

Computer Vision Education Workshop 2004

Common Syllabus Modules

Kevin Bowyer, Bruce Draper, Bob Fisher, Greg Hager, Bruce Maxwell, Sudeep Sarkar,
Daniel Scharstein, George Stockman, C. J. Taylor
Bruce Maxwell, ed.

Common Course Topics

What should be in a computer vision course? To assist educators in answering that question for their own courses, we divided the field of computer vision into fifteen common course modules, each of which corresponds roughly to a week's worth of material, give or take a few lectures. The intention is to provide educators with a set of common topics from which they can select to construct their own course. For each of the fifteen common modules we provide a more detailed definition, identify useful orderings between modules, and identify relevant mathematical skills.

The fifteen common modules we identified were:

1. Image Formation
2. Basic Image Processing
3. Camera Models & Calibration
4. Stereo and Multi-View Geometry
5. Feature Detection
6. Matching & Registration
7. Motion
8. Segmentation
9. 2D Object Recognition
10. 3D Object Recognition
11. Animate/Biological Vision Systems
12. Learning & Statistical Models
13. Tracking & Video Analysis
14. Non-visible-light Imagery
15. Applications

Common Topic Definitions

Computer Vision Paradigms

What is the purpose of computer vision?

Descriptive vision systems: look at the famous Marr Type X paper

Active vision: look at the Alymonis/Rosenfeld position paper

David Lowe's view / stimulus bound by (?)

Jain / bottom-up engineering

Relationships to human vision

Properties of human vision

- Gibson (1950) Perception of the visual world
- Irvin Rock (1983) Logic of Perception
- Gibson (1968) "Theory of Affordances"

Some want to know how does human vision work? what computational models?

How to fix blindness, color blindness, dyslexia

How to better train people in visual tasks (~Irving Biederman and sexing chickens),

Detecting NATO vs Warsaw tanks)

The primal sketch and the intrinsic image

Image Formation

In order to understand image data, we need to understand the process by which images, particularly those based on visible light, are formed. Image formation begins with light interacting with surfaces in the scene, which is described by the photometric properties of light and materials. Topics include local shape descriptions, color, texture, reflectance models, and the radiance equation.

Image formation also involves light from the scene interacting with the camera geometry, described by perspective camera models and its simplified forms, perspective geometry, and elementary geometric optics.

The final step in image formation is the transformation of light to a digital representation. This transformation involves spatial, temporal, spectral, and magnitude quantization and the resulting effects. Additional topics in image formation are: the properties and models of the sensor hardware, sensor response functions, nonlinear camera effects, image representations, and color spaces.

Basic Image Processing

The basic idea of this topic is to provide a minimal background in image processing for students who otherwise will not have a course in signal and/or image processing. This is mostly important for non-engineering students.

One of the key underlying ideas is of a frequency space representation of an image. It is important for students to have at least a limited familiarity with the (forward and inverse) Discrete Fourier Transforms. This can be taught by discussing the Nyquist rate and frequency aliasing in the context of image subsampling, smoothing or other image manipulations. Supersampling is also

important, including nearest neighbor, bilinear and cubic pixel interpolation. This topic can also include image compression, and the artifacts introduced thereby. Finally, students should become at least passingly familiar with low-pass, high-pass and band-pass filtering.

Students should also become familiar with simple, point wise image transformations. The simplest of these is thresholding. More important are transformations that remap images either geometrically (e.g. rotation, morphology) or photometrically (e.g. histogram equalization). Median filters, may also be included, especially as an example of a non-linear filter.

Camera Models & Calibration

The goal of this module is to introduce students to the essential geometric aspects of cameras, and to familiarize them with algebraic manipulations related to those models. As a related topic, the notion of projective coordinates and some essential projective concepts should be included. Another set of topics relates to the recovery of camera model parameters from data. Optionally, the problem of pose estimation from images can also be covered.

Rigid body transformations

Camera Projection Models:

- Orthographic
- Affine
- Perspective
- Projective; also homographies and other important projective concepts

Internal Parameters

Lens distortion models

Calibration Methods

- Linear (for affine cameras)
- Direct
- Indirect
- Multi-Plane method

Pose Estimation

Demo/Lab exercise is MATLAB calibration toolkit

Stereo and Multi-View Geometry

The goal of this unit this develop the concepts of multi-camera geometry, and to apply those concepts in the context of 2-camera stereo and multi-camera structure and motion. This in turn involves the development of efficient matching methods for non-verged camera pairs, the notion of rectification for general camera pairs, and finally the basic ideas of stratification and its relationship to what is known about camera geometry.

Assume known: homographies and projective coordinates, basic image processing

Basic idea of triangulation. Non-verged camera setup and correspondence search.

Introduce epipolar geometry and derive E matrix. Discuss rectification and the specific rectification issues related to stereo. General stereo (with known internal parameters) with E matrix estimation. Introduce F matrix and stratification for 2 cameras.

Extensions to multiple cameras:

- Factorization methods for orthographic and affine camera models
- Self-calibration with multiple-cameras (scale plus external parameters) e.g. Pollefeys”

Discuss other problems: occlusion, surface slant, etc.

Some solutions: robust matching, global methods such as dynamic programming, graph cuts, etc. etc. etc.

Other methods:

- Space carving
- Silhouette –based methods

Feature Detection

Pixel data is often noisy and in need of cleaning operations, such as smoothing, hole filling, or thresholding. Averaging and median filtering are common methods of reducing Gaussian noise. Gradient operations are used to detect pixels in high contrast neighborhoods that are likely to be on lines or object boundaries. Corners can be detected in neighborhoods as the intersection of edge elements or by specific template-matching. The Canny edge detector and the Frie-Chen edge detectors provide sound theory for decision-making. Image texture relates to variation in surface reflectance and provides a lot of information about both the material of the surface and also the surface orientation. Texture can be captured operationally in terms of image energy as measured by Gabor or Fourier transformations in terms of frequency and direction, or in terms of ad hoc methods such as intensity cooccurrence statistics. Texture of a region can be represented in a feature vector that may be of use in segmentation, image retrieval, or correspondence for stereo or motion detection.

- Ch 6 of Ballard and Brown
- Ch 9 of Haralick and Shapiro
- Ch 7 of Jain, Kasturi, Schunk
- Ch 6 of Nalwa
- Ch 7 of Shapiro and Stockman

Matching & Registration

The goal of this module is to introduce the notion of matching images or portions thereof subject to a given class of transformations. The transformation models should include simple translation, rigid and nonrigid 2D motion, affine deformations, homographies, and non-rigid motion. Methods for computing these given matches should be discussed. Having established this, region and feature-based matching algorithms will be introduced and various method of applying them is discussed.

Assume known: feature extraction and linear filtering, something imaging geometry

Good example: mosaicing

The notion of a deformation model:

- Simple translation
- Translation plus rotation plus scale
- Affine
- Homography
- Nonrigid

- 3D –3D

Given points, solve for model.

Region-based

- Simple parametric: SSD, SAD, NNC and discrete search
- Local refinement to subpixel through optimization.
- More complex models (rotation, affine).

Feature-based

- Review feature extraction
- Feature matching algorithms: simple matching methods (e.g. edge matching). ICP as a matching algorithm
- SIFT features as a quasi-invariant example.

Motion

The goal of this unit is to acquaint students with models for differential motion in images, methods for recovering differential motion, and finally techniques such as tracking and motion segmentation that can be developed from differential motion.

Velocities of points in 3D space; ego vs. independent motion

Motion field

Optical flow – aperture problem, regularization-based approaches pyramid/patch-based approaches

Maybe tracking

Maybe reconstruction from optical flow

Motion segmentation and/or multiple multiple motions

Segmentation

Segmentation is the process of labeling pixels in an image as being part of a group via properties. The goal is to find image parts that correspond to objects, or object parts, in the scene, such as a face, the iris of an eye, a car, etc. The two classic approaches to image segmentation are (1) region-based and (2) edge-based. Conceptually, edges are boundaries between different regions and regions are connected sets of pixels that satisfy some similarity property. In some real images some regions or pixels may not correspond to any object of interest in the scene. The simplest segmentation method is perhaps thresholding or level-slicing, which can be used when object pixel properties have known difference with the properties of the background. Segmentation can be viewed as a clustering problem, so methods, such as K-means can be used for segmentation. Other methods are based on graph cuts, mean shift, and expectation maximization. Almost all books in computer vision will have some separate treatment of image segmentation, and so do some image processing texts.

- See Chapters 4&5 of Balalrd and Brown (1982)
- Chapters 2,10,11 of Haralick and Shapiro (1992)
- Chapter 3 and parts of 6 of Jain, Kasturi, Schunk (1995)
- Chapter 3 of Nalwa (1993)
- Chapters 3&10 of Shapiro and Stockman (2001)

2D Object Recognition

2D object recognition involves matching image data to a priori representations of object instances or classes. This can be approached either in terms of image (pixel) matching, or 2D symbolic descriptors.

Image level matching is also referred to as appearance-based matching. The basic idea is to match one image of an object to another. The simplest measure is image correlation. More complex techniques include Principal Component Analysis (PCA), and mutual information measures (particularly for medical images). Although an entire course could be spent on various subspaces (ICA, LDA, ...), for an introductory course this level of detail is probably not necessary.

Symbolic matching techniques first extract features from the image, and then match feature values. One approach is to extract global features that describe the image. Image moments are probably the best known example of this. Another approach is to try to capture the statistical distributions of a feature through histogram matching. (The values being histogrammed may but do not have to be pixels.) A third approach is to measure the fit of the data to a simple parametric model, for example through a Hough transform.

3D Object Recognition

3D object recognition is similar to 2D object recognition, except that this time the objects are represented by 3D geometric models. The most important distinction within this category is between approaches that match 3D data to 3D models, and approaches that match 2D data to 3D models.

Systems that match 2D or 3D data to 3D models basically face two problems: correspondence and pose. Basic tools needed to solve these problems include iterative estimation of correspondences and pose. As part of this process one needs to know how to estimate rotation, translation, and scale from point correspondences. Another important characteristic of these methods is the choice of the features, which can be either point based, such as 2D SIFT features and 3D spin-image feature, or line based features.

Animate/Biological Vision Systems

Advancements in the study of human vision provide both motivation and analogies for computer vision. The relevant fields of human vision include psychology, neurology, neuro-anatomy and brain imaging.

Topics in early human vision include selective attention, difference of Gaussians filters, Gabor filters, stereopsis, and color spaces (XYZ in the eyes; opponent colors in LGN; HS in the visual cortex). Other topics include perceptual organization, non-accidental features, and neural models. More advanced topics include the division of the human vision system into the dorsal and ventral streams, and its implications for the definitions of visual tasks.

Learning & Statistical Models

Image-based decision processes need models and data. Models that can be hand-crafted tend to be simple, and it is hard to develop models in high dimensional spaces. Therefore, we need to learn representations for objects and processes, and parameters for those representations. This includes both probabilistic and symbolic models, how to design training and testing sets, and how to validate the models. Example topics include, but are not limited to: artificial neural networks, Ada-

boost, estimation-maximization [EM], support vector machines [SVM], and hidden Markov Models [HMM].

Tracking & Video Analysis

Given a set of features and hypothesized correspondences between at least two frames, tracking covers techniques for estimating the true state of objects being tracked given noise in the features, incorrect correspondences, and missing features. Examples of features include points, regions, lines, or wholistic properties of objects. Tracking also includes the concepts of feedback loops and differential tracking.

Video understanding can build on top of tracking and more generally covers the interpretation of activities, motion, and changes in the scene. It includes techniques, such as change detection that work directly on the video and do not rely on explicitly tracking items in the scene. Examples of tracking and video analysis tasks include recognizing behaviors or gestures, and identifying temporal relationships between scene elements. Common topics in tracking and motion understanding include: Kalman filters, condensation and particle, or Bayesian filters, hidden Markov models, snakes, change detection, and model-based tracking.

Non-visible-light Imagery

A computer vision course typically deals with images formed by visible light; that is, images available to the human eye-brain. However, computer vision techniques can be used to process many other important types of images obtained from sensors in medicine, robotics, etc.

Range images contain pixel values that record the three-dimensional shape of surfaces in the scene. Range images are also called depth images and it is often the case that the pixel value is the distance of the surface element to the depth image sensor.

Non visible light images include a variety of 2D and 3D arrays from various sensors, including the following.

- range image, where $z = R[x,y]$ is the distance to the scene surface element imaged at $[x,y]$
- thermograph, where $I[x,y]$ is related to surface temperature
- thematic, where $z=f[x,y]$ is a speed, density, or any other scalar variable
- X-ray, where $z = I[x,y]$ is the energy transmitted through an object along the ray to $[x,y]$
- computed tomograph, where voxel value $I[x,y,z]$ records the absorption of X-rays by the material element at $[x,y,z]$

We note that a conventional image is a real image, but it records X-rays transmitted through material rather than visible light reflected off material. The computed tomograph (CT) is not a real image, but rather a virtual image that is computed from a large collection of real X-ray images. Range images and other non visible light images are covered in

- Ch 11 of Jain, Kasturi, and Sunk
- Ch 2 of Shapiro and Stockman
- Ch 2 of Trucco and Verri 1998

Applications

Important applications can be found in manufacturing, medicine, the military, and many other aspects of ordinary life. Applications motivate students and exhibit theoretical problems studied and also show how a system must be situated in an operating environment. Applications given

below are commercial applications and do not include the many recreational, educational, and research applications. Examples have been selected according to their variety, interest, and availability of documentation.

Primarily 2D systems:

- Veggie Vision by IBM to automatically identify produce at the supermarket checker (S&S, Ch 16)
- Fingerprint or iris recognition system (S&S Ch 16)
- mammogram screening to detect breast cancer
- automatic reading of license plates
- chip packaging at Texas Instruments
- OCR and check reading

Primarily 3D systems:

- Industrial robot vision for A) navigation, B) bin picking, C) auto windshield installation, D) airplane rivet inspection
- monitoring a swimming pool for safety
- monitoring passengers in a car for safety and air bag control

CS Background

Computer vision also requires the use and programming of computers, and therefore relies on a certain level of basic knowledge in computer science. The most important of these skills is programming ability, although there is no specific language that is required as a vision course can be taught using C, Java, or Matlab with similar projects and outcomes.

In addition to programming experience, at least one course that covers algorithms and data structures, in particular multi-dimensional arrays, graphs and trees, is recommended as a prerequisite for computer vision.

Mathematical Background

Students taking computer vision need a certain level of mathematical background in order to appreciate and understand the basic material presented in the course. Such skill ought to be satisfied by prerequisites to a vision course. In addition, certain topics require additional mathematical skills that may be fulfilled by prerequisites, but may also need to be taught to some degree as part of the course.

We divided the math topics into five categories: linear algebra, calculus, probability & statistics, optimization, and geometry. within each category we identified relevant sub-topics, and highlighted those that are critical to almost any presentation of the computer vision material identified in the common course syllabus.

Table 1: Math Topics, Core Items Highlighted

Linear Algebra	Calculus	Prob. & Stat.	Optimization	Geometry
vectors	derivatives	normal and uniform distributions	least-squares estimators	coordinate systems
matrices	integrals	mean & standard deviation	gradient descent	trigonometry
vector product	gradient	pdfs/histograms	Newton's method	perspective
scalar product	tangent	population v. sample	constrained optimization	Euclidean geometry
metrics	partical derivatives	random variables	Levenberg-Marquardt optimization	planes
determinants	surface normals	combinatorics	evolutionary methods	lines
orthonormality	Taylor expansion	conditional probability	Estimation-Maximisation [EM]	projective transforms
inverses	Jacobian	covariance matrices	graph cuts	homogeneous coordinates
eigenvalues & eigenvectors	complex numbers	Maximum Likelihood Estimation [MLE]		invariants
groups		Maximum Mutual Information [MMI]		differential geometry
Singular Value Decomposition [SVD]		Bayes rule		matrix transformations
		robust estimation		