier.
- Higher values in our explanatory variable are associated with stronger responses.
- The alleged cause precedes the effect in time. This is just common sense.
- The alleged cause makes sense.
- If all five of these properties exist, we can usually make a good case for cause-and-effect, even without an experiment.